

УДК @@@

О ДВУХ ПОДХОДАХ К АНАЛИЗУ КИНЕТИЧЕСКИХ ДАННЫХ НА ПРИМЕРЕ ПРЕДСКАЗАНИЯ АКТИВНОСТИ АНТИОКСИДАНТОВ

© 2006 г. А. Л. Померанцев, О. Е. Родионова

*Институт химической физики им. Н.Н. Семенова РАН, Москва**E-mail: polycert@chph.ras.ru*

Поступила в редакцию 26.09.2005 г.

Для предсказания активности антиоксидантов в полиолефинах предложено использовать вместо традиционного способа ускоренного термического старения метод дифференциальной сканирующей калориметрии (ДСК) с последующей математической обработкой полученных данных. Значения температур начала окисления, полученные с помощью ДСК (данные X), связываются регрессионной зависимостью с величинами периодов индукции окисления, определяемыми в ходе термического старения (данные Y). Для построения моделей используются два подхода – содержательный и формальный. В первом случае применяется нелинейный регрессионный анализ в сочетании с последовательным байесовским оцениванием. Второй подход объединяет регрессию на латентные структуры с методом простого интервального оценивания. Использование одного и того же массива данных позволяет сравнить оба подхода и сделать выводы о том, в каком случае какой подход предпочтительнее.

Анализ данных и изучение механизмов сложных химических процессов является важнейшей задачей химической кинетики. Необходимость оценки кинетических параметров по экспериментальным данным, т.е. решение обратной задачи химической кинетики, как отдельное направление сформировалось в 70–80-х годах XX века. В разработке этого направления принимали участие как математики [1, 2], так и кинетики [3, 4]. Условно можно выделить два основных, принципиально различных, подхода для решения задач анализа кинетических данных. При использовании так называемого формального (soft) подхода [5–13] экспериментальные данные описываются линейной многофакторной моделью, справедливой в ограниченном диапазоне условий. В этом случае не нужно знать, как устроен механизм исследуемого процесса, однако такой метод иногда не обеспечивает требуемой точности. В другом подходе применяется содержательное (hard) физико-химическое моделирование [4, 14], которое базируется на основных кинетических принципах и позволяет получать оценки параметров с высокой точностью. Однако такой метод может быть применен только тогда, когда модель процесса известна априори. Каждый из подходов имеет свои сильные и слабые стороны, и у каждого из них есть свои сторонники и противники. Исторически сложилось так, что в России интенсивно развивался содержательный подход [15], тогда как на западе отдавали предпочтение формальным методам [16].

Описание экспериментальных данных, построение модели и предсказание новых значений, т.е.

то, что кратко называется градуировкой или калибровкой, – это одна из старейших, но все еще актуальных задач науки. Со времен Гаусса (1794 г.) для ее решения использовался регрессионный подход, в основе которого лежит принцип минимизации отклонений модели от эксперимента – метод наименьших квадратов [17]. Развитие этого подхода (метод главных компонент (1901 г.) [18], метод максимума правдоподобия (1912 г.) [19], ридж-регрессия (1963 г.) [20], проекция на латентные структуры (1975 г.) [21] и др.) позволило применять его для сложных, некорректно поставленных задач. Однако упомянуть методы дают результат предсказания в виде точечной оценки, тогда как на практике часто нужна интервальная оценка, учитывающая неопределенность прогноза. Построение доверительных интервалов традиционными статистическими методами невозможно из-за сложности задачи [22], а использование имитационных методов невозможно из-за большого времени расчетов [23].

В 1962 г. Канторович [24] предложил другой подход к задаче линейной калибровки – заменить целевую функцию системой неравенств, которая решается с помощью методов линейного программирования. В этом случае результат прогноза сразу имеет вид интервала, поэтому этот метод и был назван “простым интервальным оцениванием”. В свое время эта идея не получила должного признания и развития, что было связано, по-видимому, с недостаточным быстроедействием компьютеров. В 70–90-е годы с помощью этого метода было выполнено несколько прикладных работ [25–32]. Итоги этих исследований были подведены в

монографии [33], где подробно рассматривается основная задача, решаемая авторами упомянутых выше работ. Это – задача интервальной оценки параметров моделей, погружение области возможных значений этих параметров в гиперкуб, параллелепипед, эллипсоид, и т.п.

Такая постановка задачи представляется нам неплототворной и малоперспективной, что и было подтверждено практикой – за последние 10 лет новые работы в этом направлении не замечены. В то же время мы считаем, что идея Канторовича может дать интересные результаты, если применять ее для построения интервального прогноза отклика. В этом случае удастся решить две равно важные практические задачи. Во-первых, установить область неопределенности [34] для прогноза искомого отклика, т.е. оценить точность построенной модели индивидуально для каждого образца. Во-вторых, можно построить классификацию образцов [35], т.е. установить индивидуальные особенности каждого образца, определяемые по его взаимоотношениям как с моделью, так и с другими образцами. Общеизвестными примерами такой классификации являются такие понятия как выброс (образец, резко выделяющийся из общей закономерности) или экстремальный образец (находящийся в периферийной области модели и оказывающий значительное влияние на ее построение). Несмотря на широкое употребление этих понятий в различных исследованиях [36–44], не существует их общепризнанных определений и методов обнаружения. Метод ПИО может восполнить этот пробел.

Теоретические аспекты метода ПИО были опубликованы в работах [35, 45, 46], а результаты его практического применения нашли свое отражение в публикациях [47–49]. Основным предположением, на котором базируется метод ПИО, является ограниченность погрешности измерения. Такой подход к интерпретации экспериментальных данных нуждается в некотором обосновании. При анализе данных стандартным допущением является принцип нормальности погрешностей. Это стандартное предположение неоднократно подвергалось критике с самых разных позиций [49–51]. Любопытно, что исследователи не связывают с принципом нормальности факт неограниченности ошибок. На прямой вопрос о том, как часто исследователю приходилось обрабатывать данные, в которых присутствовали значения, лежащие за порогом четырех стандартных отклонений (4σ), как правило, следовал ответ, что если такие величины и встречались, то они безжалостно удалялись еще на стадии предварительной обработки. В то же время объем данных, с которым работают сейчас исследователи, часто превышает 10^6 [52], так что в них уверенно можно было бы ожидать 20–30 значений, выходящих за этот порог. По мнению авторов работы [43] “действи-

тельно, в реальных исследованиях химик часто может, в какой-то степени, отобрать образцы, что приводит к распределению, которое более равномерно, чем нормально”. Подробный анализ этого вопроса дан в работе [45].

В данной статье сопоставление содержательного и формального подходов к анализу кинетических данных проводится на примере оценки эффективности антиоксидантов (АО) в полиолефинах. Необходимо сразу оговориться, что речь не идет о предсказании возможной активности будущего соединения хорошо известными методами QSAR [53, 54] по известным значениям молекулярных дескрипторов. В рассматриваемом примере экспериментально проверяется качество уже существующих соединений, относительно которых имеются веские основания предполагать, что они могут быть использованы в качестве АО. Для этого используются кинетические данные, полученные методом дифференциальной сканирующей калориметрии (ДСК). Такой подход позволяет значительно ускорить процесс тестирования АО, отказавшись от традиционного длительного и дорогостоящего метода, который предполагает выдержку образцов в печах в течение 1–3 месяцев. Кинетические данные представляют значения температур начала окисления, полученные методом ДСК для различных скоростей нагрева. Они образуют X матрицу, которая калибруется по Y-матрице, состоящей из значений периодов индукции окисления, полученных традиционным путем. Мы использовали две калибровочные модели $Y = F(X)$: формальную линейную зависимость и содержательную нелинейную модель, в которой учитываются современные представления о механизме окисления полиолефинов [55]. Для построения первой модели мы применяли множественную линейную регрессию [5], в которой прогнозные интервалы получаются методом простого интервального оценивания. Во второй, содержательной модели мы использовали нелинейный регрессионный анализ [56] и строили традиционные доверительные интервалы с помощью метода последовательного байесовского оценивания [57, 58]. Таким образом, для одного и того же массива экспериментальных данных были применены методы формального и содержательного моделирования. Это позволило сравнить полученные результаты и понять, в каком случае какой подход предпочтительнее.

ОЦЕНКА АКТИВНОСТИ АНТИОКСИДАНТОВ

Исследование эффективности антиоксидантов в полиолефинах – это длительный и дорогой процесс. Стандартная процедура тестирования АО предполагает длительную выдержку образцов при постоянной температуре. Однако можно при-

менять и альтернативный подход, который позволяет предсказывать активность АО быстрее. Идея использовать для этого метод ДСК была предложена и исследована ранее [48, 59, 60]. В этих работах применялся как содержательный [58], так и формальный [48] подходы для построения модели. Однако использованные данные были небольшого размера. В этой статье мы исследуем довольно представительный набор образцов АО, который позволил нам решить две важные задачи. Первая – это практический вопрос о применимости предлагаемого альтернативного метода. Вторая задача состоит в сравнении содержательного и формального подходов к моделированию на одном и том же наборе экспериментальных данных.

Антиоксиданты – это специальные добавки, которые замедляют термоокислительное старение полимеров. Они защищают материал от окисления в ходе его изготовления и при эксплуатации изделий. Взаимодействуя со свободными радикалами, АО обрывает цепи, но при этом и сам расходуется. АО полностью подавляет окисление до тех пор, пока его концентрация не упадет ниже некоторого критического значения. Поэтому основной характеристикой эффективности АО является период индукции (ИП) – время, в течение которого концентрация АО достаточно велика. Чем больше период индукции (при той же температуре), тем более эффективен АО. По сложившейся практике период индукции определяется с помощью длительного термического старения образцов при температуре 120–140°C. Время выдержки зависит от качества АО и может составлять от 1 дня для плохого АО до 100 дней для хорошего АО. В течение этого времени экспериментатор обследует образцы в поисках явных признаков деградации – растрескивания, пожелтения и т.п. Разумеется, такой метод является ненадежным и требует больших затрат времени и средств.

Альтернативой этому методу является подход, использующий ДСК-измерения с последующей математической обработкой полученных данных. Дифференциальная сканирующая калориметрия – это метод, в котором измеряемым сигналом является величина теплового потока от образца, нагреваемого с постоянной скоростью. Если в образце происходит химическая реакция с ненулевым тепловым эффектом, то величина ДСК-сигнала пропорциональна скорости реакции. Применительно к рассматриваемой проблеме это означает, что пока концентрация АО в образце еще достаточна для подавления цепного окисления, ДСК-сигнал является константой, однако с некоторой температуры он начинает резко расти. Эта температура называется температурой начала окисления (ТНО). Проводя эксперименты при разной скорости нагрева (γ), можно получить несколько значений ТНО.

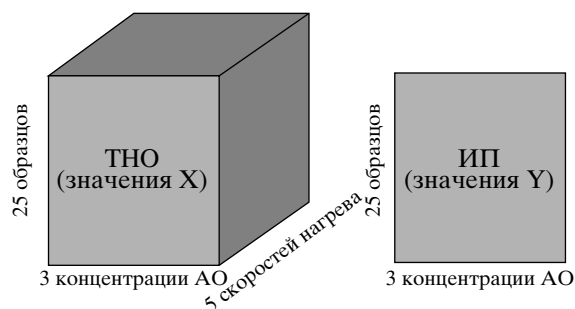


Рис. 1. Графическое представление экспериментальных данных (объяснение см. в тексте).

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

В эксперименте мы использовали 25 образцов АО, обозначенных как АО-1, ..., АО-25. Некоторые из них (например, АО-1, АО-2 и АО-3) являются стандартными добавками, используемыми в промышленности. Другие образцы – это новые вещества, для которых имеются обоснованные предположения об их эффективности. Все образцы добавляли в порошок полипропилена (ПП) в концентрациях (A_0) 0.05% (500 ppm), 0.07%, и 0.10%. После смешивания композиция была подвергнута экструзии при температуре 250°C и превращена в пленки толщиной 0.25 мм. Эти пленки и были исходным материалом для проведения ДСК-измерений и долговременного старения. Старение проводили в термошкафах при температуре 140°C. В результате этого эксперимента мы получили значения периодов индукции, выраженные в днях.

ДСК-измерения проводили в температурном диапазоне от 150 до 350°C, где наблюдается экзотермический максимум, связанный с окислением полимера. При этом использовались пять различных скоростей нагрева: 2, 5, 10, 15, 20 град/мин. Значения температур начала окисления вычисляли с использованием программы Fitter [61] способом, объясненным в [56].

В результате всех экспериментов мы получили данные, которые схематически изображены на рис. 1. Данные X являются значениями температур начала окисления (ТНО), полученными в ДСК-эксперименте. Они образуют трехмодальный (3-way) [62] блок: 25 образцов АО $\times$ 3 концентрации АО $\times$ 5 скоростей нагрева. Данные Y – это значения периодов индукции, полученные с помощью длительного термического старения. Они образуют двухмодальный блок (матрицу): 25 образцов АО $\times$ 3 концентрации АО.

Для нелинейного регрессионного анализа использовалась программа Fitter [61]. Для ПЛС-моделирования применялась программа Unscrambler [71]. Простое интервальное оценивание (ПИО) производилось с помощью программного обеспечения, выполненного как надстройка (Add-In) для

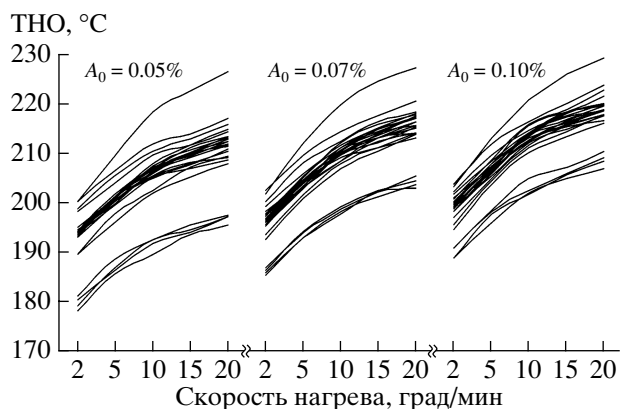


Рис. 2. Схематическое представление X-данных для ПЛС-моделирования: зависимость температуры начала окисления ПП от скорости нагрева при разных начальных концентрациях антиоксидантов.

пакета Excel. В нем использовался алгоритм NIPALS [5] для билинейного моделирования, стандартный алгоритм SIMPLEX [65] для оптимизации, а также весь необходимый набор специальных процедур для предварительной обработки данных, преобразований и т.п. Эта программа в настоящее время находится в стадии бета-тестирования.

ФОРМАЛЬНОЕ МОДЕЛИРОВАНИЕ

Полученный массив данных обрабатывали с помощью формального подхода, который объединяет методы проекций на латентные структуры (ПЛС) [5, 21] для калибровки и простого интервального оценивания (ПИО) [35] для построения прогнозных интервалов.

Исходные трехмодальные X данные раскладывали в плоскую матрицу (25×15) так, как показано на рис. 2. Важным является то обстоятельство, что ошибка в значениях периода индукции не постоянна, а растет с увеличением ТНО. Это может быть объяснено двумя обстоятельствами. Первое связано с вышеупомянутым методом визуальной инспекции, который обладает худшей точностью для более стабильных (т.е. долгоживущих) АО. Вторая причина заключается в процедуре приготовления образцов, когда малое количество АО смешивается с большим количеством ПП. Это приводит к некоторой неизбежной неоднородности распределения АО в ПП. Для того чтобы компенсировать непостоянство ошибки измерения ИП, в дальнейшем мы использовали величины квадратных корней из всех значений периодов индукции. Кроме того, при ПЛС-моделировании мы центрировали X- и Y-данные.

При построении линейной калибровочной модели

$$y = Xa + \epsilon \quad (1)$$

основная математическая трудность состоит в обращении матрицы $X^T X$, которая в нашем случае имеет размерность 15×15 . Если бы эта матрица была не вырождена (имела бы полный ранг), то такую калибровку можно было бы выполнить методом наименьших квадратов и найти оценки неизвестных параметров модели $\hat{a} = (X^T X)^{-1} X^T y$. Однако в нашем примере, как и в большинстве практических задач, эта матрица вырождена. Как мы отмечали выше, для преодоления этой трудности используются различные методы регуляризации, например, метод главных компонент, ридж-регрессия и т.д. Наш выбор пал на метод проекций на латентные структуры [5]. Суть его состоит в одновременном разложении матрицы X и вектора y и представлении их в виде

$$X = TP^t + E, \quad y = Tq + f, \quad (2)$$

где T – матрица счетов, P и q – матрица и вектор нагрузок соответственно, а E и f – матрица и вектор невязок. При этом разложении учитывают три обстоятельства. Во-первых, столбцы t_i в матрице T являются линейными комбинациями столбцов x матрицы X, т.е. $t_i = Xw_i$. Во-вторых, коэффициенты w выбираются так, чтобы максимизировать корреляцию между откликом y и t_i . В-третьих, число столбцов в матрице счетов T и матрице нагрузок P должно быть равно эффективному рангу матрицы X. Эта величина (r) называется числом главных компонент и она, естественно, меньше, чем число столбцов в X. В результате такого разложения исходная калибровочная задача (1) проецируется на пространство меньшей размерности, где новая задача (2) является уже невырожденной. Важным преимуществом метода ПЛС является возможность визуальной интерпретации данных на графиках счетов (т.е. в координатах главных компонент PC1–PC2), что мы и будем использовать далее. Ознакомиться с методом ПЛС можно по многочисленным (см., например, литературу в [5, 21]), но труднодоступным в России монографиям. За последнее время появилось, наконец, верное, но краткое изложение этого метода в учебнике [63]. Подробное изложение на русском языке ПЛС, так же как и других методов многомерного анализа данных, содержится в книге [6].

Применяя метод ПЛС к нашему примеру, мы разделили полный набор данных на две части. Калибровочный набор включал восемнадцать образцов (АО-1–АО-18); эти данные использовались для построения модели. Второй, тестовый набор данных состоял из семи образцов (АО-19–АО-25). Они использовались для проверки модели. Для каждого значения A_0 исходной концентрации АО была построена своя собственная ПЛС-модель с двумя главными компонентами. Основные характе-

Таблица 1. Главные характеристики ПЛС-моделей, построенных для различных начальных концентраций АО

$A_0, \%$	$X_{\text{expl}}, \%$	$Y_{\text{expl}}, \%$	RMSEC	r_{cal}^2	r_{test}^2	β
0.05	99	92	0.287	0.96	0.99	0.84
0.07	99	88	0.342	0.93	0.99	1.02
0.10	99	84	0.395	0.91	0.97	1.20

ристики этих моделей показаны в табл. 1. В этой таблице для каждой начальной концентрации АО представлены следующие величины, характеризующие качество построенной модели. Объясненные дисперсии в X и в y вычисляются как усредненные нормированные по калибровочному набору $(X_{\text{cal}}, y_{\text{cal}})$ невязки в модели (2).

$$X_{\text{expl}} = 1 - \frac{\|X_{\text{cal}} - TP^t\|^2}{\|X_{\text{cal}}\|^2},$$

$$Y_{\text{expl}} = 1 - \frac{\|y_{\text{cal}} - Tq\|^2}{\|y_{\text{cal}}\|^2}.$$

Они характеризуют то, насколько точно ПЛС-метод приблизил (объяснил) исходные данные. Среднеквадратичные ошибки калибровки вычисляются по данным калибровочного набора как

$$\text{RMSEC} = \frac{\|y_{\text{cal}} - Tq\|}{\sqrt{n_{\text{cal}} - r}}. \quad (3)$$

Они характеризуют точность моделирования отклика. В приведенных выше формулах n_{cal} – это число данных в калибровочном наборе, а r – число главных компонент. Норма матрицы (вектора) – это корень квадратный из суммы квадратов ее (его) компонент. Коэффициенты корреляции измеренных значений с предсказанными значениями для калибровочного (r_{cal}^2) и тестового (r_{test}^2) наборов вычисляются обычным образом. Последний столбец таблицы представляет ПИО-ошибку калибровки (β), которая объясняется далее.

Используя ПЛС-метод, мы получаем среднюю, точечную оценку периода индукции. Для того чтобы найти интервальную прогнозную оценку, мы применили новый метод простого интервального оценивания. В основе этого метода лежит единственное допущение, что все ошибки, участвующие в задаче многомерной калибровки, ограничены. Согласно этому предположению, существует такая положительная величина β , что

$$\text{Prob}\{|\varepsilon| > \beta\} = 0, \quad (4)$$

и для любого $0 < b < \beta$ $\text{Prob}\{|\varepsilon| > b\} > 0$,

где Prob обозначает вероятность события. Величина β называется максимальной погрешностью.

Она характеризует ошибку калибровки и может быть оценена традиционными статистическими методами [46]. Значения β , найденные для исследуемых данных, представлены в табл. 1.

Базируясь на постулате (4) и используя данный калибровочный набор $\{X, y\}$ с n образцами, можно построить систему неравенств относительно неизвестных регрессионных параметров a . Эта система определяет выпуклое замкнутое множество в пространстве параметров

$$A = \{a \in R^p : y^- < Xa < y^+\}. \quad (5)$$

Здесь $y_j^- = y_j - \beta$, $y_j^+ = y_j + \beta$, $j = 1, \dots, n$. Множество A называется областью возможных значений. Оно является объемным аналогом обычной точечной оценки $\hat{a}$, получаемой некоторым традиционным (например, ПЛС) регрессионным методом. Используя полученное множество A , можно предсказать значение отклика y для любого заданного вектора x . Очевидно, что если параметр a изменяется внутри A , то соответствующее предсказанное значение $\hat{y} = x^t a$ остается внутри интервала

$$V = [v^-, v^+], \quad (6)$$

в котором

$$v^- = \min_{a \in A} (x^t a), \quad v^+ = \max_{a \in A} (x^t a).$$

Эти соотношения представляют типичную задачу линейного программирования [64]. Известно, что оптимум линейной формы $x^t a$ достигается в одной из вершин выпуклого множества A . Стандартная численная процедура (симплекс-алгоритм) позволяет двигаться от одной вершины к другой в направлении максимального изменения (возрастания или убывания) формы, и это дает нам возможность провести оптимизацию таким образом, что не возникает необходимости явно строить множество A . Однако ограниченное решение задачи линейного программирования существует тогда и только тогда, когда множество A ограничено. Известно [65, 66], что множество A , заданное системой неравенств (5), ограничено тогда и только тогда, когда матрица X имеет полный ранг. Поэтому в случае неполного ранга, который является типичным случаем во многих

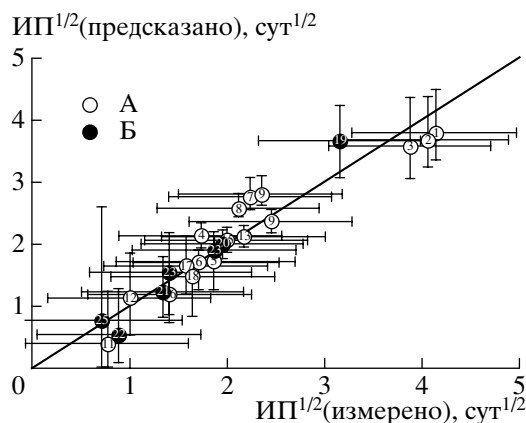


Рис. 3. Прогноз периода индукции (ИП) формальным методом для АО с начальной концентрацией 0.05%: зависимость между корнями квадратными из предсказанных и измеренных значений. А – калибровочные, Б – тестовые образцы. Горизонтальные отрезки показывают β (калибровочную ошибку), а вертикальные отрезки – ПИО (прогнозные) интервалы.

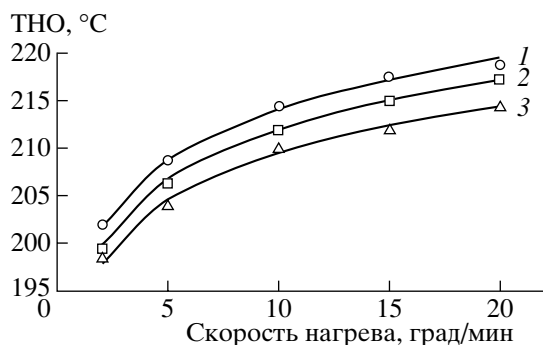


Рис. 4. Зависимость температуры начала окисления ПП в системе ПП + АО-1 (ТНО) от скорости нагрева для начальных концентраций АО, %: 0.1 (1), 0.07 (2), 0.05 (3). Точки – экспериментальные данные, кривые – калибровка.

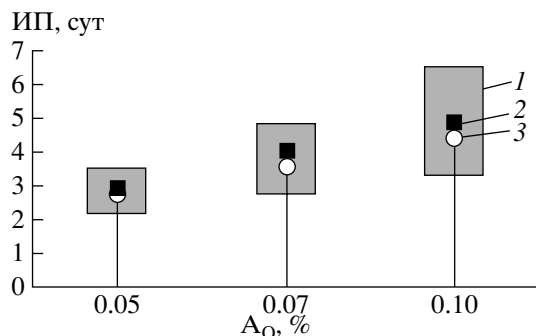


Рис. 5. Прогноз периода индукции (ИП) в системе ПП + АО-18 для различных начальных концентраций АО: 1 – доверительные интервалы для вероятности $P = 0.90$, 2 – точечные оценки ИП, 3 – измеренные значения.

практических задачах, мы должны использовать некоторую регуляризующую процедуру, например, проекцию на латентные структуры, описанную выше, и в дальнейшем использовать в ПИО-методе регулярную матрицу счетов T (2) вместо исходной матрицы X (1).

На рис. 3 показаны результаты ПЛС/ПИО-прогноза исследуемых данных для исходной концентрации АО $A_0 = 0.05\%$. Здесь предсказанные значения периода индукции изображены в зависимости от соответствующих измеренных значений, причем из обеих величин извлечен квадратный корень. Каждый образец (точки А представляются калибровочные образцы, а Б – тестовые) показан вместе с двумя интервалами ошибок. Горизонтальные отрезки представляют интервалы ошибок в измеренных величинах ИП, и они равны удвоенному значению максимальной погрешности β для всех образцов. Вертикальные отрезки соответствуют прогнозированным интервалам (6), и они различны для разных образцов. Видно, что для всех калибровочных образцов прогнозированный интервал всегда меньше или равен интервалу ошибки. Это является фундаментальным свойством, непосредственно следующим из теории ПИО-метода.

СОДЕРЖАТЕЛЬНОЕ МОДЕЛИРОВАНИЕ

Содержательные модели для прогноза периода индукции окисления строятся для каждого типа АО индивидуально. Для этого мы выделяем в полном массиве данных, показанном на рис. 1, 25 горизонтальных “срезов” и получаем, соответственно, 25 наборов данных: матрицы X_i размерностью 5×3 и векторы y_i длиной 3.

Процедура калибровки этих данных состоит из двух шагов. На первом шаге мы строим модель, которая описывает расход антиоксиданта в ходе ДСК-эксперимента – это калибровка X -данных. Регрессионная модель является неявной функцией, связывающей температуру начала окисления, начальную концентрацию АО A_0 и скорость нагрева v . Эта функция нелинейно зависит от неизвестных кинетических параметров, которые оцениваются методом нелинейного регрессионного анализа (НЛР). На втором шаге калибровки мы строим модель для описания расхода АО в ходе длительного термического старения – калибровка Y -данных. Эта регрессионная модель явно выражает период индукции как функцию температуры экспозиции (T_e) и начальной концентрации АО. В этой функции участвуют те же кинетические параметры, что и в первой модели. Их оценки мы находим на первом шаге, а на втором мы применяем специальную процедуру переноса ошибок для того, чтобы оценить неопределенность в прогнозе периода индукции.

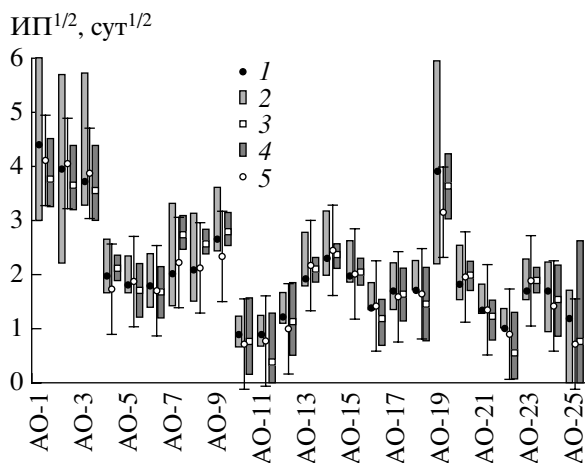


Рис. 6. Результаты прогноза периода индукции для различных образцов АО с начальной концентрацией 0.05%: 1, 2 – содержательное (НЛР) предсказание, 3, 4 – формальное (ПЛС/ПИО) моделирование, 5 – измеренные значения с вертикальными отрезками, которые показывают ошибку измерения (калибровки) β . Из всех величин извлечен квадратный корень.

Рассмотрим первый шаг калибровки. В ходе старения материала происходит расход антиоксиданта. Период индукции определяется моментом времени, когда АО израсходуется до некоторого критического значения A_c , которое зависит от температуры по закону Аррениуса [55]:

$$A_c = k_c \exp\left(-\frac{E_c}{RT}\right). \quad (7)$$

Расход АО в ходе окисления идет по закону

$$\frac{dA}{dt} = -kA, \quad (8)$$

$$A(0) = A_0,$$

где A – текущая концентрация АО; k – константа скорости, зависящая от температуры по закону Аррениуса

$$k = k_a \exp\left(-\frac{E_a}{RT}\right).$$

В ДСК-эксперименте образцы разогреваются со скоростью v от начальной температуры $T_0 = 293$ К:

$$T(t) = T_0 + vt.$$

В этом случае решение уравнения (8) можно представить в виде

$$A(t) = A_0 \exp(-k_a Z(t)), \quad (9)$$

где функция $Z(t)$ имеет вид

$$Z(t) = \int_0^t \exp\left(-\frac{E_a}{R(T_0 + vs)}\right) ds.$$

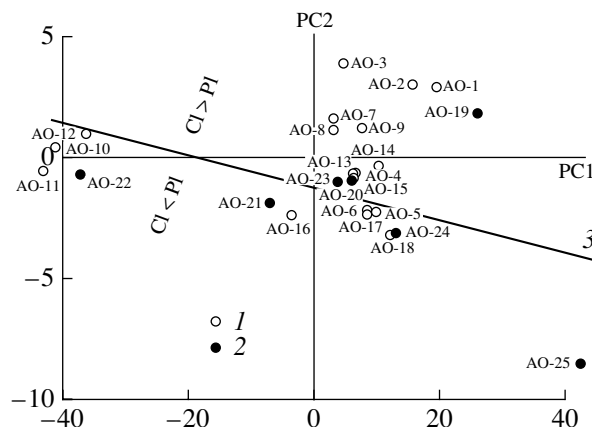


Рис. 7. График ПЛС-счетов для образцов с $A_0 = 0.05\%$: 1 – калибровочные, 2 – тестовые образцы. Линия 3 разделяет образцы, для которых доверительные интервалы (С) меньше (больше), чем прогнозные (PI).

Этот интеграл может быть представлен с использованием стандартной [67] интегральной показательной функции $E_n(z)$:

$$E_n(z) = \int_1^\infty t^{-n} e^{-zt} dt \quad (n = 0, 1, 2, \dots; z > 0),$$

в виде, зависящем не от времени t , а от температуры T :

$$Z(T) = \frac{1}{v} \left[TE_2\left(\frac{E_a}{RT}\right) - T_0 E_2\left(\frac{E_a}{RT_0}\right) \right]. \quad (10)$$

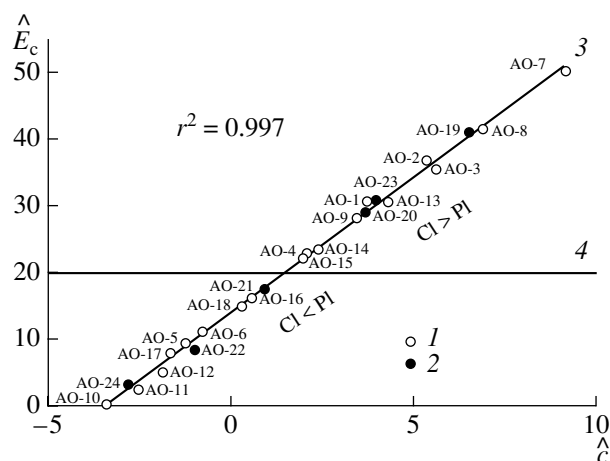


Рис. 8. Корреляция между оценками предэкспоненты ($\hat{E}_c$) и соответствующей энергией активации (E_c). 1 – калибровочные, 2 – тестовые образцы. Линия 3 показывает корреляционный тренд. Линия 4 разделяет образцы, для которых доверительные интервалы (CI) меньше (больше), чем прогнозные (PI).

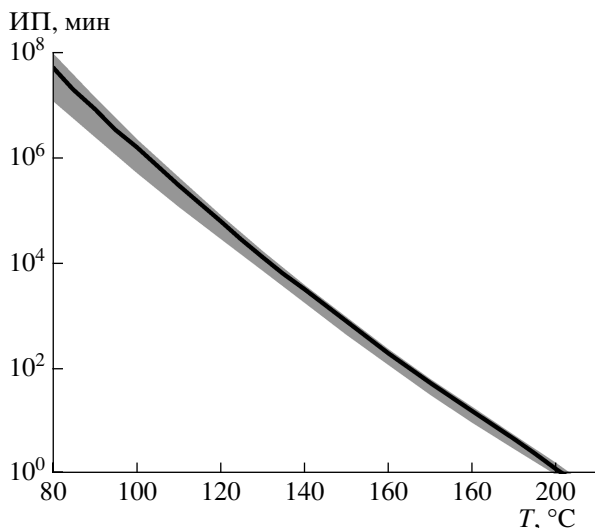


Рис. 9. Прогноз периода индукции (ИП) в зависимости от температуры экспозиции для АО-18 с начальной концентрацией 0.04%. Линия представляет среднее значение, а серый коридор показывает доверительный интервал, равный 0.95.

Подставив (10) в (9) и приравняв текущую концентрацию АО критической концентрации A_c , из уравнения (7) мы получаем уравнение

$$A_0 \exp(-k_a Z(T)) = k_c \exp\left(-\frac{E_c}{RT}\right), \quad (11)$$

которое неявно определяет зависимость температуры начала окисления T от неизвестных параметров и известных условий эксперимента. Однако практически использовать это уравнение очень неудобно. Приведем его в форму, которая более подходит для оценивания параметров. После логарифмирования и некоторых упрощений получаем

$$\exp(a) E_2\left(\frac{E_a}{RT}\right) T + v\left(c - a - \frac{E_c}{RT}\right) = 0, \quad (12)$$

где $a = \ln k_a$ и $c = \ln k_c$.

В этом уравнении температура T рассматривается как отклик, а скорость нагрева v и начальная концентрация A_0 – как предикторы (факторы эксперимента). Величины a , E_a , c и E_c – это неизвестные параметры, которые оцениваются по экспериментальным данным. При поиске оценок мы должны разрешать уравнение (12) относительно переменной T ,

$$T = T(v, A_0; a, E_a, c, E_c), \quad (13)$$

многократно, для каждого заданного набора значений предикторов v и A_0 , и искать такие значе-

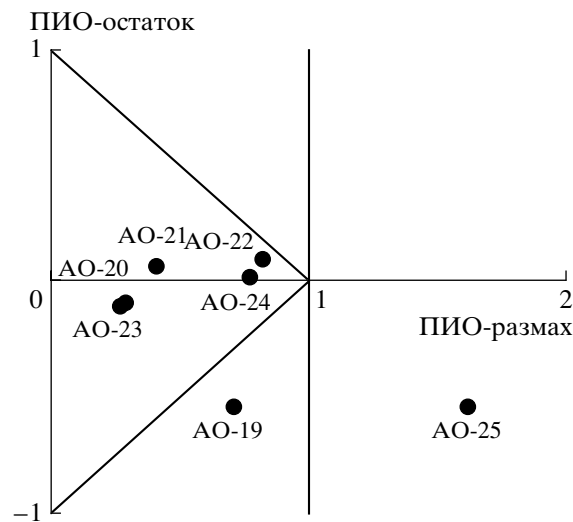


Рис. 10. График классификации PIO-статусов образцов из тестового набора с $A_0 = 0.07\%$.

ния (оценки) параметров a , E_a , c и E_c , которые минимизируют сумму квадратов отклонений

$$\min_{a, E_a, c, E_c} \sum_{ij} [Y_{ij} - T(v_j, A_{0i}; a, E_a, c, E_c)]^2. \quad (14)$$

Здесь Y_{ij} – это экспериментальные значения температуры начала окисления, индекс i соответствует трем начальным концентрациям АО, а индекс j нумерует 5 скоростей нагрева v .

При минимизации суммы (14) мы сталкиваемся со значительными трудностями. Во-первых, мы не можем представить функцию (13) явно и даже не можем неявно выразить ее через элементарные функции. Во-вторых, очевидно, что уравнение (12) задает функцию, которая нелинейно зависит от неизвестных параметров. И, наконец, минимизация суммы (14) является жесткой [3, 67] задачей. Все эти обстоятельства превращают калибровку модели в довольно сложную вычислительную проблему. Для ее решения мы использовали специальное программное обеспечение, Fitter [61], которое обладает возможностями для решения подобных задач. В этой программе реализован стабильный градиентный метод минимизации, который базируется на технике матричной экспоненты [68]. Этот метод показал высокую эффективность при решении жестких задач, превосходящую популярный метод Левенберга–Маркварда [69, 70]. Кроме того, в программе Fitter регрессионное уравнение можно задавать в естественной алгебраической форме, которая допускает и неявные выражения. Последнее и очень важное свойство программы состоит в том, что в ней вычисление производных проводится автоматически через символьное дифференцирование, что

Таблица 2. Статистические характеристики прогноза полученного содержательным (НЛР) и формальным (ПЛС/ПИО) методами

№	Статистические характеристики прогноза	НЛР ($i = 1, CI$)			ПЛС/ПИО ($i = 2, PI$)		
		Начальная концентрация АО, %					
		0.05	0.07	0.10	0.05	0.07	0.10
1.	RMSEP, сут ^{1/2}	0.242	0.246	0.272	0.239	0.251	0.336
2.	Смещение, сут ^{1/2}	0.087	0.058	0.040	0.011	0.004	0.002
3.	Корреляция ($\hat{y}_1, \hat{y}_2$)	0.953	0.934	0.916	0.953	0.934	0.916
4.	Среднее значение $(X - \hat{X})^2, K^2$	–	0.224	–	0.286	0.286	0.286
5.	Среднее значение (w_i)	1.038	1.151	1.397	0.934	1.204	1.476
6.	Корреляция (w_1, w_2)	0.202	0.007	0.028	0.202	0.007	0.028
7.	Корреляция (y, w_i)	0.815	0.846	0.836	–0.184	–0.161	–0.113

позволяет поддерживать высокую точность вычислений.

На рис. 4 представлен пример калибровки первой модели для АО-1. Показаны экспериментальные данные (точки) и соответствующие калибровочные кривые.

Рассмотрим теперь второй шаг моделирования, на котором мы строим модель для описания расхода АО в ходе длительного термического старения. Решение уравнения (8) при постоянных условиях ($T = \text{const}$) имеет вид

$$A(t) = A_0 \exp(-kt).$$

Для того чтобы найти период индукции t_i , мы должны приравнять текущую концентрацию $A(t)$ критической концентрации A_c , определенной в уравнении (7). Таким образом,

$$t_i = \left[\frac{E_c}{RT_e} + \ln A_0 - c \right] \exp\left(\frac{E_a}{RT_e} - a \right), \quad (15)$$

где T_e – температура экспозиции (старения). В нашем случае это 140°C, т.е. $T_e = 413$ К.

В уравнении (15) мы должны использовать оценки параметров $\hat{\theta} = (\hat{a}, \hat{E}_a, \hat{c}, \hat{E}_c)$, найденные на первом шаге калибровки. Для того чтобы получить не только точечную оценку периода индукции, но и определить ее доверительный интервал, мы должны принять во внимание неопределенность в этих оценках. К сожалению, мы не можем применить к уравнению (15) обычный способ переноса ошибок [17],

$$\text{var}(f) = \frac{\partial f(\hat{\theta})^T}{\partial \hat{\theta}} \text{cov}(\hat{\theta}, \hat{\theta}) \frac{\partial f(\hat{\theta})}{\partial \hat{\theta}},$$

из-за нелинейности модели. Точный доверительный интервал может быть получен с помощью

статистического моделирования, точнее – методом свободного моделирования [23]. В этом методе доверительный интервал для функции $f(\theta)$ вычисляется как процентиль выборки $\{f(\theta_1), f(\theta_2), \dots\}$, в которой значения $f(\theta_i)$ получаются в результате моделирования параметров θ_i в соответствии с нормальным распределением $N(\hat{\theta}, \text{cov}(\hat{\theta}, \hat{\theta}))$.

Пример доверительного прогноза, построенного с помощью этой техники, для образца АО-18 показан на рис. 5. Доверительная вероятность равна $P = 0.90$. Доверительные интервалы для других образцов представлены на рис. 6.

Важно, что при содержательном (НЛР) моделировании мы не использовали измеренные в ходе длительного старения значения периодов индукции. Поэтому мы можем использовать их при проверке нашей модели, вычислив среднеквадратичную ошибку прогноза (RMSEP),

$$\text{RMSEP} = \frac{\|y_{\text{test}} - Tq\|}{\sqrt{n_{\text{test}} - r}},$$

которая показана в табл. 2. В то же время мы не можем оценить такую традиционную меру точности, как среднеквадратичная ошибка калибровки RMSEC (3), поскольку на первой стадии моделирования мы калибровали не период индукции (ИП), а температуру начала окисления (ТНО) как функцию v и A_0 . Однако мы можем найти остаточную сумму квадратов для каждого АО и по ним вычислить среднее отклонение в оценках ТНО (см. № 4 в табл. 2).

СРАВНЕНИЕ МЕТОДОВ

Сравним результаты содержательного (НЛР) и формального (ПЛС/ПИО) моделирования, которые были получены на одном и том же наборе

экспериментальных данных. Эти модели используют данные различным образом. У нас имеется 25 содержательных моделей, построенных для каждого АО в отдельности, и 3 формальные модели, построенные для каждой начальной концентрации АО.

На рис. 6 мы собрали все результаты, полученные каждым методом для начальной концентрации 0.05%. На этом графике представлены результаты содержательного (НЛР) моделирования (точки 1 и прямоугольники 2 доверительных интервалов для вероятности 0.90) и результаты формального метода (ПЛС/ПИО) (точки 3 и прямоугольники 4 прогнозных интервалов). Референтные значения периодов индукции, измеренные методом длительного термического старения, обозначены точками 5 с интервалами ошибок, которые равны β – максимальной погрешности. Для того чтобы сделать сравнение более наглядным, мы извлекли квадратный корень из всех значений периодов индукции. Такие же графики могут быть построены для других наборов данных, соответствующих начальной концентрации АО 0.07 и 0.10%. Они выглядят аналогично.

В табл. 2 представлены некоторые общие статистические характеристики содержательного (НЛР) и формального (ПЛС/ПИО) методов. Здесь использованы следующие обозначения для измеренных и предсказанных X (ТНО)- и Y (ИП)-значений. X и $\hat{X}$ – это соответственно измеренные и предсказанные значения температур начала окисления (ТНО); $y = \sqrt{ИП}$ – это корень квадратный из измеренного значения периода индукции (ИП); $\hat{y}_i$ – это соответствующие векторы, полученные после извлечения квадратного корня из точечных оценок периода индукции (ИП), вычисленных содержательным (НЛР, $i = 1$) и формальным (ПЛС/ПИО, $i = 2$) методами; w_i – равны ширине доверительных (НЛР, $i = 1$) и прогнозных (ПЛС/ПИО, $i = 2$) интервалов, причем из всех интервалов также извлечен квадратный корень. Для сокращения мы будем использовать символы CI и PI для обозначения, соответственно, доверительных и прогнозных интервалов. Первые были получены методом содержательного моделирования НЛР, а вторые – формальным методом ПЛС/ПИО.

Изучая табл. 2 и рис. 6, мы можем сделать следующие выводы. Оба метода имеют близкую точность (№ 1 в табл. 2) и смещение (№ 2). Точность прогноза становится хуже, когда начальная концентрация АО увеличивается. В целом, ПЛС/ПИО-метод дает лучшие результаты для малых начальных концентраций АО, тогда как НЛР лучше для больших концентраций. Однако точечные оценки в среднем близки (см. № 3). Оба метода хорошо моделируют значения X (т.е. величины температур

начала окисления ТНО), но содержательный метод (НЛР) делает это немного лучше (№ 4).

Интервальные оценки также очень близки в среднем (№ 5), хотя доверительные интервалы (CI) могут сильно отличаться от прогнозных интервалов (PI) для отдельных образцов (см. № 6 и рис. 6). Ширина CI растет с увеличением значения периода индукции для всех начальных концентраций АО, тогда как ширина PI не зависит от y (см. № 7 в табл. 2). Это свидетельствует о том, что примененное нами преобразование откликов, $y = \sqrt{ИП}$, действительно дало ожидаемый эффект в ПЛС/ПИО-моделировании, но не смогло исправить результаты НЛР-моделирования.

Из табл. 2 (№ 5) видно, что ширина предсказанных интервалов растет с увеличением начальной концентрации АО. Это видно и из содержательной модели (15), которая представляет зависимость значения ИП t от A_0 . С другой стороны, в формальной ПЛС-модели это никак нельзя было предвидеть. По-видимому, этот факт является фундаментальным свойством исследуемой полимерной системы, а именно: чем больше добавлено АО в образец, тем хуже мы можем предсказать его период индукции. Важно, что и содержательный, и формальный методы в этом смысле дают сходные результаты. При этом результаты обоих методов могут сильно отличаться для отдельных образцов. Из рис. 6 видно, что для некоторых образцов (например, АО-5, АО-6, АО-10, АО-11) доверительные интервалы (CI) меньше, чем прогнозные интервалы (PI), тогда как для других образцов (например, АО-1, АО-2, АО-3) имеет место обратное. Было бы интересно понять, почему индивидуальные доверительные и прогнозные интервалы так отличаются.

На рис. 7 показан график ПЛС-счетов (т.е. проекций X -данных на пространство двух главных ПЛС-компонент – PC1 и PC2) для образцов с начальной концентрацией АО, равной 0.05% (АО-1, ..., АО-18 – калибровочные образцы), (АО-19, ..., АО-25 – тестовые образцы). Видно, что образец АО-25 является очевидным выбросом (или, точнее, аутсайдером по классификации образцов в ПИО-теории, о которой речь пойдет далее). Предсказанное значение для АО-25 мало отличается от измеренного значения, однако этот образец находится очень далеко от центра ПЛС-модели. Поэтому неопределенность в прогнозе АО-25 высока, и он имеет наибольший ПИО-интервал (см. рис. 6), являясь при этом единственным тестовым образцом, для которого ошибка предсказания больше, чем ошибка калибровки. Из рис. 7 также видно, что все образцы, для которых доверительные интервалы (CI с помощью НЛР) меньше, чем прогнозные интервалы (PI с помощью ПИО), расположены в левой нижней части графика, ниже прямой линии 3. Эта прямая разделя-

ет образцы на две группы: для которых $CI > PI$, и для которых $CI < PI$.

Такой результат представляется нам удивительным. Мы ожидали в образцах обнаружить некоторый порядок, структуру, которая была бы связана с шириной интервалов, т.е. с точностью предсказания. Однако мы обнаруживаем другую структуру, которая разделяет образцы по тому, какой метод – содержательный или формальный является для них лучшим. В таком случае было бы интересно изучить другой график, связанный уже с содержательным, НЛР-методом, и посмотреть, есть ли там похожие структуры. На рис. 8 представлен такой график; на нем показана корреляционная зависимость между оценками параметров E_c и c (см уравнение (12)). Важно отметить, что наблюдается значительная корреляция между оценками параметров в обеих парах (a, E_a) и (c, E_c). В уравнении Аррениуса часто корреляция между оценками предэкспоненциального множителя и энергии активации близка к 1 [7]. В нашем случае $\text{cor}(\hat{a}, \hat{E}_a) = 0.995$, $\text{cor}(\hat{c}, \hat{E}_c) = 0.997$. Рис. 8 дает возможность посмотреть на проблему с другой стороны. На нем ясно видно, что все образцы, для которых содержательный (НЛР) подход лучше ($CI < PI$), расположены в нижней части графика под линией 4. Переход соответствует критической концентрации АО, равной 0.016% при $T = 140^\circ\text{C}$.

График счетов на рис. 7 представляет только одно подмножество образцов, а именно образцы с начальной концентрацией АО, равной 0.05%. Другие графики – для $A_0 = 0.07\%$ и $A_0 = 0.10\%$ – имеют сходную структуру и здесь не приводятся. Однако оценки параметров (c, E_c) были получены с помощью содержательного моделирования, в котором все образцы с различными начальными концентрациями АО обрабатывались совместно методом НЛР, поэтому рис. 8 представляет весь набор данных. Это обстоятельство позволяет нам сделать следующий вывод, который кажется важным для выбора метода калибровки. Такое свойство, как “доверительный интервал (CI), построенный методом нелинейной регрессии, больше (меньше), чем прогнозный интервал (PI), построенный методом простого интервального оценивания” не зависит от величины начальной концентрации АО в образце, а зависит от величины критической концентрации АО, которая, в свою очередь, является свойством самого АО. Таким образом, для выбранного АО метод НЛР всегда будет лучше (хуже) метода ПЛС/ПИО вне зависимости от начальной концентрации этого АО. Для АО с низким значением критической концентрации точнее будет содержательный подход (НЛР), а для АО с высоким значением A_c – формальный (ПЛС/ПИО).

Другой важный аспект, который должен быть рассмотрен при сравнении методов моделирова-

ния, состоит в ограничениях на область применимости каждого метода. Содержательное моделирование (НЛР) имеет здесь очевидное преимущество, так как оно может использоваться для предсказания периода индукции при различных концентрациях АО и при разных температурах экспозиции. На рис. 9 показан прогноз периода индукции АО-18 при начальной концентрации 0.04% для температур экспозиции в интервале $80^\circ\text{C} < T < 200^\circ\text{C}$. Разумеется, эти условия не были исследованы в эксперименте, а представленные результаты были получены экстраполяцией уравнения (15) на эти условия. Формальный метод моделирования не может дать подобного прогноза. Между прочим данные для $T = 200^\circ\text{C}$ очень важны, так как при этой температуре происходит переработка полипропилена. С другой стороны, применение содержательной нелинейной модели к условиям, лежащим далеко от области эксперимента, довольно рискованно. В частности, эта модель не может быть использована для экстраполяции на комнатную температуру. Поэтому нельзя прогнозировать срок службы полипропилена, используя метод ДСК. Иными словами, мы не можем указать точные границы применимости содержательного подхода.

Ситуация с формальным ПЛС/ПИО-методом совершенно другая. Здесь можно точно указать соответствующие границы применимости. На рис. 10 показан график классификации ПИО-статусов тестовых образцов, который используется для установления этих ограничений. Подробное объяснение теории ПИО-классификации дано в [35, 45, 46], поэтому здесь мы дадим лишь очень краткое описание метода. Каждый образец из тестового набора (АО-18, ..., АО-25) изображается на графике статусов в следующих координатах. ПИО-остаток – это разность между центром прогнозного интервала (6) и соответствующим референтным значением, деленная на максимальную погрешность β , определенную в (4). ПИО-размах равен полуширине прогнозного интервала (6), также деленной на β . Положение тестового (или нового) образца на этом графике определяет возможность построения для него прогноза. Все образцы, расположенные внутри треугольника (АО-20, АО-21, ..., АО-24), называются инсайдерами. Они полностью соответствуют модели и прогноз для них наиболее надежен. Обратный случай означает, что такие образцы лежат вне модели, поэтому они называются аутсайдерами (образцы АО-19 и АО-25). Аутсайдеры не противоречат модели, но прогноз для них менее надежен. Тому могут быть две причины: большой размах (АО-25) и смещение (АО-19). Таким образом, используя технику ПИО-статусов, можно легко классифицировать новый образец и тем самым ограничить область применимости формального ПЛС/ПИО-метода.

Следует отметить принципиально разную природу областей применимости содержательного и формального методов. В содержательном подходе таковой является область в пространстве факторов (T и A_0), в которую модель может быть экстраполирована. При этом мы имеем дело с одним и тем же АО, который был предварительно исследован методом ДСК. При формальном подходе таковой является область новых антиоксидантов, к которым может быть применена построенная ПЛС/ПИО-модель. При этом условия эксперимента, т.е. начальная концентрация АО и скорость нагрева в ДСК, должны быть теми же самыми, что и в исходном калибровочном наборе. С этим соображением связано и еще одно обстоятельство. Допустим, мы хотим улучшить модель и достичь большей точности в прогнозе. Для этого надо применять различные планы уточняющего эксперимента. В формальной модели необходимо добавить новые образцы АО, отличающиеся от исходного калибровочного набора. В содержательной модели для каждого АО строится своя собственная модель, поэтому для ее улучшения нужно провести новые эксперименты с другими начальными концентрациями того же АО. Это повысит точность прогноза именно этого АО и не будет иметь никакого влияния на качество предсказания для других АО.

ЗАКЛЮЧЕНИЕ

Мы показали, что в задаче проверки активности АО длительный и дорогостоящий процесс термического старения может быть заменен быстрым ДСК-методом с последующей обработкой полученных данных содержательным (НЛР) или формальным (ПЛС/ПИО) методами. Оба подхода к калибровке дают удовлетворительную точность прогноза периода индукции, которая не хуже, чем ошибка в традиционном методе. Поэтому мы можем рекомендовать их для практического применения.

Каждый метод калибровки имеет свои преимущества и недостатки. Содержательный подход позволяет получать результаты прогноза, экстраполированные на условия (температура и концентрация), лежащие за областью эксперимента. Однако мы не можем ограничить область такой экстраполяции, поэтому иногда получаемый результат не надежен. С другой стороны, формальный метод имеет строгую область применимости, очерченную с помощью техники ПИО-статуса. В то же время, ПЛС/ПИО-подход не может быть применен для предсказания периода индукции в условиях, отличающихся от условий калибровочного эксперимента. Оба метода имеют близкую точность предсказания и обнаруживают сходно поведение по отношению к условиям прогноза, т.е. чем больше начальная кон-

центрация АО, тем хуже точность предсказания. Однако содержательный подход дает лучшие результаты для АО с низким уровнем критической концентрации, т.е. для АО с малым периодом индукции, тогда как формальный метод лучше в противном случае. Это является фундаментальным, внутренним свойством АО, которое сохраняется при изменении начальной концентрации АО в образце.

Таким образом, в случае, когда целью исследований является предсказание поведения некоторой данной полимерной системы, содержательный подход предпочтителен. В случае, когда исследователь желает сравнить активность различных АО, формальная модель лучше отвечает такой постановке.

Содержательный подход устанавливает связь между факторами и откликом эксперимента, учитывая современные представления о химических и физических процессах, проходящих в системе. При этом получаются довольно сложные математические модели, так что для их обработки требуется применение специальных вычислительных методов. Однако если такие модели адекватно описывают исследуемый процесс, то они позволяют делать экстраполяцию – прогноз за пределы области факторов, исследованной в эксперименте.

Формальный подход не требует глубоких знаний о природе исследуемого явления. Математические модели, используемые в этом методе, линейны относительно неизвестных параметров и поэтому просты для применения и интерпретации. Однако для построения таких моделей часто требуется угадать подходящее преобразование факторов или отклика, приводящее к линейаризации системы. В рассмотренном примере прогноза активности АО мы удачно извлекли квадратный корень из всех экспериментальных значений периода индукции и тем самым достигли нужного результата. Поэтому применение формального подхода требует от исследователя некоторого опыта решения подобных задач. В то же время, будучи правильно примененным, этот подход дает надежные результаты в самых различных областях. Он хорошо работает при решении задач классификации или интерполяции, т.е. прогнозирования на (примерно) те же условиях, в которых проходил эксперимент. Для задач экстраполяции этот подход применять нельзя.

СПИСОК ЛИТЕРАТУРЫ

1. Лаврентьев М.М., Краева А.Г., Бухгейм А.В. Обратная задача химической кинетики. Новосибирск: ВЦ СО АН СССР, 1980. Препр. С. 17.
2. Спивак С.И., Горский В.Г. // Докл. АН СССР. 1981. Т. 257. № 2. С. 412.
3. Применение вычислительной математики в химической и физической кинетике / Под ред. Полака Л.С. М.: Наука, 1969.

4. Павлов Б.В., Брин Э.Ф. // Хим. Физика. 1984. Т. 3. № 3. С. 393.
5. Esbensen K.H. // Multivariate Data Analysis – In Practice. CAMO, 2000.
6. Эсбенсен К. Анализ многомерных данных / Под ред. Родионовой О. Барнаул: Изд-во АГУ, 2003.
7. Haario H., Taavitsainen V.-M. // Chemom. Intell. Lab. Syst. 1998. V. 44. P. 77.
8. Bijlsma S., Louwerse D.J., Windig W., Smilde A.K. // Anal. Chim. Acta. 1998. V. 376. P. 339.
9. Bijlsma S., Louwerse D.J., Smilde A.K. // J. Chemom. 1999. V. 13. P. 311.
10. Taavitsainen V.-M., Haario H. // J. Chemom. 2001. V. 15. P. 215.
11. de Juan A., Maeder M., Martinez M., Tauler R. // Chemom. Intell. Lab. Syst. 2000. V. 54. P. 123.
12. de Juan A., Maeder M., Martinez M., Tauler R. // Anal. Chim. Acta. 2001. V. 442. P. 337.
13. Bezemer E., Rutan S.C. // Chemom. Intell. Lab. Syst. 2001. V. 59. P. 19.
14. Брин Э.Ф., Померанцев А.Л. // Хим. физика. 1986. Т. 5. № 12. С. 1674.
15. Pomerantsev A.L., Rodionova O.Ye. // Chemom. Intell. Lab. Syst. 1999. V. 48. P. 121.
16. Родионова О.Е., Померанцев А.Л. // Успехи химии. 2006, в печати.
17. Бард И. Нелинейное оценивание параметров. М.: Статистика, 1979.
18. Pirson K. // Phil. Mag. 1901. V. 2. № 6. P. 559.
19. Демиденко Е.З. Линейная и нелинейная регрессии. М.: Финансы и статистика, 1981.
20. Тихонов А.Н. // Докл. АН СССР. 1963. Т. 4. С. 1035.
21. Hoskuldsson A. Prediction Methods in Science and Technology. Copenhagen: Thor Publishing, 1996.
22. Faber K. // Chemom. Intell. Lab. Syst. 2000. V. 52. P. 123.
23. Pomerantsev A.L. // Chemom. Intell. Lab. Syst. 1999. V. 49. P. 41.
24. Канторович Л.В. // Сиб. мат. журн. 1962. Т. 3. № 5. С. 701.
25. Воицинин А.П., Бочков А.Ф., Сотиров Г.Р. // Завод. лаб. 1990. Т. 56. № 7. С. 76.
26. Анисимов В.М., Померанцев А.Л., Новорадовский А.Г., Карпунин О.Н. // Журн. прикл. спектроскопии. 1987. Т. 46. С. 117.
27. Ахунов И.Р., Ахмадишин З.Ш., Спивак С.И. // Хим. Физика. 1982. Т. 12. С. 1660.
28. Бахитова Р.Х., Спивак С.И. // Химия и хим. Технология. 1999. Т. 42. С. 92.
29. Слинько М.Г., Спивак С.И., Тимошенко В.И. // Кинетика и катализ. 1972. Т. 13. С. 1570.
30. Спивак С.И., Тимошенко В.И., Слинько М.Г. // Докл. АН СССР. 1970. Т. 192. С. 580.
31. Белов В.М., Карбаинов Ю.А., Суханов В.А. и др. // Журн. аналит. Химии. 1994. Т. 49. С. 370.
32. Хлебников А.И. // Журн. аналит. Химии. 1996. Т. 51. С. 347.
33. Белов В.М., Суханов В.А., Унгер Ф.Г. Теоретические и прикладные аспекты метода центра неопределенности. Новосибирск: Наука, 1995.
34. Pomerantsev A.L., Rodionova O.Ye. // Chemom. Intell. Lab. Syst. 2006, in press.
35. Rodionova O.Ye., Esbensen K.H., Pomerantsev A.L. // J. Chemom. 2004. V. 18. P. 402.
36. Westad F., Martens H. // J. Near Infrared Spectros. 2000. V. 8. P. 117.
37. Cook R.D. // Technometrics. 1977. V. 19. P. 15.
38. Cook R.D. // J. Am. Statis. Ass. 1979. V. 74. P. 169.
39. Andrews D.F., Pregibon D. // J. Royal Statis. Soc. B. 1978. V. 40. P. 84.
40. Draper N.R., John J.A. // Technometrics. 1981. V. 23. P. 21.
41. Næs T. // J. Chemom. 1989. V. 1. P. 121.
42. Hoskuldsson A. // Chemom. Intell. Lab. Syst. 2001. V. 55. P. 23.
43. Jouan-Rimbaud D., Massart D.L., Saby C.A., Puel C. // Anal. Chim. Acta. 1997. V. 350. P. 149.
44. Fernandez Pierna J.A., Wahl F., de Noord O.E., Massart D.L. // Chemom. Intell. Lab. Syst. 2002. V. 63. P. 27.
45. Померанцев А.Л., Родионова О.Е. // Журн. анал. химии. 2006, в печати.
46. Rodionova O.Ye., Pomerantsev A.L. In: Progress in Chemometrics Research / Ed. Pomerantsev A.L. N.Y.: Nova Sci. Publ., 2005. P. 43.
47. Pomerantsev A.L., Rodionova O.Ye. In: Progress in Chemometrics Research / Ed. Pomerantsev A.L. N.Y.: Nova Sci. Publ., 2005. P. 209.
48. Pomerantsev A.L., Rodionova O.Ye. In: Aging of polymers, polymer blends and polymer composites / Ed. Zaikov G.E. N.Y.: Nova Sci. Publ., 2002. V. 2. P. 19.
49. Pomerantsev A., Rodionova O., Hoskuldsson A. // Chemom. Intell. Lab. Syst. 2006, in press.
50. Clancey V.J. // Nature. 1947. V. 159. P. 339.
51. Rajko R. // Anal. Lett. 1994. V. 27. P. 215.
52. Eriksson L., Johansson E., Kettaneh-Wold N., Wold S. // Multi- and Megavariate Data Analysis, Umetrics, Umea, 2001.
53. Burden F.R., Rosewarne B.S., Winkler D. // Chemom. Intell. Lab. Syst. 1997. V. 38. P. 127.
54. Heberger K., Borosy A.P. // J. Chemom. 1999. V. 13. P. 473.
55. Shlyapnikov Yu.A. In: Development in Polymer Stabilization. L.: Appl. Sci. Publ., 1981. V. 5.
56. Bystritskaya E.V., Pomerantsev A.L., Rodionova O.Ye. // J. Chemom. 2000. V. 14. P. 667.
57. Максимова Г.А., Померанцев А.Л. // Завод. лаб. 1995. Т. 61. С. 432.
58. Pomerantsev A.L. // Chemom. Intell. Lab. Syst. 2003. V. 66. P. 127.
59. Bystritskaya E.V., Pomerantsev A.L., Rodionova O.Ye. // Conferentia chemometrica (CC-97), Budapest, 1997. <http://sol.cc.u-szeged.hu/~rajko/Pos54.gif>.
60. Bystritskaya E.V., Pomerantsev A.L., Rodionova O.Ye. // Proc. 5th Scandinavian Symposium on Chemometrics (SSC5). Lahti., 1997. P. 46.

61. *Fitter Add-Inn*. [On line], [http://polycert.chph.ras.ru / fitter.htm](http://polycert.chph.ras.ru/fitter.htm) [1 октября 2005].
62. *Smilde A., Bro R., Geladi P.* Multi-way analysis with application in the chemistry sciences. Wiley, 2004.
63. *Кельнер Р., Мерме Ж.-М., Отто М., Вудмер Г.М.* Аналитическая химия. М.: Мир, 2004.
64. *Taha H.* Operations Research. An Introduction. N.Y.: Macmillan Publ., Co., 1982.
65. *Gass S.* Linear Programming. N.Y.: McGow-Hill, 1975.
66. *Kuhn H.W., Tucker A.W.* Linear Inequalities and Related Systems. Annals of Mathematics Studies 38. Princeton, N.J.: Princeton University Press, 1956.
67. Абрамовиц М., Стиган И. Справочник по специальным функциям. М.: Наука, 1979.
68. *Павлов Б.В., Повзнер А.Я.* // Журн. вычисл. математики и мат. физики. 1973. Т. 13. С. 1056.
69. *Marquardt D.W.* // SIAM J. 1963. V. 11. P. 431.
70. *Levenberg K.* // Quart. Appl. Math. 1944. V. 2. P. 164.
71. *Родионова О.Е., Померанцев А.Л.* // Кинетика и катализ. 2005. Т. 46. № 3. С. 329.
72. Unscramber [Online] <http://www.camono.no> (1 октября 2005).