6.4.3. АДЕКВАТНОСТЬ РЕГРЕССИОННЫХ МОДЕЛЕЙ ФИЗИКО-ХИМИЧЕСКИХ СВОЙСТВ РАСТВОРОВ МАТЕМАТИЧЕСКИМ МОДЕЛЯМ ХИМИКО-ТЕХНОЛОГИЧЕСКИХ СИСТЕМ

Из концепций элементов объективности полны
выплывают перманентно адекватные челны. (ЛГ)

«Структурные» математические модели ХТС (химико-технологических систем) высокого уровня имеют сложную структуру и в качестве фрагментов включают в себя «элементарные» регрессионные модели физико-химических свойств (ФХС) растворов (к таким свойствам относятся плотность, вязкость, теплоемкость растворов и др.) [1].

В настоящей работе рассмотрены две задачи. **Прямая задача** - сформулировать условие адекватности «элементарной» математической модели ФХС структурной математической модели ХТС. **Обратная задача** - определить требования к максимальному пределу погрешности математической модели ФХС, удовлетворяющему условию адекватности математической модели ХТС.

Начнем с **решения прямой задачи**. По-видимому, правильнее проверку гипотезы об адекватности регрессионных моделей ФХС математическим моделям ХТС следовало бы формулировать в терминах моделей ХТС и проверять адекватность моделей ХТС результатам измерений параметров ХТС. Однако решение этой задачи выходит за рамки нашей темы. Во-первых, нельзя предвидеть все модели ХТС, в составе которых будут использованы модели ФХС; во-вторых, многие математические модели ХТС столь громоздки [1], что реализуются только на ЭВМ; составление таких моделей является самостоятельной задачей. Поэтому мы сформулируем заведомо более строгое, чем это необходимо в каждом случае, требование адекватности моделей ФХС: **вклад погрешности расчета параметров ФХС в погрешность расчета параметров ХТС должен быть статистически незначимым**.

Математическую модель ХТС, как правило, с помощью эксперимента сверяют с физическим объектом - некоторой реальной ХТС. Все физические параметры - факторы (условно независимые переменные) и отклики (условно зависимые переменные) математической модели ХТС измеряют на объекте, а затем сравнивают с результатами опыта [1]. Будем считать, что математическая модель ХТС адекватна реальному объекту - ХТС в том смысле, что выборочные дисперсии измерения откликов и остаточные дисперсии их вычисления однородны.

Существенной является иерархическая структура моделей ХТС: одни и те же величины могут рассматриваться как факторы структурной модели ХТС и как отклики элементарной модели ФХС. Например, в модели ХТС плотность раствора может являться фактором [1], но сама как отклик в модели ФХС вычисляться с определенной случайной ошибкой по значениям «своих» факторов - температуры и концентрации компонентов.

Назовем «теоретической» математическую модель ХТС, если она является алгебраической записью фундаментальных законов природы. Таковы, например, математические модели материальных и тепловых балансов, геометрических соотношений и другие. Теоретическая модель ХТС может иметь вид

$$U = U(\underline{\theta}, \underline{V}), \qquad (1)$$

где U - отклик (зависимая переменная), $\underline{\theta}$ - вектор известных значений коэффициентов, $\underline{V} = (\overline{v_1, v_n}) \in R^n$ - вектор факторов.

Другая разновидность моделей - регрессионная математическая модель ХТС вида

$$U = U(\underline{\theta}, \underline{V}) + \varepsilon_u. \qquad (2)$$

Эта модель аналитически отличается от модели (1) слагаемым $\varepsilon_u \sim N(0, \sigma_{0u}^2)$ - остаточной ошибкой модели с центром в нуле и дисперсией σ_{0u}^2. Кроме того, коэффициенты $\underline{\theta}$ регрессионной модели (2) являются случайными величинами, так как они определяются по экспериментальным данным.

Однако в дальнейшем для простоты будем считать их постоянными величинами.

Пусть для измерения факторов v_i используются средства (измерительные приборы), дающие случайные нормально распределенные независимые погрешности

$$\lambda_i \sim N\left(0, \sigma^2_{\lambda_i}\right), i = \overline{1,n}, \tag{3}$$

где $\sigma^2_{\lambda i}$ - дисперсия случайной погрешности, 0 - систематическая ошибка.

Таким образом, вместо факторов v_i на объекте наблюдают величины

$$\bar{v}_i = v_i + \lambda_i, i = \overline{1,n}, \tag{4}$$

которые можно подставить в уравнение (2) для вычисления величины U.

Пусть в области $D \subset R^n$ определения математической модели (2) функция $U\left(\underline{\theta}, \underline{V}\right)$ почти линейна и может быть заменена линейным преобразованием Тейлора

$$U = U\left(\underline{\theta}, \underline{V}\right) = U\left(\underline{\theta}, \underline{\bar{V}}\right) + \sum_{i=1}^{n} \left(\partial U / \partial v_i\right)_{\underline{\bar{V}}} \left(v_i - \bar{v}_i\right) + \varepsilon_p,$$

где $\underline{\bar{V}}$ - вектор средних значений факторов, ε_p - ошибка преобразования, малая величина с порядком малости, определенным ниже.

Перепишем предыдущую формулу в виде

$$U = \beta_0 + \sum_{i=1}^{n} \beta_i z_i + \varepsilon_p, \tag{5}$$

где

$$\beta_0 = U\left(\underline{\theta}, \underline{\bar{V}}\right), \beta_i = \left(\partial U / \partial v_i\right)_{\underline{\bar{V}}}, z_i = v_i - \bar{v}_i, i = \overline{1,n}.$$

Параметры распределения случайных величин λ_i обычно представляют выборочными оценками. Так, экспериментальное тарирование приборов, измеряющих величины v_i, позволяет получить оценки дисперсий измерения факторов $s^2_{\lambda i} \to \sigma^2_{\lambda i}$ с $f_{\lambda i}$ степенями свободы [2]. Путем N-кратного измерения значений $\underline{V}$ на моделируемом объекте получают средние значения факторов $\underline{\bar{V}}$.

Пусть наблюдаемые величины v_i примут некоторые значения $\bar{v}_i$. Подставляя эти значения в формулу (2), вычислим значения $\tilde{U}$ величины U:

$$\tilde{U} = U\left(\theta, \underline{\tilde{V}}\right). \tag{6}$$

Случайная ошибка вычисленного по формуле (6) значения $\tilde{U}$ обусловлена независимыми ошибками измерения v_i и, согласно допущению (5), имеет оценку дисперсии

$$s^2_{u/n} = \sum_{i=1}^{n} \beta_i^2 s^2_{\lambda i} \to \sigma^2_{u/n}. \tag{7}$$

Дисперсия отклика $s^2_u \to \sigma^2_u$ модели (2) в этом случае будет складываться из остаточной дисперсии $s^2_{0u} \to \sigma^2_{0u}$ и дисперсии $s^2_{u/n} \to \sigma^2_{u/n}$, обусловленной ошибками измерения факторов

$$s^2_u = s^2_{0z} + s^2_{u/n} \to \sigma^2_u \tag{8}$$

с числом степеней свободы, определяемым по формуле Уэлча [2]:

$$f_u = s^4_u \Big/ \left(s^4_{0u} f^{-1}_{0u} + \sum_{i=1}^{n} \beta_i^4 s^4_{\alpha i} f^{-1}_{\lambda i} \right) - 2. \tag{9}$$

Отметим теперь, что по нашему предположению ошибка ε_p преобразования (5) составляет $O(\delta_и)$.

Следует отметить, что в расчетах по формулам (7) и (9) встречаются затруднения при вычислении величин $\left(\partial U / \partial v_i\right) s_{\lambda i}, i = \overline{1,n}$. Во-первых, n раз необходимо проверить допустимость линеаризации модели (2); во-вторых, иногда сложно взять производную функции $U\left(\underline{\theta}, \underline{V}\right)$. Такие расчеты удобно выполнять на ЭВМ методом Монте-Карло (МК) [3, с.66], предназначенным для расчета параметров распределения отклика по известным параметрам распределения факторов.

Программа МК вычисляет выборочные моменты распределения отклика, связанного алгебраическим уравнением со случайными величинами - факторами, законы распределения которых известны. Исходные данные: вид алгебраического уравнения, количество, средние значения и дисперсии факторов. Результаты расчета: моменты распределения отклика (средние, дисперсии, средние отклонения, коэффициенты асимметрии и эксцесса) относительно каждого фактора.

Программа МК позволяет проверить гипотезу о нормальном распределении отклика относительно нормально распределенных факторов (методом моментов), ранжировать факторы по силе их влияния на отклик и упростить уравнение путем замены незначимых факторов их средними значениями.

Пусть для простоты первые *k* факторов модели ХТС (2) - это вычисляемые параметры ФХС. Тогда, согласно выражению (7), часть $S^2_{u/k}$ дисперсии $S^2_{u/n}$ величины *U*, обусловленная ошибкой расчета этих первых *k* факторов, будет

$$s^2_{u/k} = \sum_{i=1}^{k} \beta_i^2 s^2_{\lambda i} \to \sigma^2_{u/k}, k < n . \tag{10}$$

Другая часть $s^2_{u/m}$ дисперсии $s^2_{u/n}$, обусловленная дисперсиями $s^2_{\lambda i}, i = \overline{k+1, n}$, ошибок измерения $m = n — k$ технологических факторов $v_{k+1}, \ldots, v_n$, будет равна

$$s^2_{u/m} = \sum_{i=k+1}^{m} \beta_i s^2_{\lambda i} \to \sigma^2_{u/m} .$$

Таким образом, дисперсия отклика $S^2_u \to \sigma^2_u$ складывается из дисперсии $s^2_{u/k} \to \sigma^2_{u/k}$, обусловленной ошибками расчета параметров ФХС и «технологической» дисперсии $s^2_{uT} \to \sigma^2_{uT}$:

$$s^2_u = s^2_{u/k} + s^2_{u/m} + s^2_{0u} = s^2_{u/k} + s^2_{uT} ; s^2_{uT} = s^2_{u/m} + s^2_{0u} \to \sigma^2_{uT} . \tag{11}$$

Величина s^2_{uT} обусловлена ошибками измерения технологических факторов и остаточной ошибкой математической модели ХТС.

Утверждение о том, что математические модели ФХС адекватны требованиям математической модели ХТС (2), заключается в том, что составляющая $\sigma^2_{u/k}$ дисперсии отклика, обусловленная ошибками расчета параметров ФХС, составляет величину, ничтожно малую по сравнению с технологической составляющей σ^2_{uT} дисперсии отклика. Сформулируем это утверждение в виде следующей **статистической гипотезы**: составляющая $\sigma^2_{u/k}$ дисперсии σ^2_u отклика, обусловленная параметрами ФХС (или факторами $v_i, i = \overline{1, k}$), не превышает среднеквадратичного отклонения (СО) $\sigma_{s^2_{uT}}$ технологической составляющей дисперсии s^2_{uT}. Тогда нулевая гипотеза будет

$$H_0 : \sigma^2_{u/k} \le \sigma_{s^2_{uT}} ; \tag{12}$$

соответствующая альтернативная гипотеза имеет вид

$$H_1 : \sigma^2_{u/k} > \sigma_{s^2_{uT}} . \tag{13}$$

В качестве статистики критерия рассмотрим отношение

$$\hat{F} = s^2_{u/k} \Big/ s_{s^2_{uT}} , \tag{14}$$

где $s^2_{u/k} \to \sigma^2_{u/k}$, а $s_{s^2_{uT}} \to \sigma_{s^2_{uT}}$ определяется соотношением [4, с.207]

$$s_{s^2_{uT}} = s^2_{uT} \left(2 / f_{uT}\right)^{0,5} . \tag{15}$$

Статистика

$$\hat{F} = s^2_{u/k} / \left[s^2_{uT} \left(2 / f_{uT}\right)^{0,5} \right], \tag{16}$$

полученная после подстановки выражения (15) в (14), имеет распределение Фишера с $f_{u/k}$ и f_{uT} степенями свободы. Её можно использовать для проверки гипотезы H_0 (12) α-методом [5]. С помощью статистики $\hat{F}$ определяют оценку $\hat{\alpha}$ уровня значимости. Гипотеза (12) не отклоняется, если $\hat{\alpha} \geq \alpha_k$ где α_k - заданный исследователем критический уровень значимости.

Пример 1. Из уравнений теплового и материального балансов получили формулу для расчета температуры t жидкости, поступающей в теплообменник дистилляции содового производства [1, с.189]

$$t=(L_Bc_Bt_B + L_0c_0t_0)/[c_B(L_B - l)], \tag{17}$$

где L - расход жидкости (кг/ч), c - удельная теплоемкость [Дж/(кг·К)], t - температура жидкости (^{o}C), l - брызгоунос (кг/ч); индексы: о - отходящая флегма, в - входящая жидкость.

Средние значения входящих в уравнение (17) величин следующие: $\overline{L}_B$ = 2·10^5 кг/ч, $\overline{L}_0$= 10^4 кг/ч; $\bar{l}$ = 2·10^4 кг/ч; $\bar{t}_B$= 73 ^{o}C, $\bar{t}_0$= 83 °C; $\bar{c}_B$= 3380 Дж/(кг·К.); $\bar{c}_0$= 3300 Дж/(кг·К). СО измерения величин равны: s_{L_B} = 10^4 кг/ч, s_{L_B} = 5·10^2 кг/ч; s_{t_B} = 0,5 К; s_{t_0} = 0,5 К. СО ошибок вычисления брызгоуноса s_l = 2•l0^3 кг/ч; теплоемкости s_0 = 40 Дж/(кг·К).

Числа степеней свободы для всех среднеквадратичных отклонений равны 50. Требуется проверить гипотезу об адекватности метода расчета теплоемкости уравнению (17).

Решение. Если считать, что теплоемкость вычисляется без ошибки, то ошибка расчета по формуле (17), обусловленная ошибками технологических факторов, составляет $s_{t/5}$ = (0,69^2 +0,55^4 + 0,22^2 +0,03^2 + 0,95^2)^{0,5} = 0,42 К с числом степеней свободы, согласно формуле (9), $f_{t/5}$ = [50$S^4_{t/5}$x(0,69^4 + 0,55^4 + 0,22^4 + 0,03^4 + 0,95^4)$^{-1}$] - 2 = 131. (Слагаемые 0.69, 0.55 и т. д. в этих формулах вычислялись с помощью программы МК [3]. Методом моментов [3] установили, что отклик t в уравнении (17) почти линеен но каждому фактору во всей области определения этого уравнения.)

СО ошибки расчета t, обусловленной ошибкой расчета теплоемкостей c, составляет, согласно (10), $s_{t/2}$ = (0,053^2 + 0,054^2)^{0,5} = 0,076 К с числом степеней свободы $f_{t/2}$ = 98.

Нулевая гипотеза об адекватности расчета t имеет вид (12), альтернативная - (13).

Статистика критерия Фишера, согласно (16), равна $\hat{F}$ = 0,076^2/[0,42^{2}x x(2/131)$^{1/2}$] = 0,26. Так как $\hat{F} < 1$, то очевидно, что $\hat{\alpha} > \alpha_k$, где α_k может быть даже задан равным 0,5. Нулевая гипотеза о том, что физико-химическая составляющая дисперсии $\sigma^2_{t/2}$ отклика t, обусловленная факторами c_0 и c_B, не превышает среднеквадратичного отклонения $\sigma_{s^2_{t/5}}$ технологической составляющей дисперсии $s^2_{t/5}$, обусловленной ошибками измерения технологических факторов, не отклоняется. Метод расчета теплоемкости адекватен модели (17).

Решение обратной задачи. Практика использования математических моделей ФХС [1] показывает, что модели ФХС часто существенно более точны, чем это необходимо для расчета параметров ХТС. Однако высокая точность математических моделей ФХС достигается усложнением структуры этих моделей. В связи с этим и необходимо оценить максимальное значение «физико-химической» составляющей $s^2_{u/k}$ дисперсии отклика s^2_u модели ХТС. Для этого запишем систему уравнений (16) и $\hat{\alpha} \geq \alpha_k$ в виде эквивалентного неравенства

$$s^2_{u/k} \leq F_{f_{u/k}, f_{uT}, \alpha_k} s^2_{uT} \left(2/f_{uT}\right)^{0,5}, \tag{18}$$

где $F_{f_{u/k}, f_{uT}, \alpha_k}$ - критический верхний предел распределения Фишера с $f_{u/k}$ и f_{uT} степенями свободы. Если математическая модель ФХС включает формулы для расчета k > 1 параметров ФХС, то возникает задача оценки максимального СО ошибки расчета $s_{\lambda i}$ каждого i-го параметра ФХС (i = l, k).

После подстановки в левую часть неравенства (18) значения $s^2_{u/k}$ из формулы (10) эта задача сведется к решению полученного неравенства относительно k неизвестных $s_{\lambda i}\left(i=\overline{1,k}\right)$. Единственное решение может быть получено, если на систему уравнений (10) и (18) наложить дополнительное ограничение.

Часто в качестве такого ограничения используют «принцип равных влияний», согласно которому $\beta_i^2 s_{\lambda i}^2 = s_{u/k}^2 / k$, откуда

$$s_{\lambda i} = s_{uk} / \left(k^{0,5} \beta_i\right), i = \overline{1,k} . \qquad (19)$$

Для практического решения обратной задачи необходимо рассматривать множество конкретных «отраслевых» математических моделей ХТС, в составе которых имеются математические модели ФХС.

Пример 2. По данным примера 1 определить максимальное значение СО s_0 ошибки формулы для расчета теплоемкости жидкости.

Решение. Согласно α-методу проверки гипотез [5], зададим критический уровень значимости $\alpha_к$ = 0,1 [предпочтительна нулевая гипотеза (12), ответственность за выводы высокая]. По таблице [6, с.273] найдем $F_{98;\ 131;\ 0,1}$= 1,3. Из формулы (18): $s_{t/2}$ = $1{,}3^{0,5}\cdot 0{,}42\ (2/131)^{0,25}$ = 0,168. Производная β = $\partial t/\partial c$ = 0,053/40 = 0,001325; из формул (l9) и (18) s_c < $0{,}168/(2^{0,5}$x0,001325) = 90 Дж/(кг·К).

Таким образом, для того, чтобы формула (17) была адекватна результатам наблюдений, необходимо, чтобы СО ошибки расчета теплоемкости в этой формуле не превышало 90 Дж/кг·К).

ЛИТЕРАТУРА

1. Г. А. Ткач, В. Д. Смоляк. Моделирование десорбционных процессов содового производства. Изд. «Химия»: Л., 1973, 208 с.
2. Л. Закс. Статистическое оценивание. Изд. «Статистика», М. - 1976, 598 с.
3. Н. А. Цейтлин. Деп. рукопись № 198/74, деп., г. Черкассы, ОНИИТЭХИМ, 1974, 77 с.
4. Н. В. Смирнов, П. В. Дунин - Барковский. Курс теории вероятностей и математической статистики. Изд. «Наука», М., 1969, 430 с.
5. Н. А. Цейтлин. Элиминирующий анализ математических моделей ХТС. В кн. «Процессы и аппараты производств основной химии». Тр. НИОХИМа, 1981, 56 с.
6. Л. Н. Большев, Н. В. Смирнов. Таблицы математической статистики. ВЦ АН СССР, М., 1968, 474 с.