Повышение точности оценки значений текущих качественных показателей по их дискретным лабораторным анализам использованием алгоритмов экстраполяции

Е.А. Гребенюк, Э.Л. Ицкович (ИПУ им. В.А. Трапезникова РАН)

Рассматриваются модели статистической экстраполяции дискретных лабораторных анализов показателей качества материальных потоков непрерывного технологического производства. Анализируется точность прогноза значений лабораторных анализов разными вариантами алгоритма авторегрессии и скользящего среднего. Предлагается метод определения рационального алгоритма статистической экстраполяции и расчета его параметров.

Ключевые слова: прогноз дискретных лабораторных анализов, статистическая экстраполяция лабораторных анализов, алгоритмы авторегрессии и скользящего среднего.

Введение

Прогнозирование значения качественного показателя материального потока, полученного в его очередном лабораторном анализе, до момента оценки его значения в следующем по времени лабораторном анализе, позволяет точнее оценить текущее состояние непрерывного ТП, вырабатывающего этот материальный поток и, следовательно, эффективнее управлять этим процессом, не допуская его выхода за границы заданного технологического диапазона.

Если в качестве прогнозного значения показателя до момента времени получения результата следующего лабораторного анализа используется значение результата последнего по времени выполненного анализа, то такой прогноз будем называть «наивным». Если же в качестве прогнозного значения до момента времени получения результата следующего лабораторного анализа используется алгоритм экстраполяции, основанный на учете статистической взаимосвязи ряда последовательных лабораторных анализов, результат будем называть «статистическая экстраполяция».

В любом агрегате непрерывного технологического производства при постоянстве режима его работы изменения во времени качественных показателей, вырабатываемых агрегатом материальных потоков, близки случайному стационарному процессу. При этом повсеместно ряд последовательных лабораторных анализов качественных показателей материальных потоков непрерывного технологического производства, несмотря на невысокую частоту их проведения (обычно интервал между соседними анализами составляет 4...24 ч), статистически взаимосвязан ввиду существенной инерционности ТП, вырабатывающих материальные потоки. Это позволяет применить к такому ряду алгоритмы статистической экстраполяции, которые реализуют прогноз большей точности, чем наивный прогноз. В работе [1] было показано преимущество использования статистической экстраполяции перед наивным прогнозом при выборе рациональной частоты проведения лабораторных анализов. В настоящей статье рассматриваются методы и алгоритмы построения такого прогноза по реальным данным лабораторных анализов.

Пусть значения анализов, полученных до момента времени t, описываются временным рядом:

$$x_1, x_2 \ldots, x_{t-1} \qquad (1)$$

где x_i, i=1,2, ... — результаты лабораторного анализа пробы, взятой в момент времени t=i.

Рассмотрим метод статистической экстраполяции, основанный на описании значений временного ряда моделью авторегрессии — скользящего среднего (АРСС), учитывающей наличие статистической взаимосвязи между рядом последовательных значений лабораторных анализов. В случае стационарных рядов эта модель представляется как разность двух линейных комбинаций: прошлых значений показателя и прошлых значений ошибок его прогноза на один шаг [2]. Первая линейная комбинация известна в литературе как авторегрессия (АР), вторая — как скользящее среднее (СС). Число членов в каждой линейной комбинации и их коэффициенты определяются свойствами статистических взаимосвязей ряда (1). Таким образом, прогноз текущего значения ряда может быть вычислен на основе прошлых наблюдений. Поскольку на значения анализа в момент времени t влияют неконтролируемые случайные возмущения на ТП, то наблюдаемое истинное значение анализа отличается от его вычисленного значения прогноза АРСС.

Усредненное значение отклонений прогноза от истинных значений анализов будем оценивать среднеквадратической ошибкой прогноза σ^2. При построении прогноза статистической экстраполяцией коэффициенты линейных комбинаций выбираются таким образом, чтобы σ^2 было минимальным. Статистическая

Во всех изречениях точность приходится в известной мере приносить в жертву краткости.
Сэмюэл Джонсон

экстраполяция позволяет оценить тенденцию изменений дискретно измеряемого во времени показателя непрерывного процесса на основе значений имеющихся измерений и в зависимости от свойств процесса на 10...40% уменьшить среднеквадратическую ошибку прогноза по сравнению с «наивным прогнозом». При прогнозе методом статистической экстраполяции нельзя предсказывать будущие «резкие» скачки показателя (вызванные скачкообразными внешними возмущениями), но можно определить тенденцию его изменения и построить доверительный интервал прогноза, вероятность выхода за границы которого не превышает заданных значений.

Для одного и того же временного ряда можно построить несколько моделей, различающихся между собой числом членов, их параметрами (весами) в каждой из имеющихся линейных комбинаций, величиной среднеквадратической ошибки прогноза. Для вычисления прогноза может быть использована любая из моделей, но для получения прогноза хорошего качества следует выбирать число членов и их параметры таким образом, чтобы получить модель, наиболее точно описывающую изменения во времени рассматриваемого показателя.

Описание алгоритма статистической экстраполяции

Динамика производственных процессов в режиме нормального функционирования технологического агрегата описывается в большинстве случаев не стационарным, а кусочно-стационарным процессом, конкретный пример которого приведен на рис 1 (интервал времени наблюдения за процессом — 55 сут., шаг между соседними анализами 8 ч). Такой процесс характеризуется примерно постоянным поведением случайных колебаний относительно его среднего значения и резкого в какие-то моменты времени изменения среднего значения. Это поведение процесса (а следовательно, и определяющего его показателя) может быть, например, следствием изменения задания системе регулирования анализируемого показателя или смены хранилищ сырья, перерабатываемого в агрегате, или воздействиями на процесс каких-либо внешних не контролируемых возмущений.

По рисунку видно, что во временном графике лабораторных анализов показателя процесса выделяются два участка с различным уровнем среднего значения показателя: с первых по 30-е сутки (расчетное среднее значение — 354,1) и с 31-х по 55-е сутки (расчетное среднее значение — 343,4).

Для преобразования кусочно-стационарного процесса в стационарный процесс исходный ряд (1) следует преобразовать [2] к ряду разностей:

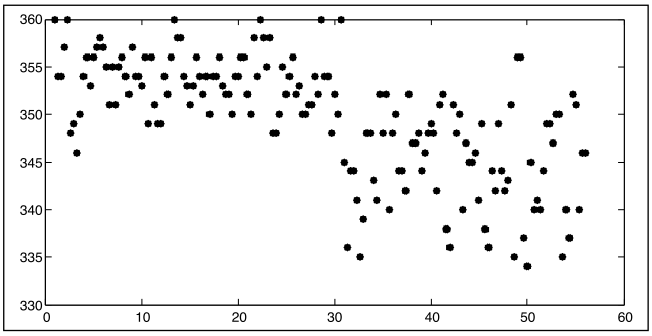

Рис. 1. Фактические значения лабораторных анализов (●). По горизонтальной оси отложено время в сутках, по вертикальной – значения лабораторных анализов

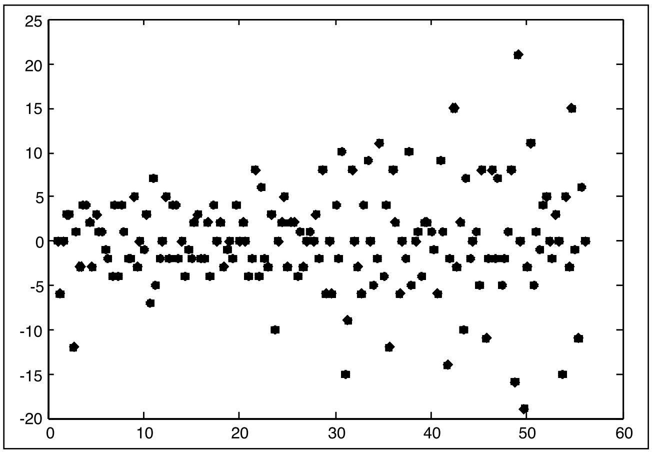

Рис. 2. Фактические значения разностей лабораторных анализов (●). По горизонтальной оси отложено время в сутках, по вертикальной – значения разностей последовательных лабораторных анализов

$$x_1, x_2, \ldots, x_n \Rightarrow 0, \Delta x_2, \ldots, \Delta x_n,$$

где $\Delta x_i = x_i - x_{i-1}$, $i = 2, \ldots, n$, n — объем исходной выборки.

В результате полученного преобразования анализов из примера на рис. 1 получаем ряд, график которого приведен на рис. 2. Этот ряд разностей соседних анализов за время наблюдения 55 сут. имеет один, общий уровень среднего значения (расчетное среднее значение ряда равно нулю).

Поскольку в дальнейшем изложении в основном будет рассматриваться не ряд (1), а его преобразование к разностям $0, \Delta x_2, \ldots, \Delta x_t, \ldots$, то обозначим преобразованные значения $\Delta x_2, \ldots, \Delta x_t, \ldots$ как

$$y_1, y_2, \ldots, y_t, \ldots. \qquad (2)$$

Вычисление статистического прогноза последовательности значений анализов на один шаг, то есть на интервал времени от момента t-1 до момента t при известных значениях прошлых анализов для ряда (2) реализуется с использованием следующей формулы [2]:

$$\begin{aligned}\hat{y}_{t|t-1,t-2,\ldots,t-m} &= \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} - \\ &- \beta_1 (y_{t-1} - \hat{y}_{t-1|t-2,t-3,\ldots,t-m-1}) - \ldots \\ &- \beta_q (y_{t-q} - \hat{y}_{t-q|t-(q+1),t-(q+2),\ldots,t-m-q}),\end{aligned} \qquad (3)$$

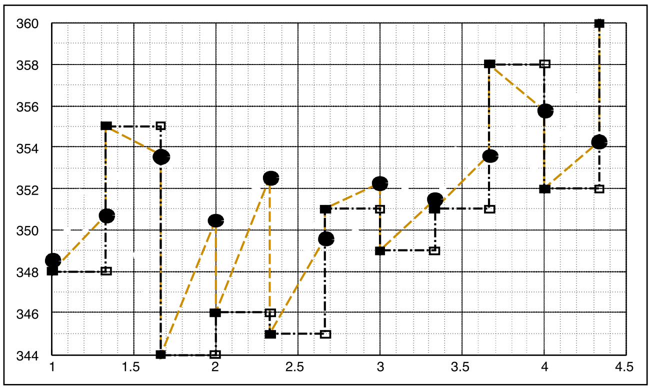

Рис. 3. Фактические значения лабораторных анализов (■), статистическая экстраполяция (– – – ●), наивный прогноз (– ▪ – □). По горизонтальной оси отложено время в сутках, по вертикальной – значения лабораторных анализов

где $\hat{y}_{t|t-1,t-2,\ldots,t-m}$ — прогноз значения ряда (2) на момент времени t по разностям значений анализов $y_{t-1},\ldots,y_{t-m}; m=max(p,q)$; p — порядок линейной комбинации авторегрессии (АР); q — порядок линейной комбинации скользящего среднего (СС); $\alpha_i,\ i=1,2,\ldots,p;$ — коэффициенты авторегрессии; $\beta_j,\ j=1,2,\ldots,q$ — коэффициенты скользящего среднего; $(y_{t-1}-\hat{y}_{t-1|t-2,t-3,\ldots,t-m-1})\ldots(y_{t-q}-\hat{y}_{t-q|t-(q+1),t-(q+2),\ldots,t-m-q})$ — ошибки прогноза в моменты времени $t-1,\ldots,t-q$.

Прогноз по формуле (3), построенный по разностям, называется прогнозом по модели проинтегрированной авторегрессии и скользящего среднего [2], которая обозначается АРПСС (p,1,q), где p — число членов авторегрессии, 1 — порядок разности (число применений операции взятия разностей к ряду (1)), q — число членов скользящего среднего.

В зависимости от свойств описанного ряда один из порядков p или q может быть равен нулю. Если q=0 (прогноз по модели АРПСС (p,1,0)), то уравнение (3) принимает вид:

$$\hat{y}_{t|t-1,t-2,\ldots,t-p}=\alpha_1 y_{t-1}+\alpha_2 y_{t-2}+\ldots+\alpha_k y_{t-p}. \quad (3a)$$

Если p=0 (прогноз по модели АРПСС (0,1,q)), то уравнение (3) принимает вид:

$$\hat{y}_{t|t-1,t-2,\ldots,t-q}=\beta_1(y_{t-1}-\hat{y}_{t-1|t-2,t-3,\ldots,t-q-1})+\ldots+ \\ +\beta_q(y_{t-q}-\hat{y}_{t-q|t-(q+1),t-(q+2),\ldots,t-2q}). \quad (3b)$$

Для вычисления прогноза на один шаг значений лабораторных анализов $\hat{x}_{t|t-1,t-2,\ldots,t-m}$ используют соотношения:
$y_t=\Delta x_t=x_t-x_{t-1},\ t=2,\ldots,n$, $\hat{y}_{t|t-1,t-2,\ldots,t-m}=\hat{x}_{t|t-1,t-2,\ldots,t-m}-x_{t-1}$, и от формул (3) — (3 б) переходят к формулам относительно $\hat{x}_{t|t-1,t-2,\ldots,t-m}$ и x_t:

$$\hat{x}_{t|t-1,t-2,\ldots,t-m}=x_{t-1}+\hat{y}_{t|t-1,t-2,\ldots,t-m}= \\ =(1+\alpha_1)x_{t-1}+(\alpha_2-\alpha_1)x_{t-2}+\ldots-\alpha_p x_{t-p-1}- \\ -\beta_1(x_{t-1}-\hat{x}_{t-1|t-2,t-3,\ldots,t-m-1})-\ldots- \\ -\beta_q(x_{t-q}-\hat{x}_{t-q|t-q-1,t-q-2,\ldots,t-m-q-1}). \quad (4)$$

$$\hat{x}_{t|t-1,t-2,\ldots,t-p}=(1+\alpha_1)x_{t-1}+(\alpha_2-\alpha_1)x_{t-2}+\ldots-\alpha_p x_{t-p-1}. \quad (4a)$$

$$\hat{x}_{t|t-1,t-2,\ldots,t-q}=\beta_1(x_{t-1}-\hat{x}_{t-1|t-2,t-3,\ldots,t-q-1})+\ldots+ \\ +\beta_q(x_{t-q}-\hat{x}_{t-q|t-q-1,t-q-2,\ldots,t-2q-1}). \quad (4b)$$

Среднеквадратическая ошибка прогноза по формулам (4) — (4 б) вычисляется по формуле:

$$\sigma^2=\frac{1}{n-k}\sum_{i=k+1}^{n}\left(x_i-\hat{x}_{i|i-1,i-2,\ldots,i-k}\right)^2, \quad (5)$$

где n — объем выборки, k=m+1 — максимальный порядок формулы для вычисления прогноза по ряду (2), $m=max(p,q)$.

На рис. 3 представлены фактические значения анализов за 4,5 сут. (данные из примера рис. 1) и их прогнозные значения, полученные существующим способом наивного прогноза и предлагаемым способом статистической экстраполяции (метод определения значений коэффициентов статистической экстраполяции изложен ниже).

Построение алгоритма статистической экстраполяции

Разработка алгоритма статистической экстраполяции показателя включает определение следующих параметров: выбор числа членов в каждой из линейных комбинаций алгоритма, оценку их коэффициентов, расчет среднеквадратической погрешности прогноза (анализ качества полученного прогноза).

Шаг I. Выборки результатов последовательных лабораторных анализов: обучающая и контрольная. Для построения алгоритма статистической экстраполяции на один шаг необходимы две указанные выборки. По обучающей выборке проверяется наличие статистической взаимосвязи, определяется ее характер, выбирается число членов авторегрессии и скользящего среднего, оцениваются их коэффициенты и ошибка прогноза. Ошибка прогноза, вычисляемая по обучающей выборке, оценивает точность прогноза, на основании которой выбирается наилучший вариант алгоритма прогноза для данной выборки. Проверка на контрольной выборке позволяет оценить устойчивость качества построенного прогноза к возможным отклонениям ТП от наблюдаемого в обучающей выборке, но при которых он находится в режиме нормального функционирования, то есть анализируемый показатель лежит в заданном для него диапазоне.

Обучающая выборка представляет собой значения лабораторных анализов показателя, числом ⩾70...100 ед. последовательных анализов качественного показателя процесса, когда он находится в режиме нормального функционирования. Длина и число контрольных выборок определяются требованиями к контролю качества алгоритма статистической экстраполяции.

Шаг II. Проверка наличия статистической взаимосвязи между последовательными анализами. Поскольку модель строится по ряду (2), то для проверки взаимосвязи используем ряд разностей (2) обучающей выборки. По этому ряду строится автокорреляционная функция, представляющая собой значения коэффициентов автокорреляции ρ_k ряда (2) с различными сдвигами k=1,2... Коэффициент автокорреляции ρ_k оценивает величину зависимости между последовательными членами ряда (2) y_t и y_{t-k} для каждого из значений k=1,2... r, где n — объем обучающей выборки n, r — число вычисляемых коэффициентов, а коэффициент автокорреляции ρ_k вычисляется по формуле:

$$\rho_k = \frac{\sum_{j=1}^{n-k} y_{j+k} y_j}{\sqrt{\sum_{j=k+1}^{n} y_j^2}\sqrt{\sum_{j=1}^{n-k} y_j^2}}, \ k = 1, 2, \ldots, r. \qquad (6)$$

Значения ρ_k лежат в интервале от [-1,1]. Если ρ_k=0, то стохастическая связь между анализами y_t и y_{t-k} отсутствует. Построение алгоритма статистической экстраполяции прогноза имеет смысл только, если $\rho_k \neq 0$ для значений k=1,2... r, где r — некоторое целое положительное число.

Поскольку вычисленные по конечной выборке значения ρ_k обладают погрешностью, то они практически никогда точно не равняются нулю. Для проверки отличия вычисляемых по конечной выборке значений ρ_k от нуля строятся доверительные интервалы G, вероятность выхода нулевого ρ_k за которые не превышает заданного значения α (обычно α выбирают равным 5%) [2]:

$$G = \pm K_{\alpha/2} \big/ \sqrt{n},$$

где n — объем обучающей выборки, $K_{\alpha/2}$ — значение, которое случайная величина не превышает с фиксированной вероятностью $\alpha/2$, если она имеет стандартное нормальное распределение.

Если α=5%, то $K_{\alpha/2} \approx 1{,}96$[3], поэтому для расчетов границы можно использовать приближенную формулу:

$$G = \pm 2 \big/ \sqrt{n}.$$

Если вычисленное по формуле (6) ρ_k не выходит за границы доверительных интервалов, то его отличие от нуля не является статистически значимым.

Шаг III. Определение порядков модели p и q в формуле (3). В зависимости от значений выбранных порядков на этом шаге получаем для прогноза одну из формул (3), (3 a), (3b). Для каждого из прогнозируемых процессов существуют такие значения порядков p и q, которые дают наименьшую среднеквадратическую ошибку прогноза на обучающей выборке, величина которой может незначительно изменяться при прогнозировании по контрольной выборке.

Рассмотрим ситуацию, когда порядки p и q в формуле для прогноза отличаются от порядков, дающих наименьшую среднеквадратическую ошибку прогноза.

Основными требованиями, предъявляемыми к построенной формуле, являются: точность прогноза и возможность применения этой формулы ко всем участкам существующего кусочно-стационарного процесса без существенного изменения точности прогноза. Лишние члены модели p и q вносят незначительный вклад в описание прогнозируемого показателя, но увеличивают погрешности оценивания коэффициентов модели. В результате ошибка прогноза увеличивается особенно при использовании этой формулы на других участках кусочно-стационарного процесса. При уменьшении порядков модели p и q относительно модели, дающей наименьшую среднеквадратическую ошибку прогноза, получают прогноз с меньшей точностью, так как из него исключаются члены, описывающие изменение прогнозируемого показателя.

Для выбора порядков p и q по обучающей выборке предлагается использовать подход, основанный на применении информационных критериев [4]. В зависимости от целей построения модели авторегрессии — скользящего среднего используются разные информационные критерии. В рассматриваемой задаче наилучшим является критерий Акаике *(AIC-Akaike information criterion)*, так как он выбирает модель, которая имеет минимальную среднеквадратическую ошибку прогноза на один шаг с минимальным «штрафом» за добавление лишних порядков p и q. Информационный критерий Акаике представляет собой взвешенную сумму двух величин, зависящих от полученного логарифма среднеквадратической ошибки прогноза и используемых порядков p и q в формулах (3) или (3 a), или (3b):

$$AIC = n \ln \sigma^2 + 2(p+q+1), \qquad (7)$$

где n — объем выборки, σ^2 — среднеквадратическая ошибка прогноза, p+q+1 — число оцениваемых параметров (число членов в АР и СС частях алгоритма, плюс среднеквадратическая ошибка прогноза).

Для выбора наилучших порядков p и q задаются максимальным числом членов в линейных комбинациях АР и СС. Если прогноз вычисляется по одной из формул (3)- (3b), то наибольшие значения порядков p и q в рассматриваемых вариантах алгоритмов обычно можно ограничить тремя. Для наиболее распространенного шага лабораторных анализов порядка 8 ч это обозначает, что существенная взаимосвязь лабораторных анализов ограничивается суточным интервалом инерционности ТП. Если число значимых коэффициентов автокорреляции r меньше трех, то максимальный порядок полагают равным m=$min(r$,3). Рассматриваются алгоритмы всевозможных комбинаций $p \leqslant m$ и $q \leqslant m$ в интервале [0,m]. Всего строится и сопоставляется $(m+1)^2$-1 вариантов алгоритмов. Например, если m=2, то следует сопоставить алгоритмы следующих вариантов значений порядков (p,q):

(0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1), (2,2).

Для каждого варианта алгоритма по разностям значений обучающей выборки оцениваются коэффициенты в одной из формул (3) — (3b) и среднеквадратическая ошибка прогноза σ^2.

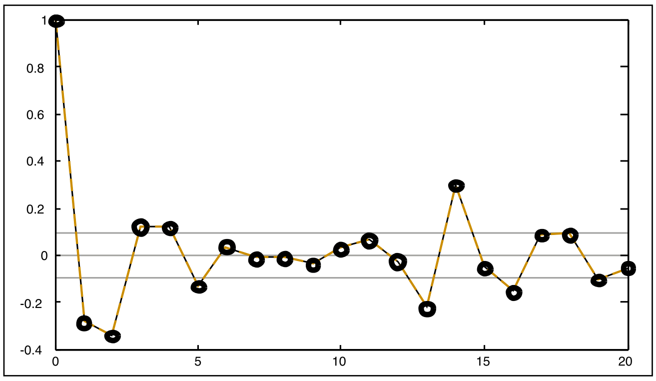

Рис.4. Автокорреляционная функция качественного показателя, построенная для ряда разностей последовательных значений лабораторных анализов (– – О) и границы значимости G коэффициентов автокорреляции (—). По горизонтальной оси отложено время в сутках, по вертикальной – значения коэффициентов автокорреляции

Процедура оценивания коэффициентов может быть выполнена специальными алгоритмами, которые реализованы во многих статистических пакетах и описаны в [5]. С использованием полученного значения σ^2 и числа оцениваемых порядков p и q для каждого варианта алгоритма вычисляется критерий Акаике по формуле (7). В отдельных статистических программных пакетах (например, МАТЛАБ, Eviews) этот критерий может быть вычислен специальной командой. Наилучшим считается алгоритм, имеющий минимальное значение критерия Акаике.

Шаг IV. Выбор формулы статистической экстраполяции и оценка ее качества. При выборе формулы для прогноза целесообразно оценивать несколько характеристик ее работы, каждая из которых оценивает свойства формулы с различных сторон. Для каждой из $(m+1)^2$-1 построенных формул вычисляются три характеристики.

Первой характеристикой является критерий Акаике, выбирающий формулу, у которой взвешенная линейная комбинация среднеквадратической ошибки прогноза и числа оцениваемых параметров минимальная. Поскольку этот критерий накладывает сравнительно небольшой штраф за увеличение порядков p и q, которое может привести к тому, что в формуле окажутся «лишние члены», то он выбирает формулы с небольшой ошибкой прогноза и несколько завышенным числом параметров.

Второй характеристикой является оценка относительного уменьшения ошибки σ статистической экстраполяции (ООУ) в% относительно σ наивного прогноза, полученная на обучающей выборке:

$$ООУ = 100\% \frac{(\sigma_N - \sigma_{se})}{\sigma_N}, \quad (8)$$

где σ_N — корень квадратный среднеквадратической ошибки наивного прогноза, σ_{se} — корень квадратный среднеквадратической ошибки статистической экстраполяции.

Эта величина имеет максимальное значение для модели с минимальным значением σ_{se}. Она дает качественную оценку улучшения прогноза за счет статистической экстраполяции и отвечает на вопрос: нужно ли применять для данного качественного показателя алгоритм статистической экстраполяции или для него достаточно ограничиться имеющимся «наивным прогнозом».

Третьей характеристикой является оценка устойчивости модели (ОУМ), которая определяется величиной относительного изменения σ статистической экстраполяции в% на обучающей и контрольной выборках:

$$ОУМ = 100\% \left| \frac{\sigma_{se}^{об} - \sigma_{se}^{конт}}{\sigma_{se}^{об}} \right|, \quad (9)$$

где $\sigma_{se}^{об}, \sigma_{se}^{конт}$ — погрешности статистической экстраполяции на обучающей и на контрольной выборках.

Окончательный выбор алгоритма по перечисленным характеристикам выполняется по следующим правилам.

- По результатам расчетов все построенные на шаге III формулы ранжируются по значению критерия Акаике и выбираются три формулы с наименьшими его значениями.
- Вычисляются среднеквадратические ошибки прогнозов по отобранным формулам на контрольной выборке. Отбираются те формулы, у которых величина ОУМ, вычисляемая по (9), не превосходит по модулю 10%. Если таких формул среди трех отобранных нет, то может быть принято решение либо о выборе по критерию Акаике больше трех формул, либо о расширении и изменении состава контрольных и обучающих выборок и повторении процедуры построения алгоритма статистической экстраполяции заново.
- Из оставшихся формул ООУ выбирается та, у которой величина, вычисляемая по (8),— максимальная. Эту формулу следует вставить в программное обеспечение ЛИМС для прогноза значения данного показателя.

Пример построения алгоритма статистической экстраполяции

Рассмотрим процесс построения алгоритма статистической экстраполяции для качественного показателя, фрагмент изменений последовательных значений лабораторных анализов которого представлен на рис. 1 (анализы берутся с интервалом 8 ч). Поскольку процесс является кусочно-стационарным, то от исходных значений переходим к разностям и строим для ряда разностей автокорреляционную функцию, которая представлена на рис. 4.

Выбираем максимальный порядок m из условия $m=min(r,3)$ по указанию шага III. Строим модели для всевозможных комбинаций порядков p и q, и вычисляем критерий Акаике, значения ООУ и ОУМ. Заранее вычисленный корень квадратный среднеквадратической ошибки наивного прогноза по обучающей выборке σ_n=5,344.

Таблица 1

Модель	Критерий Акаике	Ранг критерия Акаике
АРПСС(0,1,1)	716.06	7
АРПСС(0,1,2)	716.87	9
АРПСС(0,1,3)	717.97	14
АРПСС(1,1,0)	732.41	15
АРПСС(1,1,1)	717.03	10
АРПСС(1,1,2)	713.40	3
АРПСС(1,1,3)	714.91	6
АРПСС(2,1,0)	718.73	14
АРПСС(2,1,1)	717.87	11
АРПСС(2,1,2)	714.87	5
АРПСС(2,1,3)	714.40	4
АРПСС(3,1.0)	717.89	12
АРПСС(3,1,1)	708.23	1
АРПСС(3,1,2)	710.17	2
АРПСС(3,1,3)	716.30	8

Таблица 2

Модель	Критерий Акаике	Ранг критерия Акаике	*ООУ*	*ОУМ*
АРПСС(1,1,2)	713.40	3	18.73	5.9
АРПСС(3,1,1)	708.23	1	22.30	14.06
АРПСС(3,1,2)	710.17	2	22.34	13.72

Результаты расчетов приведены в табл. 1. В первом столбце перечислены названия моделей, по которым строятся прогнозные формулы, во втором — значения критерия Акаике, в третьем — упорядочения моделей по критерию Акаике.

Из данных, представленных в табл. 1, следует, что минимальные значения критерия Акаике имеют (в порядке возрастания) модели АРПСС (3,1,1), АРПСС (3,1,2), АРПСС (1,1,2). Вычислим для полученных моделей величины относительного уменьшения σ статистической экстраполяции в% относительно σ наивного прогноза и характеристики устойчивости модели. Результаты расчетов приведены в табл. 2.

Максимальное преимущество относительно наивного прогноза имеет модель АРПСС (3,1,2), у которой значение критерия Акаике выше, чем у модели АРПСС (3,1,1). Однако обе модели не обладают устойчивостью, поскольку ОУМ превосходит по модулю 10%. Модель АРСС (1, 1, 2) удовлетворяет обоим условиям и имеет достаточно высокое значение ООУ, поэтому реализующая ее формула выбирается для внедрения статистического прогноза рассматриваемого показателя в функционирующую на предприятии ЛИМС.

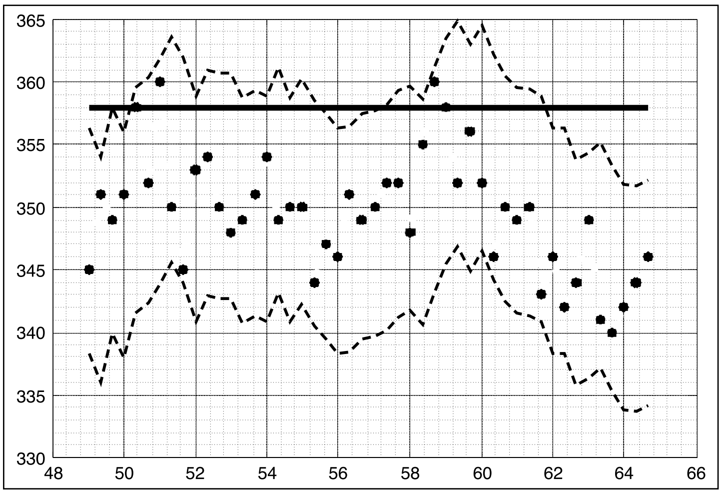

Рис.5. Доверительные интервалы прогноза. Фактические значения лабораторных анализов (●), границы доверительных интервалов прогноза при α=5% - - - , технологические ограничения ——.

Применение алгоритма статистической экстраполяции для прогноза выхода значения измеряемого показателя за границы заданного диапазона

Если имеется последовательность наблюдений (1) и используется одна из формул (4) — (4 б) для вычисления прогноза, то для каждого вновь получаемого значения прогноза можно построить доверительный интервал, выход фактического значения прогнозируемой величины за границы которого, возможен с вероятностью, не превышающей заданного значения. Величина вероятности задается при построении этого интервала и определяет его ширину. Обычно величина этой вероятности задается небольшой, в диапазоне 1...10%, то есть 0,01...0,1. Чем больше заданная вероятность выхода за доверительный интервал, тем он уже.

Границы доверительного интервала, выход фактического значения показателя за которые в момент времени t может произойти с вероятностью α, при прогнозируемом значении $\hat{x}_{t|t-1,\dots,t-m}$, вычисляются по формуле:

$$I_t^1 = \hat{x}_{t|t-1,\dots,t-m} + K_{\alpha/2}\sigma$$

$$I_t^2 = \hat{x}_{t|t-1,\dots,t-m} - K_{\alpha/2}\sigma,$$

где I_t^1, I_t^2 — верхняя и нижняя границы доверительного интервала, $\hat{x}_{t|t-1,\dots,t-m}$ — прогноз, полученный в момент времени t-1, $K_{\alpha/2}$ — квантиль стандартного нормального распределения; то есть значение, которое заданная случайная величина не превышает с фиксированной вероятностью уровня $\alpha/2$; α — вероятность выхода за границы доверительного интервала, которая обычно задается в пределах 0,01...0,1; σ — корень среднеквадратической ошибки прогноза.

Построенные таким образом границы доверительных интервалов могут служить индикатором тех участков процесса, на которых возможен его выход за технологические ограничения.

Модель статистической экстраполяции описывает стационарный или кусочно-стационарный процесс, при котором колебания значения показателя не имеют резких, кратковременных выбросов и для такого процесса она определяет с заданной вероятностью интервал времени, в котором возможен выход анализируемого показателя за заданный диапазон. Практически возможное наличие в отдельные моменты времени резких, кратковременных выбросов значения показателя, причинами которых могут являться, например, провалы напряжения у двигателей, замена единицы оборудования, задержка поступления одного из компонентов сырья естественно этой моделью не прогнозируется.

Когда формула для статистической экстраполяции построена, можно вычислять прогнозы и строить доверительные интервалы при различных значениях вероятности выхода фактических значений за их пределы. На рис. 5 для рассматриваемого в статье примера построены доверительные интервалы при значении доверительной вероятности α=5%; то есть при вероятности, что следующее значение выйдет за верхнюю границу, равной 5%.

На представленном рисунке верхняя граница прогноза, начиная с середины 50-х сут. превышает технологические ограничения, что служит сигналом о том, что будущее фактическое значение может превысить технологические ограничения с вероятностью >5%. Если участки, на которых граница доверительных интервалов превышает технологические ограничения встречаются достаточно часто и имеют значительные временные интервалы, (например, длительность их превышает несколько суток), то у такого процесса, скорее всего, следует изменить задание или параметры систем регулирования режимных величин, от значения которых зависит поведение анализируемого в лаборатории качественного показателя.

Заключение

Применение в составе прикладного математического обеспечения ЛИМС модуля статистической экстраполяции позволяет построить для любого анализируемого в лаборатории качественного показателя статистический прогноз его значения до следующего по времени анализа и оценить возможность выхода его значения за заданный технологический диапазон в ближайший интервал времени. При этом оператору соответствующего ТП должны выдаваться после каждого лабораторного анализа этого показателя не только его результат, но и прогноз его значения до следующего анализа и график доверительного интервала выхода значения показателя за границы заданного технологического диапазона. Это должно существенно облегчить оператору поддержание ТП в заданных рамках, снизить частоту появления брака, повысить качество выпускаемой продукции.

Список литературы

1. *Гребенюк Е.А., Ицкович Э.Л.* Выбор рациональной частоты проведения лабораторных анализов на непрерывном производстве технологического типа // Автоматизация в промышленности. 2014. №12. С.51.-55.
2. *Бокс Дж, Дженкинс Г.* Анализ временных рядов. Прогноз и управление, 1974.
3. *Вентцель Е.С.* Теория вероятностей: Учеб.для вузов. М.: Высш. шк., 1999. 576 с.
4. *Канторович Г.Г.* Анализ временных рядов. Лекционные и методические материалы // Экономический журнал ВШЭ. 2002. №2. С. 251 - 273.
5. *Hamilton J.D.* Time Series Analysis // Princeton, NJ: Princeton University Press, 1994.

Гребенюк Елена Алексеевна *– д-р техн. наук, главный научный сотрудник,*
Ицкович Эммануил Львович *– д-р техн. наук, проф., главный научный сотрудник*
ИПУ им. В.А. Трапезникова РАН.
Контактный телефон (495) 334-90-21.
E-mail: lgreben@ipu.ru, itskov@ipu.ru

Компания BCC построила систему управления энергоснабжением на базе концепции Smart Grid для АО «ЮРЭСК»

Завершен первый этап масштабного проекта модернизации высоковольтных подстанций Ханты-Мансийского автономного округа (ХМАО). Построенная система управления энергоснабжением на базе концепции Smart Grid позволила заказчику — Югорской региональной электросетевой компании (АО «ЮРЭСК») повысить надежность и улучшить контроль состояния оборудования и параметров сети.

АО «ЮРЭСК» — одна из ведущих электросетевых компаний ХМАО-Югры, оказывающая услуги по передаче электроэнергии и эксплуатации сетей. В ведении организации находится пять филиалов, которые обслуживают более 100 тыс. физических и юридических лиц. С 2012 г. по сетям АО «ЮРЭСК» потребителям было передано >500 млрд. кВт/ч электроэнергии.

В 2014 г. ЮРЭСК в рамках программы «Энергосбережение и повышение энергетической эффективности Ханты-Мансийского автономного округа» начала работы по модернизации и реконструкции высоковольтных подстанций региона. Основная цель программы — обеспечение бесперебойного энергоснабжения для потребителей, проживающих в удаленных районах округа, и снижение энергопотерь. Одна из ключевых задач проекта — создание современной энергосети на базе концепции Smart Grid, организация сбора и передачи данных о работе подстанций на центральный пункт управления сетями «ЮРЭСК» и в районы электрических сетей. Генеральным подрядчиком выступила компания BCC.

Специалисты BCC выполнили обследование и электросетевой аудит на объектах ЮРЭСК, проектирование, поставку оборудования и развертывание комплексной автоматизированной системы сбора, передачи данных, мониторинга качества и учета электроэнергии. На первом этапе проекта работы велись на семи подстанциях 35 – 110 кВ, две из которых расположены в Ханты-Мансийске, четыре — в Кондинском и одна — в Нижневартовском районах ХМАО.

Большинство работ велось в удаленных труднодоступных районах округа в тяжелых погодных условиях. В общей сложности было смонтировано около 200 единиц оборудования. Построенная система сбора и передачи информации дает возможность АО «ЮРЭСК» вести централизованное управление и контроль за состоянием удаленного оборудования, оперативно реагируя на внештатные ситуации.

Внедрение технологий Smart Grid позволит снизить энергопотери и сделать доставку электроэнергии в удаленные регионы ХМАО более надежной.

Http://www.bcc.ru