

**РАЗРАБОТКА ПРОГНОЗИРУЮЩЕЙ МОДЕЛИ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОГО ВРЕМЕНИ ОТБОРА ПРОБ НА ПРИМЕРЕ РЕАКЦИОННО-РЕКТИФИКАЦИОННОГО ТЕХНОЛОГИЧЕСКОГО ОБЪЕКТА**

**А.А. Гончаров, С.А. Самотылова, А.Ю. Торгашов (ИАПУ ДВО РАН),
Д.С. Ан (АО "Газпромнефть - ОНПЗ")**

На примере реакционно-ректификационного технологического процесса рассматривается задача построения модели, прогнозирующей содержание побочного компонента в выходном продукте в условиях неизвестной временной задержки отбора проб. Для определения временной задержки предлагается использовать алгоритм, основанный на итерационной процедуре EM алгоритма совместно с бутстреп-анализом и гребневой регрессией. Предлагаемый алгоритм протестирован на промышленных данных¹.

Ключевые слова: прогнозирующая модель, виртуальный анализатор, реакционно-ректификационный процесс, неточное время отбора пробы, идентификация, EM алгоритм, бутстреп-анализ, гребневая регрессия.

Введение

В настоящее время для оценки показателей качества производимой продукции широко внедряются прогнозирующие модели (виртуальные анализаторы). Они позволяют оперативно оценить необходимые показатели качества выходного продукта на основе измеряемых технологических переменных и результатов анализов проб продуктов заводскими лабораториями [1]. Недостаточное знание физико-химических особенностей технологического процесса, невозможность измерений показателей качества производимой продукции в режиме реального времени, а также неизвестные временные задержки отбора пробы затрудняют определение и снижают точность прогнозирующей модели. Разработка алгоритма, учитывающего неточное время отбора пробы, является актуальной задачей при построении виртуальных анализаторов, функционирующих в составе систем усовершенствованного управления технологическими процессами (СУУТП). Повышение точности прогнозирующих моделей в составе СУУТП в конечном итоге способствует увеличению отбора наиболее ценных продуктов требуемого качества.

В работе Li Xie и др. [2] исследуемая проблема решается с помощью EM алгоритма, который представляет собой итеративный метод для нахождения оценок максимального правдоподобия. На первом шаге E (expectation) для всех наблюдений происходит определение временной задержки на основе заранее полученных параметров модели. На втором шаге M (maximization) на основании полученной временной задержки отбора пробы происходит максимизация логарифмической функции правдоподобия с оцен-

кой новых параметров модели. Далее идет возврат к шагу E, на котором используются параметры модели, полученные на M шаге. Итерационный процесс продолжается до тех пор, пока разница между величиной оцениваемых параметров на соседних итерациях не достигнет заданного значения. Основным недостатком данного алгоритма является то, что он в ряде ситуаций выбирает первый попавшийся локальный максимум, который является ближайшим к начальному приближению [3].

В данной работе для определения параметров прогнозирующей модели при неточном времени отбора проб предлагается использовать итерационный метод, основанный на E и M шагах EM-алгоритма совместно с гребневой регрессией для оценки параметров модели на M шаге. Для решения проблемы попадания в локальный максимум предлагается использовать бутстреп-анализ для формирования нескольких начальных значений параметров модели на основе бутстреп выборок.

Постановка задачи

Рассматривается технологический объект, описываемый связью между входными и выходной переменными с помощью конечных импульсных характеристик (КИХ):

$$y_{T_i-\lambda_i} = b_0 + \sum_{m=1}^M \sum_{g=0}^{G^{(m)}-1} b_{g+1}^{(m)} u_{T_i-g}^{(m)} + e_{T_i} \quad (1)$$

где $u^{(1)}, \dots, u^{(M)}$ — входные переменные, доступные измерению в каждый период времени; $y_{T_i-\lambda_i}$ — измерения выходной переменной ($i=1, 2, \dots, M$) с неизвестной временной задержкой λ_i отбора i -ой пробы и ошиб-

¹ Работа осуществлялась при частичной финансовой поддержке Российского фонда фундаментальных исследований (проект 17-07-00235 А).

ками измерений e_{T_i} ; $G^{(m)}$ — размерность вектора КИХ по входу m ; M — количество входов; b_0 — свободный член; $b_{g+1}^{(m)}$ — параметры объекта.

Модель объекта имеет следующий вид:

$$\hat{y}_{T_i-\lambda_i} = \hat{b}_0 + \sum_{m=1}^M \sum_{g=0}^{G^{(m)}-1} \hat{b}_{g+1}^{(m)} u_{T_i-g}^{(m)} \quad (2)$$

где \hat{y}_{T_i} — оцененное значение выходной переменной объекта; $\hat{b}_{g+1}^{(m)}$ — параметры прогнозирующей модели; \hat{b}_0 — свободный член объекта модели.

Временная задержка отбора пробы является случайной величиной, заданной на вероятностном пространстве и может иметь любое дискретное распределение.

Задача заключается в оценке параметров $\hat{b}_{g+1}^{(m)}$ и временной задержки отбора пробы λ_i в уравнении (2) при известных последовательностях данных $\{u_{T_i}\}$ и $\{y_{T_i}\}$.

Описание технологического процесса

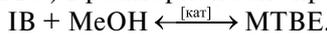
Для рассматриваемого в данной работе процесса синтеза метил трет бутилового эфира (МТБЭ) — компонента бензина, обладающего высоким октановым числом, используется изобутилен, находящийся в смеси С4-углеводородов, и метанол. МТБЭ получил широкое применение в производстве высокооктановых бензинов и способствует практически полному сгоранию топлива в двигателе, благодаря присутствующему в его молекуле кислороду, при этом является нетоксичным компонентом.

Процесс получения МТБЭ проводится в реакторе с твердым катализатором, в качестве которого выступает кислотная ионообменная смола, при температуре 50-95°С. Схема действующей технологической установки синтеза МТБЭ представлена на рис. 1. Потоки отмытой бутан-бутиленовой фракции ББФ и метанола предварительно смешиваются в смесителе и автоматически подаются на установку синтеза МТБЭ. Синтез МТБЭ из ББФ и метанола происходит в прямоточном реакторе форконтакта (P-50) и в реакционно-ректификационном аппарате, представляющем собой две ректификационные колонны (K_p-51/1 и K_p-51/2) и расположенный между ними реактор синтеза (P-51). В прямоточном реакторе P-50 происходит частичная реакция синтеза МТБЭ. Отвод тепла реакции осуществляется за счет вывода разогретой реак-

ционной смеси из P-50. Основная реакция синтеза МТБЭ протекает в реакционно-ректификационном аппарате P-51. Газовый поток с верха P-51 в дальнейшем поступает в ректификационную колонну K_p-51/2, а поток жидкости с низа P-51 в колонну K_p-51/1.

При осуществлении синтеза МТБЭ особое внимание уделяется выбору соотношения изобутилен метанол. Расход первого компонента принимается постоянным, а количество подаваемого метанола определяется в зависимости от нагрузки ББФ и содержания изобутилена в нем. Определение и поддержание соотношения компонентов осуществляется на основании опыта технолога. При неправильном задании расхода метанола относительно расхода ББФ количество непрореагировавшего метанола на выходе реактора может быть либо больше, либо меньше требуемого значения. В обоих случаях это приводит к перерасходу исходного сырья, недополучению готовой продукции и снижению ее качества. По окончании процесса выход МТБЭ должен составлять 98%.

Основная химическая реакция (образование МТБЭ) в реакторной зоне протекает по схеме:



Кроме основной химической реакции процесса производства МТБЭ в незначительной степени могут протекать побочные реакции. При наличии воды изобутилен реагирует с ней с образованием трет-бутилового спирта. Также в условиях недостаточного количества метанола изобутилен димеризуется с образованием 2,4,4-триметилпентена-1 и 2,4,4-триметилпентена 2.

Основным показателем качества данного процесса является содержание спиртов в МТБЭ. Массовая

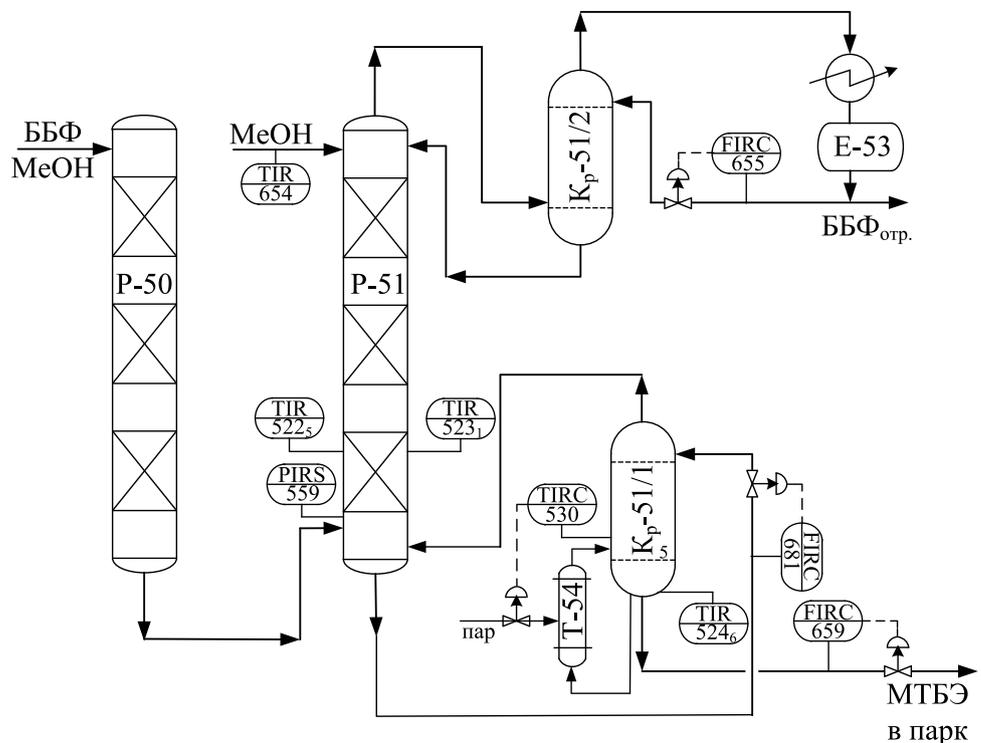


Рис. 1. Технологическая схема синтеза МТБЭ

доля спиртов (метанола и трет бутанола) не должна превышать 1,5% в товарном продукте МТБЭ.

Определение параметров модели ВА при неточном времени отбора пробы

Время отбора пробы случайная величина и не зависит от входов и параметров модели. Для определения параметров модели ВА в условиях неточного времени отбора пробы для i -го наблюдения y_{T_i} с неизвестным временем задержки предлагается $\hat{\lambda}_i = j$ следующий алгоритм.

Шаг 1. Для величины входных переменных $U = u^{(1)}, \dots, u^{(M)}$ и наблюдаемой переменной выхода $Y = y_{T_1}, \dots, y_{T_N}$ находим ошибку вычислений E_{ij} по формуле (3), используя ранее полученные значения параметров модели $\hat{b}_{0:g+1}^{(m)}$:

$$E_{ij} = (y_{T_i - \lambda_i} - \hat{y}_{T_i - j})^2 \quad (3)$$

Оценка временной задержки отбора пробы i -го наблюдения будет соответствовать минимальному значению E_{ij} среди всех задержек из заданного диапазона:

$$\hat{\lambda}_i = \arg \min_j E_{ij}, j = 0, 1, \dots, q.$$

Шаг 2. На основе полученных задержек $\hat{\lambda}_i$ вычисляем новые коэффициенты параметров модели с помощью гребневой регрессии:

$$\hat{b} = [U^T U + kI]^{-1} U^T Y, \quad (4)$$

где \hat{b} — вектор коэффициентов линейной модели; U — матрица значений независимых переменных; $k \geq 0$ параметр гребневой регрессии, задается как скалярная величина; I — единичная матрица; Y — вектор значений зависимой переменной.

Далее идет возврат к шагу 1, на котором используются оцененные на шаге 2 параметры модели. Итерационный процесс продолжается до тех пор, пока разница между величиной оцениваемых параметров

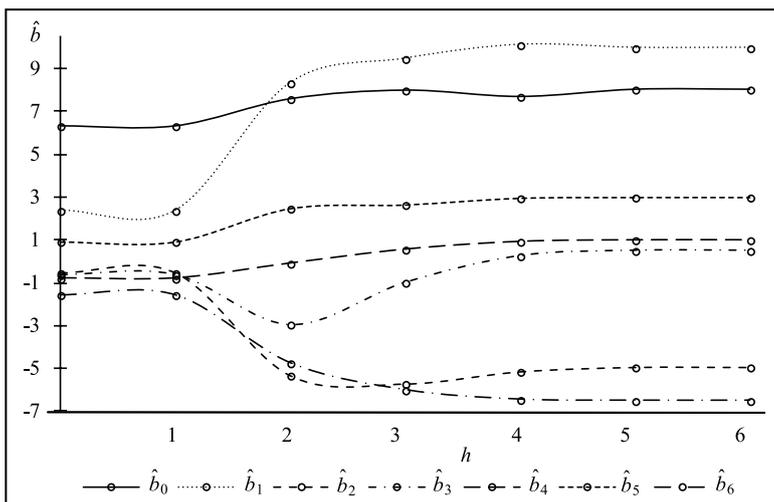


Рис. 2. Оценка параметров модели на итерации h

на соседних итерациях не достигнет заданной малой величины:

$$\delta^{(h+1)} = \frac{\|\hat{b}^{(h+1)} - \hat{b}^{(h)}\|^2}{\|\hat{b}^{(h)}\|^2}$$

В настоящей работе отличительной чертой от EM алгоритма является то, что на шаге 1 выполняется определение временной задержки $\hat{\lambda}_i$ i -ой пробы на основе уравнения (3), вместо вычисления условного математического ожидания логарифма функции правдоподобия $Q(\hat{b}, \hat{b}^{(h)})$, предложенного в работе [2]. На шаге 2 выполняется не максимизация функции $Q(\hat{b}, \hat{b}^{(h)})$ для получения новых значений параметров $\hat{b}^{(h+1)}$, а вычисление параметров модели с помощью гребневой регрессии.

Следует отметить, что кроме проблемы сходимости в локальном максимуме одной из основных проблем при построении прогнозирующих моделей является недостаточный объем данных обучающей выборки. Описанные проблемы предлагается решать с помощью бутстреп анализа. Основная идея такого метода заключается в многократном извлечении элементов из обучающей выборки и формировании

Таблица 1 Значения параметров прогнозирующей модели на каждой итерации $h = 0, \dots, 6$

| | EM-алгоритм | | | | | | | Предлагаемый алгоритм | | | | | | |
|-------------|-------------|-------|-------|-------|-------|-------|-------|-----------------------|-------|-------|-------|-------|-------|-------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| \hat{b}_0 | 6,33 | 6,33 | 7,60 | 8,00 | 7,70 | 8,04 | 8,04 | 6,33 | 6,33 | 7,60 | 8,00 | 7,70 | 8,04 | 8,04 |
| \hat{b}_1 | 2,37 | 2,37 | 8,33 | 9,47 | 10,12 | 9,97 | 9,97 | 2,37 | 2,37 | 8,33 | 9,47 | 10,12 | 9,97 | 9,97 |
| \hat{b}_3 | -0,54 | -0,54 | -5,32 | -5,71 | -5,15 | -4,94 | -4,94 | -0,54 | -0,54 | -5,32 | -5,71 | -5,15 | -4,94 | -4,94 |
| \hat{b}_3 | -0,63 | -0,63 | -2,94 | -0,98 | 0,26 | 0,50 | 0,50 | -0,63 | -0,63 | -2,94 | -0,98 | 0,26 | 0,50 | 0,50 |
| \hat{b}_4 | -0,80 | -0,80 | -0,10 | 0,56 | 0,93 | 1,00 | 1,00 | -0,80 | -0,80 | -0,10 | 0,56 | 0,93 | 1,00 | 1,00 |
| \hat{b}_5 | 0,93 | 0,93 | 2,48 | 2,65 | 2,96 | 3,00 | 3,00 | 0,93 | 0,93 | 2,48 | 2,65 | 2,96 | 3,00 | 3,00 |
| \hat{b}_6 | -1,56 | -1,56 | -4,75 | -6,02 | -6,46 | -6,51 | -6,51 | -1,56 | -1,56 | -4,75 | -6,02 | -6,46 | -6,51 | -6,51 |

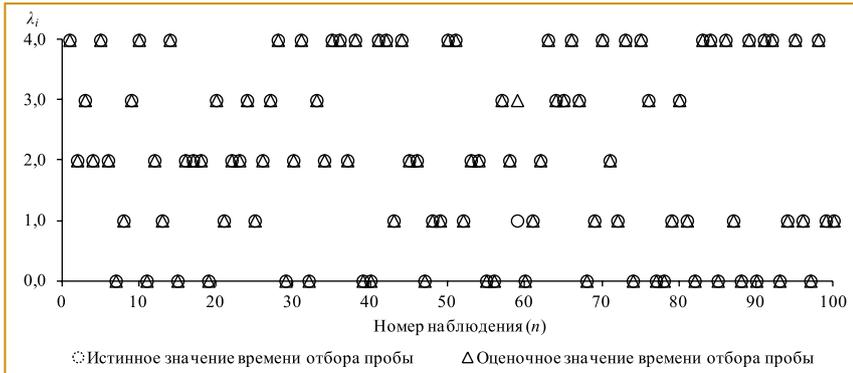


Рис. 3. Оценка неточного времени отбора проб

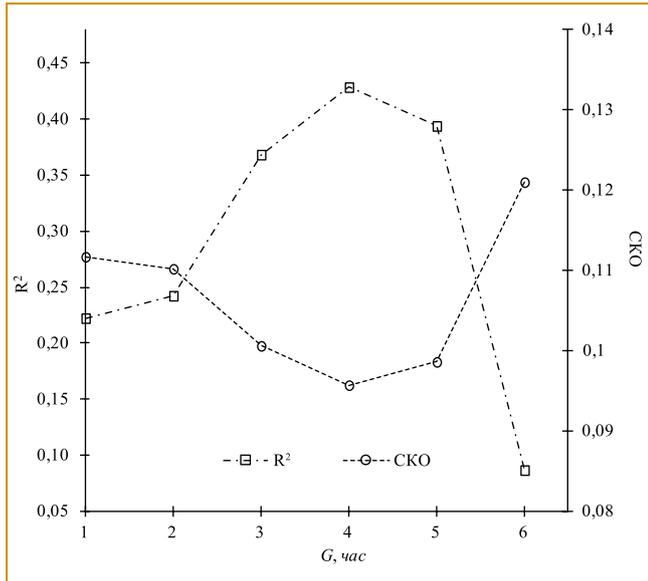


Рис. 4. Значения R^2 и SKO при различной глубине модели G

множества псевдовыборок заданного размера, состоящих из случайных комбинаций исходного набора [4, 5]. Таким образом, решались две проблемы: проблема недостаточного объема выборки и проблема сходимости в локальном максимуме при итерационном расчете $\hat{\lambda}_i$.

Для сравнения и проверки алгоритма использовали объект, описываемый следующим уравнением:

$$y_{T_i-\lambda_i} = 8 + 10u_{T_i}^{(1)} - 5u_{T_i-1}^{(1)} + 0.5u_{T_i-2}^{(1)} + 1u_{T_i}^{(2)} + 3u_{T_i-1}^{(2)} - 6.5u_{T_i-2}^{(2)} + e_{T_i}$$

где $u^{(1)}$ и $u^{(2)}$ — последовательности случайных входных переменных с нормальным законом распределением с $\mu=0$ и $\sigma=1$ для $u^{(1)}$, $\mu=0$ и $\sigma=5$ для $u^{(2)}$; — шумовая составляющая ($\mu=0$ и $\sigma=0,3$). Выход $\{y_{T_i}\}$ доступен в каждый 5-й момент времени измерений входов со случайным временем задержки $\lambda_i \in [0, 4]$. Применяя описанный выше алгоритм и основываясь на доступных данных, был определен вектор параметров исследуемой системы и проведено сравнение с EM-алгоритмом. Вектор параметров, полученный с помощью EM алгоритма и предложенного выше алгоритма, имеет следующий вид:

$$\hat{b} = [8,0401; 9,9733; -4,9417; 0,5009; 1,0017; 2,9961; -6,5123].$$

Оценка параметров модели на каждой итерации отображена на рис. 2; результаты приведены в табл. 1.

Значения параметров прогнозирующей модели на каждой итерации приведены в таблице 1.

Как видно из рис. 2 и табл. 1, использование предлагаемого алгоритма и EM алгоритма дают одинаковые результаты. Следует отметить,

что использование предлагаемого алгоритма приводит к заметному упрощению E шага EM алгоритма. Оценка временной задержки отбора пробы приведена на рис. 3.

На рис. 3 показано восстановление неточного времени отбора пробы с использованием предлагаемого алгоритма и EM алгоритма. Следует отметить, что оба метода позволяют достаточно точно определить временную задержку отбора пробы. Однако при $n=59$ временная задержка отбора пробы определена не точно в обоих случаях.

Предложенный алгоритм был применен для промышленного технологического объекта синтеза МТБЭ (рис. 1). Для построения модели, прогнозирующей содержание спиртов (y) в выходном продукте МТБЭ, в качестве регрессоров использовали измеряемые технологические переменные: $u^{(1)}$ — расход орошения Кр-51/2, м³/ч (FIRC655); $u^{(2)}$ — расход МТБЭ в парк, м³/ч (FIRC659); $u^{(3)}$ — реакционная масса в Кр-51/1, м³/ч (FIRC681); $u^{(4)}$ — давление в Р-51, кгс/см² (PIRS559); $u^{(5)}$ — температура нижнего слоя Р-51, °С (TIR5225); $u^{(6)}$ — температура нижнего слоя Р-51, °С (TIR5231); $u^{(7)}$ — температура куба Кр-51/1, °С (TIR5246); $u^{(8)}$ — температура метанола на синтез, °С (TIR654); $u^{(9)}$ — температура в Кр-51/1 на 5^{ой} тарелке, °С (TIRC530).

В качестве исходной информации для получения прогнозирующей модели использовали обучающую выборку с числом измерений выхода 65. Объем проверочной выборки — 28.

Полученные с помощью бутстреп-анализа выборки были сформированы случайным образом из обучающей выборки $Y = \{y_{T_i-\lambda_i}, \dots, y_{T_N-\lambda_N}\}$. Процесс формирования таких выборок был реализован 250 раз с последующей оценкой коэффициентов модели методом гребневой регрессии.

Табл. 2. Сравнительный анализ критериев идентификации

| Наименование | $R^2_{обуч}$ | $R^2_{пр}$ | SKO _{обуч} | SKO _{пр} |
|---|--------------|------------|---------------------|-------------------|
| Без учета задержки во времени отбора пробы | 0,9881 | 0,4287 | 0,0076 | 0,0957 |
| С учетом $\hat{\lambda}_i$ с помощью EM-алгоритма | 0,9997 | 0,1374 | 0,0012 | 0,1176 |
| С учетом $\hat{\lambda}_i$ с помощью предлагаемого метода | 0,9996 | 0,5270 | 0,0014 | 0,0871 |

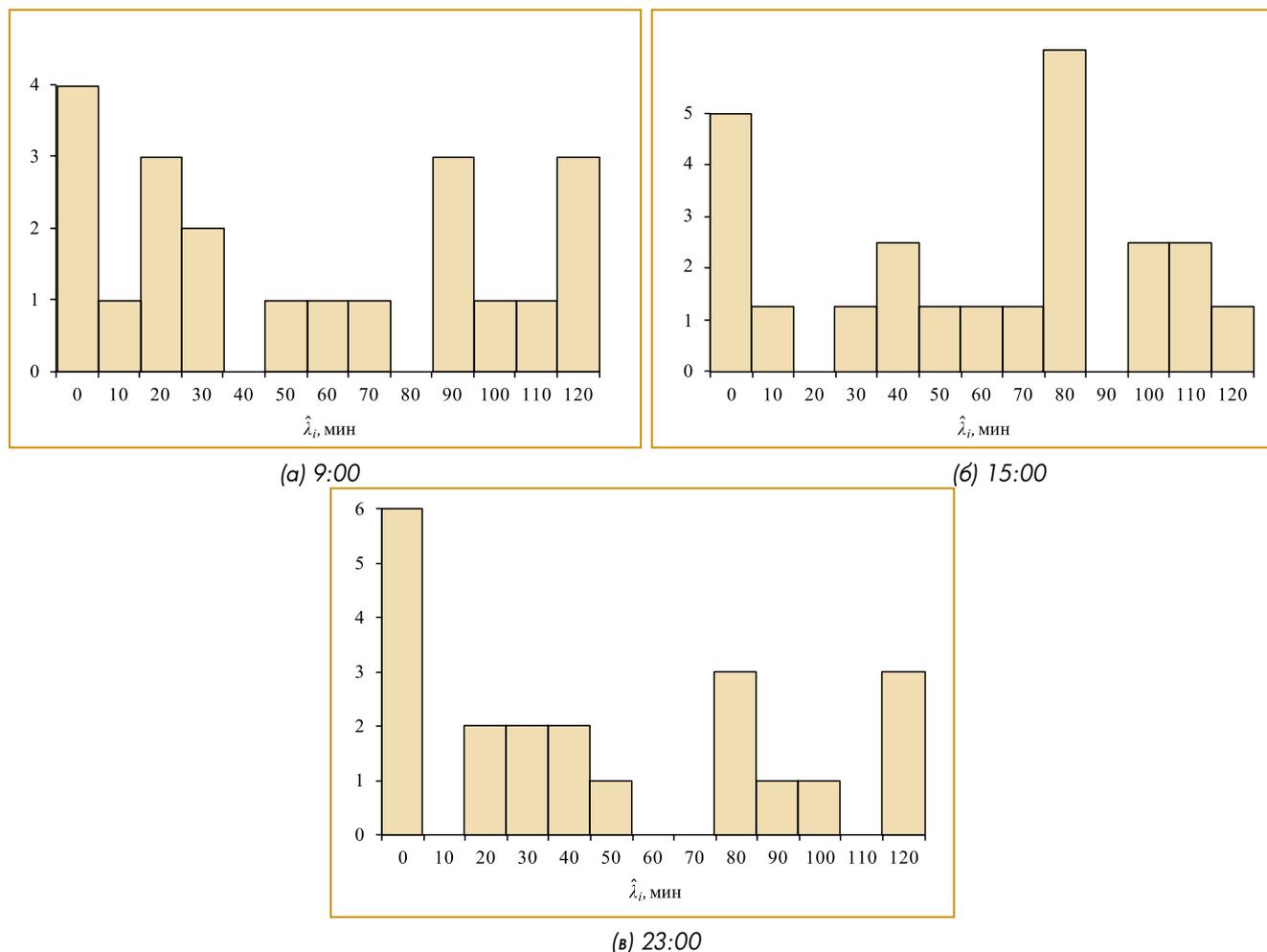


Рис. 5. Гистограммы значений $\hat{\lambda}_i$ полученные с использованием предложенного алгоритма на обучающей выборке

Для оценки точности прогнозирующей модели использовались следующие критерии: коэффициент детерминации (R^2):

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2},$$

среднеквадратическую ошибку (СКО):

$$СКО = \left(\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N} \right)^{1/2}.$$

где y_i — измеряемое значение выходной переменной, \hat{y}_i — ее значение, полученное на основе ВА, \bar{y}_i — среднее значение наблюдаемой выходной переменной. Модель тем больше соответствует исследуемому объекту, чем ближе к единице значение коэффициента детерминации R^2 , чем ближе к нулю значение среднеквадратической ошибки СКО.

Глубина модели $G^{(m)}$ (2) по каждому входу имела одинаковое значение. Перебор значений $G^{(m)}$ осуществлялся в интервале [1...6] (рис. 4).

На рис. 4 видно, что минимальное значение R^2 критерия СКО и максимальное значение достигаются при $G=4$. Восстановление времени отбора проб проводили на обучающей выборке с помощью предла-

гаемого алгоритма, EM алгоритма и без учета неточности во времени отбора пробы. Полученные результаты приведены в табл. 2 и сравнены между собой.

Из табл. 1 видно, что точность оценки показателя качества с помощью прогнозирующей модели увеличилась при использовании предложенного алгоритма.

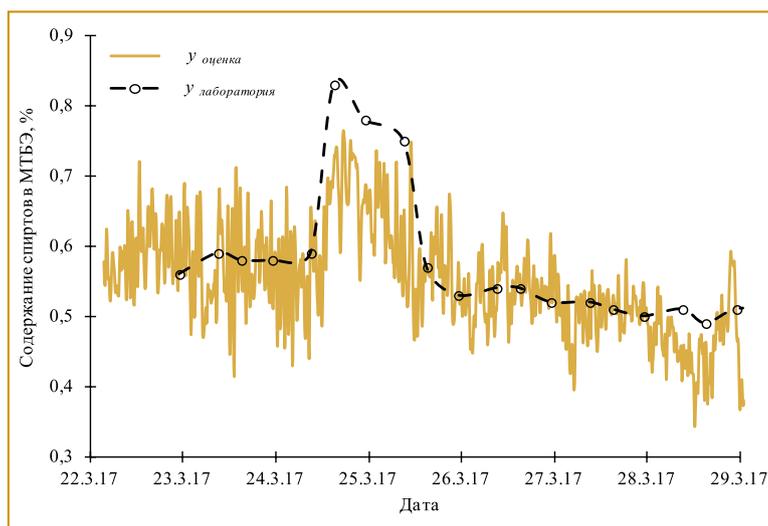


Рис. 6. Выход прогнозирующей модели и результаты лабораторных анализов (проверочная выборка)

Также следует отметить, что полученные значения критериев качества (R^2 и СКО) для модели, прогнозирующей содержание спиртов в МТБЭ с учетом восстановления времени отбора пробы с помощью EM алгоритма, существенно хуже, что свидетельствует о его сильной зависимости от выбора начального приближения значений параметров модели. Проблемы сходимости в локальном максимуме решаются с использованием бутстреп-анализа путем формирования нескольких начальных приближений оценок параметров модели на основе псевдовыборок.

Гистограммы значений $\hat{\lambda}_i$ для различного времени отбора пробы, полученных с использованием предлагаемого алгоритма, изображены на рис. 5.

По полученным гистограммам видно, что отбор пробы, соответствующий 9:00 часам, содержит в себе временную задержку от 0 до 2 часов. Временная задержка для отбора пробы в 15:00 соответствует в основном 80 минутам, а для отбора пробы в 23:00 преобладает нулевая временная задержка.

Результаты функционирования прогнозирующей модели, полученной с помощью предлагаемого алгоритма, приведены на рис. 6.

Применение предлагаемого алгоритма для определения временных задержек отбора пробы позволяет улучшить точность оценки показателя качества продукта (снижение СКО) до $((0,0957-0,0871)/0,0957) \cdot 100 \approx 10\%$ и $((0,1176-0,0871)/0,1176) \cdot 100 \approx 25\%$ в сравнении с методом без учета задержки и применением EM алгоритма на проверочной выборке.

Список литературы

1. Wu Y., Luo X. A novel calibration approach of soft sensor based on multirate data fusion technology // Journal of Process Control. 2010. V. 20. P. 1252-1260.
2. Xie Li, Yang H. Huang B. FIR Model Identification of Mutirate Processes with Random Delay Using EM Algorithm // AIChE Journal. 2013. P. 1-41.
3. Zhou Xi-Yu, Lim Joon S. Replace Missing Values with EM algorithm based on GMM and Naïve Bayesian // International Journal of Software Engineering and Its Application. 2014. V. 8. No. 5. P. 177-188.
4. Efron B. Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy // Statistical Science. 1986. V. 1. No. 5. P. 54-77.
5. Davison A. C., Hinkley D.V. Bootstrap methods and their application. USA: Cambridge University Press, 2008. 583 p.

Гончаров Антон Александрович — канд. техн. наук, младший научный сотрудник,

Самотылова Светлана Александровна — младший научный сотрудник,

Торгашов Андрей Юрьевич — д-р техн. наук, зав. лабораторией систем управления технологическими процессами ФГБУН «Институт автоматики и процессов управления Дальневосточного отделения РАН»,

ФГАОУВО Дальневосточный федеральный университет (ДВФУ),

Ан Дмитрий Сергеевич — главный метролог АО «Газпромнефть-Омский НПЗ».

Контактный телефон +7(423)231-02-02.

E-mail: torgashov@iacp.dvo.ru

ПОВЫШЕНИЕ ИНФОРМАТИВНОСТИ И ОБЪЕКТИВНОСТИ ОЦЕНКИ ПОВРЕЖДАЕМОСТИ ГЕОТЕКСТИЛЬНЫХ ПОЛОТЕН

Т.О. Гойс, А.Ю. Матрохин,

А.В. Умников (ФГБОУ ВО «Ивановский государственный политехнический университет»)

Созданы научные предпосылки для интерпретации результатов компьютерного анализа изображений геотекстильных полотен на предмет степени повреждения их структуры в привычной шкале балловых оценок, определены пути дальнейшего совершенствования количественной оценки повреждаемости геотекстильных полотен.

Ключевые слова: геотекстильные полотна, оценка повреждаемости, компьютерный анализ изображения, шкала балловых оценок.

Для оценки пригодности геотекстильных полотен (ГТП) к использованию по назначению изготовитель в ходе типовых испытаний должен определять степень повреждаемости полотна при действии различных эксплуатационных факторов (циклические нагрузки, агрессивная среда, микроорганизмы, светопогода, многократное замораживание и оттаивание).

На сегодняшний день существуют различные нормативные критерии (признаки) повреждений. Согласно ГОСТ 18976-73. «Ткани текстильные. Метод определения стойкости к истиранию», критерием повреждения при испытании тканых полотен на истирание является момент разрушения пробы, вызыва-

ющего автоматическую остановку прибора. В результате записывают число циклов воздействия (например, число оборотов головки прибора), при котором произошло разрушение испытуемой ткани. В данном подходе отсутствует оценка степени поврежденности материала, то есть самоостанов прибора может быть вызван как отдельными деформациями, так и накоплением этих деформаций на большой площади. Таким образом, невозможно с заданной точностью спрогнозировать момент разрушения материалов.

Другой подход (ГОСТ Р ИСО 12947-2-2011. «Материалы текстильные. Определение стойкости к истиранию методом Мартиндейла.» Часть 2. «Определение