

Сетевой научный журнал «Ресурсы» <https://resources.today>
Russian Journal of Resources

2026, Том 13, № 3 / 2026, Vol. 13, Iss. 3 <https://resources.today/issue-3-2026.html>

URL статьи: <https://resources.today/PDF/15INOR326.pdf>

DOI: 10.15862/15INOR326 (<https://doi.org/10.15862/15INOR326>)

2.3.3. Автоматизация и управление технологическими процессами и производствами (технические науки)

Ссылка для цитирования этой статьи:

Азизмохаммади, М.-С. М. Повышение эффективности управления технологическими процессами на основе интеллектуальной обработки производственных данных / М.-С. М. Азизмохаммади, А. В. Гончаров, А. В. Шаховской, Е. Г. Ротанов // Ресурсы. — 2026. — Т. 13. — № 3. — URL: <https://resources.today/PDF/15INOR326.pdf>.
DOI: 10.15862/15INOR326.

For citation:

Azizmohammadi M.-S.M., Goncharov A.V., Shakhovskoy A.V., Rotanov E.G. Improving the efficiency of technological process control based on intelligent processing of production data. *Russian Journal of Resources*. 2026;13(3): 15INOR326. Available at: <https://resources.today/PDF/15INOR326.pdf>. DOI: 10.15862/15INOR326. (In Russ., abstract in Eng.).

УДК 681.5:004.85

Азизмохаммади Мохаммад-сафар Мохсенович

ФГБОУ ВО «Московский государственный университет технологий и управления имени К. Г. Разумовского»,
Москва, Россия
E-mail: safar2002@yandex.ru

Гончаров Андрей Витальевич

ФГБОУ ВО «Московский государственный университет технологий и управления имени К. Г. Разумовского»,
Москва, Россия
Кандидат технических наук, доцент
E-mail: a.goncharov@mgutm.ru

Шаховской Андрей Владимирович

ФГБОУ ВО «Московский государственный университет технологий и управления имени К. Г. Разумовского»,
Москва, Россия
Кандидат технических наук, доцент
E-mail: a.shakhovskoy@mgutm.ru

Ротанов Евгений Геннадьевич

ФГБОУ ВО «Московский государственный университет технологий и управления имени К. Г. Разумовского»,
Москва, Россия
Кандидат технических наук, доцент
E-mail: e.rotanov@mgutm.ru

**Повышение эффективности
управления технологическими процессами на основе
интеллектуальной обработки производственных данных**

Аннотация. В распределённых системах управления непрерывных производств накапливаются значительные массивы производственных данных, однако показатели качества продукции по-прежнему измеряются поточными анализаторами с запаздыванием, и решения по ведению режима принимаются по устаревшей информации. В работе предложен и апробирован подход, при котором результат интеллектуальной обработки производственных данных соотносится с эффективностью управления технологическим процессом, а точность модели служит промежуточной характеристикой. Эмпирическую базу составил открытый

набор промышленных данных колонны дебутанизации нефтеперерабатывающего завода, включающий семь технологических переменных и содержание бутана в кубовом продукте. Задача поставлена как упреждающая оценка показания анализатора на заданном горизонте по информации, доступной на момент принятия решения. Сопоставлены наивная схема, одномерная авторегрессия и четыре многомерные модели, а именно гребневая регрессия, проекция на латентные структуры, градиентный бустинг и нейронная сеть, при хронологическом разделении наблюдений и перекрёстной проверке с расширяющимся окном. Установлено, что технологические переменные опережают показание анализатора, а их включение в модель сокращает ошибку упреждающей оценки тем сильнее, чем длиннее горизонт, и на средних горизонтах уменьшает её примерно наполовину относительно одномерной схемы. Регуляризованная линейная модель оказалась не хуже нелинейных, а типовая фильтрация выбросов ухудшила результат и была отклонена по перекрёстной проверке. Основной вклад в точность вносит температура на контрольной тарелке, тогда как часть переменных информационно избыточна. На основе известного результата о нижней границе дисперсии регулируемой величины ошибка упреждающей оценки интерпретирована как предел достижимого качества стабилизации. Это позволило выразить эффект через сокращение неустраняемого разброса показателя, допустимый сдвиг рабочей точки к границе спецификации и полноту раннего предупреждения о выходе показателя за порог. Результаты могут быть использованы при обосновании проектов усовершенствованного управления и при выборе состава данных для виртуальных анализаторов.

Ключевые слова: управление технологическими процессами; интеллектуальная обработка данных; производственные данные; виртуальный анализатор; искусственный интеллект; машинное обучение; запаздывание измерений; показатель качества; ректификация

Введение

Современное непрерывное производство оснащено распределёнными системами управления, которые с периодичностью в секунды и минуты регистрируют температуры, давления, расходы и уровни. Показатели качества продукции измеряются иначе. Состав потоков определяется поточными анализаторами и заводскими лабораториями, результат поступает с запаздыванием от десятков минут до нескольких часов, и на протяжении этого времени технологический процесс¹ ведётся по косвенным признакам. В связи с этим возрастает значение интеллектуальной обработки производственных данных, позволяющей оценивать показатель качества раньше, чем его покажет прибор. Потенциал такого подхода в российской промышленности используется ограниченно. По данным Института статистических исследований и экономики знаний Национального исследовательского университета «Высшая школа экономики», в 2024 году технологии сбора, обработки и анализа больших данных применяли 8,6 % крупных и средних организаций, а технологии искусственного интеллекта² лишь 4,8 %.³

¹ Термин «технологический процесс» употребляется в значении пункта 1 ГОСТ 3.1109-82 как часть производственного процесса, включающая целенаправленные действия по изменению и определению состояния предмета труда. См.: ГОСТ 3.1109-82. Единая система технологической документации. Термины и определения основных понятий: межгосударственный стандарт: введён 01.01.1983. — URL: <https://docs.cntd.ru/document/1200012103> (дата обращения: 01.10.2026).

² Термин «искусственный интеллект» употребляется в значении подпункта «а» пункта 5 Национальной стратегии развития искусственного интеллекта на период до 2030 года. См.: О развитии искусственного интеллекта в Российской Федерации: Указ Президента Российской Федерации от 10 октября 2019 г. № 490 (в редакции Указа Президента Российской Федерации от 15 февраля 2024 г. № 124). — URL: <http://www.kremlin.ru/acts/bank/44731> (дата обращения: 01.10.2026).

³ Индикаторы цифровой экономики: 2026: статистический сборник / В. Л. Абашкин, Г. И. Абдрахманова, М. Я. Бочаров [и др.]; Национальный исследовательский университет «Высшая школа экономики». — М.: ИСИЭЗ ВШЭ, 2026. — Главные выводы сборника об использовании цифровых технологий крупными и средними организациями в 2024 г. — URL: <https://issek.hse.ru/news/1122444011.html> (дата обращения: 01.10.2026).

В научной литературе сложились три направления, связанные с использованием производственных данных в управлении технологическими процессами. Первое направление образуют работы по виртуальным анализаторам (от англ. soft sensor), то есть моделям, оценивающим труднодоступный показатель по легко измеряемым переменным. П. Кадлец, Б. Габрис и С. Страндт обобщили опыт построения таких моделей на основе данных и пришли к заключению, что главные трудности связаны с пропусками, выбросами, запаздыванием измерений и сменой режимов [1]. Л. Фортуна, С. Грациани, А. Риццо и М. Г. Ксибилия изложили полный цикл проектирования виртуального анализатора на примерах установок нефтепереработки [2]. В отечественной литературе идентификационный подход к виртуальным анализаторам развит Н. Н. Бахтадзе, которая рассматривает их как составную часть систем управления производством [3]. Второе направление связано с анализом больших массивов технологических данных методами машинного обучения. С. Дж. Цинь обосновал переход от моделей, построенных на первых принципах, к моделям, извлекаемым из накопленных данных [4], а З. Гэ, Чж. Сун, С. Дин и Б. Хуан систематизировали методы машинного обучения применительно к мониторингу, прогнозированию качества и диагностике [5]. Третье направление посвящено оценке качества самого управления. Т. Харрис установил, что по данным нормальной эксплуатации определяется нижняя граница дисперсии регулируемой величины, которая задаётся запаздыванием и не зависит от настройки регулятора [6]. М. Джелали обобщил развитие этого подхода и практику его промышленного применения [7]. Экономическую сторону вопроса раскрыли М. Бауэр и И. Крейг, по обзору которых выгода от усовершенствованного управления формируется сокращением разброса показателя и последующим сдвигом рабочей точки к ограничению [8]. Отечественный опыт усовершенствованного управления технологическими процессами (от англ. advanced process control) проанализирован В. М. Дозорцевым, Э. Л. Ицковичем и Д. В. Кнеллером [9]. Перечисленные направления развиваются во многом обособленно. Точность виртуальных анализаторов оценивается по ошибке модели, качество управления по дисперсии регулируемой величины, и вопрос о том, насколько обработка производственных данных расширяет пределы достижимого качества управления, остаётся открытым.

Объектом исследования выступает технологический процесс непрерывного производства, показатель качества которого измеряется с запаздыванием, на примере колонны дебутанизации нефтеперерабатывающего завода.

Предметом исследования являются методы интеллектуальной обработки производственных данных и их влияние на достижимое качество управления технологическим процессом.

Цель исследования состоит в оценке того, в какой мере интеллектуальная обработка производственных данных повышает эффективность управления технологическим процессом при запаздывающем измерении показателя качества.

1. Материалы и методы

Производственными данными в работе названы значения технологических переменных и показателей качества, регистрируемые системой управления в ходе технологического процесса. Интеллектуальной обработкой производственных данных нами названа совокупность процедур их подготовки, преобразования и обобщения методами машинного обучения, результатом которой выступает оценка показателя, недоступного для своевременного прямого измерения. Эффективность управления рассматривается в значении, принятом в теории оценки качества регулирования, то есть как близость фактического разброса регулируемой величины к нижней границе, достижимой при данном запаздывании.

Эмпирической базой послужил открытый набор промышленных данных колонны дебутанизации, опубликованный Л. Фортуной, С. Грациани, А. Риццо и М. Г. Ксибилией в

качестве приложения к монографии.⁴ Колонна отделяет бутановую фракцию от стабильного бензина, а показателем качества служит содержание бутана в кубовом продукте. Анализатор установлен на последующем аппарате технологической цепочки, поэтому его показание запаздывает относительно состояния колонны.

2. Результаты и обсуждение

По оценке Л. Фортуны, С. Грациани и М. Г. Ксибилии, это запаздывание постоянно и составляет от 30 до 75 минут [10]. Набор содержит 2394 последовательных наблюдения семи технологических переменных и показателя качества в табличном формате, все величины приведены его составителями к шкале от нуля до единицы. Интервал дискретизации в описании набора не указан, в связи с чем запаздывание и горизонты далее выражены в отсчётах. Первые 1 500 наблюдений отнесены к обучающей части, последние 894 к контрольной. Характеристика переменных приведена в таблице 1.

Таблица 1

Характеристика технологических переменных и показателя качества колонны дебутанизации в обучающей и контрольной частях набора данных, долей единицы нормированной шкалы

Обозначение	Переменная	Среднее значение в обучающей части	Среднее значение в контрольной части	Стандартное отклонение в обучающей части	Стандартное отклонение в контрольной части
u_1	Температура верха колонны	0,262	0,320	0,090	0,114
u_2	Давление верха колонны	0,665	0,672	0,029	0,086
u_3	Расход орошения	0,572	0,643	0,216	0,156
u_4	Расход продукта на следующую стадию	0,482	0,359	0,116	0,138
u_5	Температура на шестой тарелке	0,697	0,593	0,098	0,142
u_6	Температура куба, датчик А	0,693	0,500	0,143	0,124
u_7	Температура куба, датчик Б	0,665	0,463	0,155	0,119
y	Содержание бутана в кубовом продукте	0,257	0,285	0,139	0,186

Рассчитано авторами на основе набора данных колонны дебутанизации, опубликованного в приложении к монографии Л. Фортуны, С. Грациани, А. Риццо и М. Г. Ксибилии

Данные таблицы 1 свидетельствуют о том, что обучающая и контрольная части относятся к разным режимам работы колонны. Средние значения обеих температур куба и температуры на шестой тарелке в контрольной части ниже, разброс давления верха колонны существенно выше, а разброс самого показателя качества также возрос. Такое смещение режима делает случайное перемешивание наблюдений некорректным, поскольку модель получила бы сведения о будущих режимах. По этой причине применены хронологическое разделение и перекрёстная проверка с расширяющимся окном. Порядок проверки прогнозных моделей на временных рядах рассмотрен К. Бергмейром и Х. Бенитесом [11].

Запаздывание между технологическими переменными и показанием анализатора оценено по взаимной корреляционной функции в общепринятой записи Дж. Бокса, Г. Дженкинса и соавторов [12]. По формуле (1) рассчитывается коэффициент корреляции между значением i -й переменной $u_{i,t-l}$, сдвинутым на l отсчётов назад, и показателем качества y_t , где $\bar{u}_i$ и $\bar{y}$ — средние значения, T — число наблюдений обучающей части.

⁴ Набор данных `debutanizer_data` размещён издательством в составе электронных материалов к монографии: Fortuna L., Graziani S., Rizzo A., Xibilia M.G. *Soft Sensors for Monitoring and Control of Industrial Processes*. — London: Springer, 2007. — URL: <https://doi.org/10.1007/978-1-84628-480-9> (дата обращения: 01.10.2026).

$$r_i(l) = \frac{\sum_{t=l+l}^T (u_{i,t-l} - \bar{u}_i) (y_t - \bar{y})}{\sqrt{\sum_{t=l+l}^T (u_{i,t-l} - \bar{u}_i)^2 \sum_{t=l+l}^T (y_t - \bar{y})^2}}. \quad (1)$$

Задача интеллектуальной обработки поставлена как упреждающая оценка показания анализатора на горизонте h отсчётов по информации, доступной на момент принятия решения. За основу взята нелинейная авторегрессионная структура с внешними входами, предложенная Л. Фортуной, С. Грациани и М. Г. Ксибилией, в которой показатель оценивается по текущим и запаздывающим значениям технологических переменных и по запаздывающим показаниям анализатора. Авторская модификация состоит в двух изменениях. Все регрессоры сдвинуты на горизонт h , так что модель не использует сведений, появляющихся позже момента оценки. Оцениваемой величиной принято приращение показателя относительно последнего доступного показания вместо его уровня, что снимает зависимость оценки от среднего уровня режима, смещение которого выявлено в таблице 1. Модель записывается формулой (2), где $\hat{y}_t$ — упреждающая оценка, y_{t-h} — последнее доступное показание анализатора, f — оцениваемая зависимость, $\mathbf{x}_{t-h}$ — вектор регрессоров, сформированный к моменту $t-h$. В вектор регрессоров включены значения семи технологических переменных с запаздываниями 0, 1, 2, 3, 4, 6, 8, 10, 12, 14, 16, 18 и 20 отсчётов, уровень последнего показания анализатора и четыре его последовательных приращения, всего 96 признаков.

$$\hat{y}_t = y_{t-h} + f(\mathbf{x}_{t-h}). \quad (2)$$

Вклад технологических переменных отделён от вклада собственной динамики показателя сопоставлением трёх информационных схем. В наивной схеме оценкой служит последнее доступное показание анализатора. В одномерной схеме приращение прогнозируется авторегрессией по шести предшествующим приращениям показателя в записи Дж. Бокса, Г. Дженкинса и соавторов [12] по формуле (3), где a_0 и a_j — коэффициенты, $\Delta y_t = y_t - y_{t-1}$ — приращение показателя за один отсчёт, ε_t — ошибка. В многомерной схеме используется формула (2) с полным вектором регрессоров.

$$y_t - y_{t-h} = a_0 + \sum_{j=1}^6 a_j \Delta y_{t-h-j+1} + \varepsilon_t. \quad (3)$$

Зависимость f оценивалась четырьмя методами. Гребневая регрессия (от англ. ridge regression) введена А. Хоэрлом и Р. Кеннардом для задач с коррелированными регрессорами [13]. Её коэффициенты находятся по формуле (4) в исходной записи авторов метода, где $\mathbf{X}$ — матрица стандартизованных регрессоров, $\mathbf{d}$ — вектор приращений показателя, λ — параметр регуляризации, $\mathbf{I}$ — единичная матрица. Проекция на латентные структуры (от англ. partial least squares) применена в варианте, описанном С. Волдом, М. Шёстрёмом и Л. Эрикссоном [14]. Градиентный бустинг над решающими деревьями использован в постановке Дж. Фридмана [15].

Нейронная сеть представляла собой перцептрон с одним скрытым слоем из восьми нейронов, обучаемый методом обратного распространения ошибки, предложенным Д. Румельхартом, Дж. Хинтоном и Р. Уильямсом [16], а оценка усреднялась по пяти начальным приближениям. Настраиваемые параметры всех методов выбирались по обучающей части перекрёстной проверкой с расширяющимся окном в пять шагов. Дополнительно проверена предварительная очистка технологических переменных медианным идентификатором выбросов со скользящим окном в 11 отсчётов и порогом в три масштабированных медианных отклонения, который Р. Пирсон рекомендует для технологических данных [17].

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{d}. \quad (4)$$

Точность схем измерена среднеквадратической ошибкой на контрольной части по общепринятой формуле (5), где n — число наблюдений контрольной части. Поскольку данные нормированы, ошибка, умноженная на 100, выражена в процентах диапазона шкалы. Значимость различий между одномерной и многомерными схемами проверена критерием Ф. Диболда и Р. Мариано [18] по формуле (6) в записи авторов критерия, где $\bar{d}$ — среднее разности квадратов ошибок двух схем, $\hat{V}(\bar{d})$ — оценка дисперсии этого среднего с учётом автокорреляции до порядка $h - 1$.

$$E_h = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \cdot 100. \quad (5)$$

$$DM = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}}. \quad (6)$$

Связь точности упреждающей оценки с эффективностью управления установлена на основе результата Т. Харриса [6]. Если суммарное запаздывание в контуре составляет h отсчётов, то никакой регулятор не способен сделать дисперсию регулируемой величины меньше дисперсии ошибки её прогноза на h отсчётов вперёд. По формуле (7) в записи Т. Харриса эта граница выражается через дисперсию случайного возмущения σ_a^2 и коэффициенты ψ_j импульсной характеристики модели возмущения. Доля неустранимой дисперсии в фактической дисперсии показателя σ_y^2 рассчитывается по формуле (8), отношение в которой лежит в основе индексов качества регулирования, обобщённых М. Джелали. Авторская интерпретация состоит в том, что границу определяет не только длина запаздывания, но и состав информации, по которой строится прогноз. При одномерной схеме граница равна квадрату ошибки авторегрессии, при многомерной схеме квадрату ошибки модели (2), и разность между ними измеряет вклад интеллектуальной обработки производственных данных в достижимое качество управления.

$$\sigma_h^2 = \sigma_a^2 \sum_{j=0}^{h-1} \psi_j^2. \quad (7)$$

$$\eta_h = \frac{\sigma_h^2}{\sigma_y^2} \cdot 100. \quad (8)$$

Экономический смысл сокращения разброса раскрывает правило сдвига рабочей точки, описанное П. Латуром, Дж. Шарпом и М. Делани [19] и включённое М. Бауэром и И. Крейгом в общую схему оценки выгод.

При неизменной допустимой доле нарушений спецификации среднее значение показателя может быть приближено к границе на величину, определяемую формулой (9) в общепринятой записи, где z_α — квантиль стандартного нормального распределения для допустимой доли нарушений α , $\sigma_{1,h}$ и $\sigma_{2,h}$ — границы разброса при одномерной и многомерной схемах.

$$\Delta_h = z_\alpha (\sigma_{1,h} - \sigma_{2,h}) \quad (9)$$

Практическая полезность упреждающей оценки для оператора измерена качеством раннего предупреждения о превышении порога, за который принят 90-й процентиль показателя в обучающей части. Предупреждение формируется, если оценка, полученная за h отсчётов до события, превышает порог.

Полнота Rec и точность $Prec$ предупреждений рассчитаны по общепринятым формулам (10) и (11), где TP — число верно предсказанных превышений, FN — число пропущенных превышений, FP — число ложных предупреждений. Логика исследования представлена на рисунке 1.

$$Rec = \frac{TP}{TP + FN} \cdot 100. \quad (10)$$

$$Prec = \frac{TP}{TP + FP} \cdot 100. \quad (11)$$

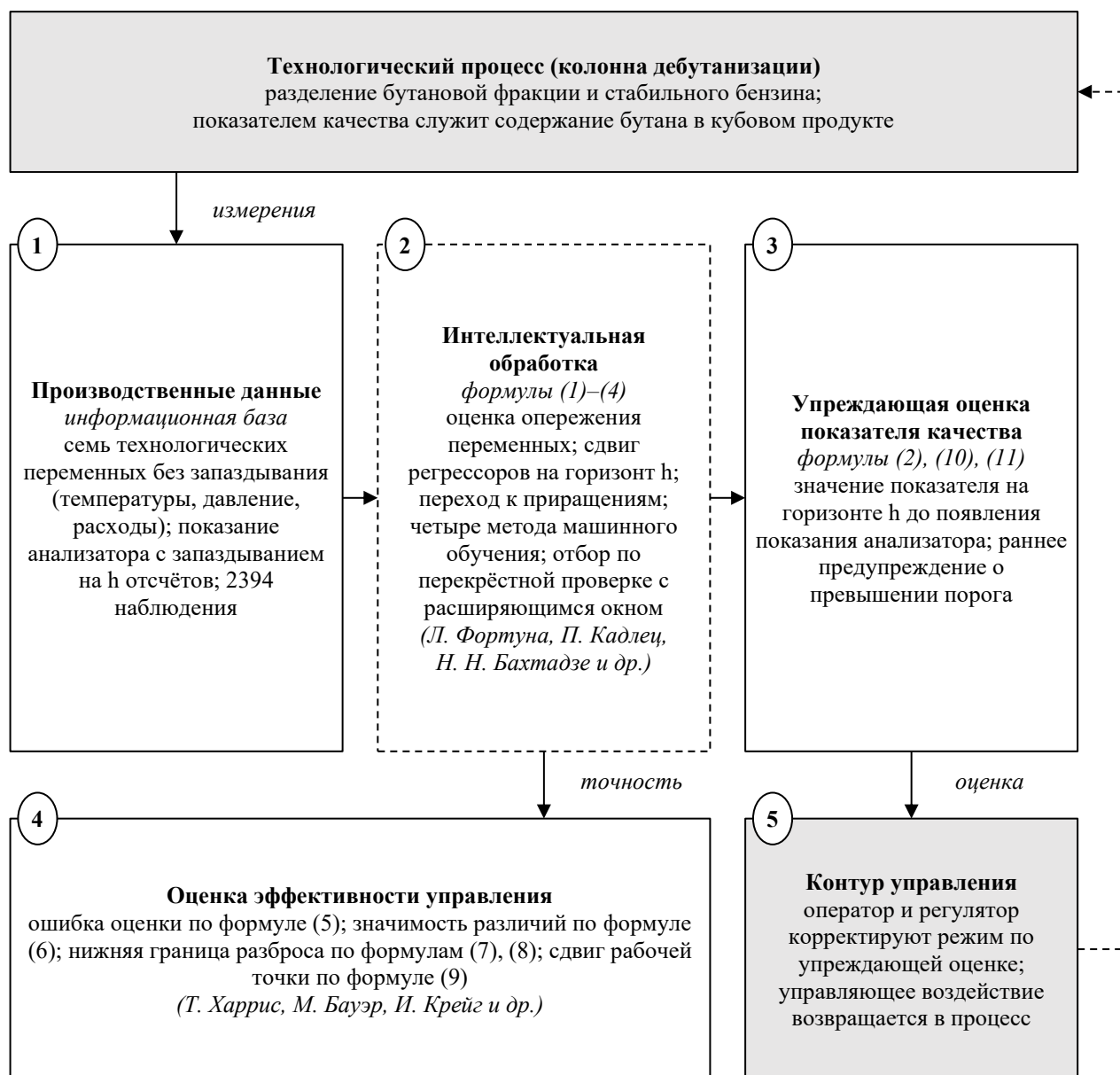


Рисунок 1. Схема повышения эффективности управления технологическим процессом на основе интеллектуальной обработки производственных данных (составлено авторами)

Рассмотрим прежде всего временную структуру производственных данных, от которой зависит сама возможность упреждающей оценки.

Взаимные корреляционные функции трёх технологических переменных, наиболее тесно связанных с показателем качества, рассчитанные по формуле (1), отражены на рисунке 2.

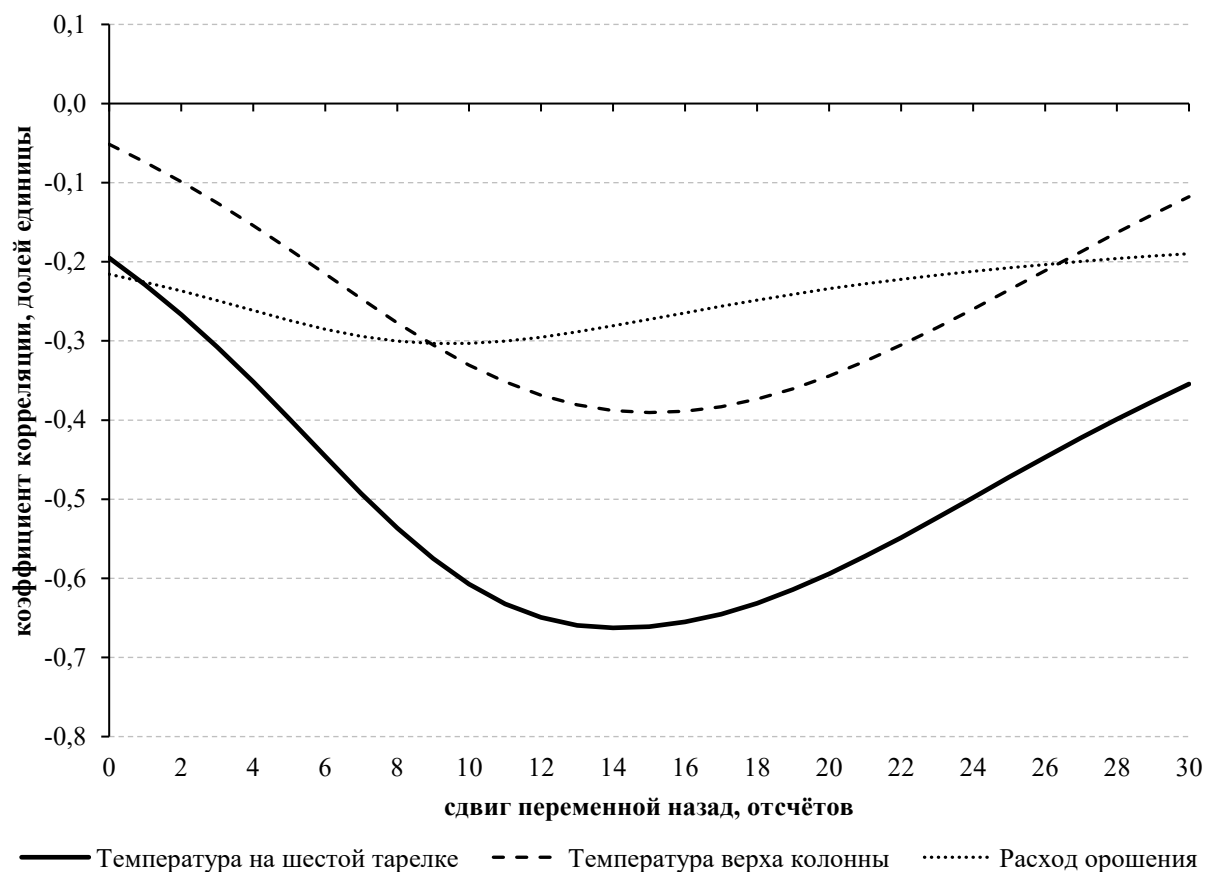


Рисунок 2. Взаимные корреляционные функции содержания бутана в кубовом продукте и технологических переменных колонны дебутанизации при сдвиге переменной назад на 0–30 отсчётов, долей единицы (рассчитано авторами на основе набора данных колонны дебутанизации, опубликованного в приложении к монографии Л. Фортуны, С. Грациани, А. Риццо и М. Г. Ксибилии)

При нулевом сдвиге связь показателя качества с технологическими переменными слабая, и для температуры на шестой тарелке коэффициент корреляции равен -0,195. По мере увеличения сдвига связь усиливается и достигает наибольшей тесноты при сдвиге в 14 отсчётов, где коэффициент составляет -0,663. Для температуры верха колонны экстремум -0,390 приходится на 15 отсчётов, для расхода орошения экстремум -0,303 на 9 отсчётов. Следовательно, технологические переменные опережают показание анализатора, и текущие значения температур несут сведения о том, что прибор покажет спустя 9–15 отсчётов. Это опережение служит источником информации для упреждающей оценки, тогда как модель, связывающая одновременные значения переменных и показателя, такой информации лишена. Проверка подтвердила сказанное. Проекция на латентные структуры, построенная по одновременным значениям семи переменных без показаний анализатора, дала на контрольной части ошибку выше стандартного отклонения самого показателя.

Перейдём к сопоставлению информационных схем. Ошибки упреждающей оценки на контрольной части, рассчитанные по формуле (5) для семи горизонтов, сведены в таблицу 2.

Данные таблицы 2 свидетельствуют о том, что каждое расширение информационной базы сокращает ошибку. Уже одномерная авторегрессия, учитывающая направление движения показателя, на горизонте в четыре отсчёта даёт ошибку 2,55 % против 6,01 % у наивной схемы. Включение технологических переменных снижает ошибку дополнительно, и выигрыш растёт вместе с горизонтом.

Таблица 2

Среднеквадратическая ошибка упреждающей оценки содержания бутана в кубовом продукте на контрольной части набора данных по информационным схемам и горизонтам, процентов диапазона нормированной шкалы

Горизонт, отсчётов	Наивная схема	Одномерная авторегрессия	Гребневая регрессия	Проекция на латентные структуры	Градиентный бустинг	Нейронная сеть	Изменение лучшей многомерной модели к одномерной авторегрессии, процентов
2	3,07	1,07	0,89	0,94	1,23	1,03	-16,4
4	6,01	2,55	1,78	1,88	2,56	1,90	-30,2
6	8,82	4,59	2,68	2,90	3,50	3,05	-41,6
8	11,42	6,93	3,48	5,18	5,23	3,94	-49,9
10	13,81	9,36	4,33	6,38	6,10	4,82	-53,7
12	15,98	11,77	6,05	7,64	7,85	5,77	-50,9
15	18,80	15,13	8,39	9,73	10,56	8,01	-47,0

Рассчитано авторами на основе набора данных колонны дебутанизации, опубликованного в приложении к монографии Л. Фортуны, С. Грациани, А. Риццо и М. Г. Ксибилии. Контрольная часть включает 894 наблюдения

На горизонте в два отсчёта он составляет 16,4 %, на горизонте в восемь отсчётов достигает 49,9 %, а наибольшее значение 53,7 % приходится на десять отсчётов. При дальнейшем увеличении горизонта выигрыш несколько снижается, что согласуется с рисунком 2, так как горизонт приближается к величине опережения и запас информации в технологических переменных исчерпывается. По критерию Ф. Диболда и Р. Мариано [18] превосходство гребневой регрессии над одномерной авторегрессией значимо на всех горизонтах. Статистика (6) равна 2,49 при горизонте в два отсчёта и находится в пределах от 4,16 до 4,59 при остальных горизонтах.

Отдельного внимания заслуживает сопоставление методов. Регуляризованная линейная модель оказалась лучшей на пяти горизонтах из семи, на двух наибольших её незначительно превзошла нейронная сеть. Градиентный бустинг уступил обоим методам, а на коротких горизонтах не дал значимого улучшения относительно одномерной схемы. Объяснение мы видим в смещении режима между частями набора. Решающие деревья не экстраполируют зависимость за пределы значений, встречавшихся при обучении, тогда как линейная модель в приращениях переносит найденную связь на новый режим. Полученный результат согласуется с выводом П. Кадлеца, Б. Габриса и С. Страндта о первостепенной роли подготовки данных при построении виртуальных анализаторов. Показателен и итог проверки фильтрации выбросов. Медианный идентификатор отнёс к выбросам от 3,7 % до 19,3 % значений отдельных переменных, то есть принял за аномалии обычные быстрые изменения режима. Ошибка гребневой регрессии на отфильтрованных данных оказалась выше при перекрёстной проверке на всех горизонтах, и фильтрация была отклонена до обращения к контрольной части. На контрольной части при горизонте в восемь отсчётов её применение увеличило бы ошибку гребневой регрессии до 4,42 %.

Различие схем наглядно на фрагменте контрольной части, включающем подъёмы и спады показателя. Фактические значения и упреждающие оценки при горизонте в восемь отсчётов сопоставлены на рисунке 3. Одномерная авторегрессия продолжает сложившуюся тенденцию и потому отклоняется от фактической траектории на участках смены направления. Оценка гребневой регрессии следует за фактическими значениями теснее, так как изменение температур в колонне сообщает о повороте раньше анализатора. На представленном фрагменте ошибка одномерной схемы составила 7,7 %, многомерной 3,8 % диапазона шкалы. Для управления существенны прежде всего повороты, поскольку ошибочное продолжение тенденции побуждает оператора к избыточному воздействию.

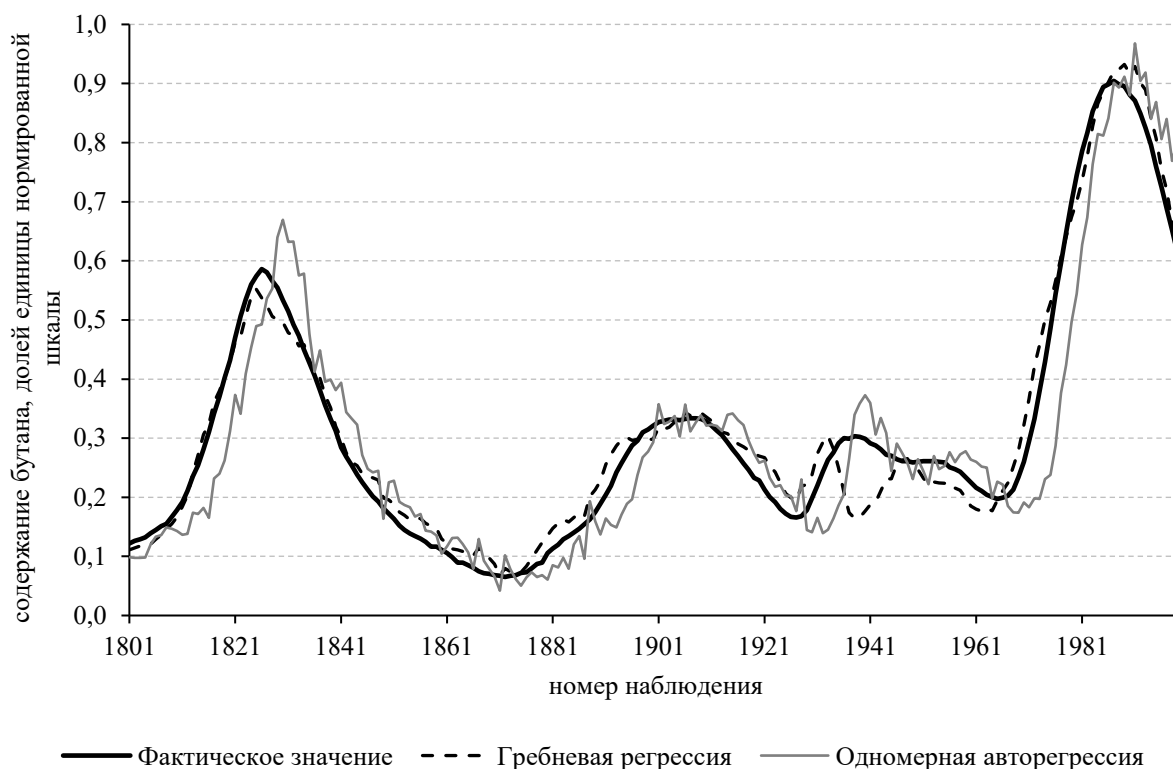


Рисунок 3. Фактическое содержание бутана в кубовом продукте и его упреждающие оценки при горизонте в восемь отсчётов на фрагменте контрольной части набора данных (наблюдения 1801–2000), долей единицы нормированной шкалы (рассчитано авторами на основе набора данных колонны дебутанизации, опубликованного в приложении к монографии Л. Фортуны, С. Грациани, А. Риццо и М. Г. Ксибилии [2])

Выясним, какие технологические переменные обеспечивают выигрыш. Для этого гребневая регрессия оценивалась повторно при поочерёдном исключении всех запаздывающих значений одной переменной. Относительный прирост ошибки приведён в таблице 3.

Таблица 3

Прирост среднеквадратической ошибки гребневой регрессии на контрольной части при исключении отдельной технологической переменной по горизонтам упреждения, процентов

Обозначение	Исключаемая переменная	Горизонт 4 отсчёта	Горизонт 8 отсчётов	Горизонт 12 отсчётов
u_1	Температура верха колонны	2,5	5,1	3,4
u_2	Давление верха колонны	-3,0	-0,5	1,1
u_3	Расход орошения	10,1	10,1	3,7
u_4	Расход продукта на следующую стадию	0,5	-1,0	-3,1
u_5	Температура на шестой тарелке	26,4	51,1	42,8
u_6	Температура куба, датчик А	-0,3	-0,7	0,1
u_7	Температура куба, датчик Б	-0,3	-0,8	-0,3

Рассчитано авторами на основе набора данных колонны дебутанизации, опубликованного в приложении к монографии Л. Фортуны, С. Грациани, А. Риццо и М. Г. Ксибилии. Отрицательное значение означает снижение ошибки при исключении переменной

Из таблицы 3 следует, что информационная ценность переменных крайне неравномерна. Решающую роль играет температура на шестой тарелке. Без неё ошибка при горизонте в восемь отсчётов возрастает на 51,1 %, и выигрыш многомерной схемы в значительной мере

утрачивается. Вторым по значимости выступает расход орошения, третьим температура верха колонны. Исключение обеих температур куба и расхода продукта на следующую стадию ошибку практически не меняет, а в ряде случаев снижает её, что указывает на информационную избыточность этих измерений для данной задачи.

Результат имеет технологическое объяснение. Температура на контрольной тарелке отражает положение фронта концентраций в колонне и реагирует на изменение состава раньше, чем оно проявится в кубовом продукте и будет зафиксировано анализатором. Для практики отсюда следует, что надёжность датчика на контрольной тарелке определяет работоспособность всей схемы и требует резервирования.

Остановимся на том, как упреждающая оценка меняет положение оператора. В контрольной части порог превышен в 232 наблюдениях из 894. Качество предупреждений, рассчитанное по формулам (10) и (11), представлено в таблице 4.

Таблица 4

**Качество раннего предупреждения о превышении
порогового содержания бутана в кубовом продукте на контрольной
части набора данных по информационным схемам, процентов**

Горизонт, отсчётов	Информационная схема	Полнота предупреждений	Точность предупреждений	Доля ложных предупреждений в наблюдениях без превышения
4	Наивная схема	86,2	86,2	4,8
4	Одномерная авторегрессия	93,5	96,0	1,4
4	Гребневая регрессия	99,1	95,8	1,5
4	Нейронная сеть	99,1	94,7	2,0
8	Наивная схема	74,1	74,1	9,1
8	Одномерная авторегрессия	85,8	86,1	4,8
8	Гребневая регрессия	98,7	90,5	3,6
8	Нейронная сеть	98,7	87,1	5,1
12	Наивная схема	62,1	62,1	13,3
12	Одномерная авторегрессия	75,4	73,8	9,4
12	Гребневая регрессия	92,7	85,0	5,7
12	Нейронная сеть	94,4	84,9	5,9

Рассчитано авторами на основе набора данных колонны дебутанизации, опубликованного в приложении к монографии Л. Фортуны, С. Грациани, А. Риццо и М. Г. Ксибили. Порог равен 90-му перцентиллю показателя в обучающей части

При горизонте в восемь отсчётов одномерная авторегрессия заблаговременно выявляет 85,8 % превышений, гребневая регрессия 98,7 %, то есть пропущенными остаются три превышения против тридцати трёх. Точность предупреждений у гребневой регрессии также выше.

На горизонте в двенадцать отсчётов многомерные схемы сохраняют полноту выше 90 %, тогда как наивная схема пропускает более трети превышений. Оператор, располагающий упреждающей оценкой, узнаёт о приближении показателя к границе почти во всех случаях и за время, достаточное для корректировки режима.

Оценим теперь вклад интеллектуальной обработки в достижимое качество управления. По формуле (8) при суммарном запаздывании в восемь отсчётов неустранимая часть дисперсии показателя составляет при одномерной схеме 13,9 %, при многомерной 3,5 % его фактической дисперсии в контрольной части.

Тем самым предел, к которому может быть приближено качество стабилизации, существенно отодвигается. Допустимый сдвиг рабочей точки, рассчитанный по формуле (9) для трёх уровней допустимой доли нарушений, приведён в таблице 5.

Таблица 5

**Допустимый сдвиг рабочей точки к границе спецификации
при переходе от одномерной схемы к лучшей многомерной по горизонтам
упреждения, процентов диапазона нормированной шкалы**

Горизонт, отсчётов	Допустимая доля нарушений 10 %	Допустимая доля нарушений 5 %	Допустимая доля нарушений 1 %
2	0,22	0,29	0,41
4	0,99	1,27	1,79
6	2,44	3,14	4,44
8	4,43	5,69	8,05
10	6,44	8,27	11,70
12	7,68	9,86	13,95
15	9,12	11,70	16,55

Рассчитано авторами на основе набора данных колонны дебутанизации, опубликованного в приложении к монографии Л. Фортуну, С. Грациани, А. Риццо и М. Г. Ксибилии. Квантили стандартного нормального распределения приняты равными 1,282, 1,645 и 2,326

Данные таблицы 5 свидетельствуют о том, что при запаздывании в два отсчёта резерв невелик, а начиная с шести отсчётов становится существенным. При запаздывании в восемь отсчётов и допустимой доле нарушений 5 % рабочая точка может быть смещена на 5,69 % диапазона шкалы. Для колонны дебутанизации это означает возможность вести процесс ближе к предельному содержанию бутана в кубовом продукте и, следовательно, с меньшим расходом тепла в кипятильнике без увеличения доли некондиционного продукта. Денежная оценка эффекта требует данных о ценах и энергозатратах конкретной установки и в настоящей работе не выполнялась. По обзору М. Бауэра и И. Крейга основная часть выгоды от усовершенствованного управления формируется сдвигом рабочей точки, и полученные оценки свидетельствуют о том, что интеллектуальная обработка производственных данных расширяет этот резерв ещё до изменения алгоритма регулирования.

Полученные оценки имеют ограничения. Граница (7) характеризует потенциал, достижимый регулятором с минимальной дисперсией, и реальный регулятор реализует его частично. Часть технологических переменных, включённых в модель, сама находится под управлением, поэтому при изменении алгоритма регулирования найденные зависимости потребуют переоценки. Расчёты выполнены на данных одной установки, и перенос количественных результатов на другие процессы возможен только после повторной идентификации. Вместе с тем качественные выводы согласуются с практикой. С. Дж. Цинь и Т. Бэджвелл в обзоре промышленных систем управления с прогнозирующей моделью отмечали, что преобладающая их часть основана на линейных моделях, идентифицированных по данным [20], а В. М. Дозорцев, Э. Л. Ицкович и Д. В. Кнеллер рассматривают виртуальные анализаторы как составную часть таких систем на отечественных предприятиях. Наш результат дополняет эти наблюдения количественной мерой, связывающей состав обрабатываемых производственных данных с пределом качества управления.

Заключение

Проведённое исследование позволяет сформулировать следующие выводы. Производственные данные колонны содержат сведения о будущем показании анализатора, поскольку технологические переменные опережают его на время, сопоставимое с запаздыванием измерения. Интеллектуальная обработка этих данных, поставленная как упреждающая оценка в приращениях со сдвигом всех регрессоров на горизонт, сокращает ошибку относительно одномерной схемы тем сильнее, чем длиннее запаздывание, и на средних горизонтах уменьшает её примерно наполовину. Выигрыш обеспечен составом и временной структурой данных в большей мере, чем сложностью метода.

Регуляризованная линейная модель не уступила нейронной сети и превзошла градиентный бустинг, типовая фильтрация выбросов ухудшила результат, а основная часть информации сосредоточена в одной переменной, температуре на контрольной тарелке. Трактовка ошибки упреждающей оценки как нижней границы разброса регулируемой величины позволила выразить эффект в терминах управления. Неустраняемая часть дисперсии показателя существенно сокращается, рабочая точка может быть приближена к границе спецификации, а оператор получает предупреждение о выходе показателя за порог почти во всех случаях.

В практическом плане результаты обосновывают три рекомендации. При проектировании виртуальных анализаторов следует начинать с оценки опережения технологических переменных относительно анализатора и формировать регрессоры с учётом этого опережения. Процедуры очистки данных и сложные модели целесообразно принимать только по итогам перекрёстной проверки с сохранением хронологии. При обосновании проектов усовершенствованного управления эффект интеллектуальной обработки данных следует измерять сокращением нижней границы разброса и сдвигом рабочей точки, что делает его сопоставимым с эффектом от изменения алгоритма регулирования. Перспективными направлениями исследования могут стать замкнутое моделирование контура с упреждающей оценкой, адаптация модели при смене режима и стоимостная оценка сдвига рабочей точки по данным конкретных установок.

ЛИТЕРАТУРА

1. Kadlec, P. Data-driven Soft Sensors in the process industry / P. Kadlec, B. Gabrys, S. Strandt — DOI: 10.1016/j.compchemeng.2008.12.012 // Computers & Chemical Engineering. — 2009. — Т. 33, № 4. — С. 795–814.
2. Fortuna, L. Soft Sensors for Monitoring and Control of Industrial Processes / L. Fortuna, S. Graziani, A. Rizzo, M. G. Xibilia. — London: Springer, 2007. — DOI: 10.1007/978-1-84628-480-9.
3. Вахтадзе, Н. Н. Виртуальные анализаторы (идентификационный подход) / Н. Н. Вахтадзе // Автоматика и телемеханика. — 2004. — № 11. — С. 3–24. — EDN: NQXARF.
4. Qin, S.J. Process data analytics in the era of big data / S.J. Qin — DOI: 10.1002/aic.14523 // AIChE Journal. — 2014. — Т. 60, №. 9. — С. 3092–3100.
5. Ge, Z. Data mining and analytics in the process industry: the role of machine learning / Z. Ge, Z. Song, S.X. Ding, B. Huang — DOI: 10.1109/ACCESS.2017.2756872 // IEEE Access. — 2017. — Т. 5. — С. 20590–20616.
6. Harris, T.J. Assessment of control loop performance / T.J. Harris — DOI: 10.1002/cjce.5450670519 // The Canadian Journal of Chemical Engineering. — 1989. — Т. 67, № 5. — С. 856–861.
7. Jelali, M. An overview of control performance assessment technology and industrial applications / M. Jelali — DOI: 10.1016/j.conengprac.2005.11.005 // Control Engineering Practice. — 2006. — Т. 14, №. 5. — С. 441–466.
8. Bauer, M. Economic assessment of advanced process control — A survey and framework / M. Bauer, I.K. Craig — DOI: 10.1016/j.jprocont.2007.05.007 // Journal of Process Control. — 2008. — Т. 18, №. 1. — С. 2–18.
9. Дозорцев, В. М. Усовершенствованное управление технологическими процессами (APC): 10 лет в России / В. М. Дозорцев, Э. Л. Ицкович, Д. В. Кнеллер // Автоматизация в промышленности. — 2013. — № 1. — С. 12–19. — EDN: PNCDID.

10. Fortuna, L. Soft sensors for product quality monitoring in debutanizer distillation columns / L. Fortuna, S. Graziani, M.G. Xibilia — DOI: 10.1016/j.conengprac.2004.04.013 // Control Engineering Practice. — 2005. — T. 13, №. 4. — C. 499–508.
11. Bergmeir, C. On the use of cross-validation for time series predictor evaluation / C. Bergmeir, J.M. Benítez — DOI: 10.1016/j.ins.2011.12.028 // Information Sciences. — 2012. — T. 191. — C. 192–213.
12. Box, G.E.P. Time Series Analysis: Forecasting and Control / G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung. — 5th ed. — Hoboken: John Wiley & Sons, 2015. — URL: <https://books.google.ru/books?id=EjXHCQAAQBAJ&hl=ru>.
13. Hoerl, A.E. Ridge regression: biased estimation for nonorthogonal problems / A.E. Hoerl, R.W. Kennard — DOI: 10.1080/00401706.1970.10488634 // Technometrics. — 1970. — T. 12, № 1. — C. 55–67.
14. Wold, S. PLS-regression: a basic tool of chemometrics / S. Wold, M. Sjöström, L. Eriksson — DOI: 10.1016/S0169-7439(01)00155-1 // Chemometrics and Intelligent Laboratory Systems. — 2001. — T. 58, №. 2. — C. 109–130.
15. Friedman, J.H. Greedy function approximation: a gradient boosting machine / J.H. Friedman — DOI: 10.1214/aos/1013203451 // The Annals of Statistics. — 2001. — T. 29, № 5. — C. 1189–1232.
16. Rumelhart, D.E. Learning representations by back-propagating errors / D.E. Rumelhart, G.E. Hinton, R.J. Williams — DOI: 10.1038/323533a0 // Nature. — 1986. — T. 323. — C. 533–536.
17. Pearson, R.K. Outliers in process modeling and identification / R.K. Pearson // IEEE Transactions on Control Systems Technology. — 2002. — T. 10, № 1. — C. 55–63.
18. Diebold, F.X. Comparing predictive accuracy / F.X. Diebold, R.S. Mariano — DOI: 10.1080/07350015.1995.10524599 // Journal of Business & Economic Statistics. — 1995. — T. 13, № 3. — C. 253–263.
19. Latour, P.L. Estimating benefits from advanced control / P.L. Latour, J.H. Sharpe, M.C. Delaney // ISA Transactions. — 1986. — T. 25. — C. 13–21.
20. Qin, S.J. A survey of industrial model predictive control technology / S.J. Qin, T.A. Badgwell — DOI: 10.1016/S0967-0661(02)00186-7 // Control Engineering Practice. — 2003. — T. 11, № 7. — C. 733–764.

Azizmohammadi Mohammad-Safar Mohsenovich

K.G. Razumovsky Moscow State University of Technologies and Management, Moscow, Russia
Email: safar2002@yandex.ru

Goncharov Andrey Vitalievich

K.G. Razumovsky Moscow State University of Technologies and Management, Moscow, Russia
E-mail: a.goncharov@mgut.ru

Shakhovskoy Andrey Vladimirovich

K.G. Razumovsky Moscow State University of Technologies and Management, Moscow, Russia
E-mail: a.shakhovskoy@mgut.ru

Rotanov Evgeny Gennadievich

K.G. Razumovsky Moscow State University of Technologies and Management, Moscow, Russia
E-mail: e.rotanov@mgut.ru

Improving the efficiency of technological process control based on intelligent processing of production data

Abstract. Distributed control systems of continuous production accumulate large volumes of production data, yet product quality indicators are still measured by on-line analysers with a delay, so operating decisions rely on outdated information. The paper proposes and tests an approach in which the result of intelligent processing of production data is related to the efficiency of technological process control, while model accuracy serves as an intermediate characteristic. The empirical base is an open industrial data set of a refinery debutanizer column comprising seven process variables and the butane content of the bottom product. The task is stated as an anticipatory estimate of the analyser reading at a given horizon from the information available at the moment of decision. A naive scheme, a univariate autoregression and four multivariate models, namely ridge regression, projection to latent structures, gradient boosting and a neural network, are compared under a chronological split of observations and expanding-window cross-validation. Process variables are shown to lead the analyser reading, and their inclusion reduces the error of the anticipatory estimate the more, the longer the horizon, cutting it by about half relative to the univariate scheme at medium horizons. The regularised linear model performs no worse than the nonlinear ones, whereas standard outlier filtering worsens the result and is rejected by cross-validation. The main contribution to accuracy comes from the temperature on the control tray, while some variables are informationally redundant. Based on the known result on the lower bound of the variance of a controlled variable, the error of the anticipatory estimate is interpreted as the limit of attainable stabilisation quality. This makes it possible to express the effect through the reduction of the irreducible variability of the indicator, the admissible shift of the operating point towards the specification limit and the completeness of early warning of threshold exceedance. The results can be used to justify advanced process control projects and to select data for soft sensors.

Keywords: technological process control; intelligent data processing; production data; soft sensor; artificial intelligence; machine learning; measurement delay; quality indicator; distillation