

УДК 66.012-52

ПРИМЕНЕНИЕ ФИЗИЧЕСКИ ОБОСНОВАННОЙ МАТЕМАТИЧЕСКОЙ МОДЕЛИ МАССООБМЕННОГО ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ОЦЕНИВАНИЯ КАЧЕСТВА КОНЕЧНОГО ПРОДУКТА

© 2022 г. С. А. Самотылова^а, *, А. Ю. Торгашов^а

^аИнститут автоматизации и процессов управления ДВО РАН, Владивосток, Россия

*e-mail: samotylova@dvo.ru

Поступила в редакцию 19.08.2021 г.

После доработки 20.12.2021 г.

Принята к публикации 30.12.2021 г.

Рассматривается задача повышения точности оценивания качества конечного продукта в условиях малого набора обучающего сегмента данных, который расширяется с использованием физически обоснованной модели (first principles model) массообменного технологического процесса производства метил-трет-бутилового эфира (МТБЭ). Показано, что предлагаемый подход превосходит другие методы построения статистических моделей (виртуальных анализаторов) для задачи оценивания качества конечных продуктов, так как позволяет учитывать физически обоснованные зависимости и характеристики технологического процесса, что в итоге приводит к повышению уровня адекватности разрабатываемой модели. Определены условия целесообразности применения физически обоснованной модели массообменного технологического процесса в алгоритме при построении статистических моделей для оценки показателя качества конечного продукта в условиях параметрической неопределенности эффективности массопереноса по Мерффри и параметров бинарного взаимодействия между димерами изобутилена и компонентами из системы производства МТБЭ. В качестве критерия эффективности расширения обучающей выборки предлагается использовать площадь области пересечения распределений выходной переменной между ее значениями в расширенной обучающей и тестовой выборках.

Ключевые слова: модель для оценки качества конечного продукта, виртуальный анализатор, массообменный технологический процесс, физически обоснованная математическая модель, малая обучающая выборка, расширение обучающей выборки, метод проекций на латентные (скрытые) структуры

DOI: 10.31857/S0040357122020117

ВВЕДЕНИЕ

Управление технологическими процессами (ТП) с целью повышения качества конечных продуктов является одной из главных задач в промышленности. Постоянный контроль качества выпускаемой продукции играет определяющую роль в управлении, минимизации издержек производства и повышении безопасности. Учитывая тенденцию к использованию систем усовершенствованного управления на основе статистических моделей для оценки показателей качества (виртуальных анализаторов) в реальном времени, совершенствование и разработка новых подходов для их построения в условиях малых обучающих выборок представляют собой актуальную тему исследования [1–5]. Под виртуальными анализаторами (ВА) понимаются различные статистические математические модели (регрессионные,

нейронные сети, нейро-нечеткие модели и т.п.), которые используются для оценки определенных физических величин или качества продукции в производственных процессах на основе доступных измерений [6, 7]. Для построения ВА используются выборки исторических данных, содержащие значения целевого показателя качества, так и измеряемых технологических переменных процесса, коррелирующих с ним [8, 9]. ВА создают информационную базу для формирования оптимального управления как отдельными химико-технологическими процессами, так и всем циклом технологического производства.

Используемые для построения ВА методы были существенно улучшены, однако, в большинстве случаев они сосредоточены на подходах, основанных на глубоком обучении [10–12] и регрессии [13]. Следует отметить, что при разработке и

внедрении ВА в системы усовершенствованного управления технологическими процессами для оперативного контроля качества конечных продуктов, одной из важнейших проблем, с которой сталкиваются, — это репрезентативность обучающей выборки исторических данных [14–17]. От размера обучающей выборки и ее характеристик будет зависеть точность модели оценки показателей качества конечных продуктов и эффективность применения алгоритмов и методов для ее разработки [18, 19].

Высокая стоимость лабораторных исследований, нелинейность технологического процесса, наличие случайного шума, отсутствие данных технологического режима на всем диапазоне изменения качества продукта приводит к неэффективности применяемых методов при разработке модели (ВА) и снижение ее работоспособности на другом диапазоне функционирования установки. В связи с этим задача построения адекватной модели для оценки качества конечных продуктов в условиях малой выборки данных остается актуальной [18–20].

При построении ВА (или моделей) для оценки показателей качества конечных продуктов в условиях малых выборок широкое распространение получили бутстрэп метод [21] и метод складного ножа [22]. Данные методы позволяют оценить параметры ВА путем формирования новых выборок из исходного набора данных. Используемые методы отличаются друг от друга подходом к формированию выборок. В первом случае формируются бутстрэп выборки путем извлечения элементов из исходного набора данных случайным образом несколько раз. При использовании концепции складного ножа формируются подвыборки складного ножа путем удаления i -ой строки из исходного набора данных.

В [23, 24] рассмотрен подход к работе с малым исходным набором данных на основе генерации виртуальной выборки методом оптимизации роя частиц. Этот метод заполняет “пропуски” между значениями данных выборки за счет добавления новых сгенерированных виртуальных наблюдений. Применение вышеуказанных методов эффективно, если обучающая выборка содержит в себе данные технологического процесса на всем диапазоне изменения качества конечного продукта.

В настоящей работе предлагается алгоритм построения модели (виртуального анализатора) для оценки качества конечного продукта (МТБЭ), отличающийся от известных подходов к построению моделей в условиях малых выборок тем, что основан на расширении обучающей выборки или повышению ее репрезентативности путем добавления данных, полученных с использованием физически обоснованной математической модели технологического процесса.

ОПИСАНИЕ ПРОМЫШЛЕННОГО ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА И ПОСТАНОВКА ЗАДАЧИ

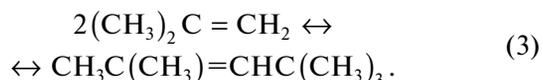
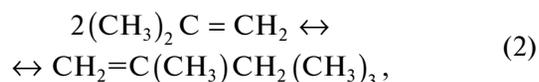
Рассматривается массообменный технологический процесс получения высокооктановой добавки бензинов — метил-трет-бутилового эфира (МТБЭ). Схема процесса производства МТБЭ представлена на рис. 1.

При производстве МТБЭ особое внимание уделяется выбору соотношения изобутилен-метанол. Количество подаваемого метанола определяется в зависимости от расхода бутан-бутиленовой фракции (ББФ) и содержания изобутилена в ней. 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 массовая доля метил-трет-бутилового эфира марки А должна быть не менее 98% в конечном продукте, а массовая доля метанола (MeOH) не должна превышать 1.5%.

Основная химическая реакция (образование МТБЭ) в реакционной зоне протекает в результате взаимодействия изобутилена (IB), содержащегося в ББФ, с метанолом на сульфокатионитном катализаторе:



В условиях недостаточного количества метанола изобутилен димеризуется с образованием побочного продукта диизобутена (DIB), который состоит из изомеров 2,4,4-триметилпентена-1 (2,4,4-ТМП-1) и 2,4,4-триметилпентена-2 (2,4,4-ТМП-2) [25, 26].



При построении моделей для оценки качества МТБЭ в обучающей выборке в реальных условиях часто наблюдается полное отсутствие данных технологического режима в области верхней границы диапазона качества продуктов (приближения параметров объекта управления к границам, обозначенным технологическим регламентом) либо частичное отсутствие данных в области всего диапазона технологического режима (рис. 2).

Отсутствие информации по качеству конечных продуктов, соответствующей текущему режиму технологического процесса, вынуждает операторов поддерживать режимы, обеспечивающие

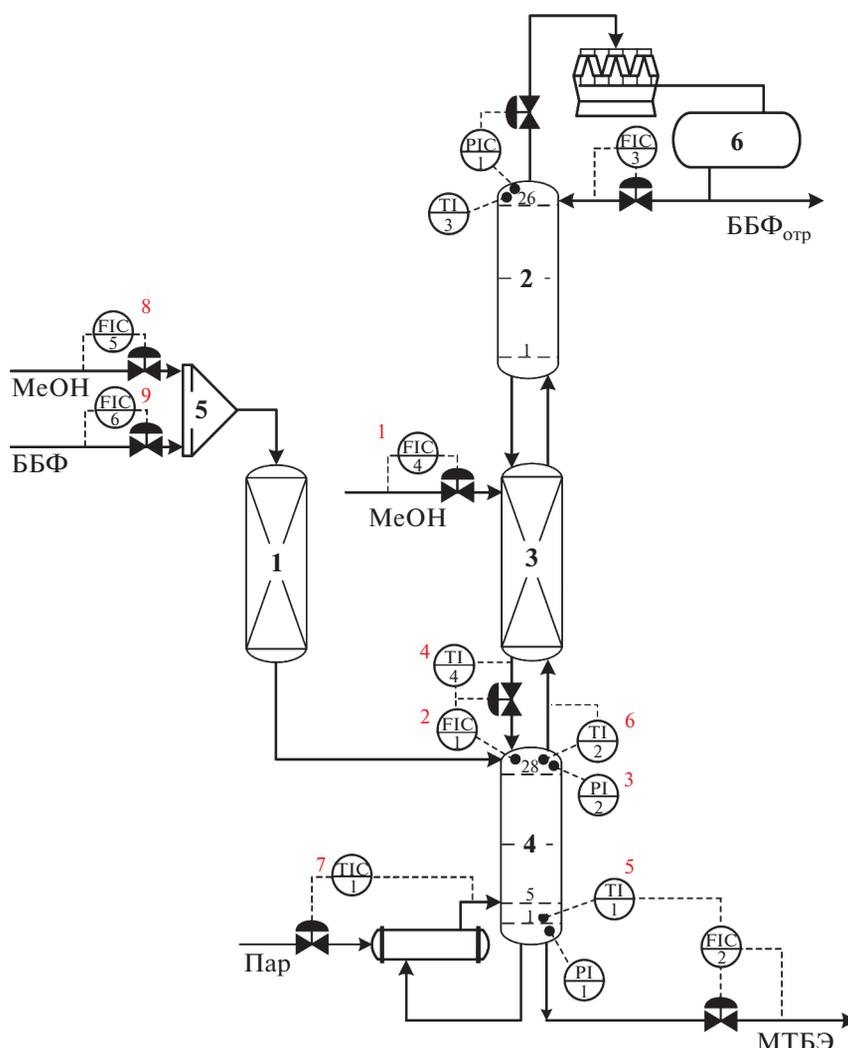


Рис. 1. Массообменный технологический процесс производства метил-трет-бутилового эфира: 1 – реактор предварительного синтеза; 2 – верхняя ректификационная колонна; 3 – реакционно-ректификационный аппарат; 4 – нижняя ректификационная колонна; 5 – смеситель; 6 – емкость; ББФ – бутан-бутиленовая фракция; ББФ_{отр} – отработанная бутан-бутиленовая фракция; MeOH – метанол; МТБЭ – метил-трет-бутиловый эфир.

большой запас по качеству продуктов. Тем самым повышается расход энергии и увеличивается стоимость конечного продукта. В связи с этим рассматривается задача построения модели для оценки содержания метанола в конечном (нижнем) продукте производства МТБЭ в условиях малой обучающей выборки с отсутствием данных технологического режима во всем диапазоне изменения качества продуктов.

Пусть N обозначает количество наблюдений, а n_U и n_Y обозначают количество входных и выходных переменных, $U \in R^{N \times n_U}$ и $Y \in R^{N \times n_Y}$. Для обучающего набора данных $U \in R^{N \times n_U}$, $Y \in R^{N \times 1}$ задача состоит в том, чтобы построить некоторую функцию f , связывающую вход и выход. Таким

образом модель для оценки показателя качества конечного продукта (\hat{Y}) в общем виде имеет вид:

$$\hat{Y} = f(U, \hat{\beta}). \tag{4}$$

В качестве входных переменных выбираются наиболее коррелируемые с выходом (концентрация метанола в МТБЭ) следующие переменные: $u_{(1)}$ – расход метанола в реактор, м³/ч (FIC4); $u_{(2)}$ – реакционная масса в нижнюю ректификационную колонну, м³/ч (FIC1); $u_{(3)}$ – давление вверху нижней ректификационной колонны, кгс/см² (PI2); $u_{(4)}$ – температура ректификационной смеси, поступающая в нижнюю ректификационную колонну, °С (TI4); $u_{(5)}$ – температура куба нижней ректификационной колонны, °С (TI1); $u_{(6)}$ – температура углеводородной смеси, отходящей свер-

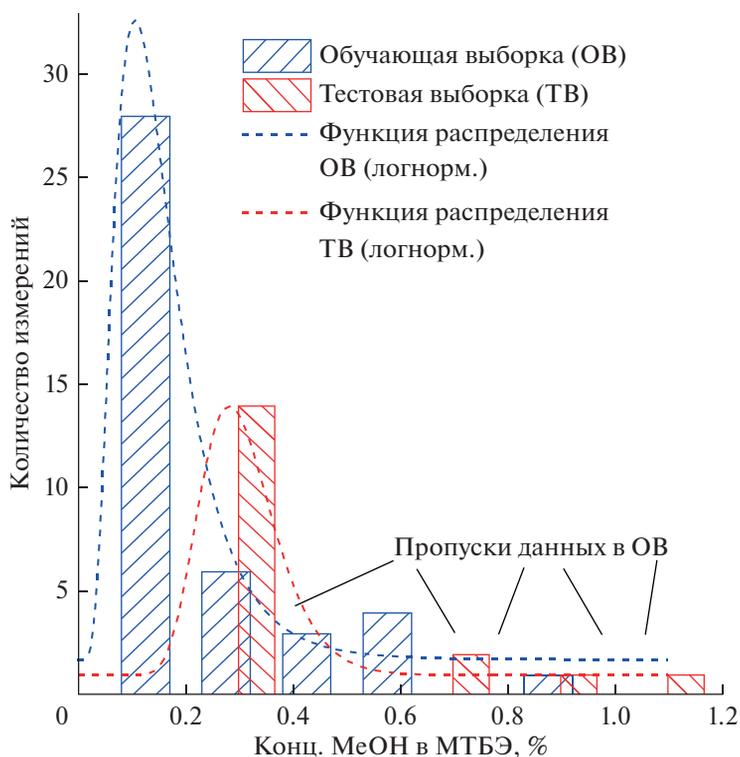


Рис. 2. Экспериментальные данные по содержанию метанола в МТБЭ.

ху нижней ректификационной колонны, °С (Т12); $u_{(7)}$ – температура на 5-ой тарелке нижней ректификационной колонны, °С (Т1С1); $u_{(8)}$ – расход метанола в смеситель, м³/ч (FIC5); $u_{(9)}$ – расход бутан-бутиленовой фракции в смеситель, м³/ч (FIC6); $\hat{\beta}$ – вектор параметров модели.

Общий размер исходной обучающей выборки составляет $N_{tr.s} = 42$. Для оценки эффективности предлагаемого алгоритма использовалась тестовая (контрольная) выборка размером $N_{te.s} = 18$.

Эффективность построенных моделей оценивается с помощью коэффициента детерминации R^2 и абсолютной средней ошибки:

$$R^2 = 1 - \frac{\sum_{n=1}^N (Y_n - \hat{Y}_n)^2}{\sum_{n=1}^N (Y_n - \bar{Y})^2}; \quad (5)$$

$$MAE = \frac{1}{N} \sum_{n=1}^N |Y_n - \hat{Y}_n|, \quad (6)$$

где Y_n – измеряемое значение выходной переменной, мас. %; \hat{Y}_n – ее значение, полученное на основе модели для оценки показателя качества конечного продукта, мас. %; \bar{Y} – среднее значение наблюдаемой выходной переменной, мас. %.

ИСПОЛЬЗУЕМЫЕ МЕТОДЫ ПОСТРОЕНИЯ МОДЕЛЕЙ ДЛЯ ОЦЕНКИ ПОКАЗАТЕЛЕЙ КАЧЕСТВА КОНЕЧНЫХ ПРОДУКТОВ

Применение того или иного метода при построении ВА в условиях малой выборки ограничено такими факторами как зашумленностью исходных данных, проблемой выбора структуры модели, наличием сильной корреляции между входами, нелинейностью технологических процессов, переобучением модели и т.д. Оценить оптимальность применения того или иного метода можно по критериям точности полученных моделей [см. (5) и (6)]. Широкое распространение в промышленности получили проекционные методы. Они позволяют выделить латентные (скрытые) переменные, определить эффективную размерность данных, удалить их шумовую составляющую и, как следствие, снизить риск переобучения моделей и тем самым улучшить их качество [27]. В связи с этим для построения моделей оценки целевого показателя качества используется метод проекций на латентные структуры [28–31].

Метод проекций на латентные структуры

Оценка $\hat{\beta}_{ПЛС}$ с помощью метода проекций на латентные структуры (ПЛС) для A главных компонент осуществляется как: $\hat{\beta}_{ПЛС} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{Q}$, где \mathbf{P} – матрица нагрузок U , \mathbf{Q} – матрица нагрузок

Y, W – матрица весов U, T – операция транспонирования, \bar{U}, \bar{Y} – векторы средних значений матриц U и Y . Расчет $\hat{\beta}_{\text{ПЛС}}$ происходит итерационно для каждой главной компоненты.

Шаг 1: для исходной выборки данных входных $U \in R^{N \times n_U}$ и выходной $Y \in R^{N \times 1}$ переменных, переменные U и Y центрируются: $U_0 = U - 1_{N \times n_U} \bar{U}$, $Y_0 = Y - 1_{N \times 1} \bar{Y}$, где $1_{N \times 1}$ единичный вектор. На этом шаге также выбирается максимальное значение главных компонент $\mathbf{a} = \overline{1, \mathbf{A}}$ ($\mathbf{A} = n_U$).

Шаг 2: вычисляется вектор взвешенных нагрузок \mathbf{w}_a по формуле:

$$\mathbf{w}_a = \mathbf{c} U_{a-1}^T Y_{a-1}, \quad (7)$$

где \mathbf{c} – это нормировочный фактор $\mathbf{c} = (Y_{a-1}^T U_{a-1} U_{a-1}^T Y_{a-1})^{-0.5}$.

Шаг 3: определяются счета \mathbf{t}_a как:

$$\mathbf{t}_a = U_{a-1} \mathbf{w}_a. \quad (8)$$

Шаг 4: вычисляются “спектральные” и “химические” нагрузки по следующим формулам [см. (9) и (10)]:

$$\mathbf{p}_a = \frac{U_{a-1}^T \mathbf{t}_a}{\mathbf{t}_{a-1}^T \mathbf{t}_a}, \quad (9)$$

$$\mathbf{q}_a = \frac{Y_{a-1}^T \mathbf{t}_{a-1}}{\mathbf{t}_{a-1}^T \mathbf{t}_{a-1}}. \quad (10)$$

Шаг 5: новые остатки U и Y получаются вычитанием вклада главной компоненты: $U_a = U_{a-1} - \mathbf{t}_a \mathbf{p}_a^T$, $Y_a = Y_{a-1} - \mathbf{t}_a \mathbf{q}_a^T$.

Величины U_{a-1} и Y_{a-1} заменяются новыми значениями U_a , Y_a и увеличиваем номер главной компоненты на единицу и возвращаемся на шаг 2. Итерационный алгоритм повторяется пока номер главной компоненты не достигнет максимальной величины.

Выход статистической модели можно получить как: $\hat{Y} = U \hat{\beta}_{\text{ПЛС}}$.

Метод проекций на латентные структуры с использованием функции ядра

Можно получить модель нелинейной регрессии в пространстве исходных входных переменных [32–34]. Для этого сделаем нелинейное преобразование входных переменных $U \in R^{N \times n_U}$ в пространстве признаков \mathbf{F} ; т.е. отображим $\Phi: U \in R^{N \times n_U} \rightarrow \Phi(U) \in \mathbf{F}$. Матрица Грама \mathbf{K} из компонент Φ вычисляется как: $\mathbf{K} = \Phi \Phi^T \in R^{N \times N}$.

Используя матрицу Φ отображенных входных данных и объединив шаги 2 и 3 предыдущего алгоритма, получаем алгоритм для (нелинейной) ядерной модели проекций на скрытые структуры – НПЛС. В общем виде алгоритм можно описать шагами, представленными ниже.

Шаг 1: вычисляется матрица входных данных Φ .

Шаг 2: определяются счета \mathbf{t}_a как:

$$\mathbf{t}_a = \mathbf{c} \Phi_{a-1} \Phi_{a-1}^T Y_{a-1}, \quad (11)$$

где \mathbf{c} – это нормировочный фактор $\mathbf{c} = (Y_{a-1}^T \Phi_{a-1} \Phi_{a-1}^T Y_{a-1})^{-0.5}$.

Шаг 3: вычисляются “спектральные” и “химические” нагрузки [см. (12) и (13)]:

$$\mathbf{p}_a = \frac{\Phi_{a-1} \Phi_{a-1}^T \mathbf{t}_a}{\mathbf{t}_{a-1}^T \mathbf{t}_a}, \quad (12)$$

$$\mathbf{q}_a = \frac{Y_{a-1}^T \mathbf{t}_{a-1}}{\mathbf{t}_{a-1}^T \mathbf{t}_{a-1}}. \quad (13)$$

Шаг 4: новые остатки U и Y получаются вычитанием вклада главной компоненты: $\Phi_a \Phi_a^T = (\Phi_{a-1} - \mathbf{t}_a \mathbf{p}_a^T)(\Phi_{a-1} - \mathbf{t}_a \mathbf{p}_a^T)^T$, $Y_a = Y_{a-1} - \mathbf{t}_a \mathbf{q}_a^T$.

Величины $\Phi_{a-1} \Phi_{a-1}^T$ и Y_{a-1} заменяются новыми значениями $\Phi_a \Phi_a^T$, Y_a и увеличиваем номер главной компоненты на единицу и возвращаемся на шаг 2. Итерационный алгоритм повторяется пока номер главной компоненты не достигнет максимальной величины.

В таком случае оценка $\hat{\beta}_{\text{НПЛС}}$ осуществляется как: $\hat{\beta}_{\text{НПЛС}} = \mathbf{Y}(\mathbf{T}^T \mathbf{K} \mathbf{Y})^{-1} \mathbf{T}^T \mathbf{Y}$, где $\mathbf{K} = \Phi(U) \Phi(U)^T = \text{ker}(u_i, u_j)$. Функция ядра $\text{ker}(u_i, u_j)$ с ее (\tilde{i}, \tilde{j}) -м элементом определяется формулами:

$$\text{ker}(u_i, u_j) = (u_i^T u_j + 1)^p, \quad (14)$$

$$\text{ker}(u_i, u_j) = \exp(-\|u_i - u_j\|^2 / 2\sigma^2). \quad (15)$$

Функция ядра [см. (14), (15)] зависит от выбора параметров p и σ , соответственно, и влияет на предсказательную способность метода на основе выбранной функции ядра.

Выход модели ВА, полученный методом НПЛС с функцией ядра, можно записать как $\hat{Y} = \mathbf{K} \hat{\beta}_{\text{НПЛС}}$. Для тестовой выборки оценка выходного параметра будет осуществляться в соответствии с функцией ядра, полученной на тестовой выборке: $\hat{Y}_{\text{te.s}} = \mathbf{K}_{\text{te.s}} \hat{\beta}_{\text{НПЛС}}$.

**АЛГОРИТМ РАСШИРЕНИЯ ВЫБОРКИ
ИСТОРИЧЕСКИХ ДАННЫХ
ПРИ ПОСТРОЕНИИ МОДЕЛЕЙ
ДЛЯ ОЦЕНКИ ЦЕЛЕВОГО ПОКАЗАТЕЛЯ
КАЧЕСТВА С ИСПОЛЬЗОВАНИЕМ
ОПТИМИЗАЦИИ РОЯ ЧАСТИЦ**

Метод оптимизации роя частиц (Particle swarm optimization algorithm – PSO) представляет собой алгоритм глобальной оптимизации, использующий имитацию социального поведения роя – группы некоторых объектов, применение которого не требует знания точного градиента оптимизируемой функции [35–37]. В алгоритме рой представляет собой множество частиц, которые на каждой итерации имеют две переменные состояния: ее текущее положение $u_{vir,d}^t$ и текущую скорость $v_{vir,d}^t$, где t – счетчик итераций, их позиции на каждой итерации обновляются для достижения решения. На каждой итерации для каждой частицы вычисляется оптимальное значение целевой функции, новое положение частицы $pbest_{vir,d}$, которое соответствует лучшему ее значению на итерации в этот момент, и $gbest_{vir,d}$ – лучшие значения, которые получили все частицы на итерации в предыдущий момент. Частицы в рое устремляются к $gbest_{vir,d}$, т.е. к лучшим значениям, обнаруженным в рое, с положением $pbest_{vir,d}$. Скорость и положение частицы можно рассчитать следующим образом [22, 23]:

$$v_{vir,d}^{t+1} = w^t v_{vir,d}^t + 2r_1 (pbest_{vir,d}^t - u_{vir,d}^t) + 2r_2 (gbest_{vir,d}^t - u_{vir,d}^t), \quad (16)$$

$$u_{vir,d}^{t+1} = u_{vir,d}^t + v_{vir,d}^t, \quad (17)$$

$$vir_i = 1, 2, \dots, n_U, \quad d = 1, 2, \dots, n_p.$$

где n_U – количество переменных, n_p – размер частиц в рое, r_1 и r_2 – два случайных числа в диапазоне $[0, 1]$, w^t – инерционный вес на итерации t , предназначенный для контроля влияния предыдущей скорости на текущую $w^t = w_{start} - t(w_{start} - w_{end}/t_{max})$, где t_{max} – максимальное количество итераций, w_{start} и w_{end} – начальное и конечное значения инерционного веса, равные 0.9 и 0.4.

Новые значения частиц вычисляются после обновления скорости и положения, при необходимости также обновляются $pbest_{vir,d}$ и $gbest_{vir,d}$. Процедура выполняется до достижения заданного количества итераций и количества частиц в рое.

Шаги работы алгоритма для расширения обучающей выборки с использованием метода PSO представлены следующим образом:

Шаг 1: строится модель методом проекций на латентные структуры на основе заданной обучающей выборки и тестируется с помощью набора данных тестирования.

Шаг 2: формируется виртуальная выборка следующим образом:

(а) инициализация параметров PSO алгоритма: максимальное количество итераций t_{max} и размер частиц в рое n_p ;

(б) случайным образом генерируются начальные скорости каждой из частиц $v_{vir,d}^1$ и устанавливаются соответствующие векторы каждой частицы $u_{vir,d}^1$;

(в) вычисление минимальной среднеквадратичной ошибки качества функции для нахождения наилучших значений – $gbest_{vir,d}$;

(г) обновление скорости роя частиц $v_{vir,d}^{t+1}$ [см. (16)] и определение нового положения частицы $u_{vir,d}^{t+1}$ [см. (17)];

(д) при достижении максимального количества итераций t_{max} вывести оптимальные значения параметров $pbest_{vir,d}$ и $gbest_{vir,d}$.

Шаг 3: после достижения заданного на шаге 1 количества итераций t_{max} исходный обучающий набор и полученный на шаге 2 (д) набор виртуальных данных $gbest_{vir,d}$ объединяются в новый обучающий набор данных.

Шаг 4: расчет параметров модели для оценки показателя качества конечного продукта с использованием нового (расширенного) набора обучающих данных (совокупность данных обучающей выборки и данных $gbest_{vir,d}$) и проверка модели на тестовой выборке.

Одной из проблем применения PSO является выбор значений параметров r_1 , r_2 , t_{max} , n_p . При их некорректных значениях могут наблюдаться значительные колебания $u_{vir,d}$, что снижает эффективность применения метода оптимизации роя частиц.

**ПРЕДЛАГАЕМЫЙ ПОДХОД К РЕШЕНИЮ
ЗАДАЧИ В УСЛОВИЯХ МАЛОЙ ВЫБОРКИ
ИСТОРИЧЕСКИХ ДАННЫХ
С ИСПОЛЬЗОВАНИЕМ ФИЗИЧЕСКИ
ОБОСНОВАННОЙ
МАТЕМАТИЧЕСКОЙ МОДЕЛИ**

Для решения проблемы малой обучающей выборки предлагается алгоритм ее расширения данными физически обоснованной математической модели массообменного технологического процесса, которая содержит уравнения фазового равновесия, материального и энерге-

тического баланса. На основе такой модели можно получить данные технологического режима во всем диапазоне изменения качества продуктов.

В общем виде структуру физически обоснованной математической модели можно представить в виде системы уравнений на каждой j -ой ступени разделения для каждого i -го компонента:

$$\left\{ \begin{array}{l} L_{j+1}x_{i,j+1} + V_{j-1}y_{i,j-1} + F_jx_{F,i,j} - L_jx_{i,j} - V_jy_{i,j} + \sum_{m=1}^r \sum_{i=1}^{nc} v_{i,m} R_{m,j} \varepsilon_j = 0, \\ L_{j+1}h_{j+1} + V_{j-1}H_{j-1} + F_jH_{F,j} - L_jh_j - V_jH_j = 0, \\ x_{i,j} \gamma_{L,i,j} \frac{p_i^0}{P} - y_{i,j}^* = 0, \\ E_j \left(y_{i,j}^* - y_{i,j+1} \right) - (y_{i,j} - y_{i,j+1}) = 0, \\ \sum_{i=1}^{nc} y_{i,j} - 1 = 0, \\ \sum_{i=1}^{nc} x_{i,j} - 1 = 0, \\ \sum_{i=1}^{nc} y_{i,j}^* - 1 = 0, \\ T_j^L = T_j^V, \\ P_j^L = P_j^V, \end{array} \right. \quad \left(\begin{array}{l} i = 1, \dots, nc \\ j = 1, \dots, nt \end{array} \right),$$

где L_{j+1} – поток жидкости, поступающий на $j + 1$ -ую тарелку, кмоль/ч; $x_{i,j+1}$ – концентрация i -го компонента, поступающего на $j + 1$ -ую тарелку в жидкой фазе, кмоль; V_{j-1} – паровой поток, покидающий $j - 1$ -ую тарелку, кмоль/ч; $y_{i,j-1}$ – концентрация i -го компонента, покидающего $j - 1$ -ую тарелку в паровой фазе, кмоль/кмоль; F_j – расход сырья, поступающего на j -ую тарелку, кмоль/ч; $x_{F,i,j}$ – концентрация i -го компонента в сырье, поступающего на j -ую тарелку, кмоль/кмоль; L_j – поток жидкости на j -ой тарелке, кмоль/ч; $x_{i,j}$ – концентрация i -го компонента на j -ой тарелке в жидкой фазе, кмоль/кмоль; V_j – паровой поток на j -ой тарелке, кмоль/ч; $y_{i,j}$ – концентрация i -го компонента на j -ой тарелке в паровой фазе, кмоль/кмоль; $\gamma_{L,i,j}$ – коэффициент активности i -го компонента в жидкой фазе на j -ой тарелке; p_i^0 – парциальное давление i -го компонента, кПа; P – общее давление в системе, кПа; E_j – эффективность массопереноса по Мерффри на j -ой тарелке; $y_{i,j}^*$ – равновесная концентрация i -го компонента на j -ой тарелке в паровой фазе, кмоль/кмоль; h_{j+1} – энтальпия жидкости, поступающей на $j + 1$ -ую тарелку, кДж/кмоль; H_{j-1} – энтальпия пара, покидающего $j - 1$ -ую тарелку, кДж/кмоль; $H_{F,j}$ –

энтальпия питания на j -ой тарелке, кДж/кмоль; h_j – энтальпия жидкости на j -ой тарелке, кДж/кмоль; H_j – энтальпия пара на j -ой тарелке, кДж/кмоль; nc – общее количество компонентов в системе; nt – общее количество тарелок в ректификационной колонне; m – номер реакции; r – количество реакций; i – номер компонента; j – номер тарелки; $v_{i,m}$ – стехиометрический коэффициент i -го компонента в реакции m ; $R_{m,j}$ – скорость реакции m , произошедшей на j -ой тарелке, кмоль/см² · ч; ε_j – объем реакционной смеси на j -ой тарелке, м³; T_j^L – температура потока жидкости на j -ой тарелке, К; T_j^V – температура потока пара на j -ой тарелке, К; P_j^L – давление потока жидкости на j -ой тарелке, кПа; P_j^V – давление потока пара на j -ой тарелке, кПа.

Скорость химической реакции m , протекающей в реакционной зоне реакционно-ректификационной колонны, на j -ой тарелке определяется как:

$$R_{m,j} = \pm \frac{\Delta n}{s \Delta t},$$

где Δn – изменение количества одного из веществ в реакции, кмоль; s – поверхность соприкосновения фаз, см²; Δt – промежуток времени реакции, ч.

Исходными данными для проведения расчетов на основе физически обоснованной матема-

Таблица 1. Параметры бинарного взаимодействия Λ_{hi} для расчета коэффициента активности в реакционной системе МТБЭ

Компонент 1	Компонент 2			
	MeOH	IB, NB	МТБЭ	DIB
MeOH	0	35.38	88.04	-7.17
IB, NB	-706.34	0	-52.2	25.51
МТБЭ	-468.76	24.63	0	-17.13
DIB	706.38	-19.69	45.75	0

тической модели являются состав перерабатываемого сырья и технологические параметры процесса, такие как температура, давление и расход подаваемого сырья, а также давление в колонне и расход нижнего продукта.

Зависимость коэффициентов активности компонентов в жидкости $\gamma_{L,i,j}$ от температуры и концентрации, согласно модели UNIQUAC [38, 39], может быть представлена в виде уравнений:

$$\ln \gamma_{L,i,j} = \ln \frac{\Phi_{i,j}}{x_{i,j}} + \frac{z}{2} \tilde{q}_{i,j} \ln \frac{\theta_{i,j}}{\Phi_{i,j}} + l_{i,j} - \frac{\Phi_{i,j}}{x_{i,j}} \sum_{h=1}^{nc} x_{h,j} l_{h,j} + \tilde{q}_{i,j} \left(1 - \ln \sum_h \theta_{h,j} \tau_{hi} - \sum_h \frac{\theta_{h,j} \tau_{hi}}{\sum_k \theta_{k,j} \tau_{kh}} \right), \quad (18)$$

$$\theta_{i,j} = \tilde{q}_{i,j} x_{i,j} / \sum_h \tilde{q}_{h,j} x_{h,j}, \quad \Phi_{i,j} = \tilde{r}_{i,j} x_{i,j} / \sum_h \tilde{r}_{h,j} x_{h,j}, \quad (19)$$

$$l_{i,j} = \frac{z}{2} (\tilde{r}_{i,j} - \tilde{q}_{i,j}) - (\tilde{r}_{i,j} - 1), \quad (20)$$

$$\tau_{hi} = \exp(-[(a_{hi} - a_{ii})/\hat{R}T]),$$

где $\Lambda_{hi} = (a_{hi} - a_{ii})$ – параметры бинарного взаимодействия пар $h-i$ и $i-i$, кДж/моль; $\tilde{r}_{i,j}$, $\tilde{q}_{i,j}$ – параметры объема и площади чистого i -го компонента на j -ой тарелке; τ_{hi} – приведенные параметры; \hat{R} – газовая постоянная (8.314 Дж/(моль · К)); T – температура, К; z – координационное число, т.е. число тесно взаимодействующих молекул вокруг центральной молекулы, устанавливается равным 10.

Параметры бинарного взаимодействия взяты из [40] и сведены в табл. 1. Следует отметить, что параметры бинарного взаимодействия димеров изобутилена – 2,4,4-триметилпентена-1 и 2,4,4-триметилпентена-2, практически отсутствуют в литературных источниках. В данной системе н-бутен (NB) выступает в качестве инерта.

Используемая физически обоснованная математическая модель технологического процесса содержит допущения и с термодинамической точки зрения является приближенной к реально-

му процессу. В связи с этим необходимо определить условия целесообразности (адекватности) применения такой модели в условиях неопределенности в алгоритме. Для оценки целесообразности применения физически обоснованной математической модели в алгоритме рассчитываются R^2 , MAE. А также предлагается рассчитывать площадь области пересечения распределений выходной переменной между ее значениями в расширенной обучающей и тестовой выборках: S – критерий эффективности расширения обучающей выборки.

Для расчета площади (S) области пересечения выходной переменной расширенной обучающей (ОБ) и тестовой выборки (ТВ) необходимо определить их функции распределения. Принадлежность данных выборок к каждому распределению проверяется по критерию согласия Колмогорова

$$K(y) [41], \quad D_{\text{ext.s}} = \sup_y \left| F_{\text{ext.s}}^*(y) - F_{\text{ext.s}}(y) \right|, \quad D_{\text{te.s}} = \sup_y \left| F_{\text{te.s}}^*(y) - F_{\text{te.s}}(y) \right|, \quad \text{где } F_{\text{ext.s}}^*(y), F_{\text{te.s}}^*(y) \text{ – эмпирические функции распределения выходной переменной в расширенной ОБ и ТВ; } F_{\text{ext.s}}(y), F_{\text{te.s}}(y) \text{ – их гипотетические функции распределения. В соответствии с этим критерием, функция распределения случайных величин } \sqrt{n_1} D_{\text{ext.s}}, \sqrt{n_2} D_{\text{te.s}} \text{ близка к } K(y). \text{ Поэтому вероятность того, что любая из случайных величин } D_{\text{ext.s}} \text{ окажется больше, чем } 1.36/5.5, \text{ а } D_{\text{te.s}} \text{ больше, чем } 1.36/3.87, \text{ равна } 1 - 0.95 = 0.05.$$

После определения вида плотностей распределения $f_{\text{ext.s}}(y)$ и $f_{\text{te.s}}(y)$ рассчитывается площадь области их пересечения. Для случая, когда $f_{\text{ext.s}}(y)$ пересекается с $f_{\text{te.s}}(y)$ в точке c (при условии, что $f_{\text{ext.s}}(y)$ и $f_{\text{te.s}}(y)$ имеют функции распределения логнормального распределения), S будет рассчитываться как [см. (21)]:

$$S \propto \int_a^c f_{\text{te.s}}(y) dy + \int_c^b f_{\text{ext.s}}(y) dy, \quad (21)$$

где c – точка пересечения $f_{\text{ext.s}}(y)$ и $f_{\text{te.s}}(y)$, $a \leq y \leq b$.

После преобразования имеем [см. (22)]:

$$S \propto \frac{1}{\sigma_{\text{te.s}} \sqrt{2\pi}} \int_a^c \exp\left[-(\ln y - \mu_{\text{te.s}})^2 / 2\sigma_{\text{te.s}}^2\right] dy + \frac{1}{\sigma_{\text{ext.s}} \sqrt{2\pi}} \int_c^b \exp\left[-(\ln y - \mu_{\text{ext.s}})^2 / 2\sigma_{\text{ext.s}}^2\right] dy, \quad (22)$$

где $\mu_{\text{te.s}}$, $\mu_{\text{ext.s}}$ – средние значения $Y_{\text{te.s}}$ и $Y_{\text{ext.s}}$; $\sigma_{\text{te.s}}$, $\sigma_{\text{ext.s}}$ – их стандартные отклонения.

Для случая, когда $f_{te.s}(y)$ содержится в $f_{ext.s}(y)$ (при условии, что $f_{ext.s}(y)$ и $f_{te.s}(y)$ имеют функции распределения логнормального распределения), S будет рассчитываться как [см. (23)]:

$$S \propto \int_a^b f_{te.s}(y) dy, \quad (23)$$

или [см. (24)]:

$$S \propto \frac{1}{\sigma_{te.s} \sqrt{2\pi}} \int_a^c \exp\left[-(\ln y - \mu_{te.s})^2 / 2\sigma_{te.s}^2\right] dy. \quad (24)$$

Если выражение для S не может быть найдено аналитически, то применяется численное интегрирование.

Предлагаемый алгоритм для построения ВА для оценки показателя качества конечного продукта с использованием физически обоснованной математической модели можно представить в виде описанных ниже шагов.

Шаг 1: для данного процесса с независимыми переменными $U \in R^{N \times n_U}$ и зависимой переменной $Y \in R^{N \times 1}$ определить верхнюю и нижнюю границы для зависимых переменных в обучающей и тестовой выборках.

Шаг 2: на основе исходных данных $U \in R^{N \times n_U}$ и $y \in R^{N \times 1}$ рассчитываются параметры модели $\hat{\beta}$ различными методами, например, методом проекций на латентные структуры.

Шаг 3: на основе физически обоснованной математической модели формируется расширенный набор данных при условии, если $\Delta \hat{E}_1 \leq \Delta \hat{E}_1^{eff.th}$, $\Delta \hat{E}_2 \leq \Delta \hat{E}_2^{eff.th}$ и $\Delta \hat{\Lambda}_{hi} \leq \Delta \hat{\Lambda}_{hi}^{eff.th}$ (где $\Delta \hat{E}_1$, $\Delta \hat{E}_2$ – отклонения эффективности массопереноса по Мерффри объекта и модели на 6–28 тарелках и на 1–5 тарелках нижней ректификационной колонны: $\Delta \hat{E}_1 = |E_1 - \hat{E}_1|$, $\Delta \hat{E}_2 = |E_2 - \hat{E}_2|$; $\Delta \hat{\Lambda}_{hi}$ – отклонения параметров бинарного взаимодействия компонентов объекта и модели: $\Delta \hat{\Lambda}_{hi} = |\Lambda_{hi} - \hat{\Lambda}_{hi}|$; $\Delta \hat{E}_1^{eff.th}$, $\Delta \hat{E}_2^{eff.th}$ и $\Delta \hat{\Lambda}_{hi}^{eff.th}$ – пороговые значения отклонений, превышение которых при формировании расширенной обучающей выборки приводит к снижению точности построенной статистической модели).

Шаг 4: в исходные данные добавляются новые данные, полученные на Шаге 3, таким образом, получается расширенная обучающая выборка – выборка, содержащая в себе промышленные данные и данные физически обоснованной математической модели технологического процесса.

Шаг 5: нахождение новых параметров модели $\hat{\beta}$.



Рис. 3. Схема построения модели для оценки показателя качества конечных продуктов на основе расширения исходной малой обучающей выборки.

Шаг 6: адекватность полученного ВА проверяется на основе критериев точности модели [см. (5), (6)].

Шаг 8: конец.

Иллюстрация предложенного подхода изображено на рис. 3.

Преимущество предлагаемого алгоритма по сравнению с известными подходами заключается в возможности учета априорных знаний об исследуемом объекте. Построенные таким образом модели близки, в сущности, к моделям на основе “серого” ящика.

Следует отметить, что использование напрямую в реальном времени (on-line) физически обоснованной математической модели для оценки показателей качества конечного продукта часто на практике оказывается невозможным, так как эффективность массопереноса по Мерффри и параметры бинарного взаимодействия некоторых компонентов в смеси известны лишь приближенно, а состав сырья на каждом периоде управления не измеряется из-за отсутствия поточных анализаторов.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Оценка эффективности массопереноса по Мерффри на основе промышленных данных

Цель калибровки физически содержательной модели заключается в уменьшении ошибки между экспериментальными данными и данными моделирования при варьировании секционными эффективностями массопереноса по Мерффри на 6–28 тарелках и на 1–5 тарелках нижней ректификационной колонны в соответствии со следующими критериями (функциями невязки):

Таблица 2. Оценка сходимости результатов расчета при $\hat{E}_1 = 0.5$ и $\hat{E}_2 = 0.8$ с промышленной данными

Технологический параметр	Экспериментальные данные	Данные моделирования	Относительная ошибка δ , %
T _{II} , °C	66.78	68.51	2.60
T _{III} , °C	138.64	136.18	1.81
MeOH, мас. %	0.28	0.29	3.45
MTBE, мас. %	98.18	99.25	1.08
DIB, мас. %	0.008	0.0084	4.76

$$\begin{cases} J_1 = (T_{\text{III}} - T'_{\text{III}})^2 + (T_{\text{II2}} - T'_{\text{II2}})^2 \\ J_2 = (x_{\text{MTBE}} - x'_{\text{MTBE}})^2 + (x_{\text{MeOH}} - x'_{\text{MeOH}})^2 + (x_{\text{DIB}} - x'_{\text{DIB}})^2, \end{cases} \quad (25)$$

где T_{III} и T'_{III} – значения экспериментальных и расчетных температур куба нижней ректификационной колонны, °C; T_{II2} и T'_{II2} – значения экспериментальных и расчетных температур углеводородной смеси, отходящей сверху нижней ректификационной колонны, °C; x_{MTBE} и x'_{MTBE} – экспериментальные и расчетные значения концентрации МТБЭ в конечном продукте, мас. %;

x_{MeOH} и x'_{MeOH} – экспериментальные и расчетные значения концентрации метанола в конечном продукте, мас. %; x_{DIB} и x'_{DIB} – экспериментальные и расчетные значения концентрации димеров изобутилена в конечном продукте МТБЭ, мас. %.

Результаты калибровки физически содержательной модели при различной эффективности массопереноса по Мерффри представлена на рис. 4.

Из представленного рис. 4 можно сделать вывод, что минимальное значение функции невязки по содержанию метанола, метил-трет-бутилового эфира и димеров изобутилена в конечном продукте достигается при $\hat{E}_1 = 0.5$ и $\hat{E}_2 = 0.8$.

Оценка сходимости результатов расчета и экспериментальных данных с промышленной установки производства МТБЭ приведена в табл. 2.

Относительная ошибка между экспериментальными данными и данными моделирования основных показателей технологического режима по качеству и расходу продуктов разделения не превышает 5%, что подтверждает пригодность физически содержательной модели (табл. 2) для ее дальнейшего применения.

Целесообразность применения алгоритма расширения обучающей выборки

На шаге 3 предлагаемого алгоритма для определения пороговых значений отклонений $\Delta\hat{E}_1$, $\Delta\hat{E}_2$ и $\Delta\hat{\lambda}_{hi}$ использовалась откалиброванная физически обоснованная математическая модель исследуемого процесса, с допущением о ее соответствии реальному объекту, и установлены границы параметрической неопределенности эффективности массопереноса по Мерффри $\Delta_{\hat{E}_1}^{\text{eff.th}} = 0.3$ и $\Delta_{\hat{E}_2}^{\text{eff.th}} = 0.1$. Для определения $\Delta_{\hat{\lambda}_{hi}}^{\text{eff.th}}$ оценивался % снижения R^2 (δR^2) для диапазона параметрической неопределенности $\Delta\hat{\lambda}_{hi} \in [0.1, 0.4]$ относительно базового случая, когда модель полностью соответствует объекту, т.е. относительно $\Delta\hat{\lambda}_{hi} = 0$ (табл. 3). Если потери δR^2 превышали 5%, то для соответствующих параметров бинарного взаимодействия компонентов системы назначалось пороговое значения отклонения $\Delta_{\hat{\lambda}_{hi}}^{\text{eff.th}}$.

В результате определены наиболее чувствительные параметры бинарного взаимодействия компонентов системы: $\Delta\hat{\lambda}_{\text{DIB,MeOH}}$, $\Delta\hat{\lambda}_{\text{MeOH,IB}}$, $\Delta\hat{\lambda}_{\text{MeOH,MTBE}}$ и $\Delta\hat{\lambda}_{\text{DIB,MTBE}}$ с $\Delta_{\hat{\lambda}_{hi}}^{\text{eff.th}} = 0$, т.е. значения этих параметров необходимо знать точно, чтобы использовать данные физически обоснованной модели для расширения обучающей выборки.

Результаты критериев точности модели для оценки показателей качества конечного продукта и критериев эффективности расширения обучающей выборки при различных отклонениях $\Delta\hat{\lambda}_{hi}$ представлены в табл. 3.

Из табл. 3 видно, что с увеличением диапазона параметрической неопределенности параметров бинарного взаимодействия точность модели для

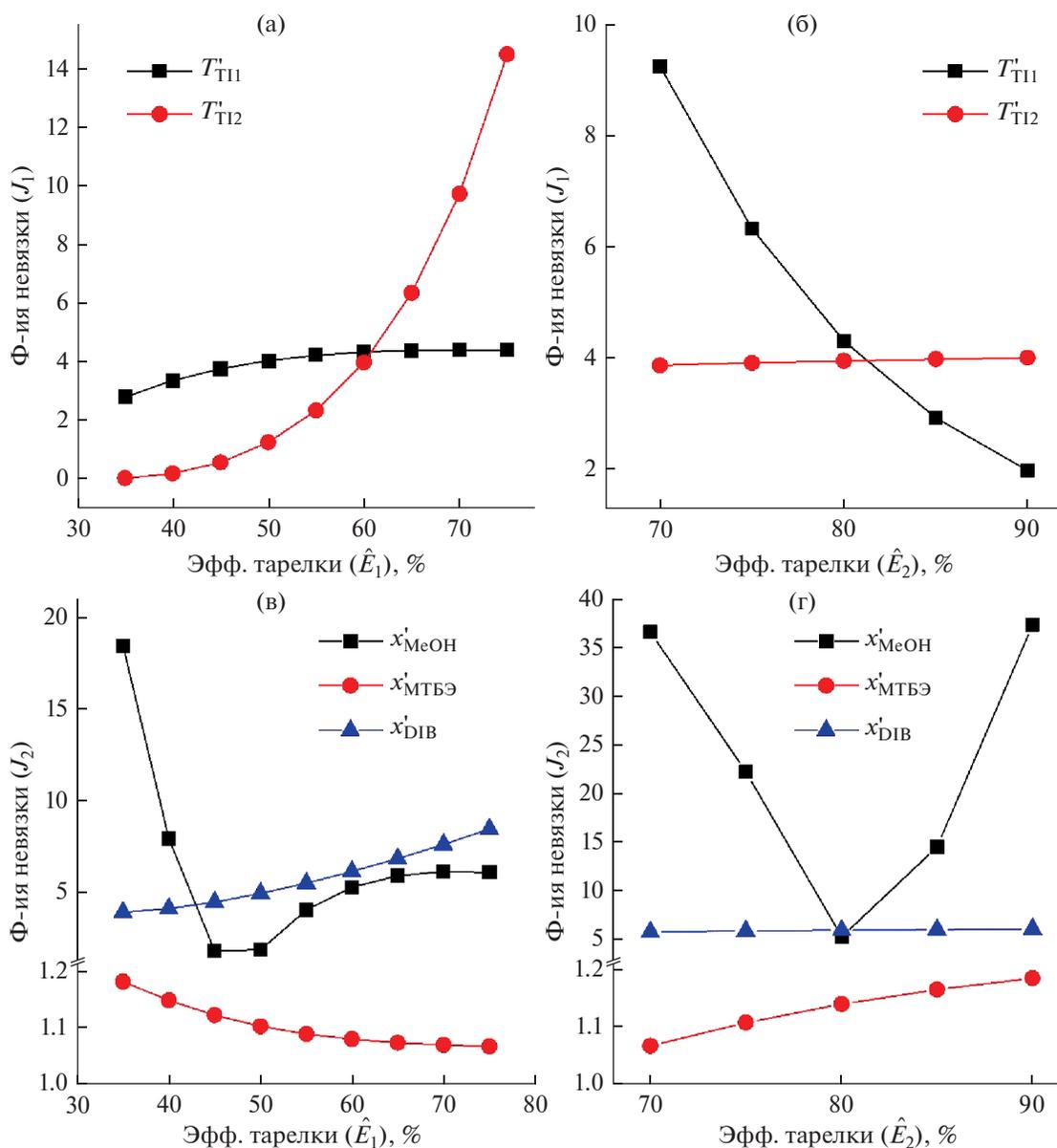


Рис. 4. Результаты определения эффективности массопереноса по Мэрфри при калибровке модели: (а) – зависимость функции невязки температур от \hat{E}_1 ; (б) – зависимость функции невязки температур от \hat{E}_2 ; (в) – зависимость функции невязки концентраций компонентов от \hat{E}_1 ; (г) – зависимость функции невязки концентраций компонентов от \hat{E}_2 .

оценки метанола в МТБЭ снижается (согласно R^2 и MAE), также снижаются и значения S . Таким образом установлено, что физически обоснованную математическую модель технологического процесса в алгоритме целесообразно применять при условии, что для $\hat{\Lambda}_{MTBE,IB}$ значение $\Delta_{\hat{\Lambda}_{hi}}^{eff.th} = 0.3$; для компонентов системы: $\hat{\Lambda}_{MTBE,MeOH}$, $\hat{\Lambda}_{MeOH,DIB}$ пороговые значения совпадают и равны $\Delta_{\hat{\Lambda}_{hi}}^{eff.th} = 0.2$; для параметров бинарного взаимодействия $\Delta_{\hat{\Lambda}_{hi}}^{eff.th}$ значение $\Delta_{\hat{\Lambda}_{hi}}^{eff.th} = 0.1$.

Сравнительный анализ методов построения модели для оценки концентрации метанола в конечном продукте в условиях малой обучающей выборки

Проводится сравнение предлагаемого алгоритма расширения обучающей выборки с известными подходами построения моделей для оценки показателей качества конечных продуктов в условиях малых обучающих выборок.

Обучение выполняется с использованием известных (часто применяемых в реальных условиях [6]) методов и их сочетаний: метода проекций на латентные структуры (ПЛС), метода проекций на латентные структуры и метода складного ножа

Таблица 3. Влияние изменения диапазона параметрической неопределенности параметров бинарного взаимодействия на R^2 , MAE и S

$\Delta\hat{\Lambda}_{hi}$	ПЛС	Граница интервала параметрической неопределенности				
		0	0.1	0.2	0.3	0.4
$\Delta\hat{\Lambda}_{\text{MTBE,IB}}$	R^2	0.5346	0.5145	0.5104	0.5081	0.4235
	MAE	0.0886	0.0929	0.0931	0.0930	0.1309
	S	0.5731	0.5339	0.5276	0.5205	0.5122
	$\delta R^2, \%$		3.76	4.53	4.96	20.78
$\Delta\hat{\Lambda}_{\text{MTBE,MeOH}}$	R^2	0.5346	0.5135	0.5085	0.5065	0.5048
	MAE	0.0886	0.0931	0.0935	0.0930	0.0936
	S	0.5731	0.5372	0.5310	0.5239	0.5157
	$\delta R^2, \%$		3.95	4.88	5.26	5.57
$\Delta\hat{\Lambda}_{\text{MTBE,DIB}}$	R^2	0.5346	0.5310	0.5074	0.5028	0.5017
	MAE	0.0886	0.0898	0.0929	0.0941	0.0939
	S	0.5731	0.5319	0.5247	0.5163	0.5060
	$\delta R^2, \%$		0.67	5.09	5.95	6.15
$\Delta\hat{\Lambda}_{\text{MeOH,DIB}}$	R^2	0.5346	0.5073	0.5063	0.5028	0.5016
	MAE	0.0886	0.0929	0.0939	0.0941	0.0939
	S	0.5731	0.5383	0.5319	0.5247	0.5163
	$\delta R^2, \%$		5.11	5.29	5.95	6.17

(ПЛС и СН), метода проекций на латентные структуры и бутстрэп метода (ПЛС и БМ) (сформировано 50 бутстрэп выборок, объем каждой выборки составил 80% от исходной), а также нелинейного метода проекций на латентные структуры (НПЛС) с функцией ядра (14) и нейронной сетью (НС) прямого распространения с двумя скрытыми слоями и пятью нейронами в каждом слое, т.е. (5 5). Следует отметить, что количество нейронов в слоях определялось таким образом, чтобы избежать переобучения.

В табл. 4 приведен сравнительный анализ наиболее известных методов на $N_{\text{т.с}} = 18$ до и после добавления данных в исходную обучающую выборку.

Следует отметить, что значения R^2 и MAE , полученные с использованием представленных в табл. 4 методов, имеют определенное сходство. До добавления данных вспомогательной выборки значения R^2 не превышают 0.53; т.е. результаты расчетов показывают низкую точность полученных моделей для оценки MeOH в конечном продукте.

В свою очередь учет априорных знаний о массообменном технологическом процессе позволяет увеличить точность моделей для оценки показателей качества конечных продуктов, а также

повысить эффективность других методов, что дает особое преимущество предлагаемого подхода в сравнении с известными методами, например с PSO. Значение MAE модели для оценки концентрации метанола в конечном продукте, построенной с использованием ПЛС и бутстрэп метода, снижается на 23% (рис. 5).

Использование метода PSO для расширения обучающей выборки на промышленных данных на первоначальном этапе (когда размер вспомогательной выборки не превышает размер обучающей выборки) позволяет увеличить R^2 на $((0.5250 - 0.4496)/0.5250) \times 100 \approx 14\%$, однако, при повышении размера вспомогательной выборки наблюдается существенное ухудшение критериев точности модели (до 36% от исходного набора данных и до 55% в сравнении с моделью, полученной ПЛС и БМ по критерию R^2). Это связано с тем, что нет учета физически-химических особенностей технологического процесса.

Расширение обучающей выборки с использованием физически обоснованной математической модели процесса позволяет учитывать данные (заполнить пропуски в данных) технологического режима на всем диапазоне изменения качества конечного продукта (рис. 6а и рис. 6б).

Таблица 4. R^2 и MAE , полученные на тестовой выборке с использованием статистических методов до и после добавления данных вспомогательной выборки

Метод	Критерий	Размер вспомогательной выборки				
		0	10	20	30	40
ПЛС	R^2	0.3992	0.4202	0.4296	0.4299	0.4303
	MAE	0.1389	0.1346	0.1309	0.1306	0.1259
ПЛС и СН	R^2	0.4574	0.4791	0.5983	0.6023	0.6062
	MAE	0.1355	0.1287	0.1038	0.1007	0.1002
ПЛС и БМ	R^2	0.5333	0.6051	0.6299	0.6397	0.6429
	MAE	0.1197	0.1035	0.0959	0.0930	0.0920
НПЛС	R^2	0.4170	0.5236	0.5275	0.5359	0.5425
	MAE	0.1245	0.1227	0.1120	0.1107	0.1108
Нейронная сеть (НС)	R^2	0.4240	0.5569	0.5140	0.5165	0.5324
	MAE	0.1556	0.1236	0.1405	0.1287	0.1085
ПЛС и PSO	R^2	0.4496	0.5250	0.5274	0.4254	0.2852
	MAE	0.1389	0.1064	0.1126	0.1377	0.1546

Изменение площади области пересечения распределений значений выходной переменной в расширенной обучающей и тестовой выборках представлены в табл. 5.

Использование физически обоснованной математической модели увеличивает площадь области пересечения значений выходной переменной

в расширенной обучающей и тестовой выборках до $((0.98 - 0.56)/0.98) \times 100 \approx 43\%$ (табл. 5).

Результаты функционирования разработанных моделей до и после расширения исходной выборки представлены на рис. 7а и рис. 7б.

Как показано на рис. 7б расширение исходной обучающей выборки с использованием физически

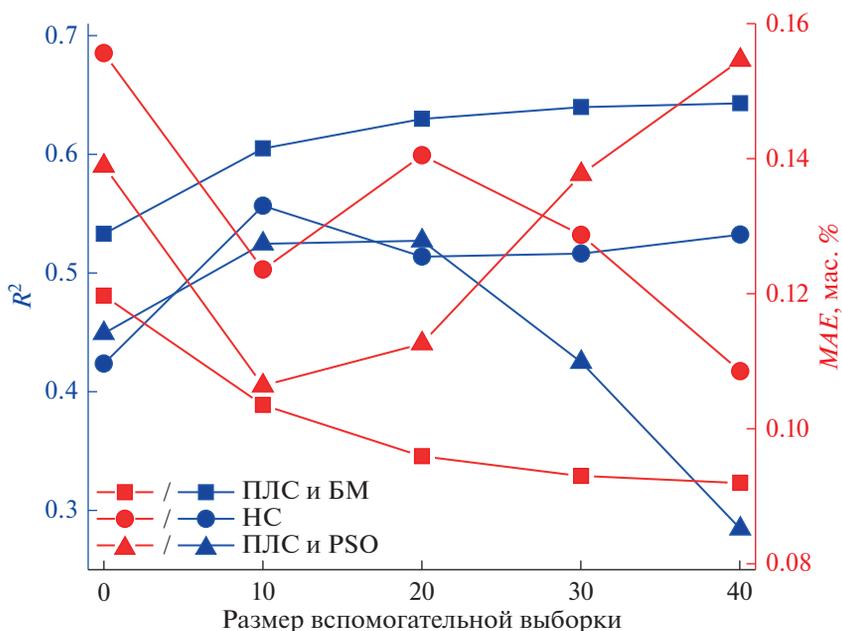


Рис. 5. Значения R^2 и MAE модели для оценки содержания концентрации метанола в МТБЭ на тестовой выборке.

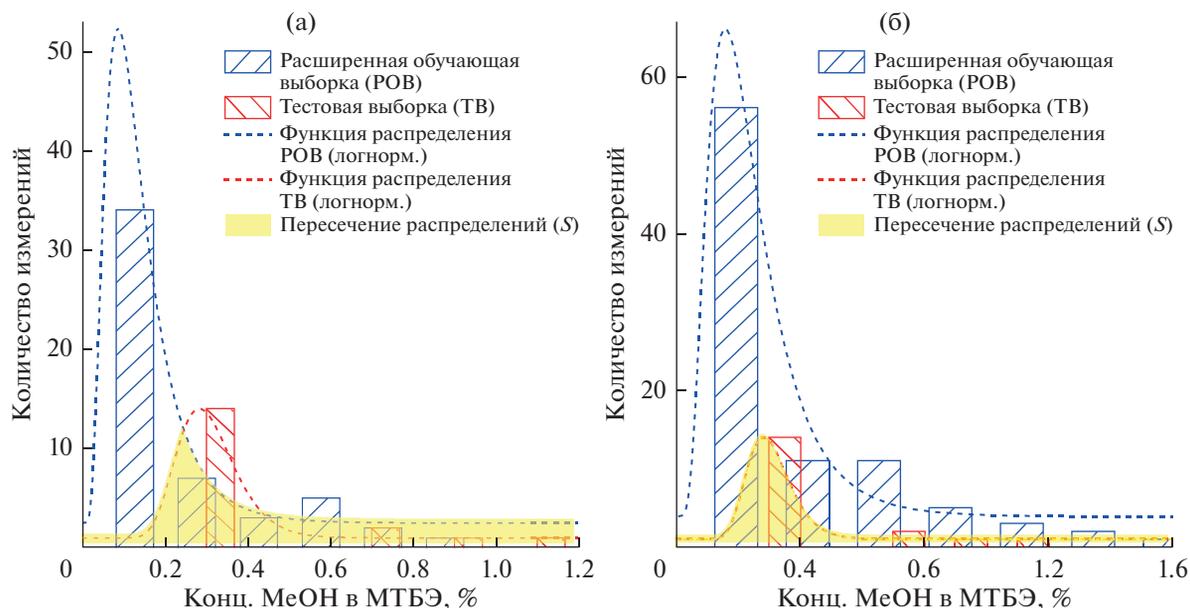


Рис. 6. Экспериментальные данные концентрации метанола в конечном продукте после расширения обучающей выборки (а) – значения выходной переменной в расширенной и тестовой выборках частично перекрываются; (б) – значения выходной переменной тестовой выборки содержатся в расширенной обучающей выборке.

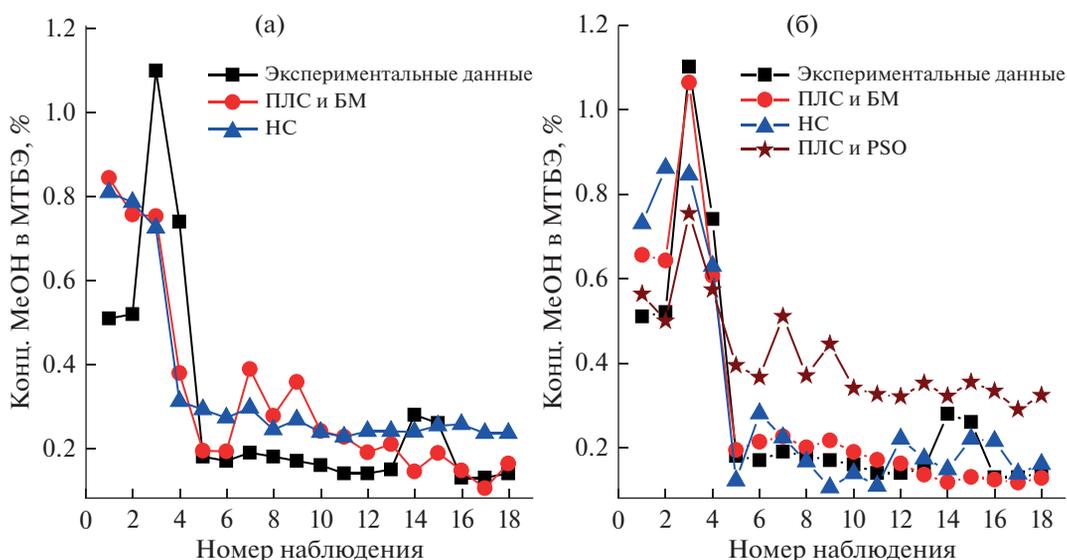


Рис. 7. Оценка метанола в конечном продукте до и после расширения обучающей выборки (а) – до расширения обучающей выборки; (б) – после расширения обучающей выборки.

Таблица 5. Площадь области пересечения обучающей и тестовой выборок до и после добавления данных вспомогательной выборки

Площадь области пересечения выборок	Расширение обучающей выборки				
	0	10	20	30	40
S	0.56	0.88	0.89	0.90	0.98

обоснованной математической модели повышает точность оценки показателя качества конечного продукта. После расширения обучающей выборки данными вспомогательной выборки размером 40 (табл. 4) эффективность использования метода проекций на латентные структуры совместно с бутстрэп методом для нахождения параметров ВА увеличилась на $((0.5333 - 0.6429)/0.5333) \times 100 \approx 21\%$ и $((0.1197 - 0.0920)/0.1197) \times 100 \approx 23\%$ по R^2 и MAE , соответственно. Использование метода ПЛС совместно с оптимизацией роя частиц для расширения исходной выборки не позволяет улучшить точность ВА, т. к. данные вспомогательной выборки не учитывают физико-химические особенности процесса (рис. 7б).

ЗАКЛЮЧЕНИЕ

Предложен алгоритм расширения обучающей выборки на основе физически обоснованной математической модели ректификационной колонны массообменного технологического процесса, позволяющий повысить точность ВА для оценки показателя качества конечного продукта и эффективность традиционных методов при построении модели (определении параметров ВА). Предлагаемый алгоритм протестирован на промышленном массообменном химико-технологическом процессе производства МТБЭ.

Определены условия применения физически обоснованной математической модели в условиях параметрической неопределенности параметров бинарного взаимодействия компонентов системы. Некоторые из них должны быть известны практически точно, т.е. $\Delta_{\hat{\lambda}_{hi}}^{eff.th} = 0$ для параметров бинарного взаимодействия: $\Delta_{\hat{\lambda}_{DIB,MeOH}}$, $\Delta_{\hat{\lambda}_{MeOH,IB}}$, $\Delta_{\hat{\lambda}_{MeOH,MTBE}}$; для $\Delta_{\hat{\lambda}_{MTBE,DIB}}$ и $\Delta_{\hat{\lambda}_{MTBE,IB}}$ пороговые значения равны $\Delta_{\hat{\lambda}_{hi}}^{eff.th} = 0.1$ и $\Delta_{\hat{\lambda}_{hi}}^{eff.th} = 0.3$, соответственно.

Применение физически обоснованной математической модели при построении ВА позволяет заполнить пропуски в данных на всем диапазоне изменения качества конечного продукта (значение критерия расширения обучающей выборки S повышается до 43%).

Точность ВА для оценки концентрации метанола в конечном продукте, полученного предложенным подходом, на 55% выше точности ВА, полученным на основе расширения выборки методом оптимизации роя частиц. Использование предлагаемого алгоритма позволяет повысить эффективность существующих методов построения статистических моделей. При разработке ВА методом проекций на латентные структуры совместно с бутстрэп методом точность модели увеличилась на 21 и 23% по критериям R^2 и MAE после расши-

рения обучающей выборки данными вспомогательной выборки размером 40. Учет априорных знаний о процессе дает существенное преимущество алгоритма перед другими методами расширения обучающей выборки.

Исследование выполнено при частичной финансовой поддержке РФФИ и Государственного фонда естественных наук Китая в рамках научного проекта № 21-57-53005

ОБОЗНАЧЕНИЯ

ББФ	бутан-бутиленовая фракция
ББФ _{отр}	отработанная бутан-бутиленовая фракция
MeOH	метанол
МТБЭ,	метил-трет-бутиловый эфир
МТВЕ	
IB	изобутилен
DIB	диизобутен
2,4,4-	2,4,4-триметилпентен-1
ТМП-1	
2,4,4-	2,4,4-триметилпентен-2
ТМП-2	
NB	n-бутен
N	количество наблюдений
n _U	количество входных переменных
n _Y	количество выходных переменных
U	матрица входных переменных $U \in R^{N \times n_U}$
Y	вектор выходной переменной $Y \in R^{N \times 1}$
β	вектор параметров модели для оценки показателя качества конечного продукта $\beta \in R^{n_U \times 1}$
$\hat{\beta}$	вектор оцененных параметров модели для оценки показателя качества конечного продукта
\hat{Y}	значение выходной переменной, полученное на основе модели для оценки показателя качества конечного продукта, мас. %
\bar{Y}	среднее значение наблюдаемой выходной переменной, мас. %
P	матрица нагрузок U
Q	матрица нагрузок Y
W	матрица весов U
\bar{U}	вектор средних значений матрицы U
a	значение главной компоненты
A	максимальное значение главной компоненты
w	вектор взвешенных нагрузок
c	это нормировочный фактор

\mathbf{t}	вектор счетов
\mathbf{p}	“спектральные” нагрузки
\mathbf{q}	“химические” нагрузки
R^2	коэффициент детерминации
MAE	абсолютная средняя ошибка
\mathbf{K}	матрица Грама, $\mathbf{K} = \Phi\Phi^T \in R^{N \times N}$
Φ	матрица отображенных входных данных
НПЛС	нелинейная ядерная модель проекций на скрытые структуры
ker	функция ядра
ρ, σ	параметры функции ядра
v	скорость движения частицы
t	счетчик итераций
pbest	новое положение частицы
gbest	лучшее значение частицы
w	инерционный вес
L, V	поток жидкости и пара, соответственно, кмоль/ч
x, y	концентрация компонента в жидкой фазе и паровой фазе, соответственно, кмоль/кмоль
y^*	равновесная концентрация компонента в паровой фазе, кмоль/кмоль
F	расход сырья, кмоль/ч
γ	коэффициент активности компонента в жидкой фазе
p^0	парциальное давление компонента, кПа
P	общее давление в системе, кПа
E	эффективность массопереноса по Мерффри
h, H	энтальпия жидкости и пара, соответственно, кДж/кмоль
nc	общее количество компонентов в системе
nt	общее количество тарелок в ректификационной колонне
m	номер реакции
r	количество реакций
v	стехиометрический коэффициент компонента в реакции
R	скорость реакции, кмоль/см ² · ч
ε	объем реакционной смеси, м ³
Δn	изменение количества одного из веществ в реакции, кмоль
Δt	промежуток времени реакции, ч
s	поверхность соприкосновения фаз, см ²
\tilde{r}, \tilde{q}	UNIQUAC параметры объема и площади чистого компонента
τ	приведенный параметр

\hat{R}	газовая постоянная, Дж/(моль · К)
z	координационное число
\hat{E}	эффективность массопереноса по Мерффри
Λ_{hi}	параметры бинарного взаимодействия
J	функция невязки
T	значения экспериментальных температур, °С
T'	значения расчетных температур, °С
$K(y)$	критерий согласия Колмогорова
$F^*(y)$	эмпирическая функция распределения выходной переменной
$F(y)$	гипотетическая функция распределения
c	точка пересечения
$f(y)$	плотность распределения выходной переменной
S	площадь области пересечения плотностей распределения

ИНДЕКСЫ

T	операция транспонирования
\tilde{i}, \tilde{j}	элементы в ker
vir_i	сгенерированное значение
start, end	начальное и конечное значения инерционного веса
i	номер компонента
j	номер тарелки
eff.th	пороговое значение
tr.s	обучающая выборка
ext.s	расширенная выборка
te.s	тестовая выборка

СПИСОК ЛИТЕРАТУРЫ

1. Vairo T., Reverberi A.P., Bragatto P.A., Milazzo M.F., Fabiano B. Predictive model and soft sensors application to dynamic process operative control // Chem. Eng. Transaction. 2021. V. 86. P. 535–540.
2. Jiang Y., Yin Sh., Dong J., Kaynak O. A Review on soft sensors for monitoring, control, and optimization of industrial processes // IEEE Sensors J. 2021. V. 21. № 11. P. 12868–12881.
3. Shokry A., Vicente P., Escudero G., Pérez-Moya M., Graells M., Espuña A. Data-driven soft-sensors for on-line monitoring of batch processes with different initial conditions // Comput. & Chem. Eng. 2018. V. 118. P. 159–179.
4. Lin B., Recke B., Knudsen J.K.H., Jørgensen S.B. A systematic approach for soft sensor development // Comput. & Chem. Eng. 2007. V. 31. I. 5–6. P. 419–425.
5. Niño-Adán I., Landa-Torres I., Manjarres D., Portillo E., Orbe L. Soft-sensor for class prediction of the percent-

- age of pentanes in butane at a debutanizer column // *Sensors*. 2021. V. 21. P. 3991.
6. *Kadlec P., Gabrys B., Strandt S.* Data-driven soft sensors for monitoring and control of industrial processes // *Comput. Chem. Eng.* 2009. V. 33. P. 759–814.
 7. *Fortuna L., Graziani S., Rizzo A., Xibilia M.G.* Soft sensors for monitoring and control of industrial processes. London, U.K.: Springer-Verlag, 2007.
 8. *Graziani S., Xibilia M.G.* On the use of correlation analysis in the estimation of finite-time delay in Soft Sensors design // *IEEE International Instrumentation and Measurement Technol. Conference*. 2021. P. 1–6.
 9. *Ge Zh., Huang B. Song Zh.* Mixture semisupervised principal component regression model and soft sensor application // *AIChE J.* 2014. V. 60. I. 2. 533–545.
 10. *Sun K., Tian P., Qi H., Ma F., Yang G.* An improved normalized mutual information variable selection algorithm for neural network based soft sensors // *Sensors*. 2019. V. 19. № 24. P. 5368.
 11. *Graziani S., Xibilia M.G.* Development and analysis of deep learning architectures (Deep learning for soft sensor design). Cham, Switzerland: Springer, ch. 2, 2020.
 12. *Sun Q., Ge Zh.* A Survey on deep learning for data-driven soft sensors // *IEEE Transactions on Ind. Informatics*. 2021. V. 17. № 9. P. 5853–5866.
 13. *Ge Zh., Song Zh., Kano M.* External analysis-based regression model for robust soft sensing of multimode chemical processes // *AIChE J.* 2014. V. 60. I. 1. P.136–147.
 14. *Bakirov R., Gabrys B., Fay D.* Multiple adaptive mechanisms for data-driven soft sensors // *Comput. & Chem. Eng.* 2017. V. 96. P. 42–54.
 15. *Zhu J., Ge Z., Song Z., Gao F.* Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data // *Annu. Rev. Control*. 2018. V. 46. P. 107–133.
 16. *Xie R., Jan N.M., Hao K., Chen L., Huang B.* Supervised variational autoencoders for soft sensor modeling with missing data // *IEEE Transactions on Ind. Informatics*. 2019. V. 16. I. 4. P. 2820–2828.
 17. *Kaneko H., Funatsu K.* Classification of the degradation of soft sensor models and discussion on adaptive models // *AIChE J.* 2013. V. 59. I. 7. P. 2339–2347.
 18. *Hsiao Y.-D., Kang J.-L., Wong D.S.-H.* Development of robust and physically interpretable soft sensor for industrial distillation column using transfer learning with small datasets // *Processes*. 2021. V. 9. P. 667.
 19. *Napoli G., Xibilia M.G.* Soft Sensor design for a topping process in the case of small datasets // *Comput. & Chem. Eng.* 2011. V. 35. I. 11. P. 2447–2456.
 20. *Chang C.-J., Li D.-C., Huang Y.-H., Chen C.-C.* A novel gray forecasting model based on the box plot for small manufacturing data sets // *Applied Math. and Comp.* 2015. V. 265. P. 400–408.
 21. *Andrijić Ž.U., Cvetnić M., Bolf N.* Soft sensor models for a fractionation reformate plant using small and bootstrapped data sets // *Brazilian J. Chem. Eng.* 2018. V. 35. P. 745–756.
 22. *Urhan A., Ince N.G., Bondy R., Alakent B.* Soft-sensor design for a crude distillation unit using statistical learning methods. In *Comp. Aided Chem. Eng.* 2018. V. 44. P. 2269–2274.
 23. *Chen Z.-S., Zhu B., He Y.-L., Yu L.-A.* A PSO based virtual sample generation method for small sample sets: Applications to regression datasets // *Eng. applications of artificial intelligence*. 2017. V. 59. P. 236–243.
 24. *He Y.L., Wang P.J., Zhang M.Q., Zhu Q.X., Xu Y.* A novel and effective nonlinear interpolation virtual sample generation method for enhancing energy prediction and analysis on small data problem: A case study of ethylene industry // *Energy*. 2018. V. 147. P. 418–427.
 25. *Di Girolamo M., Lami M., Marchionna M., Pescarollo E., Tagliabue L., Ancillotti F.* Liquid-phase etherification/dimerization of isobutene over sulfonic acid resins // *Ind. Eng. Chem. Res.* 1997. V. 36. P. 4452–4458.
 26. *Rehfinger A., Hoffmann U.* Kinetics of methyl-tertiary-butyl ether liquid phase synthesis catalyzed by ion exchange resin intrinsic rate expression in liquid phase activities // *Chem. Eng. Sci.* 1990. V. 45. № 6. P. 1605–1617.
 27. *Эбенсен К.* Анализ многомерных данных. Избранные главы. Барнаул: Изд-во Алт. ун-та, 2003.
 28. *Wold S.* PLS-regression: a Basic Tool of Chemometrics / S. Wold, M. Sjostrom, L. Eriksson // *Chemometrics and Intelligent Laboratory System*. 2001. V. 58. P. 109–130.
 29. *De Jong S.* SIMPLS: An Alternative Approach to Partial Least Squares Regression // *Chemometrics and Intelligent Laboratory Systems*. 1993. V. 18. P. 251–263.
 30. *Van Kollenburg G.H., Van Es J., Gerretzen J., Lanter H., Bouman R., Koelewijn W., Davies A.N., Buydens L.M.C., Van Manen H.-J., Jansen J.J.* Understanding chemical production processes by using PLS path model parameters as soft sensors // *Comput. & Chem. Eng.* 2020. V. 139. P. 1–8.
 31. *Kaneko H., Arakawa M., Funatsu K.* Development of a new soft sensor method using independent component analysis and partial least squares // *AIChE J.* 2009. V. 55. I. 1. P. 87–98.
 32. *Jin H., Chen X., Yang J., Wu L.* Adaptive soft sensor modeling framework based on just-in-time learning and kernel partial least squares regression for nonlinear multiphase batch processes // *Comput. & Chem. Eng.* 2014. V. 71. P. 77–93.
 33. *Si Y., Wang Y., Zhou D.* Key-performance-indicator-related process monitoring based on improved kernel partial least squares // *IEEE Transactions on Industrial Electronics*. 2020. V. 68. № 3. P. 2626–2636.
 34. *Deng X., Chen Y., Wang P., Cao Y.* Soft sensor modeling for unobserved multimode nonlinear processes based on modified kernel partial least squares with latent factor clustering. *IEEE Access*. 2020. V. 8. P. 35864–35872.
 35. *Schwaab M., Biscaia E.C., Monteiro J.L., Pinto J.C.* Nonlinear parameter estimation through particle swarm optimization // *Chem. Eng. Sci.* 2008. V. 63. I. 6. P. 1542–1552.
 36. *Ourique C.O., Biscaia E.C., Pinto J.C.* The use of particle swarm optimization for dynamical analysis in chemical processes // *Comput. & Chem. Eng.* 2002. V. 26. Is. 12. P. 1783–1793.
 37. *Prata D.M., Schwaab M., Lima E.L., Pinto J.C.* Nonlinear dynamic data reconciliation and parameter estimation through particle swarm optimization: Application

- for an industrial polypropylene reactor // *Chem. Eng. Sci.* 2009. V. 64. I. 18. P. 3953–3967.
38. *Fredenslund A., Jones R.L., Prausnitz J.M.* Group-contribution estimation of activity coefficients in nonideal liquid mixtures // *AIChE J.* 1975. V. 21. № 6. P. 1086–1099.
39. *Abrams D.S., Prausnitz J.M.* Statistical thermodynamics of liquid mixtures: a new expression for the excess Gibbs energy of partly or completely miscible systems // *AIChE J.* 1975. V. 21. № 1. P. 116–128.
40. *Sundmacher K., Uhde G., Hoffmann U.* Multiple reactions in catalytic distillation processes for the production of the fuel oxygenates MTBE and TAME: Analysis by rigorous model and experimental validation // *Chem. Eng. Sci.* 1999. V. 54. P. 2839–2847.
41. Об эмпирическом определении закона распределения / Под ред. Прохорова Ю.В. М.: Наука, 1986.
42. *Samotylova S.A., Torgashov A.Y.* Developing a soft sensor for MTBE process based on a small sample // *Automation and Remote Control.* 2020. V. 81. № 11. P. 2132–2142.