

Методы повышения точности моделей виртуальных анализаторов показателей качества фракции 80...180°С для колонны К-2 установки атмосферно-вакуумной перегонки нефти

В.А. Лаврентьев, Ю.А. Ергомышев, Е.С. Головина,
А.В. Комогоров (АО «Самаранефтехимпроект»)

Рассматривается задача повышения точности модели виртуального анализатора показателей качества фракции 80...180°С для колонны К-2 установки атмосферно-вакуумной перегонки нефти. Построены модели виртуальных анализаторов с применением регрессионных методов и искусственных нейронных сетей. Определена методика отбора данных параметров технологического процесса, позволяющая строить более точные модели, чем используемые на установке в настоящее время. Регрессионные модели оценивались по методу наименьших квадратов. Дополнительно рассмотрен альтернативный метод решения задачи с использованием нейронных сетей. Показано, что нейронные сети позволяют построить более точную модель виртуального анализатора, чем регрессионные методы, не прибегая к предварительной фильтрации и отбору данных технологического процесса для построения модели. Ввиду особенностей существующего программного обеспечения систем усовершенствованного управления технологическим процессом (СУУТП), стандартно применяемых на предприятиях, использование нейронных сетей в настоящее время не представляется возможным, но является перспективным направлением для проработки.

Ключевые слова: регрессионная модель, нейронные сети, виртуальный анализатор, усовершенствованное управление.

Введение. Современное цифровое производство

Управление технологическим процессом на современном цифровом нефтеперерабатывающем производстве осуществляется автоматизированными системами оптимизации и повышения эффективности, которые носят название систем усовершенствованного управления технологическим процессом (СУУТП). При этом качество работы установки определяется не только значением основных технологических параметров (температура, давление, расход), но

и показателями качества получаемых продуктов. Значения показателей качества учитываются при решении задачи оптимизации, и потому их получение является необходимым для функционирования СУУТП.

Однако данные лабораторных анализов или поточных анализаторов не поступают в режиме on-line, в то время как для управления процессом в границах показателей качества необходимо иметь информацию о качестве продуктов в режиме реального времени. Здесь на помощь приходят математические модели, связывающие показатели качества и технологические параметры установки, получившие название виртуальных анализаторов (ВА). Виртуальные анализаторы позволяют аналитически получать и использовать в управлении данные о трудноизмеримых технологических параметрах, позволяя оператору контролировать и вести процесс в условиях задержки данных лабораторных анализов, а также в случае выхода из строя или отсутствия поточных анализаторов. Существуют различные методы построения моделей ВА (нейронные сети, модели машинного обучения, линейные и нелинейные модели) [1].

Но какова бы ни была модель ВА, она имеет тенденцию снижения релевантности в зависимости от конструктивного изменения технологического объекта, проводимого во время капитального ремонта. Данный факт в дальнейшем приводит к снижению точности показаний, снижению корреляционной взаимосвязи между показаниями ВА и лабораторными анализами, увеличению ошибки прогноза, что в свою очередь имеет следствием снижение эффективности работы контуров управления системы СУУТП и падение экономического эффекта от их применения.

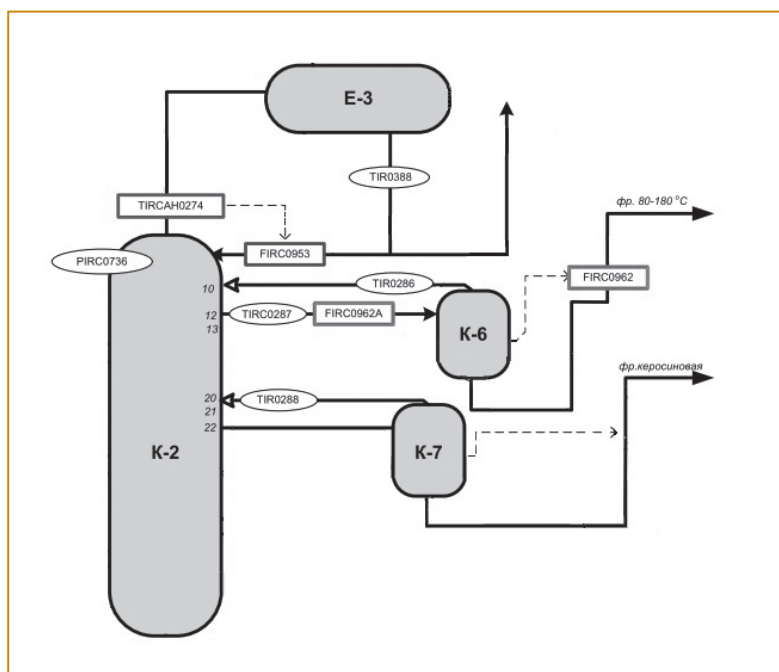


Рис. 1. Фрагмент схемы работы блока атмосферной перегонки, где К-2 – ректификационная колонна, Е-3 – буферная емкость, К-6, К-7, К-9 – отпарные колонны. Датчики: PIRC0736 – давление верха К-2, TIR0288 – температура паров керосиновой фракции из К-7 в К-2, TIRCAH0274 – температура верха К-2, TIR0388 – температура острого орошения К-2, TIRC0287 – температура перетока в К-6, TIRC0286 – температура возврата из К-6, FIRC0953 – расход острого орошения К-2, FIRC0962 – расход фракции НК-80 °С, FIRC0962A – расход фракции из К-2 в К-6.

Рассмотрим пути решения практической задачи повышения точности модели ВА показателя качества фракции 80...180° С для установки первичной переработки нефти.

Задачи и методы построения моделей ВА

Объектом исследования являлась установка атмосферно-вакуумной перегонки нефти АВТ-11 (рис.1). На объекте в 2014 г. реализована система усовершенствованного управления технологическим процессом (СУУТП).

В контуре управления ректификационной колонны К-2 атмосферного блока одним из ВА качества является температура начала кипения фракции 80...180° С. На момент исследования ВА качества не отвечал поставленной задаче в связи с конструктивными изменениями объекта, проведенными в период капитального ремонта. Среднеквадратическая ошибка (СКО) модели виртуального анализатора составляла 4,4° С; коэффициент корреляции с лабораторными данными равнялся 0,217, коэффициент детерминации R² модели равнялся 0,21. В качестве параметров технологического процесса использовались следующие технологические переменные: температура среднего циркуляционного орошения колонны К-2 (TIR0373); давление верха колонны К-2 (PIRC0736); температура паров керосиновой фракции из К-7 в колонну К-2 (TIR0288).

Проведенный с помощью модели колонны в программном обеспечении Aspen HYSYS анализ зависимости изменения показателей качества фракции от параметров технологического процесса позволил выделить технологические параметры, оказывающие наиболее существенное влияние: температура верха колонны К-2 (TIRCAN0274); температура острого орошения К-2 (TIR0388); давления верха К-2 (PIRC0736); температура перетока в К-6 (TIRC0287); температура возврата из К-6 (TIRC0286); расход острого орошения К-2 (FIRC0953); расход фракции НК-80° С (FIRC0962); расход фракции из К-2 в К-6 (FIRC0962A). Значимость приведенных параметров для температуры начала кипения фракции 80...180° С подтвердилась экспертной оценкой представителей производства.

Для указанных параметров с действующего технологического объекта были получены реальные исторические данные. Далее из этих физических параметров верха колонны путем корреляционного анализа данных приборов учета с данными лабораторных анализов было выделено три наиболее влияющих на температуру начала кипения фракции параметра, для которых абсолютное значение коэффициентов корреляции больше 0,7. Для исключения нежелательной взаимной корреляции между входными параметрами модели верха колонны был проведен взаимный корреляционный анализ имеющихся данных на идентичных отрезках времени, который подтвердил отсутствие взаимной

корреляции (абсолютное значение коэффициентов корреляции между переменными меньше 0,2). В итоге, в дальнейшем использованы следующие параметры: температура верха колонны К-2 (TIRCAN0274); температура острого орошения (TIR0388); давление верха колонны К-2 (PIRC0736). Для остальных переменных (температура перетока в К-6; температура возврата из К-6; расход острого орошения К-2; расход фракции НК-80° С; расход фракции из К-2 в К-6) получены низкие абсолютные значения коэффициентов корреляции (< 0,5), что свидетельствует о минимальном влиянии переменных на результаты лабораторных анализов. На следующем этапе подбора моделей указанные параметры не использовались.

Линейная регрессия

Наиболее распространенными, простыми в расчетах, и широко применимыми моделями ВА являются линейные регрессионные модели с коэффициентами, оцениваемыми по классическому методу наименьших квадратов. Они были испробованы для решения поставленной задачи в первую очередь.

Для построения моделей ВА, в особенности регрессионных, большое значение имеет отбор и фильтрация данных технологического процесса, позволяющих получить статистически наилучшую модель [2]. Лабораторный контроль осуществляется на рассматриваемом производстве по графику отбора проб три раза в неделю, кроме того, в 2020 г. было проведено три лабораторных пробега.

В ходе интеллектуального анализа данных отобранных параметров технологического процесса за период с июня 2020 г. по март 2021 г. было установлено, что необходимо выявлять те интервалы времени отбора технологических параметров, когда влияние (корреляция) каждого из них на лабораторные измерения рассматриваемого показателя качества максимально.

Температура верха колонны К-2 (датчик TC0274)

По результатам расчета коэффициентов корреляции для температуры верха колонны К-2 был взят временной интервал, отстоящий на 3 часа от момента отбора лабораторной пробы. Параметр температуры верха колонны К-2 оказывал наибольшее влияние (корреляцию) на показания лабораторного анализа показателя качества (начало кипения фракции 80...180° С) с временной задержкой от момента отбора пробы на период от 2 ч 32 мин до 2 ч 28 мин. Пример: отбор пробы произведен в 6:00, наиболее «коррелирующий» интервал времени с 3:28 до 3:32, с учетом периода дискретизации выборки данных, равного 1 минуте, для расчетов принималось усредненное значение за указанный период.

Температура острого орошения К-2 (датчик T0388)

Параметр температуры острого орошения К-2 оказывал наибольшее влияние (корреляцию) на по-

казания лабораторного анализа показателя качества (начало кипения фракции 80...180° С) в интервал времени со сдвигом от момента отбора пробы на величину от 3 ч 29 мин до 3 ч 18 мин.

Давление верха К-2 (датчик P0736).

Параметр оказывал наибольшее влияние (корреляцию) на показания лабораторного анализа показателя качества (начало кипения фракции 80...180° С)

в интервале взятия показаний датчика со сдвигом от момента отбора пробы на величину от 3 ч 47 мин до 3 ч 18 мин.

После описанного выше отбора переменных (температура верха колонны К-2 ТС0274; температура острого орошения колонны К-2 Т0388; давление верха колонны К-2 Р0736) и интервалов отбора данных были построены линейные регрессионные модели по методу наименьших квадратов. Для оценки качества модели брались следующие критерии [3]:

Фишера (F -критерий), коэффициент детерминации R^2 , среднеквадратическая ошибка, вычисляемые по следующим формулам:

$$F_{\text{факт}} = \frac{S_{\text{объясн}}^2}{S_{\text{ост}}^2}; \quad S_{\text{объясн}}^2 = \frac{\sum_{i=1}^N (y_i^3 - \bar{y})^2}{l};$$

$$\bar{y} = \frac{\sum_{i=1}^N y_i^3}{N}; \quad S_{\text{ост}}^2 = \frac{\sum_{i=1}^N (y_i^3 - \hat{y}_i)^2}{N - l - 1};$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i^3 - \hat{y}_i)^2}{\sum_{i=1}^N (y_i^3 - \bar{y})^2};$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}{N - 1}}; \quad \bar{\hat{y}} = \frac{\sum_{i=1}^N \hat{y}_i}{N};$$

где: $S_{\text{объясн}}^2$ - «объясненная» моделью дисперсия, $S_{\text{ост}}^2$ - остаточная дисперсия, y_i^3 - экспериментальное значение показателя качества, $\bar{y}$ - среднее выборочное значение показателя качества, $\hat{y}_i$ - расчетное по модели значение показателя качества, N - размер тестовой выборки, l - число параметров уравнения регрессии, $\bar{\hat{y}}$ - среднее расчетное значение

показателя качества, σ - среднеквадратическая ошибка значений, рассчитанных по модели. Чем меньше величина ошибки, тем выше качество модели.

Если $F_{\text{факт}} > F_{\text{табл}}$, где $F_{\text{табл}}$ это максимально возможное значение критерия Фишера F под влиянием случайных факторов при данных степенях свободы и уровне значимости α (определяется по статистическим таблицам), то коэффициент детерминации R^2 статистически значим (найденная оценка уравнения регрессии статистически надежна). Уровень значимости α - вероятность отвергнуть правильную гипотезу при условии, что она верна. В данной задаче α принимается равной 0,01.

Отметим, что модели строились в разных вариантивных подборках комбинации параметров, все модели построены по историческим данным июня и июля 2020 г. Эти месяцы выбраны из-за наличия данных лабораторных пробегов, которые увеличивают множество подтвержденных наблюдений показателя качества и позволяют построить модели, оптимально описывающие процесс. Были выделены следующие наилучшие варианты построения модели.

Модель №1

Модель построена по трем переменным (ТС0274, Т0388, Р0736), данные июня 2020 г. по выбранным временным интервалам усреднения параметров. Получено следующее регрессионное уравнение:

$$\text{ВА/НК80-180/К2М1} = 7,4691 + 1,72627 * \text{ТС0274} - 0,33641 * \text{Т0388} - 54,10544 * \text{Р0736}.$$

Модель №2

Модель построена по трем переменным, данные июня с добавлением данных пробега июля по выбранным временным интервалам усреднения параметров. Получено следующее регрессионное уравнение:

$$\text{ВА/НК80-180/К2М2} = -75,76666 + 3,272973 * \text{ТС0274} - 0,558446 * \text{Т0388} - 129,1792 * \text{Р0736}.$$

Модель №3

Модель построена по двум переменным, данные июня с добавлением данных пробега июля по выбранным временным интервалам усреднения параметров. Получено следующее регрессионное уравнение:

$$\text{ВА/НК80-180/К2М3} = 56,04316 + 1,698089 * \text{ТС0274} - 124,8201 * \text{Р0736}.$$

Критерии качества всех трех моделей представлены в табл. 1.

Таблица 1. Критерии качества моделей.

Множественный R	R^2	F	Значимость F	Число наблюдений, ед.
Модель 1				
0,83374	0,69512	12,92	0,000121	19
Модель 2				
0,689447	0,475338	6,643913	0,002309	26
Модель 3				
0,624515	0,390019	7,353057	0,003397	26

Нейросетевые модели

В качестве альтернативы традиционным моделям линейной регрессии были построены модели искусственных нейронных сетей (ИНС). Модели ИНС представляют собой отображение вектора входных переменных в вектор выходных. Характеристики данного отображения определяются топологической структурой ИНС: числом уровней, числом элементов каждого уровня, значениями синоптических связей (весов).

Обучение нейронных сетей - процесс подбора и настройки весов по определенному алгоритму (обратное распространение; Левенберга-Марквардта; сопряженных градиентов; методы переменной метрики; генетический алгоритм) [4].

После определения числа слоев и числа элементов в каждом из них ищутся значения для весов сети, которые бы минимизировали функцию ошибки прогноза. Таким образом, происходит подгонка модели к имеющимся обучающим данным. Ошибка для каждой конфигурации сети определяется посредством загрузки в сеть обучающей выборки имеющихся наблюдений и сравнения реально выдаваемых выходных значений с целевыми значениями.

На основании полученных зависимостей между физическими переменными процесса и температурой начала кипения фракции 80...180° С, выявленных в ходе построения линейных регрессионных моделей, для построения нейросетевой модели ВА были взяты те же параметры процесса (температура верха колонны К-2 TC0274; температура острого орошения колонны К-2 T0388; давление верха колонны К-2 P0736).

Моделирование производилось на выборке значений температуры начала кипения фракции 80...180° С за весь период лабораторных испытаний (июнь 2020 г. – февраль 2021 г.). Для нейросетевой модели они являются значениями целевой переменной. В качестве соответствующих им значений параметров процесса (значения входных переменных) для задач моделирования брались измеренные значения пара-

метров процесса со сдвигом на час назад относительно фиксированного времени лабораторного отбора проб.

Данные не подвергались фильтрации и преобразованиям в целях прозрачности схемы вычислений и получения прогноза на основе непосредственных производственных данных, получаемых от датчиков.

Моделирование ИНС производилось в аналитическом пакете Statistica (StatSoft). В качестве обучающей (контрольной) выборки брались случайным образом 85 % наблюдений, в качестве тестовой - оставшиеся 15 %. В результате моделирования и обучения была выбрана рассмотренная далее модель ИНС (модель MLP 3-9-1) как обладающая наилучшими показателями качества на тестовой выборке.

Критерии качества и оценки модели ИНС

Архитектура сети MLP 3-9-1:

- многослойный линейный перцептрон (МЛП);
- число скрытых нейронов 3 ед.;
- функция активации для скрытых нейронов – логистическая $\theta(s) = e^s / (1 + e^s)$;
- функция активации выходных нейронов – гиперболическая $\theta(s) = (e^{2s} - 1) / (e^{2s} + 1)$;
- функция ошибки – сумма квадратов
- алгоритм обучения – BFGS (алгоритм Бройде-на-Флетчера-Гольдфарба-Шанно для нахождения локального минимума нелинейного функционала).

Качество модели: $R^2_{обучения} = 0,41$; среднеквадратическая ошибка на обучающей выборке 2,8° С.

На тестовой выборке были получены следующие значения критерия качества модели: $R^2_{тест} = 0,51$; среднеквадратическая ошибка значений модели по тестовой выборке 2,04° С.

Обсуждение результатов

Сводные результаты по всем исследованным моделям представлены в табл. 2.

Подобранные модели линейной регрессии позволяют сделать вывод о снижении СКО и повышении корреляционной взаимосвязи между показаниями ВА и данными лабораторных анализов. Предпочтительнее выглядит модель №1, поскольку СКО построенной модели составляет 2,4° С, коэффициент детерминации R^2 составляет 0,62, что лучше используемой в настоящее время в системе управления модели, параметры которой 4,40° С и 0,21 соответственно. Для получения удовлетворительных результатов в случае с линейной регрессией необходим обоснованный выбор интервалов времени отбора технологических параметров для получения коэффициентов модели регрессии (рис.2).

Качество модели на основе искусственной нейронной сети (рис. 3) с многослойным перцептроном, тремя скрытыми нейронами и алгоритмом обучения BFGS, построенной в аналитическом пакете Statistica (StatSoft), сравнимо с качеством наилучшей из линейных моделей – СКО=2,04° С,

Таблица 2. Критерии качества построенных и текущих моделей.

Модель	Среднеквадратическая ошибка	R2
CAT_K2_80180 Old (Используемая в настоящее время)	4,40	0,21
ВА/НК80-180 K2M1 модель №1	2,4	0,62
ВА/НК80-180/ K2M2 модель №2	2,75	0,47
ВА/НК80-180/ K2M3 модель №3	2,53	0,39
Нейронная сеть MLP 3-9-1	2,04	0,51

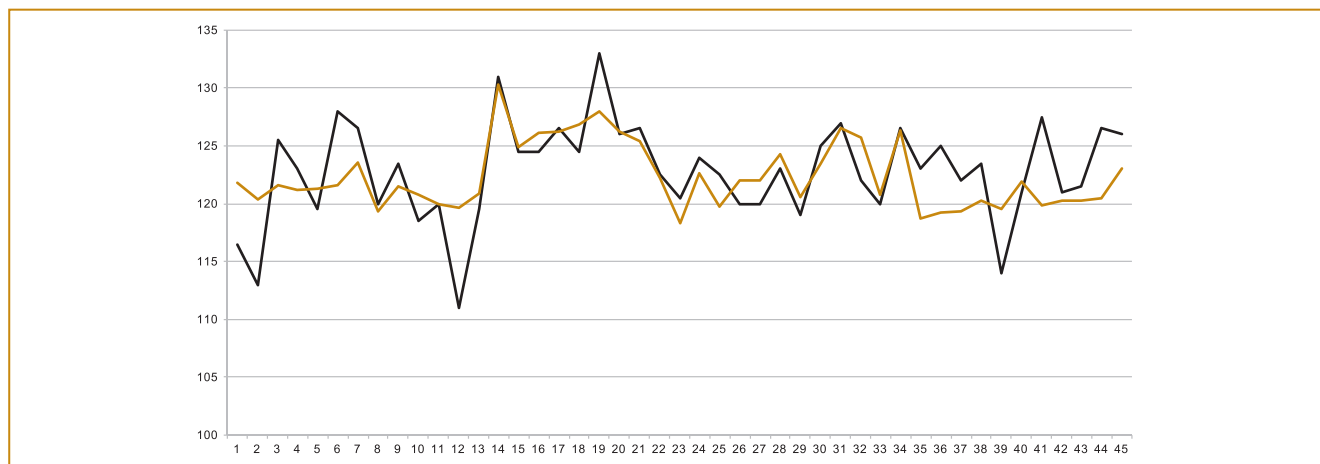


Рис. 2. Проверка показаний модели №1 на данных декабрь 2020 г. - март 2021 г. ВА/НК80-180/К2М1 (черный – лабораторные анализы; коричневый – ВА)

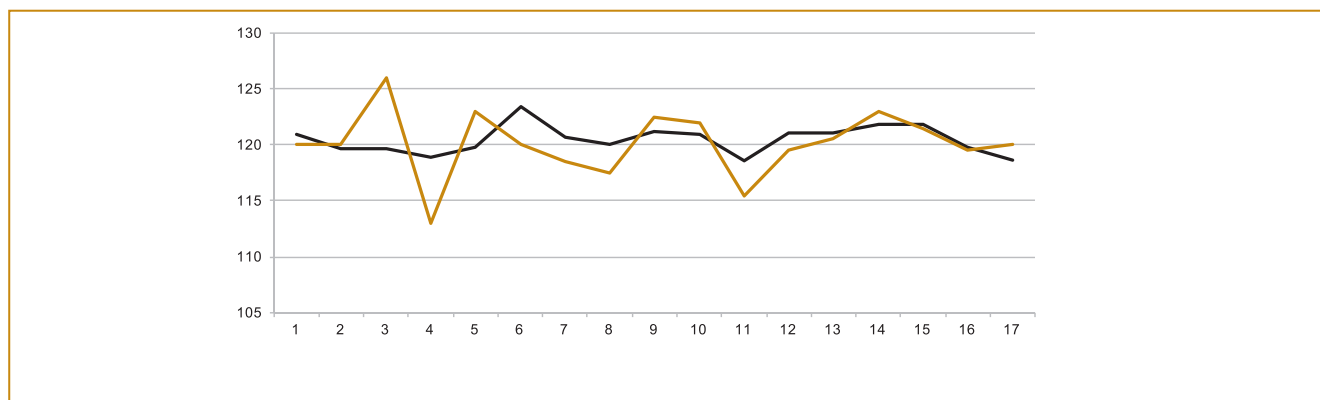


Рис. 3. Поведение модели MLP 3-9-1 для ВА на тестовой выборке (черный – лабораторные анализы; коричневый – ВА)

$R^2=0,51$, причем ошибка модели в данном случае даже ниже. При этом использование искусственных нейронных сетей позволяет избежать предварительной обработки и фильтрации данных и специального подбора временных интервалов.

Выводы и перспективы

Как показывает полученный авторами опыт и отмечается в научных публикациях [1, 5], нейронные сети предоставляют широкие возможности моделирования для построения систем ВА, однако не для всех из них возможно аналитическое представление уравнения регрессии, что делает необходимым поиск путей их адаптации для непосредственного использования в имеющихся контроллерах СУУТП. В то же время инструментальный линейных регрессионных моделей встроен в программное обеспечение систем усовершенствованного управления технологическими процессами и при своевременной актуализации моделей способен обеспечивать достаточное качество прогноза. При этом нейросетевые модели дают возможность проверять

линейные модели и исследовать потенциал увеличения качества прогноза.

Список литературы

1. Тугашова Л.Г., Виртуальные анализаторы показателей качества процесса ректификации. // Информационные комплексы и системы. 2013. № 3, Т. 9. С. 97-103.
2. Стрижов В.В., Крымова Е.А. Методы выбора регрессионных моделей - М: Вычислительный центр им. А. А. Дородницына РАН. 2010. - 60 с.
3. Прикладная статистика. Исследование зависимостей // Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Под ред. С.А. Айвазяна. - М.: Финансы и статистика, 1985. - 488 с.
4. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. М: Горячая линия-Телеком, 2002. - 382 с.
5. Таушева Е.В., Григорьева Ю.Л. Построение виртуального анализатора с помощью нейронных сетей. // 67-я научно-техническая конференция студентов, аспирантов и молодых ученых УГНТУ. Тр. конференции. — Изд-во УГНТУ (Уфа), 2016. — С. 520-521.

Лаврентьев Владимир Алексеевич — ведущий инженер, **Ергомышев Юрий Анатольевич** — ведущий инженер сектора по информатизации процессов нефтепереработки и нефтехимии, **Головина Евгения Сергеевна** — заместитель генерального директора по технологическому развитию, **Комогоров Александр Владимирович** — начальник сектора по информатизации процессов нефтепереработки и нефтехимии, АО "Самаранефтехимпроект".
E-mail: GolovinaES@snhp.ru