

<https://doi.org/10.17073/2072-1633-2024-4-1346>

Оперативное прогнозирование расхода топливного газа в газотранспортных обществах ПАО «Газпром»

А.А. Кудрявцев , С.Н. Ланин 

Санкт-Петербургский государственный экономический университет,
191023, Санкт-Петербург, наб. канала Грибоедова, д. 30-32, Российская Федерация
 s.lanin@gmail.com

Аннотация. В современных условиях экономии энергетических ресурсов и повышения эффективности газотранспортных систем актуальной задачей является разработка подходов для повышения точности прогнозирования расхода топливного газа на компрессорных станциях. В статье анализируются подходы и алгоритмы прогнозирования объемов газа, необходимого для технологических и собственных нужд компрессорных станций при сжатии (сжатию) газа в газотранспортных обществах ПАО «Газпром». Представлена классификация расходов газа на технологические нужды и потери, подчеркивается значимость управления расходом топливного газа для оптимизации себестоимости транспортировки природного газа.

Цель исследования — разработка подхода к оперативному прогнозированию расхода топливного газа на компрессорных станциях газотранспортных обществ, который позволит повысить экономическую эффективность и снизить эксплуатационные затраты газотранспортного общества. Для достижения цели решены следующие задачи: анализ существующих методов прогнозирования, исследование способов обработки данных и выявления ошибок и аномалий, сравнение различных моделей регрессии для обеспечения высокой точности прогнозов.

В исследовании использованы методы очистки и предварительной обработки данных, включая метод изоляционного леса (Isolation Forest) для обнаружения аномалий, а также различные регрессионные модели, такие как множественная линейная регрессия, RandomForestRegressor, CatBoostRegressor и XGBoost. Для сегментации данных применен кластерный анализ (KMeans), что позволило повысить точность моделей. Точность прогнозов оценивалась с помощью t -теста, F -теста и метрики средней абсолютной процентной ошибки (MAPE).

Результаты исследования подтвердили высокую точность предложенного подхода, что свидетельствует о его потенциале для оптимизации топливных затрат в газотранспортных обществах.

Ключевые слова: газовая промышленность, природный газ, компрессорные станции, топливный газ, расход газа, оперативное прогнозирование расхода газа, модели регрессии, экономическая эффективность, оптимизация себестоимости, анализ аномалий, ПАО «Газпром»

Для цитирования: Кудрявцев А.А., Ланин С.Н. Оперативное прогнозирование расхода топливного газа в газотранспортных обществах ПАО «Газпром». *Экономика промышленности*. 2024;17(4):401–423. <https://doi.org/10.17073/2072-1633-2024-4-1346>

Operative forecasting of fuel gas consumption in gas transportation companies

A.A. Kudryavtsev , S.N. Lanin 

St. Petersburg State University of Economics,
30-32 Griboedov Canal Emb., St. Petersburg 191023, Russian Federation
 s.lanin@gmail.com

Abstract. In the current context of energy resource conservation and increased efficiency of gas transportation systems, developing approaches to improve the accuracy of fuel gas consumption forecasting at compressor stations is a pressing task. This paper analyzes

approaches and algorithms for forecasting the volumes of gas needed for the technological and internal needs of compressor stations during gas compression within the gas transportation organizations of Gazprom PJSC. A classification of gas consumption for technological needs and losses is presented, emphasizing the importance of managing fuel gas consumption to optimize the cost of natural gas transportation.

The goal of this study is to develop an approach for operational forecasting of fuel gas consumption at compressor stations of gas transportation organizations, aimed at increasing economic efficiency and reducing operating costs. To achieve this goal, the following tasks were undertaken: analysis of existing forecasting methods, investigation of data processing techniques for detecting errors and anomalies, and comparison of various regression models to ensure high forecast accuracy.

The study employed data cleaning and preprocessing methods, including the Isolation Forest method for anomaly detection, as well as various regression models such as multiple linear regression, RandomForestRegressor, CatBoostRegressor, and XGBoost. Data segmentation was performed using cluster analysis (KMeans), which allowed for improved model accuracy. Forecast accuracy was assessed using t-tests, F-tests, and the mean absolute percentage error (MAPE) metric.

The results of the study confirmed the high accuracy of the proposed approach, demonstrating its potential for optimizing fuel costs in gas transportation organizations.

Keywords: gas industry, natural gas, compressor stations, fuel gas, gas consumption, operational forecasting of gas consumption, regression models, economic efficiency, cost optimization, analysis of anomalies, Gazprom PJSC

For citation: Kudryavtsev A.A., Lanin S.N. Operative forecasting of fuel gas consumption in gas transportation companies. *Russian Journal of Industrial Economics*. 2024;17(4):401–423. <https://doi.org/10.17073/2072-1633-2024-4-1346>

俄罗斯天然气工业股份公司天然气输送公司燃气消耗量的运营预测

A.A. 库德里亚夫采夫 , S.N. 拉宁 

圣彼得堡国立经济大学,

191023, 俄罗斯联邦圣彼得堡市格里博耶多夫运河沿岸街30-32号

 s.lanin@gmail.com

摘要: 在现代条件下, 节约能源和提高天然气运输系统效率的紧迫任务是开发提高预测压缩站燃气消耗量准确性的方法。本文分析了俄罗斯天然气工业股份公司天然气输送公司压缩天然气时压缩机的技术和辅助需求所需的燃气量的方法和算法。介绍了技术需求和损失所需的燃气消耗分类, 强调了燃气消耗管理对优化天然气输送成本的重要性。

该研究的目的是开发一种对天然气输送公司压缩机的燃气消耗进行运营预测的方法, 这将提高经济效率并降低天然气输送公司的运营成本。为了实现目标, 解决了以下任务: 分析现有预测方法, 研究数据处理和检测错误和异常的方法, 比较各种回归模型以确保预测的高精度。

研究采用了数据清洗与预处理方法, 包括用于检测异常的 Isolation Forest (孤立森林) 算法, 并利用各种回归模型, 如多元线性回归、RandomForestRegressor、CatBoostRegressor 和 XGBoost。聚类分析 (KMeans) 用于分割数据, 从而提高了模型的准确性。使用 T 检验、F 检验和平均绝对百分比误差 (MAPE) 评估预测的准确性。

研究结果证实了所提出方法的高准确度, 这表明该方法具有优化输气公司燃料成本的潜力。

关键词: 天然气工业; 天然气; 压缩机站; 燃气; 耗气量; 燃气量运营预测; 回归模型; 经济效益; 成本优化; 异常分析; 俄罗斯天然气工业股份公司

Введение

Природный газ имеет огромное значение для экономики России, выступая как один из главных источников доходов и гарантируя энергетическую безопасность страны. Обладая крупнейшими запасами природного газа в мире^{1,2} в размере 64,8 трлн м³, Россия является ведущим его экспортером, обеспечивая поставки в Европу, Азию и другие регионы. Экспорт природного газа вносит значительный вклад в валютные поступления и усиливает внешнеэкономические связи. Внутренний спрос на природный газ также велик, что поддерживает устойчивое развитие таких отраслей, как химическая промышленность, электроэнергетика, теплоэнергетика и жилищно-коммунальное хозяйство. Природный газ, являясь основным и важнейшим энергетическим ресурсом в России, способствует уменьшению зависимости от угля и нефти, что, в свою очередь, помогает сократить выбросы парниковых газов. Газовая отрасль не только обеспечивает рабочие места для миллионов граждан, но и играет важную роль в социальной стабильности. Кроме того, она способствует развитию передовых технологий и инноваций в сфере добычи, хранения, транспорта и переработки газа, что ускоряет общий технологический прогресс.

ПАО «Газпром» занимает практически монопольную² позицию в обеспечении природным газом потребителей в Российской Федерации, являясь крупнейшим добывающим предприятием и поставщиком газа не только в стране, но и в мире в целом³. Компания контролирует весь производственно-технологический цикл – от добычи до транспортировки, реализации и распределения газа, что гарантирует стабильные и надежные поставки. Используя современные технологии и инновационные методы разработки месторождений, ПАО «Газпром» добывает значительную часть природного газа в России. Компания управляет обширной сетью магистральных газопроводов, крупнейшей в мире и охватывающей большинство регионов страны, что позволяет эффективно и безопасно транспортировать газ. Важной частью деятельности ПАО «Газпром»

являются инвестиции в газификацию и модернизацию газовой инфраструктуры, включая новые газопроводы, газоизмерительные и газоперекачивающие станции, а также системы подземного хранения газа.

Общие подходы повышения экономической эффективности газотранспортной отрасли и оптимизации потоков газа в целом по ГТС приведены в [1, с. 665]. Расход топливного газа на компрессорных станциях существенно влияет на себестоимость транспортировки природного газа, что требует разработки точных методов прогнозирования его потребления и алгоритмов оптимизации потоков газа по ГТС на основании критерия минимизации потребления топливного газа. Это позволит не только снизить эксплуатационные затраты, но и обеспечить стабильность и надежность поставок газа. Настоящее исследование направлено на решение этих вопросов с использованием современных методов анализа данных и прогнозирования.

Пути повышения эффективности деятельности и структура себестоимости природного газа

Экспортные операции ПАО «Газпром» столкнулись с серьезными трудностями из-за нескольких факторов: сокращение поставок газа в Европу через территорию Польши после введенного запрета на транспортировку^{4,5}, прекращение экспорта через акваторию Балтийского моря вследствие терактов на газопроводах «Северный поток»⁶, а также невозможность транспортировки газа по территории Украины через ГИС «Сохранилка» в связи с отказом НАК «Нафтогаз Украины» принимать газ по этому маршруту

¹ Министерство цифрового развития, связи и массовых коммуникаций РФ. Режим доступа: <https://digital.gov.ru/ru/events/48746/> (дата обращения: 16.11.2024).

² Федеральный закон «О естественных монополиях» от 17.08.1995 № 147-ФЗ. Режим доступа: https://www.consultant.ru/document/cons_doc_LAW_7578/ (дата обращения: 16.11.2024).

³ ПАО «Газпром». Производство. Режим доступа: <https://www.gazprom.ru/about/production/> (дата обращения: 21.06.2024).

⁴ Указ Президента РФ от 03.05.2022 № 252 (ред. от 22.12.2022) «О применении ответных специальных экономических мер в связи с недружественными действиями некоторых иностранных государств и международных организаций». Режим доступа: https://www.consultant.ru/document/cons_doc_LAW_416210/?ysclid=lx520gyls6466550884 (дата обращения: 21.06.2024).

⁵ Постановление Правительства РФ от 11.05.2022 № 851 «О мерах по реализации Указа Президента Российской Федерации от 3 мая 2022 г. № 252» (с изм. и доп.) (вместе с «Перечнем юридических лиц, в отношении которых применяются специальные экономические меры», «Перечнем юридических лиц, осуществляющих деятельность в области военно-технического сотрудничества, в отношении которых применяются специальные экономические меры»). Режим доступа: <https://base.garant.ru/404594131/?ysclid=m2vmyaq0i6939420255> (дата обращения: 21.06.2024).

⁶ Смирнов В. Что известно о подрыве газопроводов «Северный поток» и «Северный поток-2». ТАСС. 7 февраля 2024. Режим доступа: <https://tass.ru/info/17258101> (дата обращения: 21.06.2024).

после начала СВО⁷. Дополнительную сложность вносит тот факт, что контракт на транспортировку газа через территорию Украины истекает в 2025 г., и его продление вызывает серьезные сомнения и неопределенность⁸. Это создает еще большую неопределенность в плане обеспечения стабильных поставок газа в Европу по этому маршруту. В условиях текущей геополитической напряженности и сложных отношений между Россией и Украиной перспектива этого транспортного соглашения остается весьма неопределенной.

Ввиду вышеизложенного, развитие внутреннего рынка газа является одной из приоритетных задач для ПАО «Газпром».

Главные направления увеличения объемов реализации природного газа следующие:

- 1) рост внутреннего потребления;
- 2) развитие СПГ-технологий для снабжения удаленных и труднодоступных регионов, где прокладка газопроводов технически невозможна или экономически нецелесообразна;
- 3) создание новых газоперерабатывающих мощностей.

Увеличение экономической эффективной газификации для внутреннего потребления газа

⁷ Дятел Т. Украина отключила треть транзитного газа. Страна останавливает подачу российского топлива через ГИС «Сохрановка». ТАСС. 10.05.2022. Режим доступа: <https://www.kommersant.ru/doc/5347383> (дата обращения: 21.06.2024).

⁸ Макарычев М. Bloomberg: ЕС ищет способ продлить транзит газа из РФ через Украину в 2025 году. 11.06.2024. Режим доступа: <https://rg.ru/2024/06/11/bloomberg-es-ishchet-sposob-prodlit-tranzit-gaza-iz-rf-cherez-ukrainu-v-2025-godu.html> (дата обращения: 21.06.2024).

в Российской Федерации практически достигло 100 %, два следующих направления использования природного газа сталкиваются с рядом проблем, и принимая во внимание санкционное давление, очень затруднительны. В связи с этим достаточно перспективным направлением повышения эффективности деятельности являются меры по оптимизации управления и сокращению затрат на транспортировку газа. Одним из ключевых элементов себестоимости являются расходы на природный газ, используемые для собственных технологических нужд (СТН). Этот газ необходим для работы газоперерабатывающих агрегатов компрессорных станций, которые обеспечивают его движение по трубопроводам.

На рис. 1 представлена структура бухгалтерских расходов ПАО «Газпром» в 2022 г. Затраты на СТН составили примерно 12 % от всех расходов на транспортировку природного газа по магистральным газопроводам.

На примере одного из дочерних газотранспортных обществ ПАО «Газпром» (ООО «Газпром трансгаз Чайковский») продемонстрирована доля затрат на приобретение газа для СТН, которая варьировалась от 12,9 % в 2023 г. до 23,5 % в 2021 г., составляя значительную часть от общей выручки, полученной газотранспортным обществом от основного вида деятельности. На рис. 2 представлена структура выручки по обычным видам деятельности дочернего газотранспортного общества (ГТО) ПАО «Газпром» за период 2020–2023 гг.

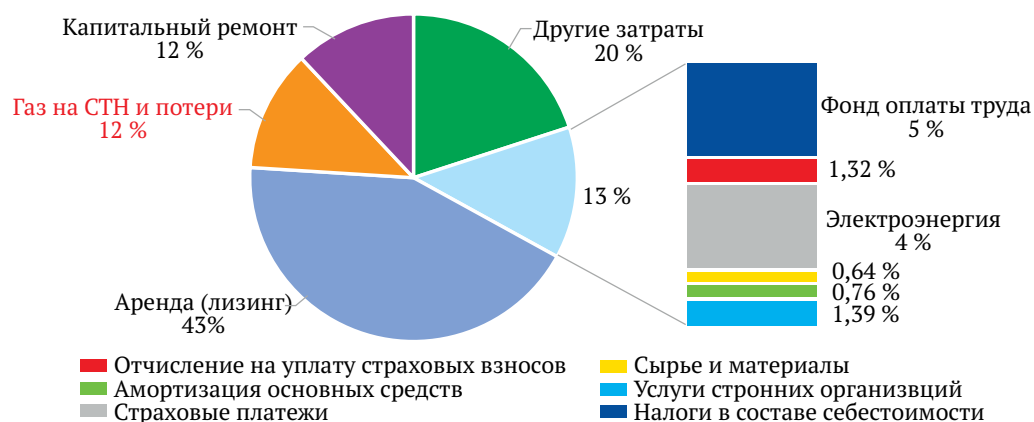


Рис. 1. Структура расходов на транспортировку газа в 2022 г.

Источник: составлено авторами с использованием данных бухгалтерской отчетности ПАО «Газпром» за 2022 г. Информация об услугах по транспортировке. Режим доступа: <https://www.gazprom.ru/investors/disclosure/transportation-services/> (дата обращения: 21.06.2024).

Fig. 1. Accounting expenses for gas transportation in 2022

Source: compiled by the authors using data from the financial statements of PJSC Gazprom for 2022. Available at: <https://www.gazprom.ru/investors/disclosure/transportation-services/> (accessed on 21.06.2024).

На рис. 3 приведены данные суточного расхода топливного газа и газа на СТН ООО «Газпром трансгаз Чайковский» по данным подсистемы хранения (ПХ) модернизированной (М) автоматизированной системы диспетчерского управления (АСДУ) Единой Системы Газоснабжения (ЕСГ) Российской Федерации (ПХ М АСДУ ЕСГ). Данные показывают, что в 2023 г. доля топливного газа, использованного для СТН на компрес-

сорных станциях (КС), составила 76 % от общего объема расходов газа на СТН.

На основании вышеизложенного можно сделать вывод, что топливный газ играет ключевую роль в формировании эксплуатационных затрат газотранспортного общества, а точное прогнозирование расхода топливного газа является важнейшей задачей для обеспечения эффективного управления ресурсами.

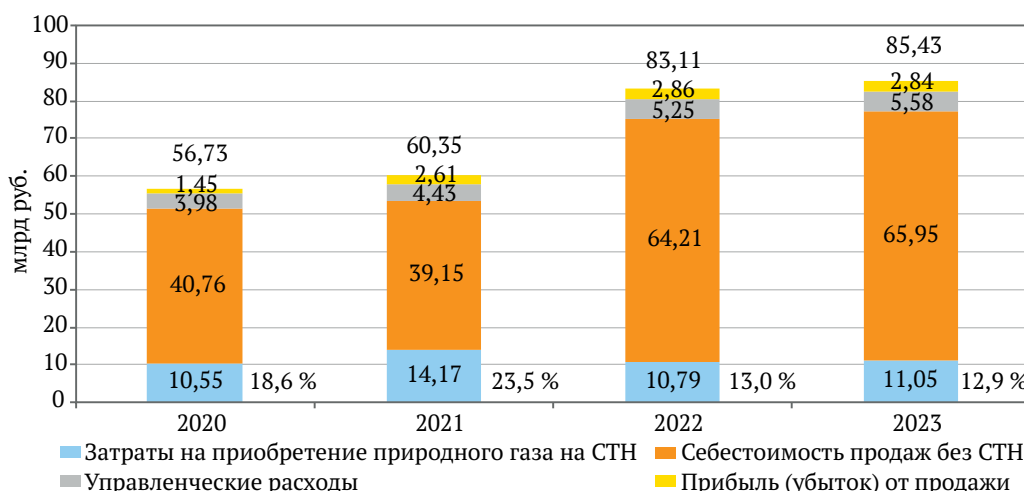


Рис. 2. Структура выручки по обычным видам деятельности рассматриваемого ГТО

Источник: составлено авторами на основе бухгалтерской отчетности дочернего общества ПАО «Газпром» (Интерфакс Спарк. Карточка компании ООО «Газпром трансгаз Чайковский». Доступ для зарегистрированных пользователей. Режим доступа: <https://spark-interfax.ru/system/home/card#/company/F2ED28267A8B4A70B44481BC4D7D16FB/508> (дата обращения: 07.06.2024))

Fig. 2. Revenue structure for ordinary activities of the gas transportation company under consideration

Source: compiled by the authors based on the financial statements of a subsidiary of PJSC Gazprom (<https://spark-interfax.ru/system/home/card#/company/F2ED28267A8B4A70B44481BC4D7D16FB/508> (available: 07.06.2024))

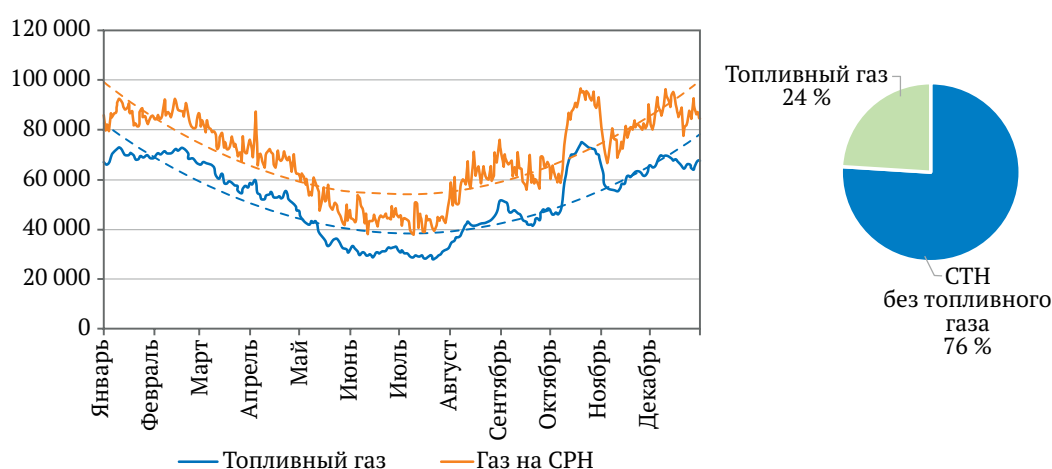


Рис. 3. Объем (тыс. м³) и доля (%) топливного газа и газа на СТН в 2023 г.

Источник: составлено авторами по данным подсистемы хранения модернизированной Автоматизированной системы диспетчерского управления Единой системы газоснабжения Российской Федерации

Fig. 3. Volume (thousand m³) and share (%) of fuel gas and gas for own technological needs in 2023

Source: compiled by the authors based on data from the storage subsystem of the modernized Automated Dispatch Control System of the Unified Gas Supply System of the Russian Federation

Значимость и роль точного прогнозирования расхода топливного газа

Точность прогнозирования расхода топливного газа в рамках СТН является критически важным аспектом для эффективного управления газотранспортными системами. Это способствует оптимизации эксплуатационных процессов, снижению издержек и повышению общей экономической эффективности. Точные прогнозы позволяют эффективно распределять ресурсы, планировать техническое обслуживание и ремонты, своевременно выявлять потенциальные проблемы. Это, в свою очередь, будет способствовать оптимизации эксплуатационных расходов, уменьшению потребности в избыточных топливных резервах и сокращению затрат на закупку газа.

Управление диспетчеризацией в ПАО «Газпром» построено по многоуровневой иерархической структуре⁹. Центральный производственно-диспетчерский департамент (ЦПДД) осуществляет контроль за газовыми потоками на уровне всей газотранспортной системы (ГТС) ЕСГ, обеспечивая бесперебойные поставки газа на внутреннем рынке, его транспортировку на экспорт в ближнее и дальнее зарубежье, а также управление потоками в условиях проведения ремонтных и профилактических работ при пиковом потреблении и изменении маршрутов транспортировки. Управление потоками природного газа последовательно транслируется на уровень газотранспортных обществ, затем на линейные управления и КС, где устанавливаются режимы работы газоперекачивающего оборудования. Каждый уровень управления сосредоточен на обеспечении эффективного и надежного функционирования ГТС ЕСГ. Для оптимизации энергозатрат в рамках всей ГТС необходимо применять единые подходы к расчетам распределения потоков, учитывая как топливный газ, так и электроэнергию, используемую для компримирования. Поскольку неэффективное распределение потоков не может быть исправлено на уровне дочерних обществ, на верхнем уровне управления должны применяться алгоритмы оптимизации потоков по всей системе, основывающиеся на энергетических критериях. Для таких алгоритмов оптимизации критически важен механизм расчета расхода топливного газа в зависимости от конфигурации газотранспортных потоков по всей ГТС.

⁹ СТО Газпром 8-003-2013 Диспетчерское управление. Общие положения: стандарт организации. Введ. 2014-03-24. М.: ОАО «Газпром»; 2014. IV. 17 с.

Обзор существующих подходов к прогнозированию расхода топливного газа

Сегодня в ПАО «Газпром» уже известны методические подходы, которые позволяют достаточно точно рассчитать расход топливного газа в зависимости от режима работы оборудования ГТС на уровне газотранспортного общества. Эти методы базируются на детальном гидравлическом моделировании режимов работы ГТС при оптимальной эксплуатации газоперекачивающего оборудования. Однако такой детализированный подход не может быть применен на верхнем уровне управления потоками, поскольку он сложен реализуем для задач, связанных с размерностью ГТС ЕСГ. ЦПДД не имеет доступа к отдельным газоперекачивающим агрегатам и не может прогнозировать их состояние и рабочие режимы.

Для того чтобы решить задачу определения затрат топливного газа на уровне ЕСГ, необходимо разработать обобщенные зависимости энергетических затрат на компримирование газа от объемов транспортировки по участкам ГТС. Эти зависимости должны исключать параметры режимов работы оборудования, такие как давления на входе и выходе компрессорных цехов (КЦ), степень сжатия и температуры газа. Благодаря этому подходу станет возможным проводить численный расчет энергетических затрат на уровне всей ГТС ЕСГ за приемлемое время. Такой подход позволит определить расход топливного газа и оптимизировать энергетические затраты на высоком уровне управления без необходимости детального моделирования работы каждого газоперекачивающего агрегата.

Хотя в мировой литературе можно встретить ссылки на теоретические исследования, касающиеся зависимости энергозатрат от потока газа, публикаций, посвященных решению прикладных задач на схемах ГТС с использованием таких зависимостей, найти не удалось.

В статье [2] подробно и детально описана ситуация с подходами, применяемыми в ПАО «Газпром» к учету затрат топливно-энергетических ресурсов в транспортировке газа газотранспортными обществами. В статье приведены локальные нормативные документы^{10,11}, при-

¹⁰ Методика определения норм расхода и нормативной потребности в природном газе на собственные технологические нужды магистрального транспорта газа РД 153-39.0-112-2001. Режим доступа: <https://meganorm.ru/Data1/39/39282/index.htm> (дата обращения: 21.06.2024).

¹¹ СТО Газпром 2-1.20-122-2007. Методика проведения энергоаудита компрессорной станции, компрессорных цехов с газотурбинными и электроприводными ГПА. Введен Распоряжением ОАО «Газпром» от 5 марта 2007 г. № 31

меняемые в процессе нормирования затрат на транспортировку природного газа по ГТС, в том числе и действующий стандарт организации (СТО)¹². Согласно данному стандарту, расчет топливно-энергетических затрат осуществляется с использованием специализированных программно-вычислительных комплексов, которые учитывают технические характеристики компрессорных станций, параметры установленного газоперекачивающего оборудования, включая количество агрегатов, их производительность и коэффициент полезного действия (КПД). При этом допускаются погрешности в измерениях или расчетах параметров энергоэффективности оборудования, которые могут варьироваться от 5,3 до 10,1% из-за инструментальных или методических неточностей¹³.

Кроме того, расчеты объемов топливно-энергетических ресурсов, необходимых на транспортировку, определяются на основании проектных значений, заложенных при строительстве газотранспортных объектов. В технической документации не всегда содержится полная информация по параметрам ненормальных режимов используемого и смонтированного оборудования. Программно-вычислительные комплексы оперируют именно проектными параметрами, которые на практике значительно отличаются от фактических параметров работы КС и газоперекачивающего оборудования из-за проведения ремонтов или замены узлов агрегатов на более современные.

Зависимость энергетических затрат на транспортировку газа по ГТС также описана в СТО Газпром 3.3-2-044-2016¹⁴. Этапы расчета затрат на транспортировку газа по ГТС в соответствии с СТО Газпром 3.3-2-044-2016 выглядят следующим образом:

Необходимо провести разбиение всей ГТС на участки. Разбиение должно прежде всего удовлетворять следующим требованиям:

- участки должны покрывать все основные газотранспортные коридоры, так как цель про-

ведения расчетов с использованием энергокритерия заключается в перераспределении потоков на уровне всей ГТС ЕСГ по коридорам;

- участки не должны быть слишком короткими, либо их число должно быть невелико;

- в рамках одного участка должно сохраняться неизменное рабочее давление за исключением временного снижения из-за ремонта или неудовлетворительного состояния оборудования;

- участок ГТС ЕСГ должен быть неразветвленным, а ветвления и межсистемные перемены моделируются в виде притоков и отборов.

Рассчитать коэффициенты энергокритерия для каждого участка. Необходимо определить ежемесячные значения коэффициентов, которые рассчитываются на основе приведенной в методике формулы. Для этого для каждого участка ГТС выполняется расчет в специализированном программно-вычислительном комплексе с использованием усредненных данных по месяцу температур окружающего воздуха и температур грунта. В результате рассчитываются значения энергозатрат, на основе которых определялся коэффициент энергокритерия. При этом на коэффициенты наложены следующие ограничения:

- не учитываются фактор ремонтных работ, фиксированные давления на входе и выходе в систему, а также фиксированная температура на входе в систему; фиксированные значения притоков и отборов по участку;

- фиксированные калорийность и относительная плотность газа;

- фиксированные температуры на выходе аппаратов воздушного охлаждения (АВО).

Суммарные энергозатраты на транспортировку газа по участку ГТС моделируются, как куб суммарного потока, взятый с некоторым коэффициентом энергокритерия¹⁵:

$$N = k \cdot q^3. \quad (1)$$

Утверждение, что суммарные энергозатраты (N) на участке ГТС моделируются как куб суммарного потока (q) с некоторым коэффициентом (k), имеет несколько слабых сторон. Во-первых, оно предполагает простую зависимость, которая может не учитывать сложные физические и технические аспекты работы. Во-вторых, такая модель может игнорировать влияние на энергозатраты температуры, давления и других факторов. В-третьих, использование единственного коэффициента может привести к неточ-

¹² Оптимизация потоков газа по ГТС ЕСГ с учетом энергозатрат на транспортировку при диспетчерском управлении : рекомендации организации. Введ. 2021-11-02. СПб.: Газпром экспо; 2021. V. 31 с. Режим доступа: https://rusneb.ru/catalog/000199_000009_010776655/

¹³ Там же, табл. 9.1

¹⁴ СТО Газпром 3.3-2-044-2016. Система норм и нормативов расхода ресурсов, использования оборудования и формирования производственных запасов ПАО «Газпром». Методика нормирования расхода природного газа на собственные технологические нужды и технологические потери магистрального транспорта газа (с изм. № 1). Введен Распоряжением ПАО «Газпром» от 22.12.2016 № 430.

¹⁵ Оптимизация потоков газа по ГТС ЕСГ с учетом энергозатрат на транспортировку при диспетчерском управлении: рекомендации организации. Введ. 2021-11-02. СПб.: Газпром экспо; 2021. 31 с.

ностям при моделировании реальных условий эксплуатации, что снижает точность прогнозов и эффективность управления энергозатратами. Частый перерасчет этого коэффициента приводит к дополнительным сложностям и нестабильности модели, затрудняя долгосрочное планирование и управление.

Разбиение на участки носит условный характер. В топологии реальной ГТС присутствуют короткие участки, участки без КС или с КС, работающими на проход. Это означает, что не все участки будут иметь однородные эксплуатационные условия. Участки могут находиться в зоне эксплуатации как одного ГТО, так и нескольких. В таких случаях отнесение энергозатрат к конкретному ГТО становится затруднительным, что усложняет учет и оптимизацию затрат на транспортировку газа. Кроме того, сама по себе кубическая зависимость не ограничена справа, и результаты, полученные по формуле (1) при росте расхода газа через участок, могут принимать значения, выпадающие за пределы эксплуатационных характеристик оборудования.

ЦПДД оперативно прогнозирует расход газа на собственные нужды КС в рамках всего ГТО или ГТС, однако предложенная методика¹⁶ не способна обеспечить такой оперативный результат. Разбиение ГТС дочернего газотранспортного общества на участки, расчет энергетических затрат по этим участкам на основе имеющихся коэффициентов, а затем обратная консолидация данных для всей ГТС представляет собой достаточно сложную задачу. Кроме того, рассчитанные коэффициенты, определенные на основе фактических режимов работы ГТС в прошлом, не подходят для прогнозирования текущих расходов топливного газа, так как они могут не учитывать изменений в конфигурации ГТС.

Предлагаемый подход к прогнозированию расхода топливного газа на СТН

Как уже указывалось, вышеописанные подходы основаны на детальном гидравлическом моделировании, направлены на подбор оптимальных режимов работы газоперекачивающего оборудования и зависят от корректности расчетов коэффициентов энергокритерия, которые рассчитывает подрядная организация. Такие де-

тализованные методы неприменимы на верхнем уровне управления потоками, так как они труднореализуемы для задач размерности ГТС ЕСГ. Однако ЦПДД обычно не оперирует режимами работы отдельных газоперекачивающих агрегатов. В связи с этим остро встает вопрос разработки простой, быстрой и достаточно точной модели расчета суммарных энергозатрат ГТО на транспортировку газа.

Кроме того, достаточно простую с точки зрения исходных данных модель предполагается использовать в оптимизационных моделях при расчете оптимальных маршрутов транспортировки газа или управления качеством и компонентным составом газа при газоснабжении потребителей и поставках газа на экспорт. Необходимо построить обобщенные зависимости энергетических затрат на компримирование газа от объемов транспортировки в целом по ГТО, в которых не будут явно присутствовать параметры режимов работы оборудования. Такие зависимости должны давать возможность численно проводить прогноз энергетических затрат на уровне всей ГТС ЕСГ за приемлемое время. Кроме того, модель должна быть достаточно простой для использования в работе и внесения в нее необходимых корректировок сотрудниками ЦПДД без привлечения сторонних и подрядных организаций. Внедрение такой модели в существующие системы управления и планирования ПАО «Газпром» позволит быстро реагировать на изменения эксплуатационных условий и эффективно использовать возможности ГТС.

Целью данного исследования является разработка подхода оперативного прогнозирования расхода природного газа на СТН КС на основе укрупненных балансовых данных. Исходной информацией для модели служат данные о входящих и выходящих потоках газа в зоне деятельности ГТО. Успешностью работы модели считается преодоление погрешности, установленной СТО¹⁷ (от 5,3 до 10,1 %).

Описание работы газотранспортного общества. Укрупненная схема потоков газа любого газотранспортного общества изображена на рис. 4.

На рисунке фигурами зеленого цвета показаны источники поступления газа как в зону деятельности ГТО, так и внутри зоны, фигуры желтого цвета – распределение газа.

¹⁶ СТО Газпром 3.3-2-044-2016. Система норм и нормативов расхода ресурсов, использования оборудования и формирования производственных запасов ПАО «Газпром». Методика нормирования расхода природного газа на собственные технологические нужды и технологические потери магистрального транспорта газа (с изм. № 1). Введен Распоряжением ПАО «Газпром» от 22.12.2016 № 430.

¹⁷ Методика оценки энергоэффективности газотранспортных объектов и систем. СТО Газпром 2-3.5-113-2007. Введ. 2007-11-15. Режим доступа: <https://files.stroyinf.ru/Data1/54/54561/?ysclid=m2vpf8gytc635534984> (дата обращения: 21.06.2024).

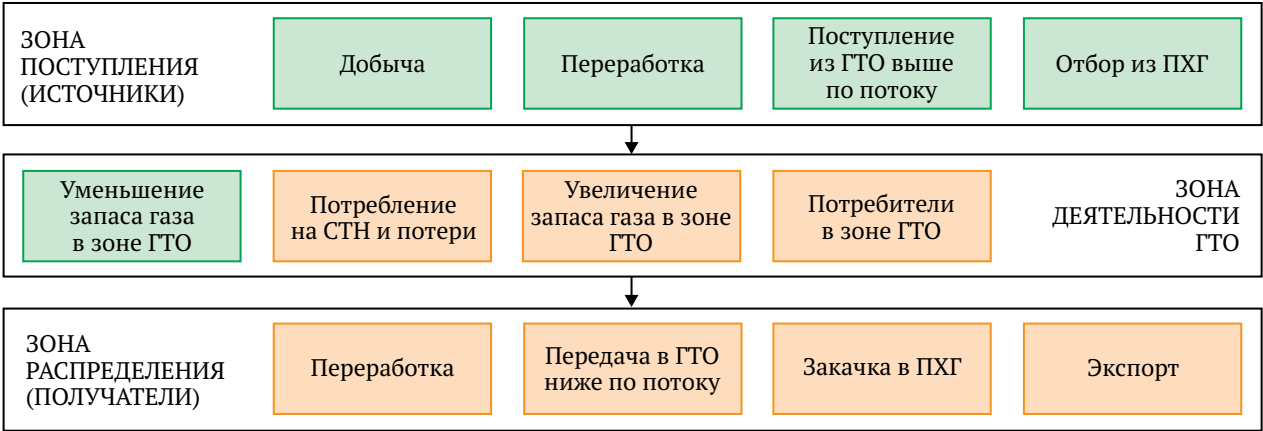


Рис. 4. Схема потоков газа газотранспортного общества

Источник: составлено авторами по данным подсистемы хранения модернизированной Автоматизированной системы диспетчерского управления Единой системы газоснабжения Российской Федерации

Fig. 4. Gas flow diagram of the gas transportation company

Source: compiled by the authors based on the data of the storage subsystem of the modernized Automated Dispatch Control System of the Unified Gas Supply System of the Russian Federation

Таблица 1 / Table 1

Первичный состав набора данных для расчетов

Primary composition of the data set for calculations

Зона	Количество параметров	Количество значений параметра	Тип значения
Поступления	9	3700 non-null	float64
Деятельности ГТО	5		
Распределения	3		

Источник: составлено авторами по данным подсистемы хранения модернизированной Автоматизированной системы диспетчерского управления Единой системы газоснабжения Российской Федерации

Source: compiled by the authors based on data from the storage subsystem of the modernized Automated Dispatch Control System of the Unified Gas Supply System of the Russian Federation

Описание и первичная обработка набора данных. Для создания модели прогнозирования расхода газа на СТН КС использовались данные балансового учета за период более 10 лет модельного газотранспортного общества ПАО «Газпром» (ООО «Газпром трансгаз Краснодар»). В наборе данных присутствуют исключительно потоковые параметры работы ГТО. Выбор такого набора параметров связан с тем, что при моделировании схемы потоков в целом по ГТС для управления калорийностью или качеством газа в заданных точках или оптимизации расхода топливного газа на КС по всей ГТС диспетчер или технолог ЦПДД оперирует именно потоками, а не режимами работы отдельных КС или агрегатов. Диспетчер в рамках заданий ГТО может изменить поток по одному из направлений, установить уровень отбора или закачки в подземные хранилища газа (ПХГ). Другие параметры, такие как внутреннее потребление промышленности

и населения в зоне ГТО, объем газа на экспорт в ближнее и дальнее зарубежье, диспетчером в основном изменяться не могут. Структура первичного набора данных для расчетов модельного газотранспортного общества ПАО «Газпром» представлена в **табл. 1**.

Первым шагом создания моделей является анализ выборки на наличие аномалий, выбросов и пропусков. Необходимо установить допустимые интервалы, в пределах которых данные могут существовать, а также проверить соответствие значений параметров по отношению к соседним параметрам. В настоящее время существует множество методов и подходов для решения проблемы идентификации и устранения ошибок, аномалий и пропусков во временных рядах данных. В [3–6] достаточно подробно описаны такие методы и подходы, как использование автокодировщиков или ансамблей алгоритмов DBSCAN. В рамках данного исследования про-

цесс обнаружения аномалий и выбросов будет осуществляться в два этапа:

Этап 1. Объемы запасов газа в ГТС, топливного газа, потребление внутри ГТО по своей сущности не могут быть меньше нуля. Данные, содержащие значения этих параметров равными 0, являются недостоверными. В выборке было обнаружено 304 дня с нулевыми значениями этих параметров, они были удалены.

Этап 2. Для окончательной очистки данных от аномалий необходимо применить один или комбинацию методов очистки данных от аномалий. Автором был применен метод изоляционного леса (Isolation Forest)¹⁸. Метод изоляционного леса – это алгоритм, предназначенный для обнаружения аномалий и выбросов данных. Он работает по принципу изоляции точек данных, предполагая, что аномалии отличаются от нормальных точек и встречаются реже. По данному методу строится множество деревьев решений (изоляционных деревьев), каждое из которых обучается на случайном подмножестве данных. Узлы в этих деревьях случайным образом выбирают признаки и значения для разбиения данных, что позволяет деревьям быстро изолировать выбросы.

Основная идея алгоритма заключается в том, что аномалии легче изолировать, чем нормальные точки данных. В процессе построения деревьев аномалии будут изолированы на меньшей глубине (меньшее количество разбиений), что отражает их редкость и отличие от основной массы данных. Оценка аномальности каждой точки данных основывается на средней глубине изоляции среди всех деревьев в лесу. Чем короче путь до изоляции точки, тем выше ее оценка аномальности. Isolation Forest эффективен на больших наборах данных, не требует предположений о распределении данных и способен обрабатывать многомерные данные, что делает его важным инструментом в таких областях, как обнаружение мошенничества, кибербезопасность и мониторинг производственных процессов.

Параметры используемой модели Isolation Forest, кроме доли выбросов (0,2), установлены равными дефолтным библиотеки scikit-learn¹⁹ и представлены в табл. 2.

¹⁸ Liu F.T., Ting K.M., Zhou Z.H. Isolation forest. In: 2008 8th IEEE Inter. conf. on data mining. IEEE. December 15–19, 2008. Piza, Italy. IEEE; 2008. P. 413–422. Available at: <https://doi.org/10.1109/ICDM.2008.17>

¹⁹ Scikit-learn: библиотека машинного обучения для языка программирования Python. Режим доступа: <https://scikit-learn.org> (дата обращения: 16.11.2024).

Таблица 2 / Table 2

Параметры модели Isolation Forest

Isolation Forest model parameters

Параметр	Описание	Значение
contamination (Доля выбросов)	Доля выбросов в данных, которую модель должна искать	2,0 %
n_estimators (Количество деревьев)	Количество деревьев в лесу	100
max_samples (Количество выборок)	Количество выборок, используемых для построения каждого дерева	auto
max_features (Количество признаков)	Количество признаков, рассматриваемых при каждом разбиении	1,0
random_state (Начальное значение)	Устанавливает начальное значение генератора случайных чисел для обеспечения воспроизводимости результатов	None

Источник: документация библиотеки scikit-learn, версия 1.5. Режим доступа: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html> (дата обращения: 15.07.2024).

Source: scikit-learn library documentation, version 1.5. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html> (accessed on 15.07.2024).

После применения метода Isolation Forest и удаления нулевых значений в выборке осталось 3328 элементов.

Для проверки работы моделей данные были разбиты на 2 части – обучающая выборка, 3298 первых значений и выборка для тестирования модели (30 последних значений).

Критерии оценки моделей. Для оценки качества моделей использовались *t*-тест (тест Стьюдента), который определяет наличие значимых различий между средними значениями двух групп и описанного в [7], и *F*-тест [8], применяемый для сравнения дисперсий двух или более групп, что позволяет выявить значимые различия в дисперсии между ними. Оба теста помогают оценить точность и надежность моделей прогнозирования.

Для оценки качества различных моделей прогнозирования использовалась метрика средней абсолютной процентной ошибки MAPE (*Mean Absolute Percentage Error*). Метрика представляет собой показатель, который измеряет среднее значение абсолютной процентной ошибки между прогнозируемыми и фактическими значениями, служа основным критерием для сравнения моделей. Она рассчитывается как среднее значение

абсолютных ошибок, выраженных в процентах, что делает ее легко интерпретируемой и удобной для сравнения моделей. Формула MAPE²⁰ выглядит следующим образом:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%, \tag{2}$$

где y_i – фактическое значение; $\hat{y}_i$ – предсказанное значение; n – число наблюдений.

Модель линейной регрессии. В качестве первой из протестированных модели была использована многофакторная линейная регрессия, также известная как множественная линейная регрессия и описанная в [9]. Она представляет собой расширение простой линейной регрессии, которое позволяет моделировать зависимость целевой переменной от нескольких независимых переменных. В этом подходе модель пытается установить линейную связь между целевой переменной и множеством признаков, что позволяет более точно предсказывать значения целевой переменной на основе нескольких факторов.

Основное преимущество многофакторной линейной регрессии заключается в ее способности учитывать влияние нескольких факторов одновременно, что делает ее полезной в более сложных ситуациях, где множество признаков может влиять на результат. Модель легко интерпретировать, что позволяет понять вклад каждого признака в предсказание целевой переменной. Однако,

²⁰ Средняя абсолютная ошибка (MAPE): практика применения. Режим доступа: <https://4analytics.ru/metodicheskaya-analiza/mape-srednyaya-absolyutnaya-oshibka-praktika-primeneniya.html> (дата обращения: 16.11.2024).

как и в случае с простой линейной регрессией, многофакторная линейная регрессия предполагает наличие линейной связи между переменными, и ее точность может снижаться при наличии нелинейных взаимосвязей или выбросов в данных.

Результаты расчета расхода топливного газа за 30 дней с помощью модели многофакторной линейной регрессии приведены на **рис. 5**.

Результаты оценки модели многофакторной линейной регрессии приведены в **табл. 3**.

Поскольку p -значение больше 0,05, то нет оснований отвергнуть нулевую гипотезу. Это означает, что на уровне значимости 0,05 нет статистически значимых различий между истинными данными и данными, предсказанными моделью. Таким образом, можно заключить, что предсказанные моделью данные не отличаются существенно от истинных данных с точки зрения их средних значений.

Параметры модели линейной регрессии приведены в **табл. 4**.

Таблица 3 / Table 3
Оценки качества модели линейной регрессии
Quality assessment of the linear regression model

Параметр оценки качества модели	Значение	
	Обучающая выборка	Тестовая выборка
t-статистика	0,000	0,820
p-значение	1,000	0,415
F_test	0,076	0,546
MAPE	23,22 %	4,73 %

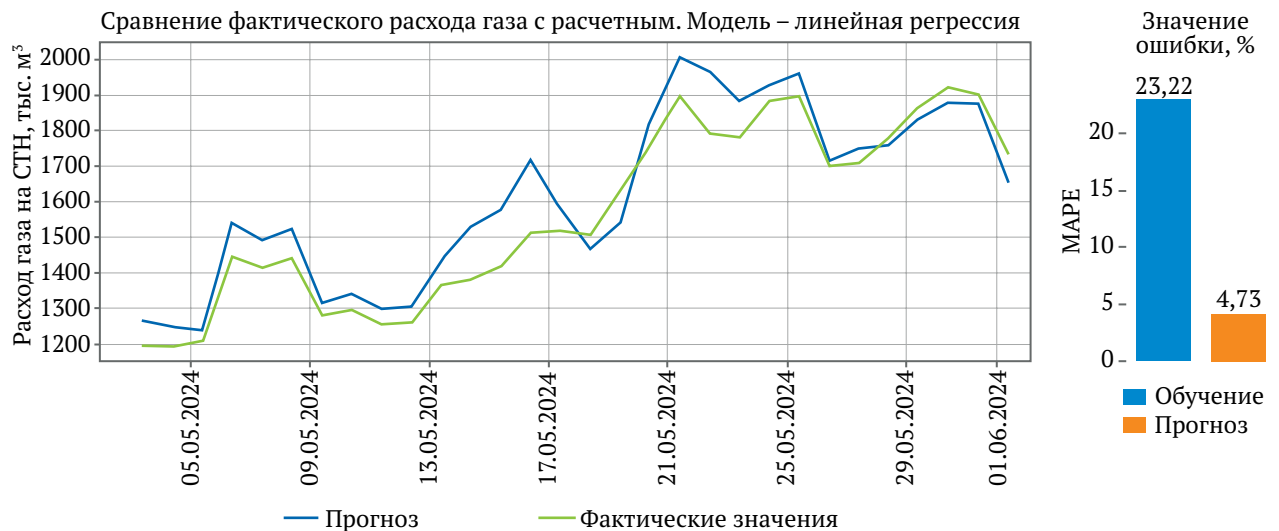


Рис. 5. Результаты работы модели линейной регрессии
Fig. 5. Results of the linear regression model

Модель RandomForestRegressor. Описанный в [10] метод RandomForestRegressor является мощным и универсальным методом машинного обучения, широко применяемым для решения задач регрессии. Данный метод относится к ансамблевым, так как основывается на объединении предсказаний нескольких деревьев решений для повышения точности и устойчивости результатов. Основной принцип RandomForest заключается в создании множества независимых деревьев решений на различных подмножествах данных и признаков, после чего их предсказания усредняются для формирования окончательного результата. Такой подход позволяет существенно снизить вероятность переобучения и улучшить способность модели к обобщению.

Каждое дерево в случайном лесу строится на основе случайной выборки данных (с использованием метода бутстреп (bootstrap)) и случайного подмножества признаков, что придает каждому дереву уникальность. Это способствует повышению производительности модели, особенно при работе с данными, имеющими высокую дисперсию или большое количество признаков. Кроме того, RandomForestRegressor обладает функцией оценки важности признаков, что позволяет определить, какие из них оказывают наибольшее влияние на предсказания модели. Благодаря этому метод становится полезным не только для построения предсказательных моделей, но и для глубокого анализа данных, позволяя лучше понять структуру и значимость различных факторов.

Ключевыми преимуществами RandomForestRegressor являются его высокая точность, устойчивость к выбросам и шуму в данных, а также

способность эффективно обрабатывать большие и сложные наборы данных. Тем не менее, как и другие сложные модели, случайный лес может быть ресурсоемким, особенно при работе с большим количеством деревьев или очень объемными наборами данных, что может потребовать значительных вычислительных мощностей.

Параметры использованной модели представлены в табл. 5.

Таблица 4 / Table 4
Значения коэффициентов уравнения линейной регрессии
Values of the coefficients of the linear regression equation

Параметр	Значение регрессора
Месторождение (1)	–0,011626
ГПЗ 1	0,014428
Месторождение (2)	–0,000169
Месторождение (3)	17,705003
От ГТО выше по потоку (1)	–0,002336
От ГТО выше по потоку (2)	–0,003127
От ГТО выше по потоку (3)	–0,003528
От ГТО выше по потоку (4)	0,003474
Экспорт (1)	0,011254
Экспорт (2)	0,023112
Собственным потребителям	0,008708
В ГТО ниже по потоку (1)	0,010855
В ГТО ниже по потоку (2)	0,011245
ПХГ, отбор	–0,004037
ПХГ, закачка	0,019217
Запас газа	–0,000083
Коэффициент e	–478,886080

Таблица 5 / Table 5
Параметры модели RandomForestRegressor
RandomForestRegressor model parameters

Параметр	Описание	Значение
n_estimators (Количество деревьев)	Количество деревьев в лесу влияет на точность и стабильность модели: большее количество деревьев обычно приводит к более надежным результатам	20
max_features (Количество признаков)	Количество признаков, рассматриваемых при каждом разбиении, определяет разнообразие деревьев в лесу, что способствует снижению корреляции между ними	1
max_depth (Максимальная глубина)	Максимальная глубина деревьев ограничивает их рост, что помогает избежать переобучения и улучшить способность модели к обобщению	–
min_samples_leaf (Минимальное количество выборок листа)	Минимальное количество выборок, необходимое для каждого листа, влияет на размер листьев деревьев, контролируя уровень детализации, на котором модель принимает решения	1
min_samples_split (Минимальное количество выборок для деления)	Минимальное количество выборок, необходимое для разделения узла, определяет сложность деревьев, влияя на их способность к разделению данных на более мелкие категории	2

Источник: документация библиотеки scikit-learn, версия 1.5. Режим доступа: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html> (дата обращения: 15.07.2024).
Source: scikit-learn library documentation, version 1.5. Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html> (accessed on 15.07.2024).

Таблица 6 / Table 6

Оценка модели RandomForestRegressor
Results of the RandomForestRegressor model

Параметр оценки качества модели	Значение	
	Обучающая выборка	Тестовая выборка
t -статистика	0,007512	-1,31867
p -значение	0,994006	0,192462
F_{test}	0,309581	0,163173
MAPE	4,61 %	7,68 %

Результаты расчета расхода топливного газа за 30 дней с помощью модели RandomForestRegressor приведены на **рис. 6**.

Результаты оценки модели RandomForestRegressor представлены в **табл. 6**.

Поскольку p -значение больше 0,05, нет оснований отвергнуть нулевую гипотезу при использовании этой модели.

Модель показала достаточно высокую ошибку на тестовых данных, хотя на данных тренировочной выборки ошибка была на уровне модели линейной регрессии. Вероятно, модель «переобучилась» и «запомнила» тренировочные данные, вследствие этого сильно ошибалась на тестовых данных. Для повышения точности работы модели необходимо провести отдельные исследования, подобрать параметры модели для ее более качественной работы.

Модель CatBoostRegressor. Разработанная компанией Яндекс модель CatBoostRegressor [11] – современный и эффективный алгоритм гради-

ентного бустинга, специально оптимизированный для работы с категориальными признаками. CatBoost (Categorical Boosting) способен автоматически обрабатывать категориальные данные, что избавляет пользователя от необходимости предварительно кодировать их вручную. Этот регрессор хорошо справляется с задачами, где есть смешанные типы данных, и обеспечивает высокую точность и производительность даже при наличии большого количества категориальных признаков.

Одной из основных особенностей CatBoost является способность минимизировать переобучение за счет использования схемы случайного перемешивания данных и специальной обработки категориальных признаков во время обучения. Это делает модель более устойчивой и надежной, особенно при работе с небольшими и средними по размеру наборами данных. CatBoost также выделяется высокой скоростью обучения и предсказания, что делает его идеальным для использования в продуктивных системах и приложениях реального времени.

Модель CatBoostRegressor допускает изменение только одного параметра²¹ – функцию потерь (loss_function). По умолчанию функция потерь задана как среднеквадратичная ошибка (Root Mean Square Error, RMS Error, RMSE).

Результаты расчета расхода топливного газа за 30 дней моделью CatBoostRegressor приведены на **рис. 7**.

²¹ CatBoostRegressor, версия 1.1. Режим доступа: https://catboost.ai/en/docs/concepts/python-reference_catboostregressor (дата обращения: 15.07.2024).

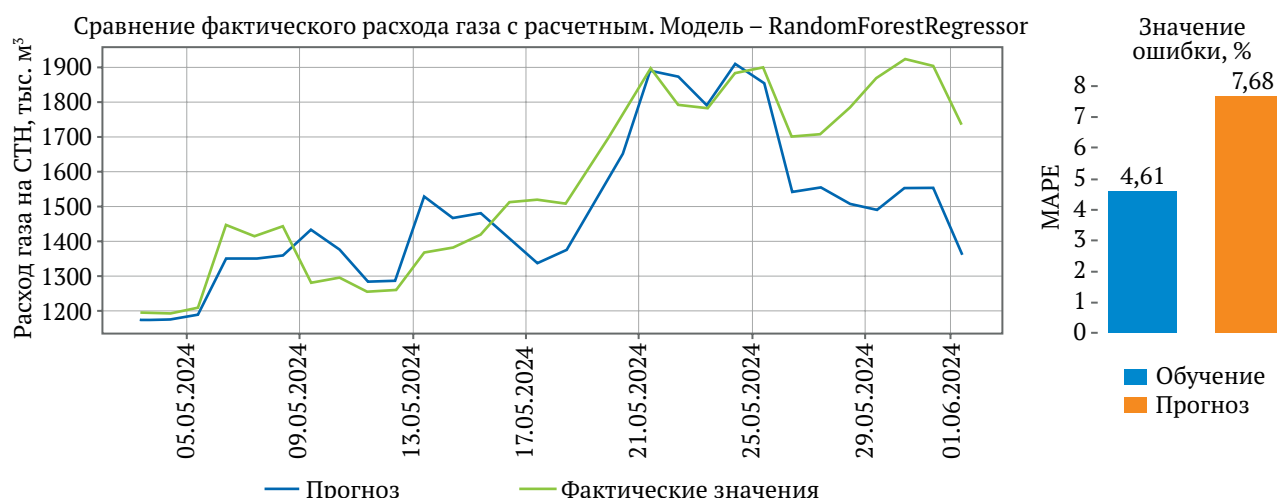


Рис. 6. Результаты работы модели RandomForestRegressor

Fig. 6. Results of the RandomForestRegressor model

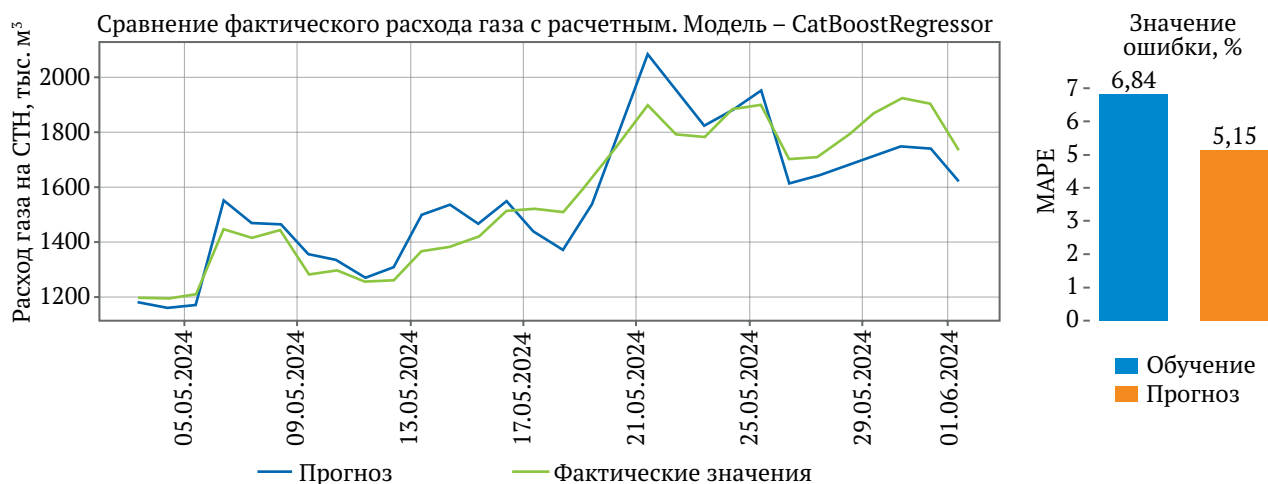


Рис. 7. Результаты работы модели CatBoostRegressor

Fig. 7. Results of the CatBoostRegressor model

Таблица 7 / Table 7

Оценка модели CatBoostRegressor

Evaluation of the CatBoostRegressor model

Параметр оценки качества модели	Значение	
	Обучающая выборка	Тестовая выборка
<i>t</i> -статистика	–0,00105	–0,04925
<i>p</i> -значение	0,999162	0,960892
<i>F</i> -test	0,412757	0,439485
MAPE	6,84 %	5,15 %

Результаты оценки модели CatBoostRegressor приведены в табл. 7.

Как видно из табл. 7, модель на основе CatBoostRegressor так же, как и модель линейной регрессии, показала достаточно высокую точность прогнозирования (MAPE = 5,15 %), что ниже предела погрешности, установленной СТО ПАО «Газпром» 2-3.5-113-2007²².

Модель XGBRegressor. Алгоритм XGBRegressor, описание которого приведено в [12] – это мощный и эффективный алгоритм регрессии, основанный на методе градиентного бустинга. Разработанный в рамках библиотеки XGBoost, он выделяется высокой производительностью и гибкими возможностями настройки, что делает его популярным среди специалистов по анализу данных и машинному обучению. XGBRegressor сочетает преиму-

щества бустинга и деревьев решений, формируя ансамбль слабых моделей, которые последовательно корректируют ошибки предыдущих в целях минимизации функции потерь.

Одной из ключевых особенностей XGBRegressor является способность обрабатывать большие и сложные наборы данных с высокой скоростью и эффективностью. Он включает множество параметров для настройки, таких как глубина деревьев, скорость обучения, регуляризация и частота подвыборок, что позволяет пользователям оптимизировать модель для достижения наилучших результатов. Кроме того, XGBoost поддерживает параллельные вычисления и оптимизации на уровне аппаратных средств, что ускоряет процесс обучения и делает его подходящим для использования в приложениях, требующих высокой производительности.

XGBRegressor предлагает мощные средства для анализа важности признаков, что позволяет выявить, какие факторы оказывают наибольшее влияние на результаты модели. Это делает его не только полезным инструментом для построения предсказательных моделей, но и ценным для глубокого анализа данных и получения важных инсайтов.

Для построения модели были использованы предлагаемыми библиотекой scikit-learn параметры по-умолчанию²³.

Результаты расчета расхода топливного газа за 30 дней моделью XGBRegressor приведены на рис. 8.

²² Методика оценки энергоэффективности газотранспортных объектов и систем. СТО Газпром 2-3.5-113-2007. Введ. 2007-11-15. Режим доступа: <https://files.stroyinf.ru/Data1/54/54561/?ysclid=m2vpf8gvtc635534984> (дата обращения: 21.06.2024).

²³ Документация библиотеки XGBoost, версия 1.7. Режим доступа: https://xgboost.readthedocs.io/en/stable/python/python_api.html#xgboost.XGBRegressor (дата обращения: 15.07.2024).

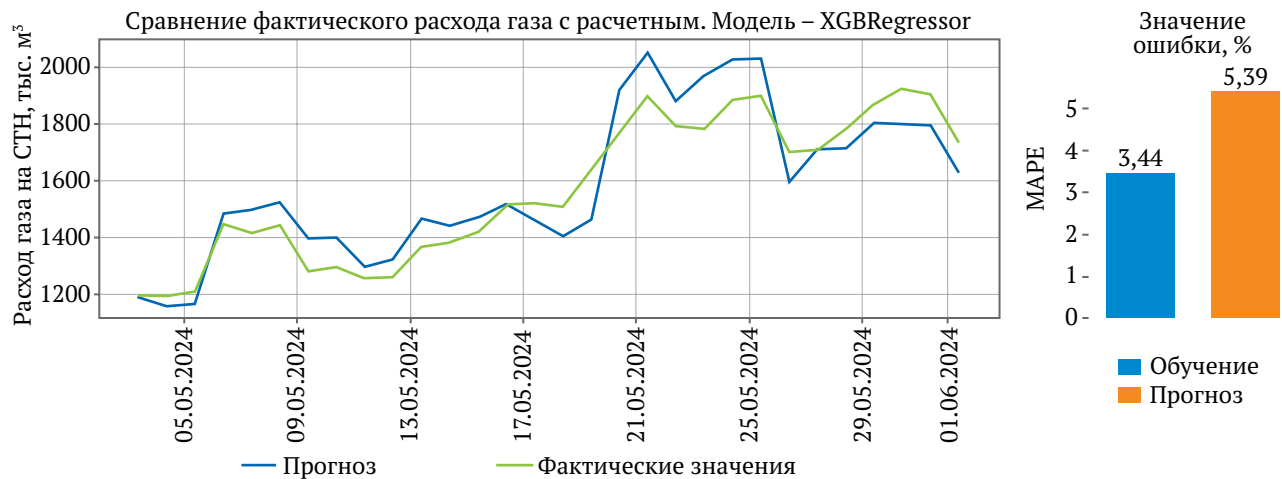


Рис. 8. Результаты работы модели XGBRegressor

Fig. 8. Results of the XGBRegressor model

Таблица 8 / Table 8

Оценка модели XGBRegressor

Evaluation of the XGBRegressor model

Параметр оценки качества модели	Значение	
	Обучающая выборка	Тестовая выборка
t -статистика	0,000546	0,302916
p -значение	0,999564	0,763038
F -test	0,472783	0,603488
MAPE	3,44 %	5,39 %

Результаты оценки модели XGBRegressor приведены в табл. 8.

Модель показала достаточно неплохие результаты. Стоит отметить наименьшую ошибку модели **XGBRegressor** из всех протестированных моделей на обучающей выборке. Вероятно, здесь тоже имеет место переобучение и запоминание моделью тренировочных данных.

Модель множественной регрессии по трем кластерам данных. В выборке данных за 10 лет присутствуют значительные структурные изменения самих данных. За 10 лет в рассматриваемом ГТО произошли вводы в эксплуатацию новых и модернизация существующих компрессорных станций, подключение крупных потребителей, появились новые экспортные маршруты, выводились участки магистральных газопроводов из эксплуатации. Учет этих факторов был реализован в рамках разделения всей выборки на кластеры. Существует множество методов классификации и кластеризации данных. В данном

исследовании был использован один из самых распространенных и простых методов – метод k -ближайших соседей.

Алгоритм KMeans – это популярный алгоритм кластеризации, используемый для разбиения набора данных на k -кластеров и описанный в [13]. Алгоритм работает, распределяя данные в кластеры таким образом, чтобы минимизировать сумму квадратов расстояний между точками данных и центрами их кластеров (центроидами). Начинается расчет с произвольного размещения центроидов, затем итеративно алгоритм обновляет их, перемещая центроиды в центр масс своих кластеров и переназначая точки данных к ближайшим центроидам. Этот процесс продолжается до тех пор, пока центроиды не перестанут значительно изменяться. KMeans прост в реализации и эффективно работает на больших наборах данных, что делает его популярным выбором для задач сегментации, группировки и поиска скрытых структур в данных.

В качестве основного параметра модели кластеризации экспертно был установлен параметр «количество кластеров» ($n_clusters$) = 3.

Визуализация многомерных результатов работы метода KNN представлена на рис. 9 и выполнена с помощью t -SNE, описание которого приведено в [14]. Метод t -SNE (t -Distributed Stochastic Neighbor Embedding) – это метод нелинейного уменьшения размерности, используемый для визуализации многомерных данных в двух- или трехмерном пространстве. Алгоритм работает, моделируя вероятность сходства точек данных в исходном высокоразмерном пространстве, и затем пытается найти отображение в низкоразмерное

пространство, которое сохраняет эти вероятности как можно ближе. В результате получается визуализация, где точки, находящиеся близко друг к другу в высокоразмерном пространстве, также находятся близко друг к другу в низкоразмерном пространстве. *t*-SNE особенно полезен для визуализации сложных данных, таких как результаты глубокого обучения, и для обнаружения скрытых структур и кластеров в данных. На рис. 9, а представлено визуальное отображение нелинейного уменьшения размерности данных, использованных для обучения моделей.

Кроме того, на рис. 9, б изображены кластеры (изображены разными цветами) ряда «потребле-

ние внутри ГТО». На графике видно, что в период с 2014 по 2016 г. и с 2018 по 2021 г. наблюдалось схожее состояние ГТС (1423 набора данных). Второй кластер расположен между 2016 и 2018 гг. и составляет 551 набор данных. Третий кластер начинается с 2021 г. и содержит 1264 набора данных.

На основе разделения исходных данных на три кластера построены три модели многофакторной линейной регрессии. К каждому набору тестовых данных были применены соответствующие модели линейной регрессии.

Результаты расчета расхода топливного газа за 30 дней модели линейной регрессии по трем кластерам данных приведены на рис. 10.

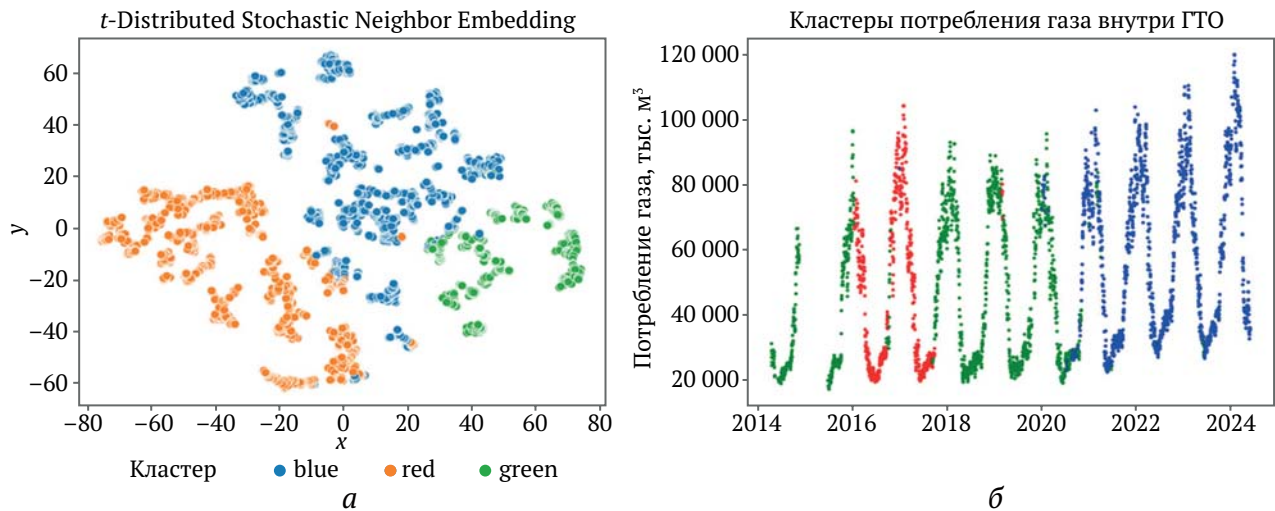


Рис. 9. Визуальное представление результатов кластеризации набора данных алгоритмом KMeans: а – *t*-Distributed Stochastic Neighbor Embedding; б – кластеры потребления внутри ГТО

Fig. 9. Visual representation of the results of clustering the data set using the KMeans algorithm: (a) *t*-Distributed Stochastic Neighbor Embedding; (b) consumption clusters within the gas transportation company

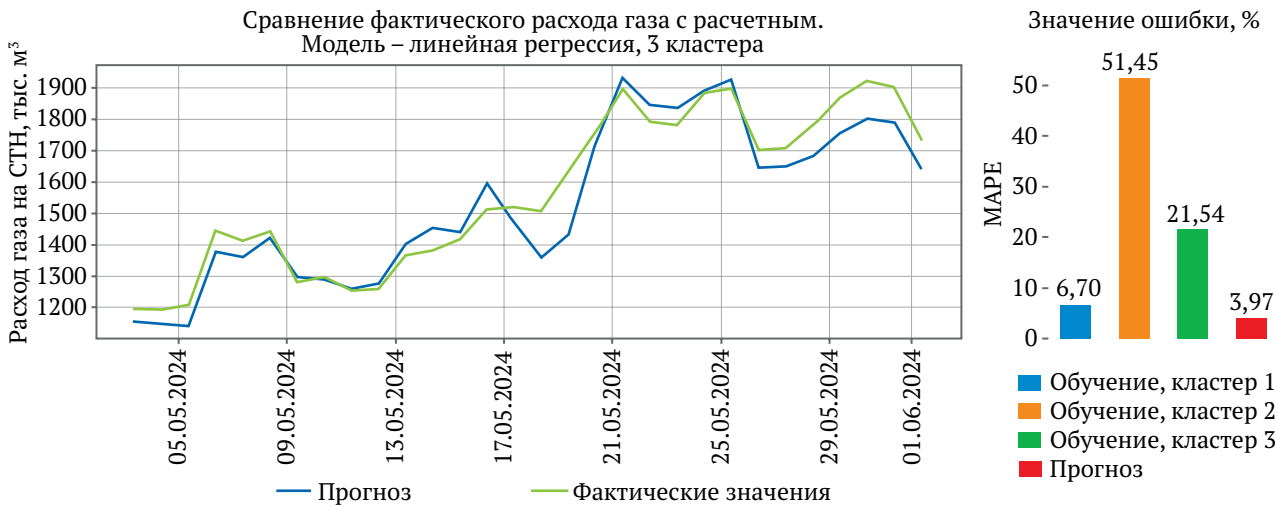


Рис. 10. Результаты работы модели линейной регрессии по трем кластерам данных

Fig. 10. Results of the linear regression model on three data clusters

Результаты оценки модели многофакторной линейной регрессии, построенной по трем кластерам, приведены в **табл. 9**.

Поскольку p -значение больше 0,05, оснований отвергнуть нулевую гипотезу при использовании этой модели нет.

Параметры всех моделей линейной регрессии приведены в **табл. 10**.

Предлагаемая модель многофакторной линейной регрессии по кластерам данных показала наилучший результат среди всех протестированных моделей. Средняя процентная ошибка на тестовых данных составила 4,12%, что уже ниже установленной СТО Газпром 2-3.5-113-2007²⁴.

Определение экономической эффективности точного прогнозирования расхода газа на СТН

Экономический эффект от разработки модели оперативного прогнозирования объема топливного газа в зависимости от потоков природного газа в ГТС становится заметным при создании оптимизационных моделей, ориентированных на минимизацию расхода топливного

газа в рамках всей ГТС. Чем точнее модель прогнозирует расход топливного газа, тем выше будет общая точность и эффективность глобальной оптимизационной модели.

Авторами предложен подход к определению экономической эффективности оптимизационной модели по принципу сравнения «без проекта» и «с проектом». В качестве проекта рассмотрено внедрение глобальной оптимизационной модели, ядром которой является подход к оперативному прогнозированию расхода топливного газа. В качестве подхода «без проекта» использованы фактические показатели работы газотранспортного общества в 2023 г.

Таблица 9 / Table 9

Оценка модели линейной регрессии по трем кластерам данных

Estimating a linear regression model on three clusters of data

Параметр оценки качества модели	Значение параметра тестовой выборки
t -статистика	–0,49215
p -значение	0,624474
F _test	0,474186
MAPE	3,97 %

²⁴ Методика оценки энергоэффективности газотранспортных объектов и систем. СТО Газпром 2-3.5-113-2007. Введ. 2007-11-15. Режим доступа: <https://files.stroyinf.ru/Data1/54/54561/> (дата обращения: 21.06.2024).

Таблица 10 / Table 10

Значения коэффициентов линейной регрессии трех моделей

Values of linear regression coefficients of three models

Параметр	Значение регрессора модели		
	Первый кластер	Второй кластер	Третий кластер
Месторождение (1)	–0,097991	0,004205	–0,366329
ГПЗ 1	0,003551	–0,003489	0,106615
Месторождение (2)	0,115896	0,069554	–0,217679
Месторождение (3)	–1,264403	30,817546	134,768121
От ГТО выше по потоку (1)	0,000704	–0,003659	0,005182
От ГТО выше по потоку (2)	–0,000960	–0,003133	0,004849
От ГТО выше по потоку (3)	–0,001685	–0,003273	0,004026
От ГТО выше по потоку (4)	0,005736	0,002032	0,011754
Экспорт (1)	0,010325	0,010564	0,002559
Экспорт (2)	0,000000	0,021962	0,016226
Собственным потребителям	0,004347	0,009193	0,005089
В ГТО ниже по потоку (1)	0,010156	0,014354	0,000213
В ГТО ниже по потоку (2)	0,007083	0,010303	0,006102
ПХГ, отбор	–0,001258	–0,005857	0,001260
ПХГ, закачка	0,013715	0,018647	0,014650
Запас газа	–0,000539	0,000078	0,000178
Коэффициент ϵ	–135,621	–444,806	–573,906

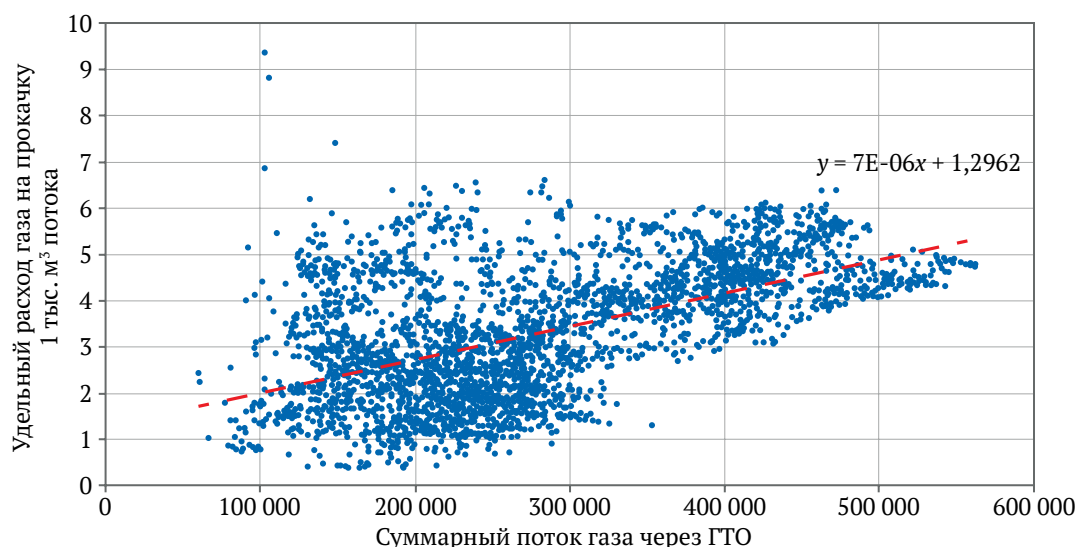


Рис. 11. Зависимость удельного расхода газа на СТН от объема потока газа через ГТО

Источник: составлено авторами по данным подсистемы хранения модернизированной Автоматизированной системы диспетчерского управления Единой системы газоснабжения Российской Федерации

Fig. 11. Dependence of specific gas consumption for own technological needs on the volume of gas flow through the gas transportation company

Source: compiled by the authors based on data from the storage subsystem of the modernized Automated Dispatch Control System of the Unified Gas Supply System of the Russian Federation

На **рис. 11** показана зависимость удельного расхода топливного газа на транспортировку природного газа по ГТС от суммарного потока газа, проходящего через газотранспортное оборудование.

Как видно на рис. 11, удельный расход топливного газа имеет зависимость, близкую к линейной от суммарного потока газа, проходящего через ГТО по магистральным газопроводам. Это означает, что с увеличением транспортного потока требуется больше топлива для транспортировки 1 тыс. м³ природного газа. Данная зависимость также косвенно подтверждается в СТО Газпром 3.3-2-044-2016²⁵. Анализ зависимости удельного расхода топливного газа от суммарного транспортного потока проведен для того, чтобы показать, что типовые значения расхода топливного газа одного интервала могут являться аномально высокими для другого интервала. Эти выводы потребуются далее.

На следующем этапе была проведена разбивка объемов запаса газа на 10 интервалов в зависи-

мости от удельного расхода топливного газа на СТН КС и расчет статистических показателей для каждого из интервалов. На **рис. 12** изображена такая разбивка, численные значения интервалов приведены в **табл. 11**.

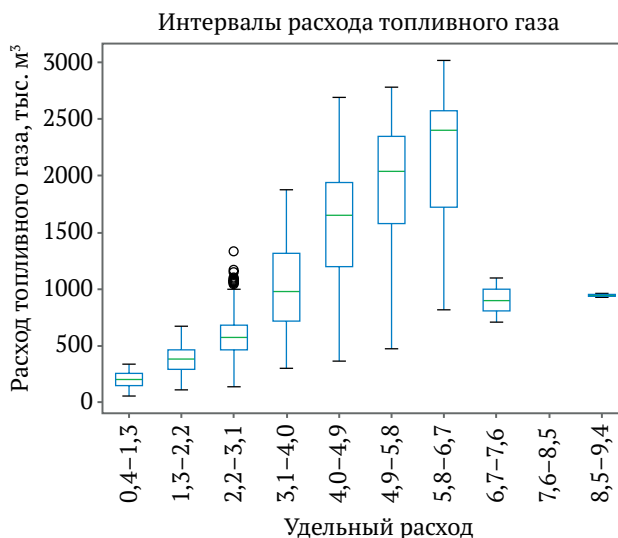


Рис. 12. Статистическая информация результатов разбиения удельного расхода топливного газа на 10 интервалов

Fig. 12. Statistical information of the results of dividing the specific fuel gas consumption into 10 intervals

²⁵ СТО Газпром 3.3-2-044-2016. Система норм и нормативов расхода ресурсов, использования оборудования и формирования производственных запасов ПАО «Газпром». Методика нормирования расхода природного газа на собственные технологические нужды и технологические потери магистрального транспорта газа (с изм. № 1). Введен Распоряжением ПАО «Газпром» от 22.12.2016 № 430.

Таблица 11 предоставляет статистическую информацию о результатах разбиения удельного расхода топливного газа на 10 интервалов. В таблице представлены значения статистических показателей, которые отображают распределение объема топливного газа по этим интервалам в рамках всей ГТС.

Для расчета эффективности оптимизации потоков газа по ГТС дочернего общества было введено следующее предположение: с вероятностью 95,4 % (интервала в 2 сигмы (± 2 стандартных отклонения)) режим работы КС газотранспортного общества установлен оптимально.

В нормальном распределении интервал ± 2 сигмы охватывает примерно 95,4 % всех значений. Это означает, что значения, выходящие за пределы 2 сигм, встречаются примерно в 4,6 % случаев, и схема потоков газа, проходящих че-

рез ГТО, является неоптимальной из-за резких колебаний потребления или изменений заявок экспортных покупателей. В таком случае оборудование КС настроено неоптимально и расход топливного газа был превышен.

Для расчета перерасхода топливного газа в разрезе интервалов удельного расхода газа введем предположение, что режимы работы КС, при которых расход топливного газа, как указывалось выше, входит в диапазон 94,5 % значений, указывает на оптимальный режим работы оборудования. Расход газа, находящийся за пределами этого интервала, говорит о перерасходе топливного газа. Суммарный перерасход за 2023 г. приведен в табл. 12.

Таким образом, перерасход топливного газа за 2023 г. составил 14,5 млн м³, или 0,43 % от всего использованного топливного газа.

Таблица 11 / Table 11

Значения статистических показателей разбиения удельного расхода топливного газа на 10 интервалов
Values of statistical indicators for dividing the specific fuel gas consumption into 10 intervals

Интервал удельного расхода ТГ (м ³ /тыс. м ³)	Медиана	25 % квантиль	75 % квантиль	95,4 % квантиль
(0.38, 1.288]	201,57	148,90	256,17	328,38
(1.288, 2.186]	382,96	288,86	463,14	577,47
(2.186, 3.084]	574,30	460,95	677,77	953,42
(3.084, 3.982]	981,40	717,81	1317,33	1619,20
(3.982, 4.88]	1653,78	1198,91	1938,92	2336,81
(4.88, 5.779]	2039,38	1579,20	2349,87	2652,65
(5.779, 6.677]	2401,73	1717,24	2575,97	2762,59
(6.677, 7.575]	901,32	803,94	998,70	1078,17
(7.575, 8.473]	NaN	NaN	NaN	NaN
(8.473, 9.371]	944,82	937,70	951,94	957,75

Таблица 12 / Table 12

Значения перерасхода топливного газа по интервалам
Fuel gas overconsumption values by intervals

Интервал	95,4 % квантиль	Перерасход топливного газа на СТН КС, тыс. м ³
(0.38, 1.288]	328,381	62,986
(1.288, 2.186]	577,465	817,253
(2.186, 3.084]	953,422	3 193,379
(3.084, 3.982]	1 619,199	3 011,937
(3.982, 4.88]	2 336,810	5 985,408
(4.88, 5.779]	2 652,646	861,120
(5.779, 6.677]	2 762,586	513,270
(6.677, 7.575]	1 078,168	17,919
(7.575, 8.473]	NaN	0,000
(8.473, 9.371]	957,754	1,310
Итого		14 464,581

Согласно Приказу ФАС²⁶, для потребителей, находящихся на территории Краснодарского края утверждена оптовая цена, равная 5873 руб./тыс. м³. Стоимость перерасхода природного газа в 2023 г. в ценах 2024 г. составит 84,95 млн руб.

²⁶ Приказ ФАС России от 28.11.2023 N 910/23 «Об утверждении оптовых цен на газ, добываемый ПАО «Газпром» и его аффилированными лицами, реализуемый организациям, осуществляющим по состоянию на 1 декабря 2023 г. в качестве основного вида экономической деятельности производство электроэнергии тепловыми электростанциями, производство, передачу и распределение пара и горячей воды тепловыми электростанциями и (или) котельными с кодами ОКВЭД 35.11.1, 35.30.11, 35.30.14, 35.30.2, 35.30.3, организациям, осуществляющим деятельность с указанными кодами ОКВЭД в качестве дополнительного вида экономической деятельности при условии осуществления ими в качестве основного вида экономической деятельности с кодами ОКВЭД 35.11, 35.3, 35.30, 35.30.1, 68.32.1, организациям, выручка от реализации тепловой и (или) электрической энергии (мощности) которых составляет не менее 75 процентов общей выручки, организациям, выручка от реализации тепловой и (или) электрической энергии (мощности) которых составляет менее 75 процентов общей выручки, приобретающим газ в объемах, необходимых для производства и реализации тепловой и (или) электрической энергии (мощности) по регулируемым ценам (тарифам), либо организациям-недропользователям (коды ОКВЭД 06.10.1, 06.20) и (или) их аффилированным лицам, не входящим с такими недропользователями в одну группу компаний, или организациям, осуществляющим в качестве основного вида экономической деятельности распределение газообразного топлива по газораспределительным сетям и торговлю твердым, жидким и газообразным топливом и подобными продуктами с кодами ОКВЭД 35.22, 46.71, приобретающим газ для перепродажи потребителям, относящимся к организациям, осуществляющим по состоянию на 1 декабря 2023 г. в качестве основного вида экономической деятельности производство электроэнергии тепловыми электростанциями, производство, передачу и распределение пара и горячей воды тепловыми электростанциями и (или) котельными с кодами ОКВЭД 35.11.1, 35.30.11, 35.30.14, 35.30.2, 35.30.3, организациям, осуществляющим деятельность с указанными кодами ОКВЭД в качестве дополнительного вида экономической деятельности при условии осуществления ими в качестве основного вида экономической деятельности с кодами ОКВЭД 35.11, 35.3, 35.30, 35.30.1, 68.32.1, организациям, выручка от реализации тепловой и (или) электрической энергии (мощности) которых составляет не менее 75 процентов общей выручки, организациям, выручка от реализации тепловой и (или) электрической энергии (мощности) которых составляет менее 75 процентов общей выручки, приобретающим газ в объемах, необходимых для производства и реализации тепловой и (или) электрической энергии (мощности) по регулируемым ценам (тарифам), кроме потребителей Российской Федерации, указанных в пунктах 15.1 и 15.1.1 Основных положений формирования и государственного регулирования цен на газ, тарифов на услуги по его транспортировке, платы за технологическое присоединение газоиспользующего оборудования к газораспределительным сетям на территории Российской Федерации и платы за технологическое присоединение к магистральным газопроводам строящихся и реконструируемых газопроводов, предназначенных для транспортировки газа от магистральных газопроводов до объектов капитального строительства, и газопроводов, предназначенных для транспортировки газа от месторождений природного газа до магистрального газопровода, утвержденных постановлением Правительства Российской Федерации от 29 декабря 2000 г. № 1021» (Зарегистрировано в Минюсте России 01.12.2023 № 76205). Режим доступа: <http://publication.pravo.gov.ru/document/0001202312010032> (дата обращения: 15.07.2024).

Оценка экономической эффективности проектов обычно выполняется в соответствии с методическими документами^{27,28}.

Основными эффектами от внедрения проекта, согласно СТО Газпром РД 1.12-096-2004²⁹, являются коммерческий и ресурсный эффекты. В связи с этим оценка экономической эффективности была выполнена с учетом следующих предположений и допущений, сделанных авторами:

- оценка проекта проводится в рамках расчетного периода, составляющего 5 лет;
- используется принцип сравнения сценариев «без проекта» и «с проектом», т.е. эффективность оценивается путем сопоставления денежных потоков, связанных с реализацией проекта и использованием его результатов, с денежными потоками, которые имели бы место в случае, если бы оптимизация потоков газа не была внедрена.

Для выполнения оценки экономической эффективности внедрения в модельное дочернее общество ПАО «Газпром» описанного подхода приняты следующие исходные параметры:

- горизонт расчета установлен на период с 2023 по 2027 г. (5 лет);
- все расчеты выполнены в постоянных ценах;
- валюта расчетов – российские рубли;
- дисконтирование денежных потоков производится к дате 01.01.2023;
- норма дисконта принята на уровне 12%.

Чистый доход (ЧД), млн руб., рассчитывается по следующей формуле:

$$\text{ЧД} = \sum_{t=T_n}^{T_N} \text{ДП}(t), \quad (3)$$

где ДП(t) – денежный поток в t -м году, млн руб.; t – индекс расчетного периода T_n .

Чистый дисконтированный доход (ЧДД), млн руб., рассчитывается по следующей формуле:

$$\text{ЧДД} = \sum_{t=T_n}^{T_N} \text{ДП}(t) \cdot \text{Кд}(t), \quad (4)$$

где ДП(t) – денежный поток в t -м году, млн руб.; Кд(t) – коэффициент дисконтирования в t -м году; t – индекс расчетного периода (T_n ; T_N).

²⁷ Методические рекомендации по оценке эффективности инвестиционных проектов (вторая редакция), утвержденные Министерством экономики РФ, Министерством финансов РФ, Государственным комитетом РФ по строительной, архитектурной и жилищной политике от 21.06.1999 № ВК 477.

²⁸ Методика оценки экономической эффективности инвестиционных проектов в форме капитальных вложений. Введ.: 09.09.2009 № 01/07-99.

²⁹ СТО Газпром РД 1.12-096-2004. Внутрикorporативные правила оценки эффективности НИОКР. Введ. Приказом ОАО «Газпром» от 16 августа 2004 г. № 70 с 2004-09-01. М.: ООО «ИРЦ Газпром».

Коэффициент дисконтирования $K_d(t)$ определяется по следующей формуле:

$$K_d(t) = \frac{1}{(1+d)^{t-1-(T_{\text{год}}-1)}}, \quad (5)$$

где d – ставка дисконтирования, %; $T_{\text{год}}$ – расчетный год, в котором коэффициент дисконтирования равен 1; d – ставка дисконтирования, которая является одним из основных параметров проекта и вводится вручную.

На сегодняшний день в ПАО «Газпром» применяются ставки дисконтирования в размере 10 или 15 %. Для настоящего расчета, как указывалось выше, использовалась ставка 12 %

Расчет экономического эффекта приведен в табл. 13.

Заключение

В данной статье рассмотрены различные статистические подходы к оперативному прогнозированию расхода природного газа на собственные технологические нужды в масштабах газотранспортного общества. Особое внимание уделено сильным и слабым сторонам существующих подходов к расчету расхода топливного газа, применяемых в ПАО «Газпром».

Предложенный в статье подход к оперативному прогнозированию расхода природного газа на нужды компрессорных станций состоит из нескольких этапов и не требует сбора данных с газоизмерительных станций, а также детального анализа режимов работы отдельных компрессорных цехов, газоперекачивающих агрегатов. Кроме того, предложенный подход не зависит от внешних факторов, таких как температура, скорость ветра или влажность воздуха. Этот подход продемонстрировал стабильные и высокие результаты на тестовой выборке данных за 10-летний период для одного из газотранспортных обществ, обеспечив среднюю процентную ошибку

4,12 %, что ниже допустимых пределов согласно нормативам ПАО «Газпром».

Предложенный подход прост в реализации, не требует сложных входных данных и может быть успешно применен для оперативного прогнозирования в масштабах всей газотранспортной системы. Учитывая доступность современных вычислительных мощностей, простоту моделей и независимость от внешних источников данных, предложенный подход обладает значительным потенциалом для применения. В то же время экономический эффект от внедрения предложенного подхода может составить несколько сотен миллионов рублей на горизонте 5 лет.

Тем не менее предложенный подход имеет потенциал для дальнейшего улучшения и оптимизации. В статье рассматривались одни из самых простых подходов, таких как метод классификации состояний ГТС с использованием K-means и линейная регрессия для прогнозирования временных рядов. Однако дальнейшее усовершенствование метода может быть достигнуто за счет применения более сложных и продвинутых методов кластеризации и классификации, таких как Gaussian Mixture Models (GMM), DBSCAN (Density-Based Spatial Clustering of Applications with Noise), описанный в [15] или спектральная кластеризация (Spectral Clustering) [16]. После разделения данных на кластеры с помощью этих методов могут быть использованы более мощные модели регрессии, такие как градиентный бустинг (Gradient Boosting Regression), описанный в [17], метод опорных векторов (Support Vector Regression, SVR) и «случайный лес» (Random Forest Regression).

Кроме того, для регрессионного анализа и разбиения на кластеры можно рассмотреть использование нейронных сетей (Neural Networks), которые предлагают огромные возможности для

Таблица 13 / Table 13

Расчет экономического эффекта

Calculation of economic effect

Параметр	Год				
	2023	2024	2025	2026	2027
Денежный поток от экономии топливного газа, млн руб.	84 950	84 950	84 950	84 950	84 950
Фактор дисконтирования	1	0,89286	0,7972	0,7118	0,6355
Приведенный денежный поток, млн руб.	84 950	75 849	67 722	60 466	53 988
Итого, млн руб.					342 975

настройки и обладают различными архитектурными вариациями.

В дальнейшем целесообразно провести тестирование предложенных методов и архитек-

тур, чтобы определить наиболее точные и устойчивые подходы для каждого ГТО, что позволит улучшить оперативный расчет расхода топливного газа в рамках всей ГТС ПАО «Газпром».

Список литературы / References

1. Посягин Б.С., Герке В.Г. *Справочное пособие для работников диспетчерских служб газотранспортных систем*. М.: ООО «Газпром экспо»; 2015. 796 с.
2. Халикова Э.Р. Управление затратами топливного газа в дочерних газотранспортных обществах ПАО «Газпром». *Технико-технологические проблемы сервиса*. 2021;(2((56))):55–62.
Khalikova E.R. Fuel gas costs management in gas transmission subsidiaries of PJSC Gazprom. *Tekhniko-tehnologicheskie problemy servisa*. 2021;(2((56))):55–62. (In Russ.)
3. Чернышов Ю.Ю. Применение автокодировщиков для выявления аномалий в киберфизических системах. *Вестник Пермского университета. Математика. Механика. Информатика*. 2022;(4(59)):89–94. <https://doi.org/10.17072/1993-0550-2022-4-89-94>
Chernyshov Y.Yu. About using of autoencoders for anomaly detection in cyber-physical systems. *Bulletin of Perm University. Mathematics. Mechanics. Computer Science*. 2022;(4(59)):89–94. (In Russ.). <https://doi.org/10.17072/1993-0550-2022-4-89-94>
4. Чесноков М.Ю. Поиск аномалий во временных рядах на основе ансамблей алгоритмов DBSCAN. *Искусственный интеллект и принятие решений*. 2018;(1):98–106.
Chesnokov M.Y. Time series anomaly detection based on DBSCAN ensembles. *Iskusstvenniy Intellekt i Prinyatie Resheniy = Artificial Intelligence and Decision Making*. 2018;(1):98–106. (In Russ.)
5. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. *ACM Computing Surveys*. 2009;41(3):1–58. <https://doi.org/10.1145/1541880.1541882>
6. Dau H.A., Ciesielski V., Song A. *Anomaly detection using replicator neural networks trained on examples of one class*. In: Dick G. (ed.). *Simulated evolution and learning. SEAL 2014. Lecture notes in computer science*. Cham: Springer; 2014. Vol. 8886. P. 311–322. https://doi.org/10.1007/978-3-319-13563-2_27
7. Шаталов К.В., Кириллова А.В. Применение критерия Стьюдента для оценки результатов межлабораторных сравнительных испытаний. *Стандартные образцы*. 2016;(1):42–49. <https://doi.org/10.20915/2077-1177-2016-0-1-42-49>
Shatalov K.V., Kirillova A.V. Application of Student t-test for evaluation of interlaboratory comparative tests results. *Measurement Standards. Reference Materials*. 2016;(1):42–49. (In Russ.). <https://doi.org/10.20915/2077-1177-2016-0-1-42-49>
8. NIST/SEMATECH e-Handbook of Statistical Methods. Exploratory Data Analysis: Measures of Skewness and Kurtosis. <https://doi.org/10.18434/M32189>
9. Sallehuddin R., Shamsuddin S.M., Hashim S.Z.M. Hybridization model of linear and nonlinear time series data for forecasting. In: *Second Asia inter. conf. on modelling and simulation (AMS 2008)*. Kuala Lumpur, Malaysia, 13–15 May, 2008. P. 597–602. <https://doi.org/10.1109/AMS.2008.142>
10. Breiman L. Random Forest. *Machine Learning*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
11. Dorogush A.V., Ershov V., Gulin A. CatBoost: gradient boosting with categorical features support. 2018. <https://doi.org/10.48550/arXiv.1810.11363>
12. Friedman J. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*. 2000;29(5):1189–1232. <https://doi.org/10.1214/aos/1013203451>
13. Subramaniaswamy V., Logesh R. Adaptive KNN based recommender system through mining of user preferences. *Wireless Personal Communications*. 2017;97(2):2229–2247. <https://doi.org/10.1007/s11277-017-4605-5>
14. Van der Maaten L., Hinton G. Visualizing data using t-sne. *Journal of Machine Learning Research*. 2008;9(2605):2579–2605.
15. Ester M., Kriegel H.P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proc. of the 2nd Intern. Conf. on Knowledge Discovery and Data Mining (KDD-96)*; 1996. P. 226–231.
16. Arias-Castro E., Chen G., Lerman G. Spectral clustering based on local linear approximations. *Electronic Journal of Statistics*. 2011;(5):1537–1587. <https://doi.org/10.1214/11-ejs651>
17. Дружков П.Н., Золотых Н.Ю., Половинкин А.Н. Реализация параллельного алгоритма предсказания в методе градиентного бустинга деревьев решений. *Вестник Южно-Уральского государственного университета. Серия: Математическое моделирование и программирование*. 2011;(37(254)):82–89.
Druzhkov P.N., Zolotykh N.Yu., Polovinkin A.N. Parallel implementation of prediction algorithm in gradient boosting trees method. *Vestnik Yuzhno-ural'skogo gosudarstvennogo universiteta. Seriya: Matematicheskoe modelirovanie i programirovanie = Bulletin of the South Ural State University. Series: Mathematical Modeling and Programming*. 2011;(37(254)):82–89. (In Russ.)

Информация об авторах

Андрей Алексеевич Кудрявцев – д-р. экон. наук, профессор кафедры статистики и эконометрики, Санкт-Петербургский государственный экономический университет, 191023, Санкт-Петербург, наб. канала Грибоедова, д. 30-32, Российская Федерация; ORCID: <https://orcid.org/0000-0001-7722-8044>; e-mail: kudr2007@inbox.ru

Сергей Николаевич Ланин – аспирант, Санкт-Петербургский государственный экономический университет, 191023, Санкт-Петербург, наб. канала Грибоедова, д. 30-32, Российская Федерация; e-mail: s.lanin@gmail.com

Information about the authors

Andrey A. Kudryavtsev – Dr.Sci. (Econ.), Professor of the Chair for Statistics and Econometrics, St. Petersburg State University of Economics, 30-32 Griboedov Canal Emb., St. Petersburg 191023, Russian Federation; ORCID: <https://orcid.org/0000-0001-7722-8044>; e-mail: kudr2007@inbox.ru

Sergey N. Lanin – Postgraduate Student, Graduate School of the Chair for Statistics and Econometrics, St. Petersburg State University of Economics, 30-32 Griboedov Canal Emb., St. Petersburg 191023, Russian Federation; e-mail: s.lanin@gmail.com

Поступила в редакцию 02.10.2024; поступила после доработки 18.11.2024; принята к публикации 20.11.2024
Received 02.10.2024; Revised 18.11.2024; Accepted 20.11.2024