

DOI:

## ПРОГНОЗИРОВАНИЕ ТЕМПЕРАТУРЫ ГРУНТОВ ТРАССЫ НЕФТЕПРОВОДА

**Владова А.Ю.**

*Институт проблем управления им. В.А. Трапезникова РАН, Россия, г. Москва*

*ул. Профсоюзная д.65*

*Финансовый университет при Правительстве РФ*

*avladova@ipu.ru*

*Аннотация: Предлагаемый подход к прогнозу температурного режима вдольтрассовых грунтов в системах геотехнического мониторинга позволяет комплексировать геолого-геофизическую, геохимическую и технологическую информацию и принимать обоснованные производственные решения по компенсации влияния линейных протяженных объектов на несущий грунт.*

Ключевые слова: Временные ряды, эконометрика, геотехническая система мониторинга

### **Введение**

Специализированные программные комплексы Cripe, WARM, Frost 3D Universal, Freezer включающие модули для теплотехнических расчетов, при прогнозе температурного режима многолетнемерзлых пород ориентированы на численное решение нестационарного уравнения теплопроводности для среды, вмещающей подвижную границу раздела фаз [1, 2]. Необходимость выполнять учащенную дискретизацию участков с экзогенными геологическими процессами требует, во-первых, значительных вычислительных и временных ресурсов, а во-вторых, правильное задание граничных условий. Невыполнение этих условий приводит к резкому возрастанию погрешности численного решения [3]. Таким образом, контроль твердомерзлого состояния грунтов и прогнозирование срока перехода их в талое состояние является комплексной проблемой [4].

### **1 Состояние вопроса**

В настоящее время разработан класс методов машинного обучения, позволяющий обработать многомерные температурные данные вдольтрассовых грунтов. Например, в работе [5] предложена система мониторинга отложений на стенках трубопровода, основанная на активной термометрии, в которой экспериментальные температурные профили нелинейно подгоняются к теоретическим моделям. Признаки, извлеченные с помощью нелинейной подгонки, составляют основу модели. В статье [6] проведен многомерный прочностной анализ подземных трубопроводов с использованием метода машинного обучения, называемой регрессией Лассо. Самообучающийся алгоритм реализован в работе [7] для характеристики взаимосвязи между деформацией трубы и факторами окружающей среды. В статье [8] на основании данных, собранных ультразвуковым, литологическим, геофизическим методами, сформирована база о физических параметрах пород и физико-химических параметрах флюида и проведена бинарная классификация пород месторождения нефти с выделением сильных коллекторов.

### **2 Постановка задачи**

По имеющимся данным, собранным за полтора года, провести прогнозирование срока перехода твердомерзлого состояния грунтов в талое. Предлагаемый алгоритм решения задачи приведен на рисунке 1. На первом шаге проведен статистический анализ для температурных измерений грунтов каждой термометрической скважины. На втором и третьем шагах выполнены визуализация данных и синтез имеющихся признаков с целью получения более информативных. Результатом четвертого шага является прогнозная модель температур вдольтрассовых грунтов.

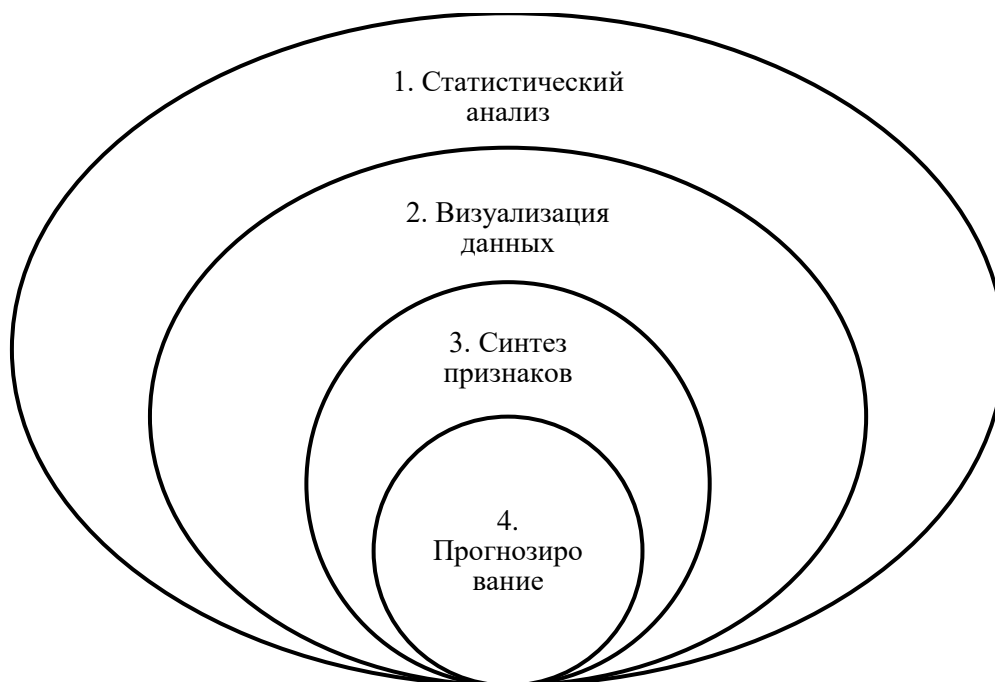


Рисунок 1 – Алгоритм решения задачи прогнозирования

### 3 Статистический анализ

На рисунке 2 представлена посезонная выборка температурных измерений грунтов подземного нефтепровода, произведенных одной из термоскважин.

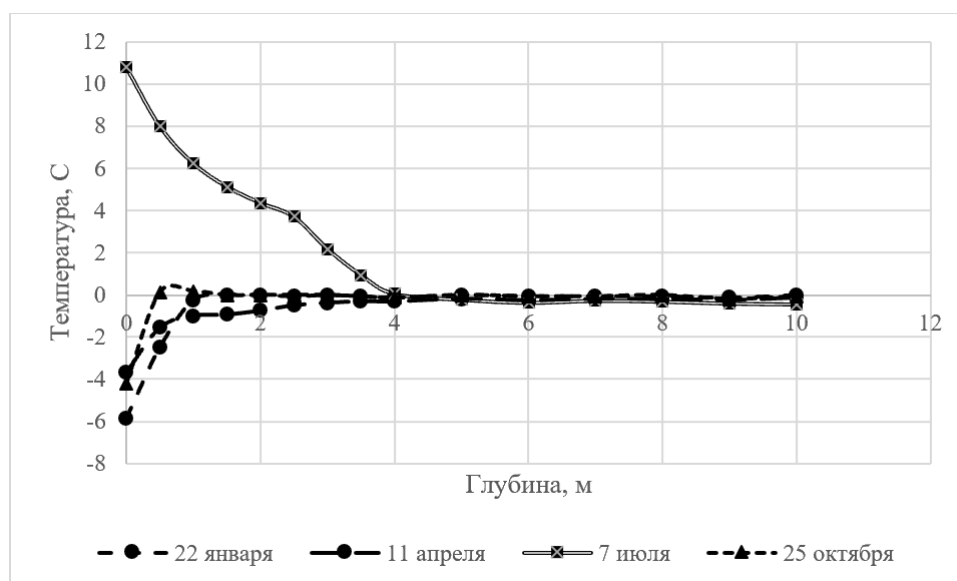


Рисунок 2 – Температурные ряды

40% измерений представленной выборки лежит в положительном диапазоне, что говорит о возможном снижении несущей способности грунта на этом участке трассы. Отрицательные значения температуры в весенний сезон свидетельствуют о запаздывании (памяти) процесса растепления.

Проверка на стационарность временных рядов и единичные корни проведена дополненным тестом Дики-Фуллера, реализованным в библиотеке statsmodels. Для AR-модели, числа наблюдений 4,4 тыс. и 5% значимости, равной -2.8621 стационарными (в смысле данного теста) являются три временных ряда, расположенных на глубине 10, 11, 13 м. Понижающая передискретизация [9] временного ряда (переход от дней к месяцам с вычислением средней температуры в интервале) увеличивает число срезов по глубине (5-13 метров), на которых температурные ряды можно считать стационарными, что говорит о постоянном техногенном фоне, создаваемым нефтепроводом.

Таким образом, предварительный анализ данных показал, что сформированные временные ряды, представляющие температуры на разных глубинах, образуют по крайней мере две группы, которые должны быть аппроксимированы разными моделями.

Дальнейший анализ проведен для обобщенной матрицы, полученной слиянием температур разных термокос по времени наблюдения. Анализ с помощью библиотеки `pandas-profiling` установил значительное количество послойно коррелирующих измерений, например, на глубине 0.5 – 3.5 метра, 6.0 - 7.0 метра, 9.0 - 13.0 метров.

Гистограммы рисунков 3-4 показывают, что приповерхностные температуры распределены слева и справа от нуля, в то время как температуры на глубине смещены в отрицательную область.

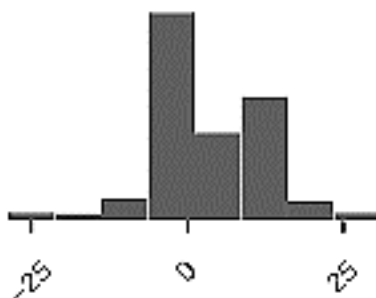


Рисунок 3 – Распределение приповерхностных температур

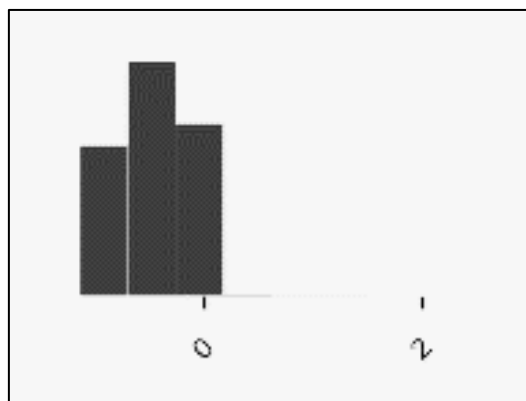


Рисунок 4 – Распределение глубинных температур

#### 4 Визуализация данных

Визуализация многомерного пространства признаков относительно двух целевых классов (грунт в мерзлом состоянии, грунт в растепленном состоянии), усредненных для каждой скважины положительных и отрицательных температур выполнена методом главных компонент (рисунок 5). Метод главных компонент уменьшает размерность  $d$ -мерного набора данных, проецируя его на  $k$ -мерное подпространство ( $k < d$ , выбрав наиболее важные признаки, несущие максимальную информацию о наборе данных). Выбор осуществляется по дисперсии, которую признаки вызывают в целевой переменной. Признак с максимальной дисперсией является первой главной компонентой. Признак со вторым по величине отклонением считают второй главной компонентой и так далее. Важно заметить, что главные компоненты не имеют корреляции друг с другом.

Перед применением метода главных компонент набор признаков был нормализован с помощью метода `Normalizer` библиотеки `sklearn`. Если нормализацию не провести, то главные компоненты будут смещены по отношению к признакам с высокой дисперсией, что приведет к ложным результатам [10].

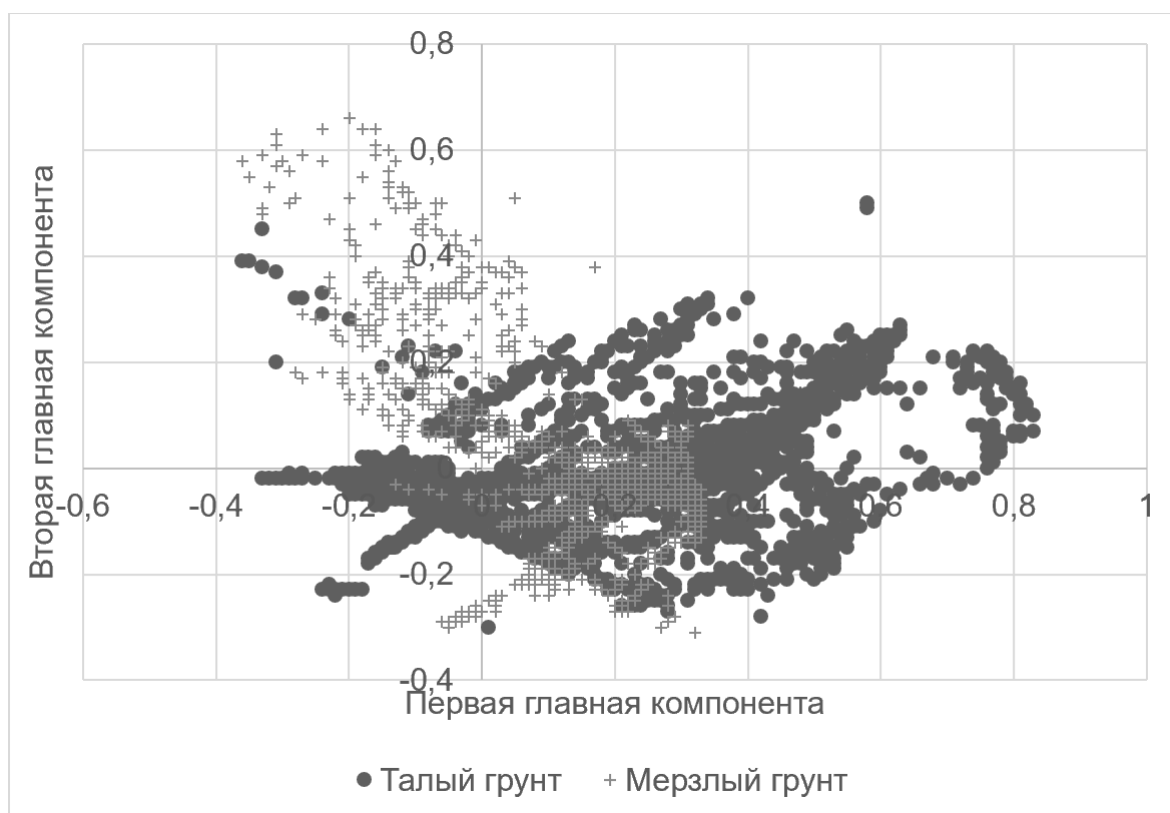


Рисунок 2 – Распределение целевой переменной: + – грунт в мерзлом состоянии, • – грунт в талом состоянии

Первая компонента объясняет 72.91% дисперсии данных, а вторая 9.61%, что в сумме дает 82.52%. Третья компонента позволяет увеличить охват дисперсии данных на 8,42%. Таким образом, при дальнейших исследованиях достаточно использовать три главных компоненты.

На этапе подготовки данных выявлены корреляции между послойными измерениями термоскважин на одном и на разных удалениях от нефтепровода. Поэтому необходимо оценить влияние признаков на целевую переменную и выбрать наиболее значимые из них. Для этого использованы подходы на основе корреляционных зависимостей, а также деревьев решений и случайных перестановок признаков, реализованные в библиотеках `gfrimp` и `eli5` (таблица 1).

Таблица 1 – Значимость признаков

Метод	Значимые признаки (номера датчиков)
Корреляционный анализ	1.1, 1.10, 1.13
Случайный лес	1.1, 1.5, 1.10, 1.18
Rfrimp	1.1,
Eli5	1.1, 1.9, 1.2

Пересечением полученных множеств является признак 1.1, содержащий измерения на уровне дневной поверхности.

## 5 Синтетические признаки

Синтетические признаки (наклон прямой, точка пересечения с осью  $y$ , коэффициент корреляции  $r$  value, двустороннее значение  $p$ -value для гипотезы о нулевом наклоне прямой и стандартное отклонение `stderr`) получены построчным построением модели линейной регрессии по дням измерений с помощью библиотеки научных и инженерных расчётов (таблица 2).

Таблица 2 - Синтетические признаки, объясняющие характер изменения температуры

Индекс	Наклон	Пересечение	Индекс	Наклон	Пересечение
204	0.9996	0.9680	2453	1.0687	0.5729

Индекс	Наклон	Пересечение	Индекс	Наклон	Пересечение
1081	1.0184	-0.2266	3214	1.0711	0.4756
1993	1.0417	1.7956	3717	1.0715	1.6598

Анализ признаков наклон и пересечение выявил, что исследуемые температурные ряды похожи по характеру тренда. Признаки  $r$ -value и  $p$ -value были исключены как слабовариативные.

## 6 Прогнозирование температуры

Прогноз температуры выполнен внутри кластеров для каждого временного ряда с помощью моделей экспоненциального сглаживания, реализованных в библиотеке statsmodels: модели Хольта, учитывающей тренд в данных и модели Хольта-Винтера, учитывающей тренд и сезонность (рисунки 6-7).

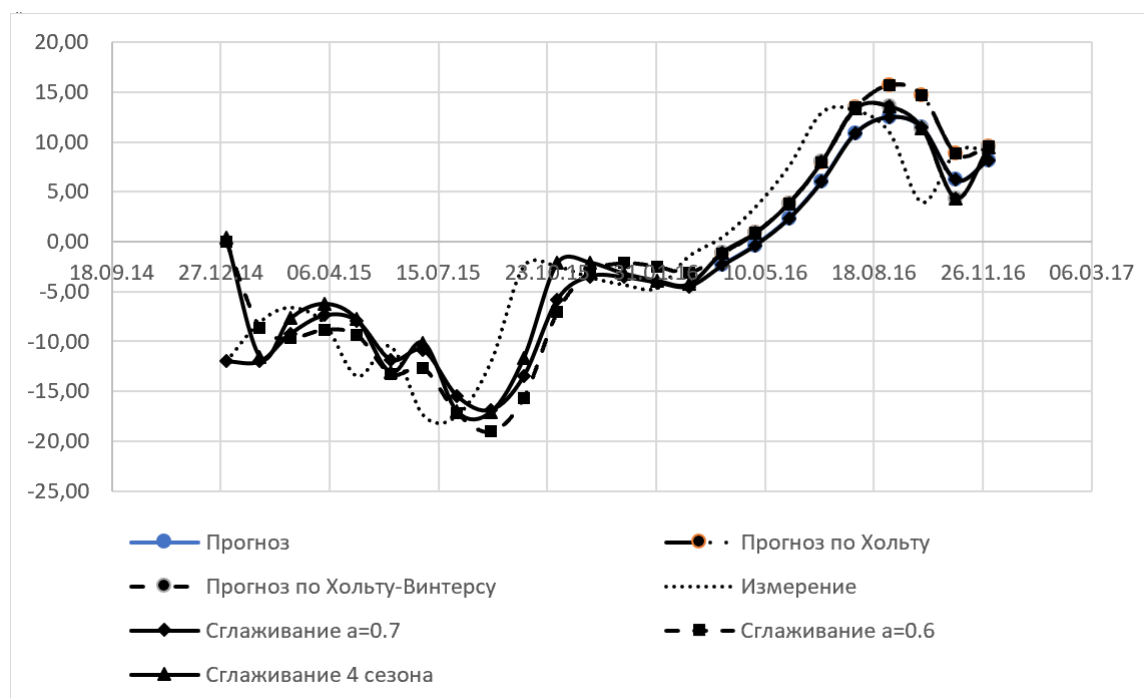


Рисунок 6 – Прогнозирование температуры на уровне дневной поверхности

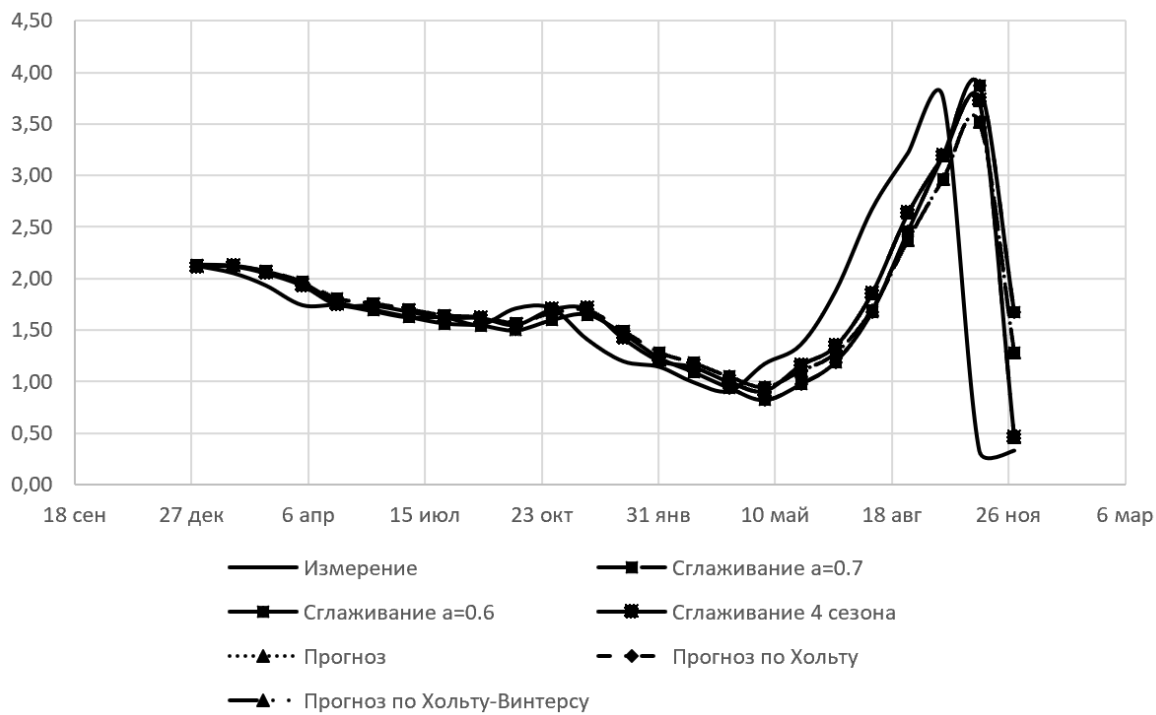


Рисунок 7– Прогнозирование температуры на глубине 3 метра (h1.7)

Параметры тренд и сезонность определены декомпозицией временных рядов. Рисунок 8 представляет декомпозицию исходного временного ряда (observed) на глубине 5 м контрольной ТС на три составляющие: тренд (trend), сезонность (seasonal) и остаток (residual).

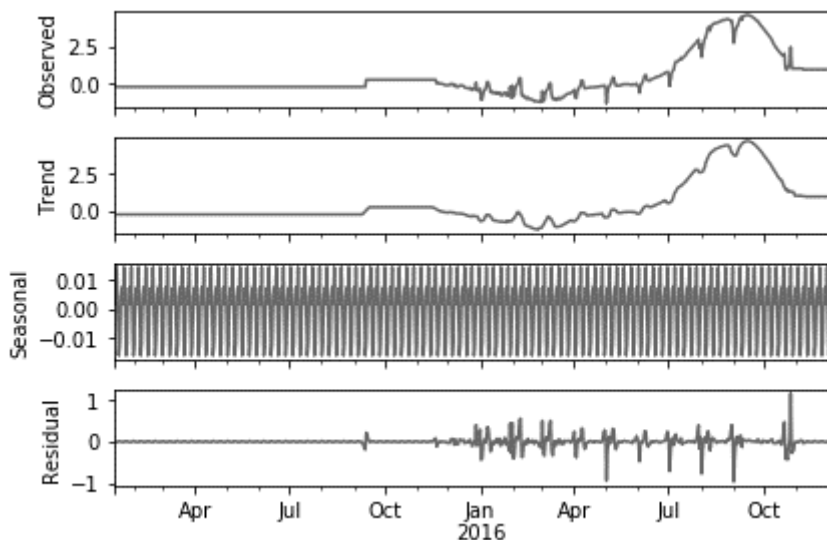


Рисунок 8 - Декомпозиция временного ряда

У температур большинства слоев грунта исходного набора данных трендовый компонент изменяется нелинейно и имеет мультипликативный характер. Амплитуда сезонных колебаний постоянна и имеет аддитивное поведение. Остаток меняется с течением времени, поэтому носит мультипликативный характер.

### Заключение

В результате исследования подготовлен набор данных по температурным измерениям вдольтрассовых грунтов; размечены данные, синтезированы дополнительные признаки временных рядов, выполнены массовые автоматизированные проверки работоспособности датчиков, проведено прогнозирование температур внутри выделенных кластеров данных.

В результате анализа и обработки экспериментальных данных, полученных в течение 2014 - 2016 гг. в процессе полевых замеров температур грунтов трассы магистрального трубопровода, получены следующие результаты о температурном поле исследуемого объекта:

- определена фактическая глубина зоны нулевых годовых колебаний температуры;
- рассчитаны среднемесячные температуры грунтов на разной глубине;
- получено семейство моделей, прогнозирующих циклическое изменение температуры грунта.

Результаты исследования позволили предложить методику, повышающую точность прогнозирования изменений температурного поля грунтов вдоль нефтепровода с 76 до 87 %.

Настоящая работа выполнена в рамках комплексного проекта Минобрнауки 10.331-17 «Моделирование и интеллектуальное управление производственными процессами в промышленности и электроэнергетике».

## Литература

1. *Реутских Н.В., Бережной М.А., Дуденко И.А.* Геотехнический мониторинг для магистральных трубопроводов в различных типах многолетнемерзлых пород // Научный журнал российского газового общества, 2016. с.22-26.
2. *Владова А.Ю.* Алгоритмическое обеспечение информационной системы геотехнического мониторинга транспорта углеводородов в условиях многолетней мерзлоты // Информационные технологии. - 2017. - Т 23. - № 3. - С. 205-212.
3. *Гишкелюк И.А., Станиловская Ю. В., Евланов Д. В.* Прогнозирование оттаивания многолетнемерзлых грунтов вокруг подземного трубопровода большой протяженности // Наука и технологии трубопроводного транспорта нефти и нефтепродуктов. 2015. № 1 (17). С. 20-25.
4. *Богданов А.И.* Система температурного мониторинга состояния многолетних грунтов в основании земляного полотна северного широтного хода. // Проектирование развития региональной сети железных дорог: сб. науч. тр.; под ред. В.С. Шварцфельда. - Хабаровск: Изд-во ДВГУПС, 2014. - Вып. 2. - С. 155-164.
5. *Zhao, X., Li, W., Zhou, L., Song, G., Ba, Q., Ho, S. C. M., & Ou, J.* // Application of support vector machine for pattern classification of active thermometry-based pipeline scour monitoring. Structural Control and Health Monitoring, 2015, 22(6),- p. 903-918.
6. *Ni, P., Mangalathu, S., Liu, K.* Enhanced fragility analysis of buried pipelines through Lasso regression // Acta Geotechnica, 2018.- p. 1-17.
7. *Shi, F., Peng, X., Liu, Z., Li, E., Hu, Y.* A data-driven approach for pipe deformation prediction based on soil properties and weather conditions // Sustainable Cities and Society, 2020. - p. 102012.
8. *Vladova A.Yu., Vladov Y.R.* Machine Classification of Pore Space for Hydrocarbon Reservoir Characterization // 2019 IEEE 21st Conference on Business Informatics (CBI), 2019. V. 1, - p. 391-396.
9. *Флах П.* Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. — М.: ДМК Пресс, 2015. — 400 с.
10. *Skiena S.S.* The data science design manual. - Springer, 2017, 445 p.