

# ФОРМИРОВАНИЕ ПРОСТРАНСТВА ПРИЗНАКОВ И АВТОРЕГРЕССИОННЫХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗА ОТСТУПЛЕНИЙ ЖЕЛЕЗНОДОРОЖНОГО ПОЛОТНА

А.Ю. Владова

**Аннотация.** Диагностика железнодорожного полотна выявляет отступления параметров рельс в плане и профиле от нормативных значений. Приближение размеров отступлений к предельным значениям требует ограничения скоростей движения поездов. Поэтому прогноз изменений размеров отступлений является актуальной проблемой. Несмотря на значительный объем собираемых данных, эксплуатирующие организации недостаточно применяют методы машинного обучения для повышения качества прогноза параметров отступлений. Предлагаемый метод отличается от известных тем, что позволяет: увеличить размерность пространства признаков путем вычисления отклонений амплитуд отступлений от нормативных значений, площадей отступлений по длинам и амплитудам или отклонениям амплитуд с последующим представлением пространства признаков в виде 3D-матрицы; сформировать множество управляющих параметров (шаг дискретизации по времени, вид сезонных колебаний, количество точек изменения тренда и др.); прогнозировать размеры отступлений в группах, отличающихся видом и положением вдоль железнодорожного полотна. Прогноз ведут в рамках оптимизационной задачи по критерию минимума эмпирического риска и в результате получают семейство авторегрессионных моделей для каждого интервала дискретизации по длине полотна.

**Ключевые слова:** временные ряды, диагностика, программный модуль, дискретный технологический процесс.

## ВВЕДЕНИЕ

Железные дороги работают в сложных условиях: свое влияние оказывают природно-климатические факторы и нагрузки от проезжающих составов. Во время диагностики железнодорожного полотна измеряют геометрию пути и выявляют отступления параметров рельс в плане и профиле от нормативных значений. Выявленные отступления разделяют следующим образом:

- ширина колеи (сужения, уширения),
- положение рельс по уровню (перекосы, плавные отклонения уровня),
- просадки по правому и левому рельсу в вертикальной плоскости (просадки правая и левая),
- положение рельс в плане (рихтовка).

Выявленные отступления привязаны к дискретным моментам времени и дистанции пути и распределены неравномерно. Приближение размеров отступлений к предельным значениям требует ограничения скоростей движения поездов. Поэтому

прогноз изменений размеров отступлений является актуальной проблемой.

Таким образом, сформулирована цель работы – повышение точности прогнозирования размеров отступлений железнодорожного полотна. Для достижения цели поставлены следующие задачи исследования: провести обзор современных методов прогнозирования размеров дефектов протяженных объектов, выполнить статистический анализ параметров отступлений как многомерных данных; сформировать множества входных и выходных параметров и комбинаций их значений; спрогнозировать размеры отступлений железнодорожного полотна.

## 1. ОБЗОР ЛИТЕРАТУРЫ

### 1.1. Общая аналитика

Анализ литературы выполнен с помощью реферативно-аналитической базы данных Dimen-

sions, объединяющей свыше 100 млн публикаций журналов, патентов и грантов. Соотношение печатных работ (в тыс. шт.) по научным категориям по запросу «*railway degradation*» (деградация железнодорожного полотна) представлено на мозаичной диаграмме рис. 1.

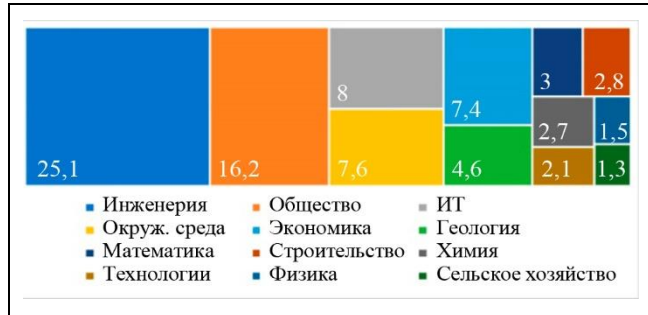


Рис. 1. Распределение работ (в тыс. шт.) по научным категориям за 2012–2022 гг.

С ростом протяженности и длительности эксплуатации железных дорог остро встают вопросы идентификации и прогнозирования их технического состояния. Этот процесс отражен в увеличении числа публикаций и их цитирований (рис. 2) за последние 11 лет. Характерный рост публикаций и цитирований наблюдается в отраслях «Инженерия» и «ИТ» и, по-видимому, связан с появлением новых сенсоров и методов обработки больших данных, ими генерируемых.

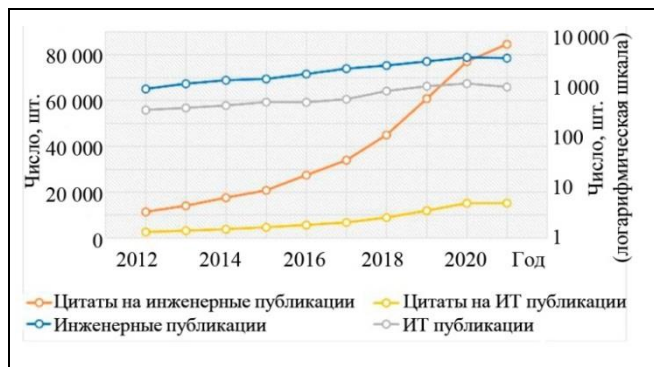


Рис. 2. Рост числа публикаций и их цитирования по запросу «*railway degradation*» в отраслях «Инженерия» и «ИТ»

## 1.2. Прогноз размеров дефектов железнодорожного полотна

Эволюция подходов, применяемых при прогнозировании технического состояния железнодорожных систем, связана с применением различных типов датчиков, а также методов, ориентированных на автоматизацию обработки собираемых ими данных. Обзорная статья [1], посвященная мони-

торингу состояния железнодорожных систем, приводит список основных направлений развития, среди которых создание бортовых систем, основанных на различных типах датчиков и методов обработки собираемых данных. В публикации [2] разработан метод классификации признаков вибрации подвижного состава для автоматического обнаружения неисправностей пути. В статье [3] представлено исследование причин быстрой деградации железнодорожного переезда. Динамические характеристики переезда оценивались с использованием сенсорной аппаратуры и было выявлено, что быстрое ухудшение характеристик переезда вызвано значительными силами удара колеса о рельс, связанными с рывком проезжающих поездов. В статье [4] выполнен статистический анализ данных диагностических вагонов, построены плотности распределений рельс по типу, новизне, назначению, классу и на этой основе созданы кластеры с рекомендациями по объемам транзитного тоннажа. В докладе [5] представлена модель оценки интенсивности износа железнодорожных путей, основанная на углах наклонов переходных кривых. Анализ литературы показал, что прогноз по многомерным временным рядам, характеризующим состояние железнодорожного полотна, преимущественно основан на двух видах моделей [6]:

- авторегрессионные, предсказывающие значения временного ряда  $\hat{z}(t)$  с учетом прошлых значений  $z(t), \dots, z(t-p)$ , например, по аддитивной модели  $\hat{z}(t) = a + \varphi_1 z(t-1) + \dots + \varphi_p z(t-p) + \epsilon(t)$ , где  $\varphi_p$  – веса,  $\epsilon(t)$  – ошибка;

- регрессионные  $\hat{z}(t) = f(x_1(t), \dots, x_n(t)) + \epsilon(t)$ , предсказывающие значения временного ряда по набору признаков  $x_1(t), \dots, x_n(t)$ .

Основным недостатком рассмотренных подходов является ограниченное использование больших данных и современных методов их обработки для повышения качества прогноза технического состояния железнодорожного полотна. В то же время в литературе, посвященной анализу климатических, геотехнических [7, 8], сейсмических данных [9], а также астрономических наблюдений широко применяются как классические многомерные параметрические модели и преобразование Фурье, так и сравнительно недавно созданные вейвлет-преобразования, канонические корреляции, траекторные матрицы. Кроме того, перечисленные модели можно положить в основу оптимизационных задач и выполнить дополнительные исследования по поиску оптимального набора параметров [10].

### 1.3. Модели прогноза временных рядов, реализованные в библиотеках машинного обучения

При анализе методов машинного обучения для прогноза временных рядов рассмотрены свободно распространяемые библиотеки Darts и GreyKite. Библиотека Darts содержит реализации регрессионных моделей ARIMA, Exponential Smoothing, Prophet, прогноз на основе быстрого преобразования Фурье и др. [11]. Библиотека GreyKite обеспечивает прогнозирование временных рядов на моделях Silverkite, Prophet и ARIMA и особенно подходит для рядов, у которых есть точки изменения тренда или сезонности [12]. В табл. 1 проведен качественный анализ трех современных моделей Silverkite, Prophet и ARIMA.

Таблица 1

#### Качественный анализ моделей

Критерий	Silverkite	Prophet	ARIMA
Скорость вычисления	Высокая	Низкая	Средняя
Точность прогноза	Очень хорошая	Хорошая	Удовлетворительная
Интерпретируемость	Хорошая	Очень хорошая	Хорошая
Простота использования	Удовлетворительная	Хорошая	Очень хорошая
Наглядность	Хорошая	Очень хорошая	Приличная

В результате анализа выбрана модель аддитивной регрессии с настраиваемыми компонентами Prophet [13]

$$\hat{z}(t) = g(t) + s(t) + h(t) + \epsilon(t), \quad (1)$$

где  $t$  – время;  $z(t)$  – фактическое значение временного ряда;  $\hat{z}(t)$  – прогнозное значение временного ряда;  $g(t)$  – компонент тренда, смоделированный с помощью кусочно-линейной, кусочно-логистической или гладкой функции;  $s(t)$  – сезонная составляющая, отвечающая за моделирование периодических изменений (сезонность оценивается с использованием частичной суммы Фурье);  $h(t)$  – компонент, отвечающий за специфические моменты перелома тренда, например график ремонтов;  $\epsilon(t)$  – ошибка, содержащая сведения, не учитываемые моделью.

Модель позволяет управлять параметрами, характеризующими сезонность временного ряда, количество точек изменения тренда, шаг по времени, а также устойчива к отсутствующим данным и

сдвигам в тренде и, как правило, хорошо справляется с выбросами.

Тонкая настройка компонентов, входящих в уравнение (1) позволит получить приемлемое качество прогноза на выходе модели, оцененное через среднюю абсолютную ошибку в процентах [14]

$$MAE \% = \frac{\sum_{t=1}^N |z(t) - \hat{z}(t)|}{\hat{z}(t)}, \quad (2)$$

где  $N$  – количество точек временного ряда.

Кроме того, для установления области, значений выходного признака модель Prophet визуализирует доверительный интервал [8].

Таким образом, ставится задача на обучающей выборке  $\{x_i : t = 1, \dots, h\}$  найти вектор управляющих параметров  $U$  модели  $A(X, U)$  по критерию минимума эмпирического риска (2). Векторы  $X$  и  $U$  могут принимать значения из заданных подмножеств евклидового пространства  $E^n$  и  $E^r$ :  $X \subset E^n, U \subset E^r$ . На векторы  $X$  и  $U$  могут быть наложены ограничения:  $g_i(X, U) \geq 0, i = 1, \dots, m$ .

## 2. ПРОГНОЗИРОВАНИЕ ПАРАМЕТРОВ ОТСТУПЛЕНИЙ

### 2.1. Статистический анализ исходных данных

Исходные данные занимают 8,9 Мб, распределены по десяти столбцам-признакам и содержат 101 тысячу строк, из которых 91 строка имеет пропуски в данных (табл. 2).

Таблица 2

#### Формат исходных данных

Присвоенный индекс	Признак	Интервал значений	Число уникальных значений
1D	Километр	[1; 650]	645
	Пикет, м	[1; 1105]	1076
2D	Год	[2018; 2021]	4
	Месяц	[1; 12]	12
	День	[1; 31]	29
3D	Код отступления	[2065; 2161]	17
	Отступление	[Отв.у; Уш]	18
Признаки	Амплитуда, мм	[6; 1543]	144
	Нормативная амплитуда, мм	[10; 1535]	156
	Длина, мм	[1; 308]	151
	Степень опасности	[1; 4]	4

Дальнейший анализ показал, что данные несбалансированы:

- По годам (рис. 3, а). В 2020 г. отступлений зарегистрировано столько же сколько в 2018-м, 2019-м и первых двух месяцах 2021 г. вместе взятых.

- По видам отступлений (рис. 3, б). Разбивка по годам и видам дана на рис. 3, г. В 2020 г. значительно возросло число диагностируемых отступлений, а в 2021 г. увеличилось число видов отступлений: с 11 в 2018 г. до 18 в 2021 г. Причем от года к году происходила смена лидеров среди

наиболее часто встречающихся отступлений: просадки (Пр.Л, Пр.П), перекосы (П), сужения (С), уширения (Уш) и плавные отклонения по уровню (У).

- По длине железнодорожного полотна – на железнодорожных узлах зарегистрировано значительно больше отступлений, чем в среднем по дистанции пути (рис. 3, в).

Вследствие малочисленности отступлений некоторых видов возможно либо их объединение в группу, либо использование bootstrap-методов для синтеза дополнительных данных.

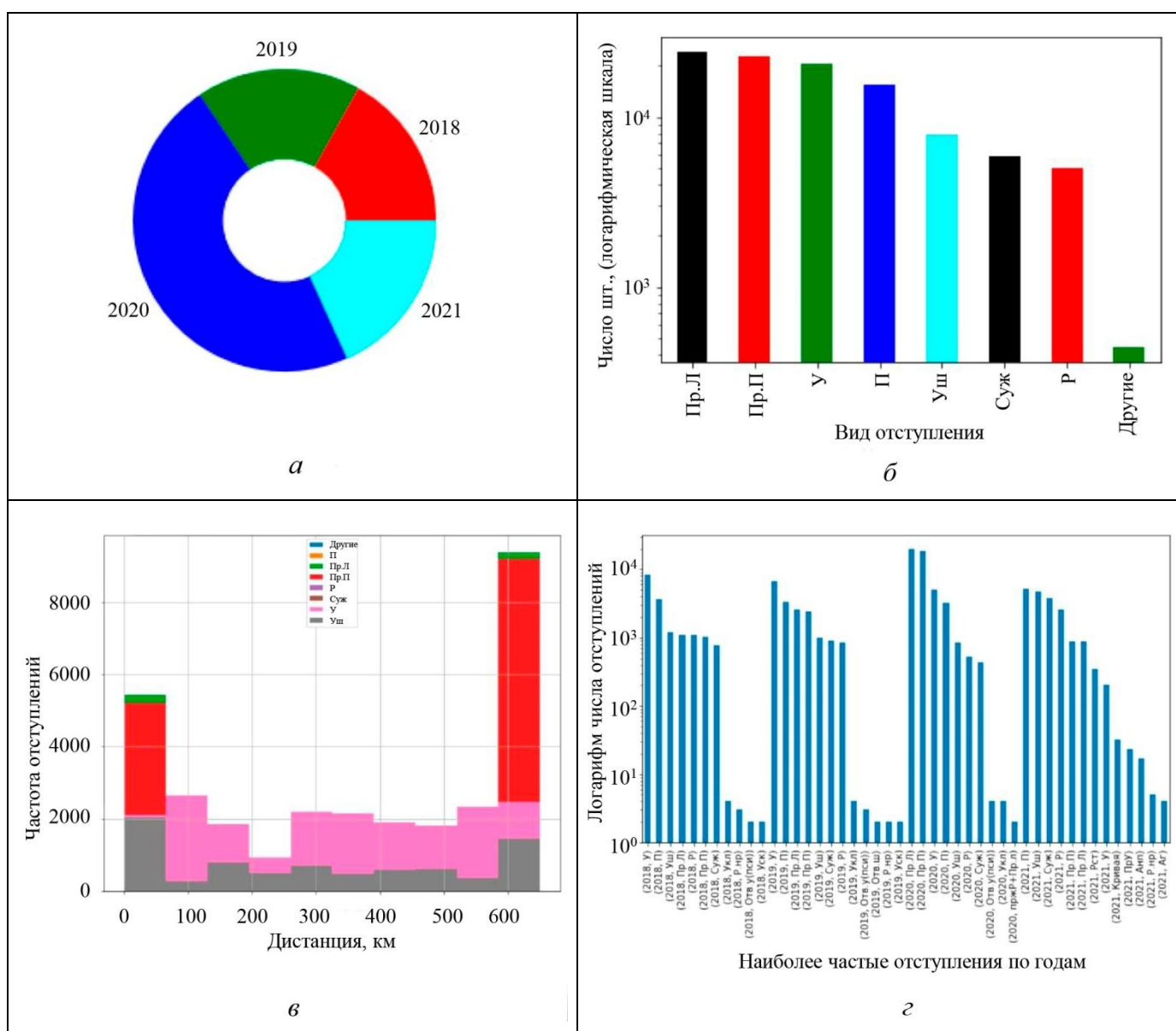


Рис. 3. Несбалансированность данных: а – по годам; б – по видам отступлений; в – по дистанции; г – по годам и видам отступлений

### 2.1.1. Тренд и сезонность

Проверка наличия тренда и сезонности проведена с помощью операций повышения и понижения дискретизации данных при изменении шага временной сетки:

- В то время как тренд на уменьшение амплитуд отклонений прослеживается при вычислении их медианных значений за каждый год, сначала уменьшаются, а затем резко растут длины отступлений (рис. 4, а). Причем точка изменения тренда соответствует 2020 г. Отметим, что агрегированный параметр «Степень опасности» индифферентен к изменениям медианной амплитуды и длины отступлений.

- Понижение дискретизации по времени (двухмесячная шкала) и логарифмирование показало (рис. 4, б), что наиболее сильно амплитуды отступлений отклоняются в июле. Эти изменения говорят о наличии сезонности в данных. Локальные экстремумы длин отступлений в какой-то мере следуют за экстремумами амплитуд с временным лагом, но не имеют выраженной сезонности. У признака «Степень опасности» также не удалось графически выделить сезонные изменения.

### 2.1.2. Увеличение размерности пространства признаков

Поскольку время представлено тремя целочисленными признаками: «Год», «Месяц», «День», на их основе сформирован вычисляемый признак «Время». Совокупность двух признаков «Километр» и «Пикет» характеризует местонахождение отступления и формирует новый признак «Дистанция». Для повышения стационарности временного ряда вычислены разницы между признаками «Амплитуда» и «Нормативная амплитуда» и создан признак «Отклонение». Для более полного

описания признака «Отклонение» с использованием плоскостного, а не точечного значения, введены два агрегированных признака «Площадь\_А» и «Площадь\_О», вычисленные как произведения длин отступлений на амплитуды и отклонения соответственно. Кроме того, для увеличения количества значений временного ряда 11 редко встречающихся отступлений агрегированы в признак «Другие». На рис. 5 продемонстрированы особенности новых признаков:

- отличия в разбросе отклонений: только положительные значения у «Уширений», «Перекосов» и «Просадок»; только отрицательные у «Сужений» со значимой степенью опасности, большая дисперсия у «Плавных отклонений по уровню» (рис. 5, а);

- разделение значений признака «Площадь\_А» на кластеры по видам отступлений при переходе на логарифмическую шкалу (рис. 5, б).

Эти отличия говорят о необходимости использования различных моделей прогнозирования [15, 16] для каждого вида отступлений либо о нормировке их значений.

### 2.1.3. Преобразование данных в 3D-матрицу

Поскольку изменение размеров каждого вида отступлений может подчиняться своему закону, а каждый километр железнодорожного полотна имеет свои геотехнические особенности, в прогнозных моделях длины и амплитуды отступлений, их нормативные значения, а также агрегированные признаки представлены в виде 3D-матрицы с осями:

- местонахождение вдоль железнодорожного полотна,
- время, как момент фиксации отступления,
- код отступления.

Фрагмент преобразованных данных представлен на рис. 6.

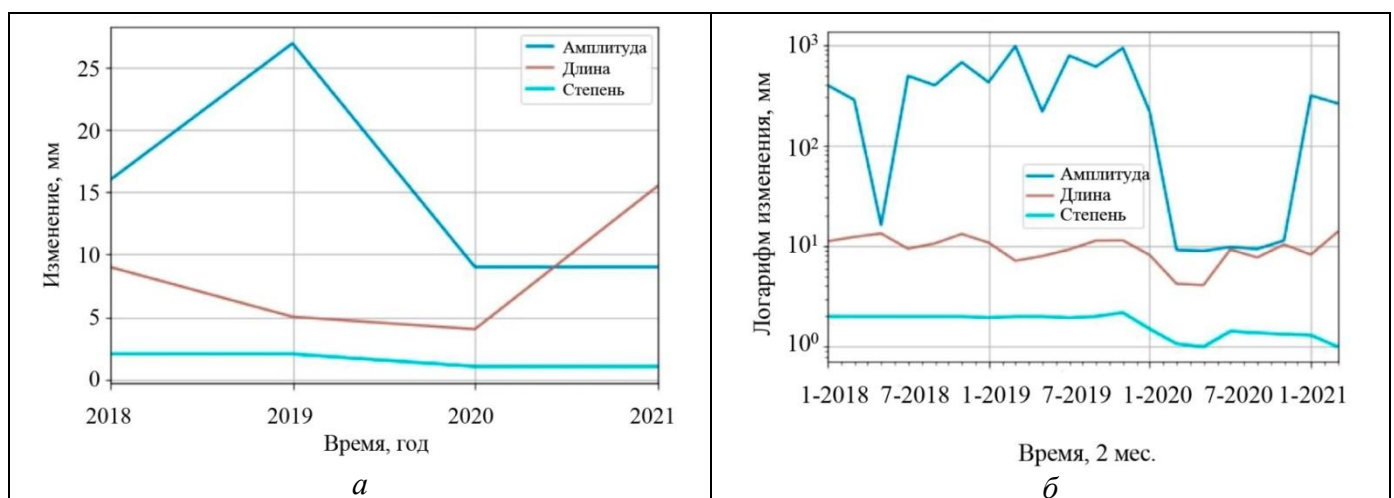


Рис. 4. Выявление: а – тренда, б – сезонности у признаков «Амплитуда», «Длина отступления», «Степень опасности»

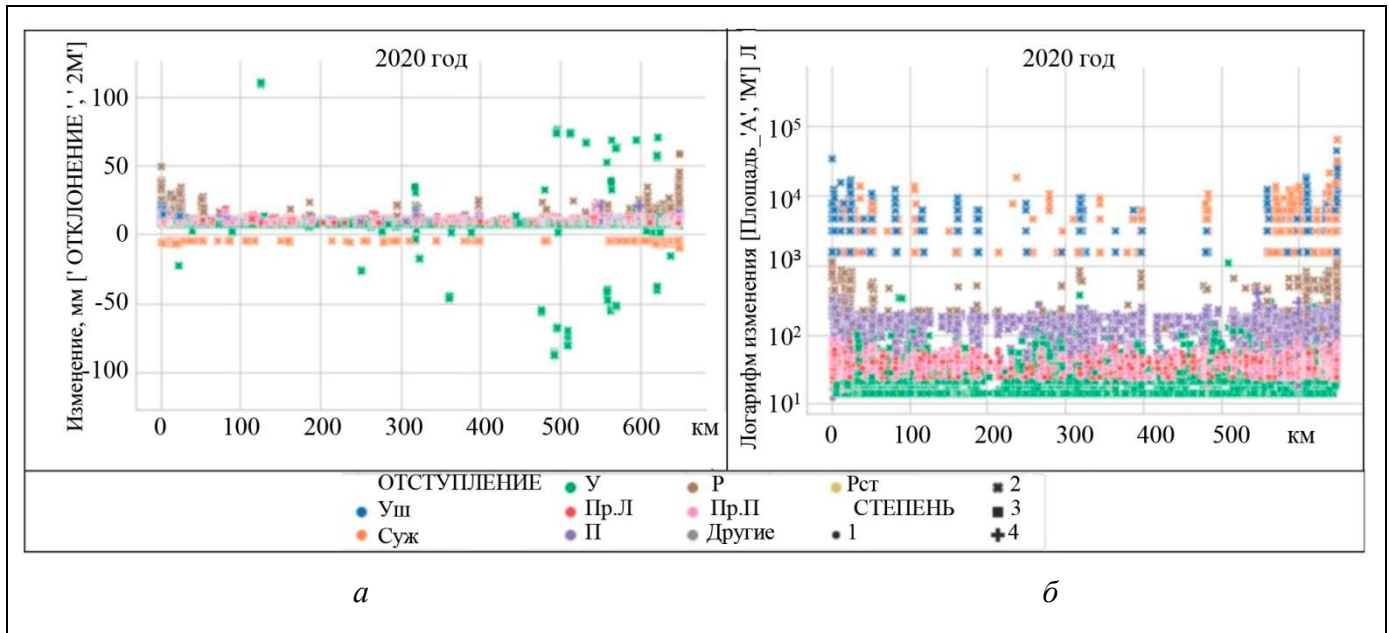


Рис. 5. Анализ дисперсии значений по видам отступлений: а – отличия в разбросе, б – кластеризация значений

КОДОТСТУП	ДИСТАНЦИЯ	ВРЕМЯ	КМ	АМПЛИТУДА	НОРМА	ДЛИНА	ОТСТУПЛЕНИЕ	ГОД	СТЕПЕНЬ	ОТКЛОНЕНИЕ	ПЛОЩАДЬ_А	ПЛОЩАДЬ_О
2081	1469	2018-01-29	1	1529.0	1520	4	Уш	2018	2	9.0	6116.0	36.0
	1614	2018-01-29	1	1530.0	1520	2	Уш	2018	2	10.0	3060.0	20.0

Рис. 6. Фрагмент преобразованных данных

#### 2.1.4. Плотности распределения вероятностей

Плотности распределения вероятностей построены на разных временных сетках (с усреднением от нескольких недель до квартала) для исходных значений длин, амплитуд, отклонений амплитуд и площадей каждого вида отступлений, а также степени отступлений. Наиболее близко к нормальному закону распределения оказались значения признаков «Отклонение» и «Площадь\_О» (рис. 7).

Поскольку плотности распределений значений оставшихся признаков имеют двухвершинное распределение, нормирование их значений проведено в интервал  $[0; 1]$  по формуле

$$x_{i, \text{norm}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (3)$$

где  $x_{i, \text{norm}}$  – нормализованное значение признака;

$x_{\min}$ ,  $x_{\max}$  – наименьшее и наибольшее значение признака;  $x_i$  –  $i$ -е исходное значение признака.

Таким образом, проведен статистический анализ параметров отступлений как многомерных данных, который выявил несбалансированность данных по годам и типам отступлений, наличие тенденций и локальных минимумов при передискретизации данных по годам, месяцам и кварталам, позволил сгруппировать наименее часто встречающиеся виды отступлений и увеличить размерность пространства признаков путем добавления положения отступлений и времени их диагностирования, вычисления отклонений амплитуд от нормативных значений, площадей отступлений по длине, амплитуде и отклонению амплитуды с последующим представлением пространства признаков в виде трехмерной матрицы.

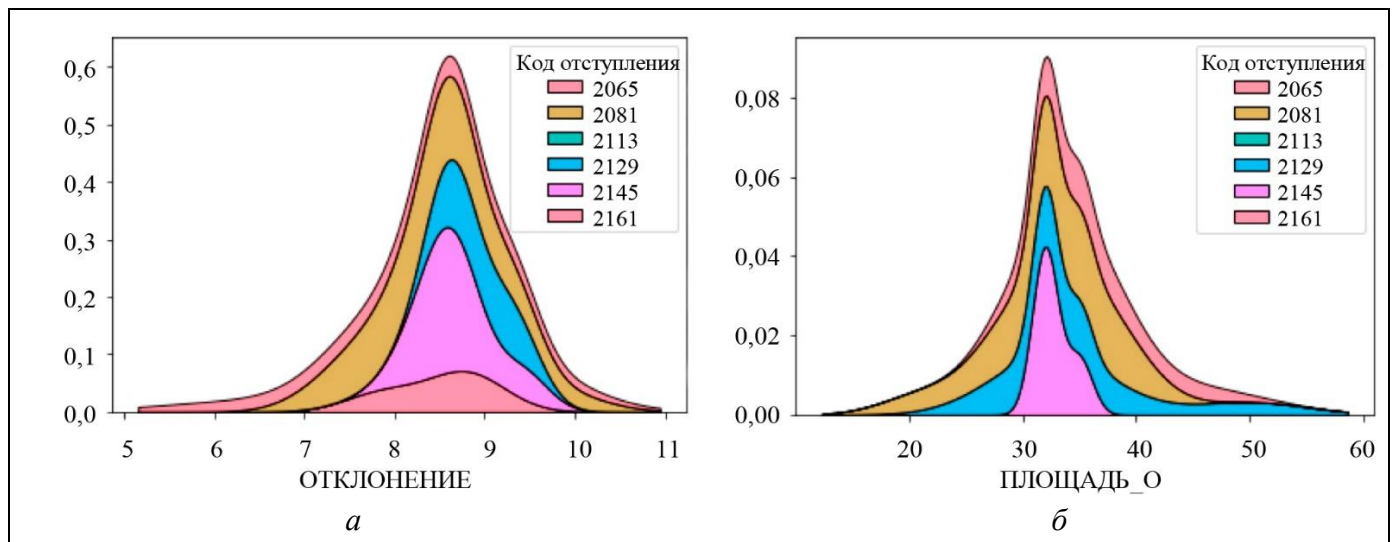


Рис. 7. Плотность распределения значений признаков: а – «Отклонение», усредненного на шестинедельной шкале; б – «Площадь\_О», усредненного на семинедельной шкале

## 2.2. Метод прогнозирования параметров отступлений железнодорожного полотна

Предлагаемый метод прогнозирования параметров отступлений железнодорожного полотна включает три этапа. На первом этапе проводят статистический анализ, во время которого объединяют редко встречающиеся виды отступлений в группу, синтезируют вектор  $X$  параметров модели: дистанцию и время, отклонение амплитуд от нормативных значений, площади отступлений. Затем индексируют значения признаков по трем осям – дистанция, код отступления и время – и тем самым формируют временные ряды для каждого километра и вида отступления.

На втором этапе увеличивают пространство признаков путем добавления нормированных и логарифмированных длин, амплитуд, отклонений амплитуд, площадей, степеней опасности. Затем задают вектор управляющих параметров  $U$ , включающий шаг дискретизации по времени, учет сезонности, а также количество точек изменения тренда. После формируют ограничения  $g_i(U)$ , учитывающие:

- глубину данных и шаг прогноза: один год; три, два и один квартал; два и один месяц; пять, три и две недели;
- тип сезонности: годовая, квартальная, месячная, недельная;
- количество точек изменения тренда, согласно числу ремонтов;
- тип учета сезонности в модели (аддитивный,

мультипликативный);

- скорость изменения сезонности (количество слагаемых в частичной сумме Фурье).

На третьем этапе, в случае наличия достаточного числа отступлений в группе, охарактеризованной видом отступления и километром полотна, обучают модель  $A$  (1) в каждой группе с оценкой эмпирического риска  $MAE\%$  (2). В результате определяют оптимальные значения вектора параметров  $U$  каждой группы по критерию минимума эмпирического риска.

Практической значимостью предлагаемого метода прогнозирования отступлений железнодорожного полотна является возможность его применения для получения новых знаний, способствующих оптимизации транспортного процесса.

## 2.3. Реализация

В табл. 3 приведены наборы параметров модели, позволившие получить наилучшее качество прогноза.

Средняя абсолютная ошибка  $MAE\%$  лежит в диапазоне от 2,6 до 8,4 % для рекомендуемых параметров модели. По результатам расчетов установлено, что:

- квартальная и годовая сетки значительно ухудшают качество прогноза;
- лучшие прогнозы выполнены для отступлений «Перекас», «Рихтовка», «Просадка левая», «Уширение» и «Другие» (агрегированный показатель из 11 редко встречающихся отступлений);

## Результаты прогнозирования

MAE%	Шаг по времени	Отступление	Отклонение	Число точек изменения
2,68	1 месяц	Перекос	ПЛОЩАДЬ_А	0
3,03	3 недели	Рихтовка	ДЛИНА	0
3,27	3 недели	Уширение	ДЛИНА	2
3,89	3 недели	Просадка левая	ОТКЛОНЕНИЕ	0
4,36	1 месяц	Перекос	АМПЛИТУДА	1
4,45	3 недели	Другие	ДЛИНА	2
4,60	5 недель	Перекос	ОТКЛОНЕНИЕ	2
5,34	1 месяц	Перекос	АМПЛИТУДА	1
5,34	1 месяц	Просадка правая	АМПЛИТУДА	2
5,74	3 недели	Другие	ПЛОЩАДЬ_О	0
5,74	3 недели	Уширение	ПЛОЩАДЬ_О	0
6,59	1 месяц	Перекос	СТЕПЕНЬ	2
7,90	5 недель	Рихтовка	ДЛИНА	2
8,39	3 недели	Другие	ПЛОЩАДЬ_А	0
8.39	3 недели	Уширение	ПЛОЩАДЬ_А	0

– следует выбирать не более двух точек изменения тренда;

– для пятидневной сетки лучшие значения показателя MAE% достигаются при количестве точек изменения тренда, равном двум;

– наименьшие значения MAE% достигаются при учете недельных, месячных и квартальных сезонностей;

– логарифмирование значений признаков ухудшает качество прогноза;

– нормирование значений признаков в интервал [0; 1] улучшает качество прогноза;

– центрирование значений путем вычитания математического ожидания и их нормирование путем деления на стандартное отклонение, а также использование мультипликативных регрессоров, отвечающих за сезонность, ухудшает качество прогнозирования.

Рис. 8, а отражает прогноз на месяц вперед, выполненный для признака «Перекос». Рис. 8, б демонстрирует прогноз на три недели вперед для признака «Уширение». Черные точки являются фактическими значениями, синяя линия представляет модельные значения, а голубая область является коридором неопределенности. Рис. 8, в, г отображают выделенные тренды, а рис. 8, д – з из-

менение сезонной составляющей: месячной и квартальной (рис. 8, д, ж), недельной и месячной (рис. 8, е, з). Для обеих моделей добавлен расчет коридора неопределенности для сезонной и трендовой составляющей.

Выполненный прогноз на один шаг временной сетки показал, что длина и амплитуда отступлений на этих участках растут. Последняя точка обучающей выборки у приведенных результатов является точкой локального экстремума. Рост амплитуд для признака «Перекос» приходится на последний месяц в квартале. Возможно, это связано с характером ремонтных работ.

Практическая значимость реализации метода заключается в том, что разработана программа на языке Python, формирующая семейство прогнозных моделей для групп по видам отступлений и километру полотна по сформированным признакам и позволяющая оптимизировать транспортный процесс.

Для манипулирования, визуализации и изучения данных использован язык программирования Python и платформа Google Colaboratory. Визуализация результатов статистического анализа выполнена с помощью функций библиотек Matplotlib, Pandas и Seaborn. Для обучения моделей использованы библиотеки SciPy, Sklearn и Prophet.



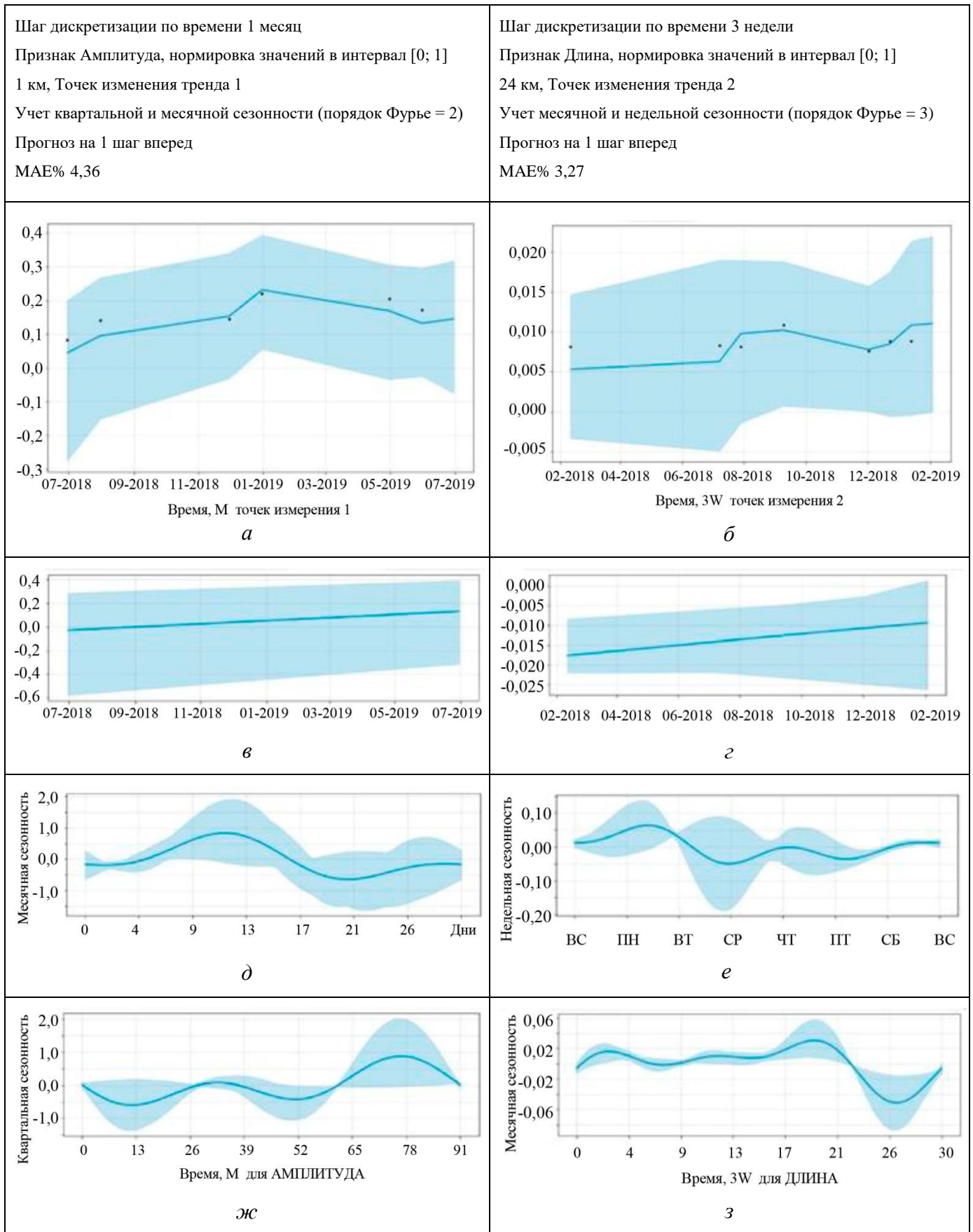


Рис. 8. Результаты прогнозирования: *а, б* – по модели, *в, г* – тренд, *д-з* – сезонность



## ЗАКЛЮЧЕНИЕ

Анализ литературы показал, что методы машинного обучения хорошо себя зарекомендовали в задачах идентификации и прогноза технического состояния объектов при наличии довольно длительной истории наблюдений.

Проведенный статистический анализ позволяет:

- выявлять количество отступлений в любой заданный промежуток времени для любого участка полотна;

- сравнивать суммы отклонений отступлений от нормированных значений для любого участка полотна в различные периоды времени;

- нормировать значения признаков в определенный диапазон;

- установить несбалансированность данных по разным периодам времени и типам отступлений;

- синтезировать управляющие параметры и признаки (время и дистанция, нормированные и логарифмированные длины, амплитуды, отклонения амплитуд, площадей, степени опасности).

В результате статистического анализа параметров отступлений железнодорожного полотна пространство признаков дополнено за счет отклонения амплитуд отступлений от нормативных значений, площадей отступлений и отклонений.

Метод прогнозирования параметров отступлений железнодорожного полотна позволяет:

- задать вектор управляющих параметров  $U$ ,

- сформировать ограничения  $g_i(U)$ ,

- получить семейство прогнозных моделей по видам отступлений и интервалам железнодорожного полотна по минимальной ошибке.

Вектор управляющих параметров прогнозной модели учитывает вид сезонных колебаний (от недельных до годовых), тип их учета в модели (аддитивный, мультипликативный), скорость изменения сезонности (количество слагаемых в частичной сумме Фурье), количество точек изменения тренда (в общем случае определено как количество точек временного ряда минус одна).

Для повышения точности прогнозирования, в том числе отступлений с неглубокой историей регистрации, предложено применить повышающую дискретизацию вдоль железнодорожного полотна (полтора километра, два километра и т. д.).

**Доступность данных.** Фрагмент кода и фрагмент данных (три года однотипных измерений на двух участках трассы), подтверждающие изложенные результаты,

можно найти на GitHub автора: <https://github.com/avladova/Railway-track-deviations> или на сайте <http://vladova.ru/About>.

## ЛИТЕРАТУРА

1. *Kostrzewski, M., Melnik, R.* Condition Monitoring of Rail Transport Systems: A Bibliometric Performance Analysis and Systematic Literature Review // *Sensors*. – 2021. – Vol. 21, no. 14. – Art. no. 4710.
2. *Tsunashima, H.* Condition Monitoring of Railway Tracks from Car-Body Vibration Using a Machine Learning Technique // *Applied Sciences*. – 2019. – Vol. 9. – Art. no. 2734.
3. *Liu, X., Markine, V.L.* Train Hunting Related Fast Degradation of a Railway Crossing-Condition Monitoring and Numerical Verification // *Sensors*. – 2020. – Vol. 20, no. 8. – Art. no. 2278.
4. *Vladova, A.Yu.* Identification of the Railway Track Technical State // 2021 14th International Conference Management of large-scale system development (MLSD). – Moscow: IEEE, 2021. – P. 1–5.
5. *Дубицкий И.С., Енин А.В., Владова А.Ю.* Анализ динамики износа железнодорожных путей // Материалы 14-й междунар. конф. «Управление развитием крупномасштабных систем» (MLSD'2021). – Москва, 2021. – С. 979–985. [*Dubitskii I.S., Enin A.V., Vladova A.Yu.* Analiz dinamiki iznosa zheleznodorozhnykh putei // Materialy 14-i mezhdunar. konf. «Upravlenie razvitiem krupnomasshtabnykh sistem» (MLSD'2021). – Moskva, 2021. – С. 979–985. (In Russian)]
6. *Лукашин Ю.П.* Адаптивные методы краткосрочного прогнозирования временных рядов. – М: Финансы и статистика (ФС), 2003. [*Lukashin, Yu.P.* Adaptivnyye metody kratkosrochnogo prognozirovaniya vremennykh ryadov. – M: Finansy i statistika (FS), 2003. (In Russian)]
7. *Владова А.Ю.* Гармонический анализ температур мерзлого грунта полосы отведения линейного объекта // Безопасность труда в промышленности. – 2017. – № 7. – С. 25–30. [*Vladova, A.Yu.* Garmonicheskij analiz temperatur merzlogo grunta polosa otvedeniya linejnogo objekta // Bezopasnost' truda v promyshlennosti. – 2017. – No. 7. – P. 25–30. (In Russian)]
8. *Vladova, A.Yu.* Remote Geotechnical Monitoring of a Buried Oil Pipeline // *Mathematics*. – 2022. – Vol. 10, no. 11. – Art. no. 1813.
9. *Любушин А.А.* Анализ данных систем геофизического и экологического мониторинга. – М.: Наука, 2007. [*Lyubushin, A.A.* Analiz dannykh sistem geofizicheskogo i ekologicheskogo monitoringa. – M.: Nauka, 2007. (In Russian)]
10. *Валеев С.Г.* Регрессионное моделирование при обработке данных. – Казань: Изд-во ФЭН Академии наук Республики Татарстан, 2001. [*Valeev, S.G.* Regressionnoe modelirovanie pri obrabotke dannykh. – Kazan': Izd-vo FEN Akademii nauk Respubliki Tatarstan, 2001. (In Russian)]
11. *Herzen, J., Lässig, F., Piazzetta, S.G., et al.* Darts: User-Friendly Modern Machine Learning for Time Series // *Journal of Machine Learning Research*. – 2022. – Vol. 23. – P. 1–6.
12. *Hosseini, R., Yang, K., Chen, A., Patra, S.* A flexible forecasting model for production systems // *ArXiv*: 2105.01098v1. – 2021. – DOI: <https://doi.org/10.48550/arXiv.2105.01098>.
13. *Taylor, S., Letham, B.* Forecasting at scale // *The American Statistician*. – 2018. – Vol. 72, no. 1. – P. 37–45.

14. Vandeput, N. Forecast KPI: RMSE, MAE, MAPE & Bias Towards Data Science. – URL: <https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d> (Дата обращения 04.02.2023) [Accessed February 4, 2023].
15. Hartomo, K., Nataliani, Y. A new model for learning-based forecasting procedure by combining k-means clustering and time series forecasting algorithms // PeerJ Computer Science. – 2021. – Vol. 7, no. 2. – Art. no. e534. – DOI: <https://doi.org/10.7717/peerj-cs.534>.
16. Vivas E., Allende-Cid H., Salas R. A Systematic Review of Statistical and Machine Learning Methods for Electrical Power Forecasting with Reported MAPE Score // Entropy. – 2020. – Vol. 22, no. 12. – Art. no. 1412.

*Статья представлена к публикации членом редколлегии  
А.С. Манделем.*

*Поступила в редакцию 30.11.2021,  
после доработки 27.03.2023.  
Принята к публикации 13.04.2023.*

**Владова Алла Юрьевна** – д-р техн. наук, Институт проблем управления им. В.А. Трапезникова РАН; Финансовый университет при Правительстве РФ, г. Москва,  
✉ [avladova@ipu.ru](mailto:avladova@ipu.ru)

## CREATING FEATURE SPACES AND AUTOREGRESSIVE MODELS TO FORECAST RAILWAY TRACK DEVIATIONS

A.Yu. Vladova

Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia  
Financial University under the Government of the Russian Federation, Moscow, Russia

✉ [avladova@ipu.ru](mailto:avladova@ipu.ru)

**Abstract.** Diagnosis of railway tracks reveals the deviations of rail parameters in the plan and profile from their nominal values. If the deviations approach the limit values, the speeds of trains must be reduced. Therefore, forecasting changes in the deviations is a topical problem. Despite the significant amount of diagnostic data collected, railway operators underuse machine learning methods to improve the quality of prediction. The proposed approach differs from known counterparts as follows. First, the dimensionality of the feature space is increased by calculating the variation of the amplitudes of deviations from the nominal values and two types of areas (the deviation length multiplied by the amplitude and the deviation length multiplied by the variation of the amplitude); subsequently, this space is represented in the 3D matrix form. Second, a set of control parameters is formed; it includes the time and space discretization step, the type of seasonal fluctuations, the number of trend change points, etc. Third, the deviations are predicted in groups differing in type and position along the track. Forecasting is based on minimizing the empirical risk criterion. As a result, a family of autoregressive models is obtained for each discretization interval along the length of the railway track.

**Keywords:** time series, diagnosis, software package, discrete technological process.