

ПРИМЕНЕНИЕ СИСТЕМНОГО АНАЛИЗА И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В СРЕДЕ LOGINOM ДЛЯ КЛАСТЕРИЗАЦИИ ДАННЫХ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА АТМОСФЕРНОГО ВОЗДУХА МЕГАПОЛИСА

APPLICATION OF SYSTEM ANALYSIS AND MACHINE LEARNING METHODS IN THE LOGINOM ENVIRONMENT FOR CLUSTERING ENVIRONMENTAL MONITORING DATA OF THE MEGACITY

*A. Krynetskaya
V. Trushina*

Summary. The paper considers the use of system analysis and machine learning tools implemented in the domestic Loginom analytical platform for the study of environmental monitoring data for atmospheric air in Moscow. The data obtained from the open data portal of the Moscow Government, containing average monthly concentrations of pollutants measured by automatic air pollution control stations, was downloaded, preprocessed, filtered and clustered. The correlation between the concentrations of pollutants and the types of observation zones (highways, residential areas, natural areas, etc.) was estimated. Cluster analysis methods and the Pearson correlation coefficient were used to test the hypothesis that the spatial distribution of pollutants depends on the nature of anthropogenic load. It is shown that the Loginom platform allows you to implement a full cycle of analytical data processing without programming, ensuring reproducibility and visualization of results. The work confirms the possibility of using Russian data mining tools for environmental monitoring tasks.

Keywords: system analysis, machine learning, Loginom, clustering, Pearson correlation, air pollution, environmental monitoring, data processing.

Крынецкая Анастасия Дмитриевна

Ассистент, Российский технологический университет МИРЭА, г. Москва
krynetskaya@mirea.ru

Трушина Вероника Игоревна

Ассистент, Российский технологический университет МИРЭА, г. Москва
trushina_v@mirea.ru

Аннотация. В работе рассматривается применение инструментов системного анализа и машинного обучения, реализованных в отечественной аналитической платформе Loginom, для исследования экологических данных мониторинга атмосферного воздуха города Москвы. Проведена загрузка, предобработка, фильтрация и кластеризация данных, полученных с портала открытых данных Правительства Москвы, содержащих среднемесячные показатели концентраций загрязняющих веществ, измеренных автоматическими станциями контроля загрязнения атмосферы. Оценена корреляция между концентрациями загрязняющих веществ и типами зон наблюдения (автомагистрали, жилые территории, природные территории и др.). Используются методы кластерного анализа и коэффициент корреляции Пирсона для проверки гипотезы о зависимости пространственного распределения загрязнений от характера антропогенной нагрузки. Показано, что платформа Loginom позволяет реализовать полный цикл аналитической обработки данных без программирования, обеспечивая воспроизводимость и визуализацию результатов. Работа подтверждает возможность применения российских инструментов интеллектуального анализа данных для задач экологического мониторинга.

Ключевые слова: системный анализ, машинное обучение, Loginom, кластеризация, корреляция Пирсона, загрязнение воздуха, экологический мониторинг, обработка данных.

Введение

Современные мегаполисы сталкиваются с комплексной проблемой загрязнения атмосферного воздуха. Интенсивное развитие транспорта, промышленности и коммунального хозяйства приводит к росту концентраций оксидов азота, угарного газа, сероводорода и других веществ, негативно влияющих на здоровье населения и состояние экосистем.

Для выработки эффективных мер управления качеством воздуха необходим системный подход, позволяющий анализировать большие массивы экологических данных.

Системный анализ обеспечивает методологическую основу для исследования сложных взаимосвязанных процессов и принятия решений в условиях неопределённости. В сочетании с методами машинного обучения он позволяет не только выявлять закономерности, но и строить модели прогнозирования состояния среды.

Особую актуальность приобретает применение российских low code-платформ анализа данных, таких как Loginom, которые позволяют проводить статистический и интеллектуальный анализ без написания кода, обеспечивая прозрачность, воспроизводимость и визуализацию этапов обработки.

Целью настоящей работы является разработка и апробация методики кластерного анализа данных экологического мониторинга атмосферного воздуха Москвы в среде LogiNot и оценка эффективности методов системного анализа для выявления пространственных закономерностей загрязнения.

Обзор литературы и существующих методов

В последние годы наблюдается рост интереса к применению машинного обучения в задачах экологического анализа. Методы кластеризации активно используются для классификации регионов по уровню загрязнения воздуха [1]. Исследователи применяют алгоритмы *k-means* [2], DBSCAN [3], иерархическую кластеризацию и метод главных компонент (PCA) для выявления закономерностей между показателями загрязнения и факторами окружающей среды.

Работы показали, что использование комбинированных подходов, включающих корреляционный анализ и машинное обучение, повышает точность прогнозирования концентраций загрязняющих веществ [4–6]. Отечественные исследователи подчёркивают необходимость создания отечественных аналитических инструментов, способных решать задачи Data Mining без привлечения зарубежного ПО [7–9].

Платформа LogiNot представляет собой универсальную среду визуального проектирования аналитических процессов [10]. Она объединяет средства импорта данных, статистического анализа, машинного обучения и визуализации. Применение LogiNot в экологических задачах описано в отдельных кейсах, однако систематические исследования её потенциала для анализа экологических данных пока ограничены.

Таким образом, настоящая работа восполняет этот пробел, демонстрируя практическое использование LogiNot в рамках методологии системного анализа.

Источник данных и структура набора

Для анализа использовались данные с Портала открытых данных Правительства Москвы, содержащие среднемесячные показатели концентраций загрязняющих веществ, измеренные автоматическими станциями контроля загрязнения атмосферы (АКЗА) [11].

Каждая запись включает:

- дату измерения;
- тип зоны наблюдения (автомагистраль, жилая территория, природная зона, промышленная зона, Новая Москва и др.);
- значения концентраций загрязняющих веществ: оксид углерода (CO), диоксид азота (NO₂), ок-

сид азота (NO), сероводород (H₂S), аммиак (NH₃), сернистый ангидрид (SO₂), взвешенные частицы (PM₁₀, PM_{2.5}) и др.

Набор данных охватывает более 35 000 записей, полученных в течение нескольких месяцев с различных станций мониторинга. Часть данных представлена на рисунке 1.

Импорт и преобработка

Импорт данных осуществляется через компонент «Узел импорта», позволяющий выбрать Excel-файл и определить схему таблицы.

Далее используется компонент «Калькулятор» для преобразования формата дат и замены символов десятичных знаков (Рисунок 2).

Для удобства анализа столбцы переименовываются в компоненте «Параметры полей». После этого выполняется очистка данных — удаление или заполнение пропусков.

Фильтрация и заполнение пропусков

Рассмотрены два подхода:

1. Удаление строк с пропусками.
2. Заполнение пропусков медианными, средними или наиболее вероятными значениями с помощью компонента «Заполнение пропусков».

Условие фильтрации: строка не должна содержать пустых значений в ключевых полях (Рисунок 3). После фильтрации из 35 522 записей осталось 20 229 (Рисунок 4).

Кластеризация данных

Для анализа закономерностей применён компонент «Кластеризация». Использовался алгоритм *k-means*, реализованный в LogiNot [12]. Количество кластеров *k* выбиралось на основе метода локтя (Elbow Method), при котором анализируется зависимость внутрикластерной суммы квадратов $W(k)$ от числа кластеров:

$$W(k) = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|,$$

где C_i — кластер, μ_i — его центр.

Оптимальным считается значение *k*, при котором дальнейшее увеличение числа кластеров не приводит к существенному снижению $W(k)$.

Оценка качества кластеризации

Для оценки применялись два подхода:

№	Период измерения	Название станции	Характеристика зоны наблюдения	Адресный ориентир	Исчерпывающий параметр	Среднее значение в абсолютных единицах	Среднее значение в долях ПДВ/СС
1	02.2016	Лесной остров	Природные территории	улица Роплага, дом 4, строение 5	Оксид углерода	0.19	0.06
2	02.2016	Глебовская	Смешанные антропогенные источники загрязнений	Глебовская улица, дом 3, корпус 1	Оксид углерода	0.32	0.11
3	02.2016	Глебовская	Смешанные антропогенные источники загрязнений	Глебовская улица, дом 3, корпус 1	Сервоаэрозоль	0.001	
4	02.2016	Глебовская	Смешанные антропогенные источники загрязнений	Глебовская улица, дом 3, корпус 1	Оксид азота	0.016	0.23
5	02.2016	Глебовская	Смешанные антропогенные источники загрязнений	Глебовская улица, дом 3, корпус 1	Диоксид азота	0.024	0.59
6	02.2016	Кожухово	Смешанные антропогенные источники загрязнений	Лушановская улица, дом 34	Сумма углеводородных соединений за вычетом метана	0.12	
7	02.2016	Кожухово	Смешанные антропогенные источники загрязнений	Лушановская улица, дом 34	Сумма углеводородных соединений	1.7	
8	02.2016	Кожухово	Смешанные антропогенные источники загрязнений	Лушановская улица, дом 34	Метан	1.58	
9	02.2016	Кожухово	Смешанные антропогенные источники загрязнений	Лушановская улица, дом 34	Оксид углерода	0.22	0.07
10	02.2016	Кожухово	Смешанные антропогенные источники загрязнений	Лушановская улица, дом 34	Сервоаэрозоль	0.0007	

Рис. 1. Пример данных

Калькулятор

Имя выражения: **Выражение1**

9.0 Expr1

9.0 Expr2

9.0 Expr3

Метка: **Выражение1**

Выражение1

Выражение2

Выражение3

Предпросмотр... | AND OR NOT XOR | = < > <= >= | 9.0 " " | FALSE TRUE

StrToDate(Concat("01.",Period), "DD.MM.YYYY")

Рис. 2. Компонент «Калькулятор» и редактирование выражений

Фильтрация данных

Состояние входа: Не активировано Активировать

9.0 **Выражение3** не пустое × И.И.

9.0 **Выражение4** > 0.00 ×

+

9.0 **Выражение3** > 0.00 ×

И.И.

Фильтр строк

Параметры полей

Соответствуют условию

Порт активен, данные есть

Рис. 3. Фильтрация данных по портам

#	Выражение3	Выражение4	Выражение2	ID Код	global_id global_id	StationName Название станции	SurveillanceZoneCharacteristics Характеристика зоны наблюдения	AdmArea Админ
1	0,32	0,11	01.02.2016	2	272777249	Глебовская	Смешанные антропогенные источники загрязнений	Восточный адми
2	0,34	0,11	01.02.2016	18	272777265	Новокозино	Смешанные антропогенные источники загрязнений	Восточный адми
3	0,29	0,1	01.02.2016	24	272777271	МГУ	Природные территории	Западный админ
4	0,3	0,1	01.02.2016	33	272777280	Толбухина	Жилые территории	Западный админ
5	0,31	0,1	01.02.2016	37	272777284	Зеленоград 6	Жилые территории	Зеленоградский
6	0,49	0,16	01.02.2016	47	272777294	Зеленоград 16	Жилые территории	Зеленоградский
7	0,45	0,15	01.02.2016	60	272777307	Щербинка	Территория Новой Москвы	Новомосковский
8	0,47	0,16	01.02.2016	67	272777314	МАДИ	Вблизи автомагистралей	Северный админ
9	0,48	0,16	01.02.2016	76	272777323	Долгопрудная	Смешанные антропогенные источники загрязнений	Северный админ
10	0,32	0,11	01.02.2016	82	272777329	Колтевский	Жилые территории	Северный админ
11	0,84	0,28	01.02.2016	85	272777332	Нижняя Масловка	Вблизи автомагистралей	Северный админ
12	0,31	0,1	01.02.2016	92	272777339	Светлый проезд	Вблизи автомагистралей	Северный админ
13	0,5	0,17	01.02.2016	98	272777345	Полярная	Смешанные антропогенные источники загрязнений	Северо-Восточн
14	0,32	0,11	01.02.2016	106	272777353	Останкино 0	Жилые территории	Северо-Восточн
15	0,3	0,1	01.02.2016	111	272777358	Народного ополчения	Вблизи автомагистралей	Северо-Западн
16	0,31	0,1	01.02.2016	120	272777367	Туристская	Жилые территории	Северо-Западн
17	0,4	0,13	01.02.2016	151	272777398	Сухаревская площадь	Вблизи автомагистралей 1	Центральный ад
18	0,31	0,1	01.02.2016	179	272777426	Марьино	Смешанные антропогенные источники загрязнений	Юго-Восточный
19	0,55	0,18	01.02.2016	186	272777433	Люблино	Смешанные антропогенные источники загрязнений	Юго-Восточный
20	0,4	0,13	01.02.2016	206	272777453	Гурьянова	Смешанные антропогенные источники загрязнений	Юго-Восточный
21	0,37	0,12	01.02.2016	217	272777464	Черемушки	Жилые территории	Юго-Западный а
22	0,33	0,11	01.02.2016	223	272777470	Мелитопольская	Смешанные антропогенные источники загрязнений	Юго-Западный а
23	0,44	0,15	01.02.2016	239	272777486	Бирюлево	Смешанные антропогенные источники загрязнений	Южный админис
24	0,42	0,14	01.02.2016	246	272777493	Кожуховский проезд	Вблизи автомагистралей	Южный админис
25	0,29	0,1	01.02.2016	264	272777511	Гурьевский проезд	Смешанные антропогенные источники загрязнений	Южный админис
26	0,47	0,16	01.02.2016	269	272777516	Пролетарский проспект	Смешанные антропогенные источники загрязнений	Южный админис
27	0,37	0,12	01.03.2016	277	272777541	Глебовская	Смешанные антропогенные источники загрязнений	Восточный адми
2978	0,36	0,12	01.03.2016	293	272777557	Новокозино	Смешанные антропогенные источники загрязнений	Восточный адми

Рис. 4. Таблица с очищенными данными

№	Метка	Доля	Итого
5	Смешанные антро...		7183
3	Жилые территории		5116
1	Вблизи автомагис...		5052
7	Территория Ново...		1915
4	Природные терри...		766
2	Вблизи автомагис...		168
6	Специальная		29

Рис. 5. Распределение кластеров по характеристикам зон наблюдений

- коэффициент корреляции Пирсона между концентрациями веществ и типами зон наблюдения;
- индекс силуэта S , характеризующий качество разбиения:

$$S = \frac{b - a}{\max(a, b)},$$

где a — среднее внутрикластерное расстояние, b — минимальное расстояние до ближайшего кластера.

Коэффициент Пирсона рассчитывается по формуле:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)\left(\sum_{i=1}^n (y_i - \bar{y})^2\right)}}$$

Результаты анализа

В Таблице 1 представлено сравнение способов обработки пропусков.

Таблица 1.

Таблица корреляционного анализа

Метод	Кол-во записей	Средний r (Пирсон)	Комментарий
Удаление строк	20 229	0,53	Связь слабая, но прослеживается зависимость между зонами и концентрациями
Заполнение средними	35 522	0,44	Снижение силы связи, вероятно из-за усреднения данных

Слабая, но статистически значимая корреляция указывает на существование пространственной зависимости загрязнений от типа зоны.

Пространственное распределение загрязнений

Кластеризация выделила 4 основных кластера, примерно соответствующих:

- зонам с высокой транспортной нагрузкой;
- промышленным зонам;
- жилым районам;
- природным и пригородным территориям.

Наибольшие значения концентраций CO и NO_2 наблюдаются в кластере 1, соответствующем автомаги-

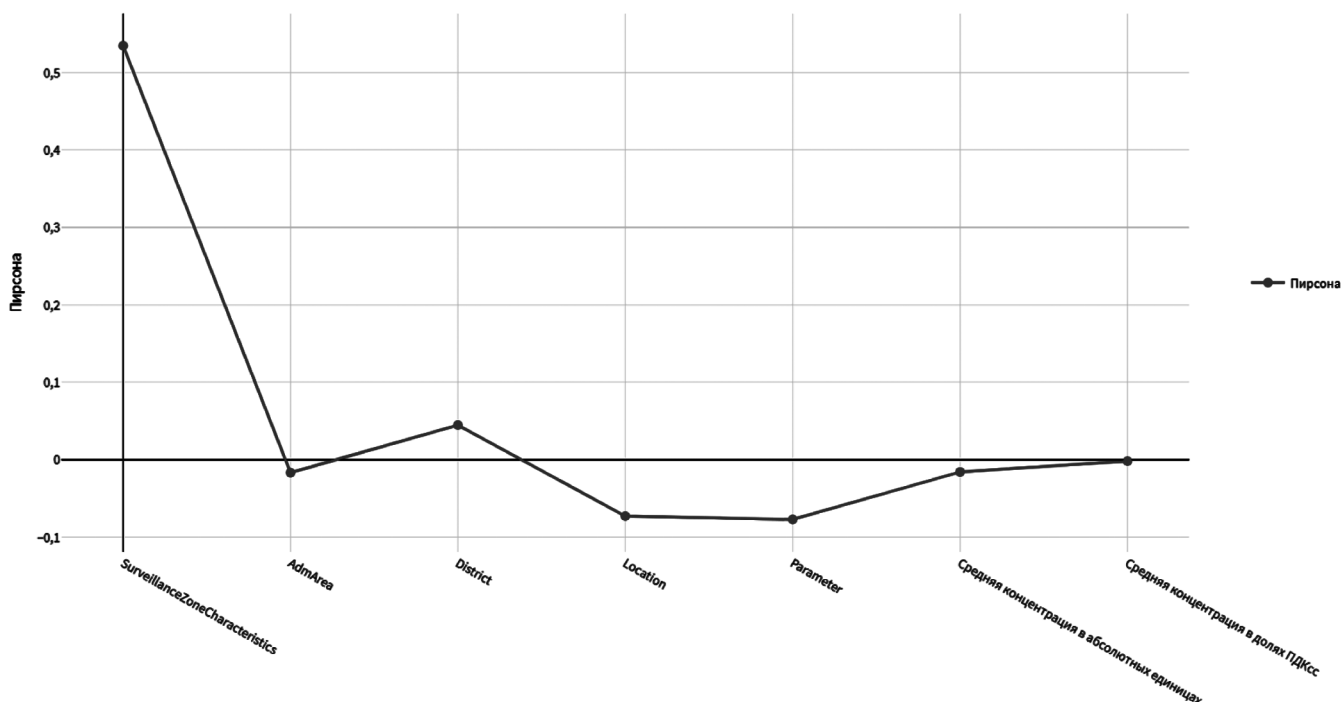


Рис. 6. Диаграмма распределения кластеров

стралям. В кластере 4 концентрации минимальны, что подтверждает адекватность модели.

Обсуждение результатов

Полученные результаты подтверждают, что даже при слабой линейной корреляции между параметрами, алгоритмы машинного обучения способны выявлять значимые закономерности.

Коэффициент Пирсона 0,53 свидетельствует о наличии умеренной зависимости между типом зоны и уровнем загрязнения, что объясняется влиянием метеоусловий, сезонности и различий в количестве измерений по станциям. При заполнении пропусков медианными значениями наблюдается снижение корреляции, так как уменьшается дисперсия выборки.

Применение Loginom позволило реализовать полный аналитический цикл — от импорта до визуализации — в графической форме, что делает процесс воспроизводимым и доступным для специалистов без опыта программирования.

Использование системного подхода позволило рассмотреть процесс мониторинга как совокупность взаимосвязанных подсистем: источник данных, процесс измерения, метод анализа и управленческие решения. Это соответствует принципам системного анализа, где каждая подсистема имеет входы, выходы и связи с другими элементами.

Заключение

В ходе исследования реализована методика анализа данных экологического мониторинга в среде Loginom.

Основные результаты:

1. Разработан алгоритм загрузки, предобработки и кластеризации данных мониторинга атмосферного воздуха.
2. Выявлено четыре кластера, отражающих пространственные различия загрязнения.
3. Коэффициент корреляции Пирсона между зонами наблюдений и концентрациями веществ составил 0,53, что подтверждает наличие зависимости.
4. Показано, что использование Loginom обеспечивает удобство визуализации и позволяет воспроизводить аналитические сценарии.

Практическая значимость работы заключается в возможности использования предложенного подхода для оперативного анализа и классификации зон по уровню загрязнения, что способствует повышению эффективности систем экологического мониторинга.

В дальнейшем планируется:

- использовать временные ряды для анализа динамики загрязнений;
- применять методы прогнозирования (регрессия, нейронные сети);
- интегрировать результаты в систему поддержки принятия решений по управлению качеством воздуха.

ЛИТЕРАТУРА

1. Садовникова Н.А. Кластеризация регионов по масштабам загрязнения атмосферного воздуха / Н.А. Садовникова // Повышение открытости отечественной статистики: материалы Международной научно-практической конференции, посвященной профессиональному празднику — Дню работника статистики, Москва, 24 июня 2016 года / Российский экономический университет имени Г.В. Плеханова, Федеральная служба государственной статистики. — Москва: Российский экономический университет имени Г.В. Плеханова, 2016. — С. 154–158. — EDN WALDZX.
2. Aliguliyev R. Performance comparison of K-means, parallel K-means, and K-means++ / R. Aliguliyev, Sh. F. Tahirzada // Reliability: Theory & Applications. — 2025. — Vol. 20, No. S7(83). — P. 169–176. — DOI 10.24412/1932–2321-2025-783-169-176. — EDN DBQWCO.
3. Correlation Analysis of PMS Data of Incheon Using DBSCAN Clustering / S.H. Lee, Ja.H. Lee, Ye.T. Kim, J.H. Jeong // International Journal of Highway Engineering. — 2022. — Vol. 24, No. 6. — P. 181–188. — DOI 10.7855/ijhe.2022.24.6.181. — EDN GHAJBH.
4. Зупарова В.В. Прогноз динамики загрязнения воздуха с использованием машинного обучения / В.В. Зупарова // Интеллектуальный потенциал России 2025 : Сборник статей по итогам Всероссийской научно-практической конференции с международным участием, Калуга, 04 августа 2025 года. — Стрелитамак: ООО «Агентство международных исследований», 2025. — С. 71–73. — EDN QWXPUP.
5. Ведерникова С.Д. Использование методов машинного обучения для прогнозирования загрязнения атмосферного воздуха / С.Д. Ведерникова // Современные проблемы географии и геоэкологии: сборник научных статей. — Смоленск: Смоленский государственный университет, 2024. — С. 5–11. — EDN NYTXHL.
6. Прогнозирование уровня загрязнения воздуха при помощи машинного обучения / А.И. Сулов, М.А. Криницкий, Ш. Стаке, С.А. Алымкулов // Известия Кыргызского государственного технического университета им. И. Раззакова. — 2022. — № 4(64). — С. 490–498. — DOI 10.56634/16948335.2022.4.490-498. — EDN ZNWEUJ.
7. Свидетельство о государственной регистрации программы для ЭВМ № 2022661079 Российская Федерация. Программа прогнозирования загрязнения атмосферного воздуха на основе методов машинного обучения: № 2022660029: заявл. 01.06.2022: опубл. 15.06.2022 / А.А. Толстых, А.И. Ситников, В.А. Никитенко. — EDN DWXDVP.
8. Капленкова П.А. Предсказывание загрязнения атмосферного воздуха с помощью машинного обучения и PySpark / П.А. Капленкова, А.Н. Сивова // Наука и бизнес: пути развития. — 2020. — № 10(112). — С. 54–56. — EDN LTZEYE.
9. Тагиева Я.Р. Использование методов машинного обучения для прогнозирования уровня загрязнения воздуха / Я.Р. Тагиева // Устойчивое развитие науки и образования. — 2020. — № 3(42). — С. 93–97. — EDN UTOINH.
10. Колесников А.А. Loginom: основные возможности / А.А. Колесников, С.В. Пальмов // Форум молодых ученых. — 2018. — № 10(26). — С. 582–587. — EDN YVSMKL.
11. Правительство Москвы. Данные автоматических станций контроля загрязнения атмосферного воздуха [Электронный ресурс]. — Режим доступа: <https://data.mos.ru/opendata/2453> (дата обращения: 01.10.2025).
12. Яковлев А.В. Оценка качества кластеризации данных с использованием алгоритмов «к-средних», «g-средних» и «EM» / А.В. Яковлев, А.А. Мазниченко, М.О. Кожанчиков // Фундаментальные и прикладные научные исследования: актуальные вопросы, достижения и инновации: сборник статей LVIII Международной научно-практической конференции, Пенза, 27 июля 2022 года. — Пенза: Наука и Просвещение (ИП Гуляев Г.Ю.), 2022. — С. 111–115. — EDN JYIDOA.

© Крынецкая Анастасия Дмитриевна (krynetskaya@mirea.ru); Трушина Вероника Игоревна (trushina_v@mirea.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»