

Интернет-журнал «Отходы и ресурсы» <https://resources.today>
Russian Journal of Resources, Conservation and Recycling

2025, Том 12, № 2 / 2025, Vol. 12, Iss. 2 <https://resources.today/issue-2-2025.html>

URL статьи: <https://resources.today/PDF/13INOR225.pdf>

DOI: 10.15862/13INOR225 (<https://doi.org/10.15862/13INOR225>)

1.6.20. Геоинформатика, картография (технические науки)

Ссылка для цитирования этой статьи:

Капустина, О. А. Первичный анализ пространственно-временных данных для оценки геодинамического состояния месторождений полезных ископаемых / О. А. Капустина, М. Ю. Нестеренко, О. В. Антонова // Отходы и ресурсы. — 2025. — Т. 12. — № 2. — URL: <https://resources.today/PDF/13INOR225.pdf>. DOI: 10.15862/13INOR225.

For citation:

Kapustina O.A., Nesterenko M.Yu., Antonova O.V. Primary analysis of spatiotemporal data to assess the geodynamic state of mineral deposits. *Russian Journal of Resources, Conservation and Recycling*. 2025;12(2): 13INOR225. Available at: <https://resources.today/PDF/13INOR225.pdf>. DOI: 10.15862/13INOR225. (In Russ., abstract in Eng.).

УДК 528.9:004

Капустина Оксана Александровна

ФГБУН «Оренбургский федеральный исследовательский центр
Уральского отделения Российской академии наук», Оренбург, Россия
Старший научный сотрудник
Кандидат технических наук, доцент
E-mail: onica1@yandex.ru

ORCID: <https://orcid.org/0000-0001-8981-6599>

РИНЦ: https://elibrary.ru/author_profile.asp?id=531209

Нестеренко Максим Юрьевич

ФГБУН «Оренбургский федеральный исследовательский центр
Уральского отделения Российской академии наук», Оренбург, Россия
Заведующий отделом геоэкологии
Доктор геолого-минералогических наук, доцент
E-mail: n_mu@mail.ru

ORCID: <https://orcid.org/0000-0003-1465-0752>

РИНЦ: https://elibrary.ru/author_profile.asp?id=632852

Антонова Ольга Викторовна

ФГБУН «Оренбургский федеральный исследовательский центр
Уральского отделения Российской академии наук», Оренбург, Россия
Аспирант
E-mail: olga-oren@rambler.ru

РИНЦ: https://elibrary.ru/author_profile.asp?id=1132788

**Первичный анализ пространственно-временных
данных для оценки геодинамического состояния
месторождений полезных ископаемых**

Аннотация. В статье исследуются исходные геоданные необходимые для оценки геодинамического состояния месторождений углеводородов. В качестве основных, авторы рассматривают пространственно-временные данные о сейсмической активности на месторождениях углеводородов Южного Предуралья, полученные из Центра обработки данных сети сейсмологических станций «Нефтегаз-сейсмика» (отдел геоэкологии,

федеральный исследовательский центр Уральского отделения Российской академии наук, Оренбург, Россия) и Казахстанского национального центра данных (г. Алматы, Казахстан). Также исследуется закрытая исходная информация о параметрах разрабатываемых и эксплуатируемых месторождений полезных ископаемых Оренбургской области, предоставленные недропользователями. Географические, геологические, гидрологические и другие данные по исследуемым объектам предлагается получить из открытого источника OpenStreetMap путем формирования запросов на языке Overpass. Авторами выделены основные характеристики геоданных, полученных из различных источников и определены основные этапы их предобработки, включающие: загрузку пространственной информации из различных источников, проверку их целостности и очистку, преобразование категориальных признаков, визуализацию и оценку качества. Средствами языка программирования Python проведен первичный анализ исходной геоинформации согласно предложенным этапам по каждому набору данных. Визуализация данных позволила наглядно оценить качество представленных данных и обосновать необходимость дальнейшего их анализа. В результате работы сформирован новый набор данных для более глубокого анализа или использования в геоинформационных системах при решении задачи оценки геодинамического состояния месторождений полезных ископаемых.

Ключевые слова: геоданные; предобработка; анализ пространственных данных; аномалии пространственных данных; дублирование геоданных; геодинамическое состояние; месторождения

Введение

Оценка геодинамического состояния месторождений полезных ископаемых на начальном этапе предполагает проведение работ по сбору, обработке и анализу исходных пространственно-временных геоданных из множества различных источников в разных форматах, структурах и системах координат. В качестве исходной пространственной информации могут выступать данные о расположении месторождений, их геологической и тектонической структуре, представленные в виде карт на бумажных носителях, растровых или векторных изображений, готовых цифровых моделей, полученных из открытых или закрытых источников. А также наборы датасетов, в состав которых входят пространственные, временные компоненты, описываемые числовым и/или форматом дат, статистические данные о параметрах разработки и эксплуатации месторождений, результаты мониторинга геодинамических и сейсмических процессов, и компоненты описательного содержания.

Эффективность моделей оценки геодинамического состояния зависит от качества исходных данных, точности выделения значимых признаков и адекватности выбранных методов анализа. Поэтому обработка исходной многомерной пространственно-временной информации предполагает применение комплексного геоинформационного подхода, важнейшей частью которого является этап предобработки геоданных и их первичного анализа, включающий очистку, нормализацию, заполнение пропусков и выделение наиболее информативных характеристик. В связи с этим возникает необходимость проведения первичного анализа полученных геоданных и оценки качества исходной геопрограммной информации, что позволит подготовить исходные данные для выявления их корректности, совместимости и сопоставимости, определения формата данных из разных источников, тем самым расширив возможности анализа геодинамического состояния месторождений полезных ископаемых. Цель настоящей статьи — рассмотреть современные подходы к предобработке пространственно-временных данных и провести первичный анализ исходных геоданных о месторождениях Оренбургской области, на которых ведется мониторинг сейсмической активности.

Методы

Современные подходы сбора и первичного анализа цифровых геоданных в задачах геофизики ориентированы на применение алгоритмов анализа данных, построенных на основе машинного обучения и искусственных нейронных сетей [1; 2].

В зависимости от формы представления исходной пространственно-временной информации о сейсмических событиях (запись сейсмограммы или каталоги и бюллетени) можно выделить методы предобработки и первичного анализа для автоматической обработки сейсмических записей [3] и методы анализа каталогов сейсмических событий, которые в свою очередь включают методы оценки представительности каталогов [4]. В данном случае рассматриваем предобработку и первичный анализ геоданных, включающих пространственно-временную информацию о сейсмических событиях, представленную в виде представительных каталогов. Предобработку и первичный анализ полученных геоданных проводится в несколько основных этапов.

Импорт данных. Загрузка и импорт геоданных из различных источников, таких как каталоги и сейсмологические бюллетени, GPS устройства, базы данных географических информационных систем, онлайн сервисы карт и т. д.

Таблица 1

Фрагмент каталога сейсмических событий

№ п/п	Дата	Время в очаге	Широта	Долгота	Глубина (км)	Магнитуда Мl	Магнитуда Ms	Энергетический класс
1	02.01.2008	21:36:18	51,72	54,47	5,00	1,40	1,90	6,20
2	11.01.2008	1:27:42:	51,46	54,77	5,00	1,80	2,10	5,90
3	03.02.2008	3:47:24	51,76	54,90	5,00	2,20	2,80	5,70
4	13.02.2008	3:23:58	51,47	54,25	6,00	1,40	2,10	6,30
5	26.03.2008	7:54:15	51,96	53,95	0,00	0,10	1,50	5,20
6	27.03.2008	7:24:23	51,74	54,39	0,00	2,10	2,60	5,10
7	02.04.2008	10:35:27	51,68	54,52	8,00	1,90	2,05	5,40
8	23.04.2008	17:16:42	51,69	54,46	10,00	1,70	2,00	6,00
9	01.05.2008	7:59:30	51,80	54,62	0,00	1,20	1,80	5,80
10	04.05.2008	17:39:11	51,72	54,56	5,00	0,70	1,30	5,50
11	17.06.2008	9:56:18	51,68	55,14	5,00	1,10	1,50	6,10
12	05.07.2008	10:22:06	51,88	54,80	5,00	2,00	2,30	7,20
13	28.07.2008	3:13:56	51,76	54,65	6,00	1,10	1,60	6,80
14	19.08.2008	3:52:11	51,39	55,39	5,00	1,20	1,60	6,20
15	20.08.2008	2:34:24	51,49	55,43	5,00	1,10	1,50	5,90
16	11.09.2008	16:53:18	51,71	54,89	0,00	0,50	1,10	4,40
17	13.09.2008	18:23:40	51,56	54,93	2,00	0,50	1,10	4,10
18	01.10.2008	15:32:57	51,93	53,65	5,00	1,40	2,00	7,20

Составлено авторами

Основной исходный материал для проведения анализа и получения объективного информационного отражения состояния геодинамических и сейсмических процессов на территории Оренбургской области получен по результатам мониторинга территорий сейсмической сети «Нефтегаз-сейсмика» для месторождений углеводородов Южного Предуралья [5]. Также в отделе геоэкологии ОФИЦ УрО РАН составляются каталоги сейсмических событий в формате электронных таблиц для месторождений твердых полезных ископаемых восточной части Оренбургской области по данным сети сейсмостанций Казахстана Казахстанского национального центра данных (г. Алматы, Казахстан). Каталоги сейсмических событий по каждому изучаемому месторождению полезных ископаемых содержат на сегодняшний день сотни и более записей сейсмических событий, описываемых координатами очагов, глубиной гипоцентров, магнитудами и временем возникновения сейсмических

событий, пример каталога представлен в таблице 1. Данные по параметрам разработки некоторых месторождений в формате электронных таблиц подготовлены недропользователями и содержат по каждой залежи следующую информацию: средний дебит газ, тыс. м³/сут.; средний дебит нестабильного конденсата, т; средний дебит вода, м³/сут.; годовая добыча газ сепарации, млрд м³; годовая добыча нестабильного конденсата, млн т.; годовая добыча вода, млн м³; накопленная добыча газ сепарации, млрд м³; накопленная добыча нестабильного конденсата, млн т.; накопленная добыча вода, млн м³; пластовое давление по залежи, Мпа; пластовое давление в зоне дренирования, Мпа; общее количество действующих скважин в течение года, ед.; количество действующих скважин в течение года с водой, ед.; коэффициент эксплуатации действующего фонда добывающих скважин, доли ед.; темп отбора от остаточных геологических запасов газа, %; степень выработки геологических запасов газа, %.

Географические, геологические, гидрологические и другие данные можно получить из открытых источников, например OpenStreetMap путем формирования запросов на языке Overpass на интересующие нас объекты [6; 7]. Например, запрос на выборку карьеров, расположенных в Оренбургской области, составленный авторами можно представить следующим образом рисунок 1.

```
[out:json][timeout:90];
{{geocodeArea:Orenburg oblast}}->.searchArea;
(
  nwr["landuse"="quarry"](.searchArea);
);
out geom;
```

Рисунок 1. Пример запроса на карьеры Оренбургской области (составлено авторами)

Результаты запроса геоданные экспортируются в формат для хранения и обмена географическими данными, основанный на стандарте JavaScript Object Notation.

Проверка целостности данных и их очистка: необходимо убедиться, что набор данных целостен и готов к дальнейшему анализу. Проверка данных на наличие ошибок, некорректных или невозможных значений (например, отрицательные величины глубин или неправильные форматы дат), выявление пропусков или несоответствий. Данный этап также включает в себя удаление дубликатов, коррекцию аномальных значений координат и других параметров сейсмических событий.

Эффективные методы очистки данных играют ключевую роль в повышении качества и надежности результатов анализа пространственно-временных данных [8], в том числе о сейсмическом событии. Рассмотрим основные техники, применяемые на этапе предварительной обработки:

1. Обнаружение и удаление выбросов.

Выбросы — аномальные значения, значительно отличающиеся от основной массы наблюдений. Они могут возникать вследствие ошибок измерений или случайных факторов. Методы обнаружения включают статистические критерии (например, метод IQR, трех сигм), визуальное исследование распределений (графики типа Box Plot). После выявления выбросы либо удаляются, либо заменяются средними значениями, медианой или специальными методами восстановления значений (интерполяция) [9].

2. Заполнение пропущенных значений.

Отсутствие части данных («пропуски») часто встречается в реальных наборах данных. Для заполнения пустых полей используют следующие методы:

- Замещение средней величиной: простая замена недостающих значений средним значением соответствующего признака.
- Интерполяция: восстановление отсутствующих значений путем аппроксимации соседних точек, используя линейную, полиномиальную или сплайн-интерполяцию.
- Метод k ближайших соседей: расчет неизвестных значений на основании значений ближайшего окружения.

3. Нормализация и стандартизация

Нормализация необходима для приведения всех переменных к единой шкале измерения, исключая влияние масштаба отдельных признаков на итоговые модели. Существуют различные виды нормализации:

- Min-max нормализатор.
- Z-нормализатор (стандартизация): приведение данных к нулевому среднему и единичной дисперсии, формула.

4. Удаление дубликатов

Дублирование записей также негативно влияет на качество анализа. Простым способом удаления повторяющихся строк является фильтрация уникальных комбинаций признаков.

Преобразование и структурирование: преобразование категориальных признаков. Пространственно-временные данные иногда содержат качественные признаки (например, наименование месторождения, категория сейсмического события импульс или взрыв). Такие признаки требуют преобразования в числовой вид перед применением большинства методов машинного обучения. Используются такие способы, как one-hot-кодирование или label encoding. Приведение геоданных к удобному для работы формату геоинформационных систем, если они не были в таком формате изначально. Для обработки и анализа современными средствами языка программирования Python используем библиотеки GeoPandas & GeoPlot library. Также данный этап включает преобразование координатных систем, если данные представлены в разных системах координат.

Извлечение основных характеристик: анализ общего количества сейсмических событий, базовых статистических характеристик пространственно-временных геоданных, таких как средние, минимальные и максимальные значения параметров сейсмических событий (координаты, время и дата возникновения, магнитуда, глубина и т. д.).

Визуализация: построение графиков и карт для визуализации геоданных. Это помогает лучше понять распределение точек, выявить возможные паттерны или аномалии.

Оценка качества пространственных данных: оценка точности и достоверности данных, идентификация потенциальных проблем или искажений. Оценка качества геоданных играет ключевую роль в обеспечении точности и достоверности получаемых результатов первичного анализа исходных геопро пространственных данных [10]. Согласно ГОСТ Р 57773-2017 (ИСО 19157:2013) «Пространственные данные. Качество данных», можно выделить основные аспекты, которые следует учитывать при оценке качества геоданных:

Точность данных: это оценка того, насколько близко измеренные координаты отражают реальные местоположения объектов или событий. Точность может зависеть от используемых GPS устройств, методов съемки данных, условий окружающей среды и других факторов.

Полнота данных: это оценка того, насколько полными являются исходные геоданные относительно охвата рассматриваемых территорий (месторождений полезных ископаемых)

или временного промежутка. Недостаток данных может привести к искаженным результатам анализа геодинамического состояния разрабатываемых месторождений полезных ископаемых.

Согласованность данных: это проверка согласованности пространственных данных внутри одного набора или между различными источниками. Например, проверка соответствия координат объектов на разных картах или в разных базах геоданных.

Актуальность данных: для некоторых аналитических задач требуется использовать только актуальные данные.

Однозначность и уникальность идентификации: проверка наличия уникальных и однозначных идентификаторов для объектов или событий в геоданных. Это важно для исключения дублирования или ошибок при их слиянии из различных источников.

Защищенность данных: оценка мер защиты данных от несанкционированного доступа, модификации или утраты.

Для проведения оценки качества геоданных можно использовать различные методы и инструменты, включая визуальные анализы, сравнение с эталонными данными (если они доступны), статистические тесты и экспертные оценки. Важно также учитывать контекст использования данных и специфику задачи, для которой они предназначены.

Эти шаги помогают провести предварительную обработку пространственно-временных данных, подготовить их для более глубокого анализа или использования в геоинформационных системах для решения задачи оценки геодинамического состояния месторождений полезных ископаемых.

Результаты

На первом этапе происходит загрузка исходного набора данных из файла формата Excel. Таблица содержит параметры землетрясений, включая координаты, магнитуды, дату и время события, глубину и энергетический класс.

Используя современные средства языка программирования Python исходные данные о сейсмичности рассматриваемого региона объединены и загружены. Пространственно-временные данные не имеют однородной структуры, поэтому требуется ряд преобразований:

- проведена обработка временных значений, путём перевода времени очага в секунды;
- корректировка типов ряда параметров сейсмических событий путем очистки от символов и преобразование типа данных к числовому;
- замена пропущенных значений медианным значением исследуемого параметра.

Проведен первичный анализ исходных геопространственных данных о сейсмических событиях, в результате которого выявлены:

- аномалии в значениях координат сейсмических событий, представленные на рисунке 3;
- отсутствие пространственных координат у 0,03 % сейсмических событий от общей выборки;
- пропуски по значениям магнитуды у 0,2 % сейсмических событий от общей выборки;
- дублирующие записи сейсмических событий отсутствуют.

Результаты очистки данных. Коррекцию аномальных значений координат представлены на разработанном авторами рисунке 2.

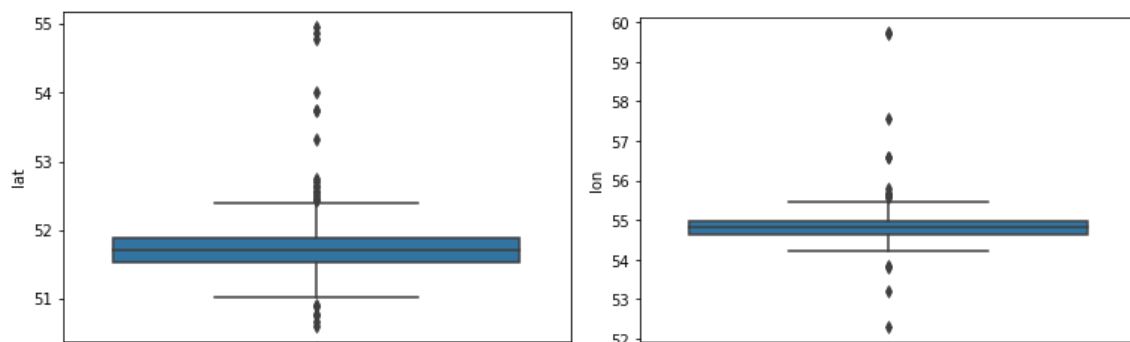
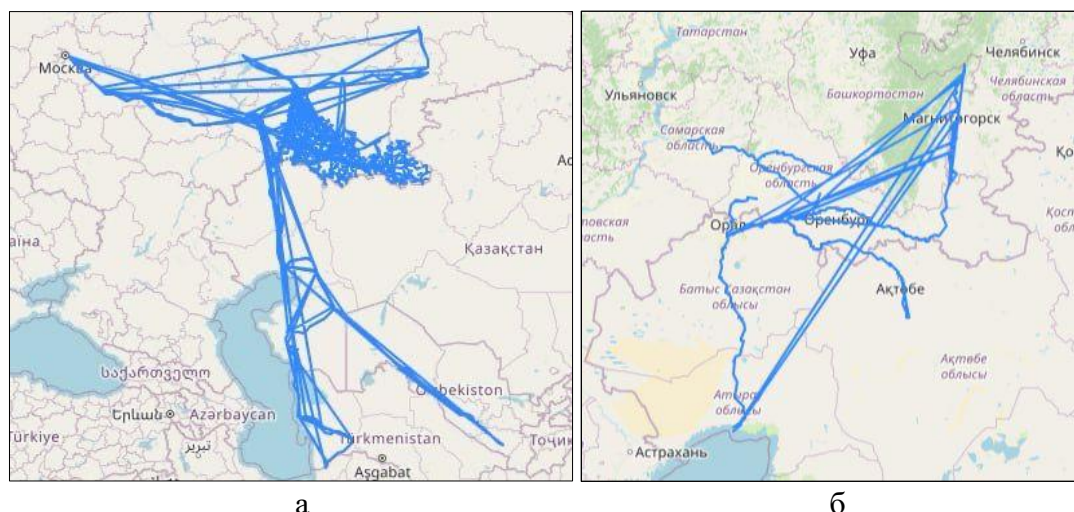


Рисунок 2. Проверка данных на аномальность значений координат (составлено авторами)

Методом визуализации средствами библиотек языка программирования Python получены результаты отображения на карте Оренбургской области слоев, полученных в результате запросов к открытым источникам данных. На разработанном авторами рисунке 3 представлены наихудшие результаты выборки исходных данных по дорогам и рекам.



Рисунки 3. Результаты визуализации исходных данных по дорогам и рекам: а — данные по дорогам; б — данные по рекам (составлено авторами)

В результате отображения геопространственных данных делается вывод о некорректности исходных данных, наличие повторяемости данных.

Представленные слои подчёркивают необходимость подготовки исходных геопространственных данных для дальнейшего анализа. Визуализация данных позволяет наглядно оценить качество представленных данных и обосновать необходимость дальнейшего анализа данных. В связи с этим сформирован новый набор данных, состоящий только из необходимых параметров. Строки с отсутствующими значениями удаляются, обеспечивая чистоту входных данных для последующего анализа.

Обсуждение

Первичный анализ геоданных позволил понять их структуру, устранить очевидные проблемы и сформировать четкое представление о дальнейших действиях. Первичная обработка важна для выбора правильных подходов к моделированию и обеспечивает успех

последующей геоаналитики и разработки моделей оценки геодинамического состояния месторождений полезных ископаемых.

От грамотной подготовки данных зависят точность оценки геодинамического состояния и надежность построенных моделей, что особенно актуально в области мониторинга природно-техногенной сейсмичности месторождений полезных ископаемых, где минимизация ошибок критична для принятия своевременных мер безопасности.

ЛИТЕРАТУРА

- 1 Соловьев, А.А. Обзор и перспективы использования современных подходов комплексного анализа геоданных для прогноза пространственного распределения геолого-геофизических параметров / А.А. Соловьев, И.А. Лисенков — DOI 10.21455/gr2024.2-2. // Геофизические исследования. — 2024. — Т. 25, № 2. — С. 20–45.
- 2 Использование хранилища параметров машинного обучения при геоинформационном картографировании техногенно нарушенных территорий / А.А. Колесников, П.М. Кикин, Т.А. Платонов, Е.В. Печенкина — DOI 10.15372/FPVGN2024110412 // Фундаментальные и прикладные вопросы горных наук. — 2024. — Т. 11, № 4. — С. 88–94.
- 3 Кувайскова, Ю.Е. Предобработка данных при построении модели нейронной сети для прогнозирования состояния технического объекта / Ю.Е. Кувайскова, А.А. Немыкин — DOI 10.14357/20718594250106. // Искусственный интеллект и принятие решений. — 2025. — № 1. — С. 67–81.
- 4 Гитис, В.Г. Анализ влияния очистки каталогов от афтершоков на эффективность систематического прогноза землетрясений / В.Г. Гитис, А.Б. Дерендяев, К.Н. Петров // Информационные процессы. — 2019. — Т. 19, № 4. — С. 399–407.
- 5 Оренбургская региональная сеть “Нефтегаз-сейсмика” / М.Ю. Нестеренко, И.М. Алешин, А.Г. Гоев [и др.] — DOI 10.21455/si2023.2-1. // Сейсмические приборы. — 2023. — Т. 59, № 2. — С. 5–17.
- 6 Jacobs, K.T. OpenStreetMap quality assessment using unsupervised machine learning methods / K.T. Jacobs, S.W. Mitchell // Transactions in GIS. — 2020. — Т. 24, № 5. — С. 1280–1298.
- 7 Колесников, А.А. Возможности NOSQL СУБД для обработки пространственных данных / А.А. Колесников — DOI 10.33764/2411-1759-2022-27-3-95-106. // Вестник СГУГиТ (Сибирского государственного университета геосистем и технологий). — 2022. — Т. 27, № 3. — С. 95–106.
- 8 Кислов, К.В. Возможности предподготовки сейсмических данных для анализа глубокой нейронной сетью / К.В. Кислов, В.В. Гравиров, Ф.Э. Винберг // Физика Земли. — 2020. — № 1. — С. 150–162. — DOI 10.31857/S0002333720010056.
- 9 Концепция построения цифрового двойника города / С.А. Иванов, К.Ю. Никольская, Г. И. Радченко [и др.] — DOI 10.14529/cmse200401 // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. — 2020. — Т. 9, № 4. — С. 5–23.
- 10 Дышленко, С.Г. Анализ и разработка характеристик качества геоданных / С.Г. Дышленко // Перспективы науки и образования. — 2016. — № 2(20). — С. 23–27.

Kapustina Oksana Aleksandrovna

Orenburg Federal Research Center of the Ural Branch of the Russian Academy of Sciences, Orenburg, Russia
E-mail: onical@yandex.ru
ORCID: <https://orcid.org/0000-0001-8981-6599>
RSCI: https://elibrary.ru/author_profile.asp?id=531209

Nesterenko Maksim Yuryevich

Orenburg Federal Research Center of the Ural Branch of the Russian Academy of Sciences, Orenburg, Russia
E-mail: n_mu@mail.ru
ORCID: <https://orcid.org/0000-0003-1465-0752>
RSCI: https://elibrary.ru/author_profile.asp?id=632852

Antonova Olga Viktorovna

Orenburg Federal Research Center of the Ural Branch of the Russian Academy of Sciences, Orenburg, Russia
E-mail: olga-oren@rambler.ru
RSCI: https://elibrary.ru/author_profile.asp?id=1132788

Primary analysis of spatiotemporal data to assess the geodynamic state of mineral deposits

Abstract. The article examines the initial geodata necessary to assess the geodynamic state of hydrocarbon deposits. As the main ones, the authors consider spatial and temporal data on seismic activity in hydrocarbon deposits of the Southern Urals obtained from the Data Processing Center of the network of seismological stations «Neftegaz-Seismika» (Department of Geoecology, Federal Research Center of the Ural Branch of the Russian Academy of Sciences, Orenburg, Russia) and the Kazakh National Data Center (Almaty, Kazakhstan). The closed source information on the parameters of the developed and exploited mineral deposits of the Orenburg region, provided by subsurface users, is also being investigated. The article proposes to obtain geographical, geological, hydrological and other data on the studied objects from the open source OpenStreetMap by forming queries in the Overpass language. The authors have highlighted the main characteristics of geodata obtained from various sources and identified the main stages of their preprocessing, including: uploading geodata from various sources, checking their integrity and cleaning, converting categorical features, visualization and quality assessment. Using the Python programming language, a primary analysis of the initial geoinformation was carried out according to the proposed stages for each data set. Data visualization made it possible to visually assess the quality of the data presented and justify the need for further data analysis. As a result of the work, a new set of data has been formed for deeper analysis or use in geoinformation systems when solving the problem of assessing the geodynamic state of mineral deposits.

Keywords: geodata; preprocessing; analysis of spatial data; anomalies of spatial data; duplication of geodata; geodynamic condition; deposits