Технология обработки больших данных сейсмических станций и ее применение*

А.В. Смагличенко[I], Т.А. Смагличенко[II], М.К. Саянкина[II]

[I] Крымский федеральный университет им. В.И. Вернадского, г. Симферополь, Россия
[II] Институт проблем нефти и газа Российской академии наук, г. Москва, Россия

Аннотация. Данные локальной сейсмичности, содержащие наборы информации о станциях, зарегистрировавших сигналы, используются для диагностирования неоднородностей при разведке энергетических ресурсов. Для оценки детальности результата, необходимо визуализировать рассредоточение станций в провинциях практического интереса. Для этого требуется обработка больших данных, предоставленных в открытом доступе на сайтах Единой Геофизической Службы Российской Академии Наук. Данные включают множество значений географических координат станций, функционирующих по всему миру. В статье разработана технология выборки информации, относящейся к любым регионам России. Технология протестирована на произвольном наборе данных Восточной Сибири и применена для области действующих месторождений Западной Сибири. Впервые построенная карта расстановки станций показывает крайне низкое их число и большую удаленность от области добыч. Тем не менее, просвечивание участков, соседних с запасами, может дать ценнейшие сведения о присутствии флюидов, скоплении газов.

Ключевые слова: *экологически чистые технологии; месторождения Сибири; распределение сейсмических станций; большие данные.*

DOI: 10.14357/20790279260306 **EDN:** QRHCAQ

Введение

Сейсмическая информация является необходимым инструментом комплексного просвечивания сложной глубинной геологической среды в районах добычи источников тепловой энергии. В России функционирует Единая Геофизическая Служба Российской Академии Наук (ЕГС РАН) по сбору данных о произошедших землетрясениях [1]. База данных представляет собой наборы бюллетеней для определенного временного интервала (типичная длина интервала - 10 дней). Текстовой файл каждого бюллетеня начинается с набора данных о станциях, зарегистрировавших события в течение временного периода, и включает аббревиатуру и наименование каждой станции, ее географические координаты и высоту над уровнем моря. Вслед за этой информацией идет перечисление сведений о событиях, которые были зафиксированы станциями из искомого набора. А именно, для каждого произошедшего землетрясения сформирована строка, содержащая числовые значения времени возникновения очага, географические координаты эпицентра, глубины, магнитуды. Далее в другом формате одна за другой даны строки, каждая из которых включает текстовые значения, идентифицирующие станцию, численные значения времен вступлений объемных продольной P- и поперечной S- волн. В случае если станция зарегистрировала поверхностную волну, присутствует строка из символов, относящихся к деталям записи этой волны. Форматы строк для объемных волн и поверхностной волны сильно разнятся. Поперечная волна не всегда определена, ее вступление трудно выделить. Поэтому вместо численного значения вступления может быть пробел. Выделить поверхностную волну чрезвычайно сложно, поэтому для станции, где она не была установлена, соответствующая строка часто вовсе отсутствует.

* Исследование выполнено в рамках государственного задания (тема № FZEG-2024-0016 и в рамках темы «Фундаментальный базис инновационных, цифровых технологий прогноза, поиска, разведки и освоения нефтегазовых ресурсов (фундаментальные, поисковые, прикладные, экономические и междисциплинарные исследования до 2030 года)» № 125021302095-2).

Таким образом, набор данных о каждом событии имеет сложно – структурированный формат, отличающийся от общепринятого мирового стандарта. Создание метода, позволяющего преобразовать описанные выше наборы данных о землетрясениях, в строго структурированный формат – задача, требующая разработки пакета прикладных программ.

Основной целью настоящей работы является создание структурированного файла выборки данных о станциях, расположенных в локальном районе интереса. Важность решения этой проблемы, прежде всего, обусловлена отсутствием реляционной базы данных для малозаселенных или незаселенных территорий Западной и Восточной Сибири, где необходимы комплексные исследования возможностей добычи. Известно, что традиционные места с высокой плотностью заселения, где в течение многих лет ведутся работы по извлечению газа, отличаются экологическими рисками для людей, проживающих там и для существующих инфраструктур. Чтобы понять текущие возможности сейсмического просвечивания предполагаемых залежей провинций необходимо визуализировать сейсмическую сеть постоянно действующих станций России в локальных районах, а также расположение станций в соседних странах, которые можно было бы использовать как дополнительный источник информации при условии их близости к изучаемому району. Представленный в статье алгоритм и его результаты на примере провинций Сибири являются вкладом в развитие Географической Информационной Системы (ГИС).

1. Специфика распределения географических данных. Постановка задачи создания перечня действующих станций

Данные каждого бюллетеня имеют ограничения, связанные с имеющейся информацией о сейсмической активности для рассматриваемого промежутка времени. В бюллетень входят события, которые произошли в различных зонах большой территории, и, соответственно, географический разброс расположения станций довольно широк. Географические координаты составляют конечное множество. Таким образом, информация, содержащаяся в бюллетенях для значимого периода времени, даст объединение таких множеств. Построенное множество будет представлять совокупность пересекающихся подмножеств, относящихся к каждому бюллетеню, так как координаты станций, принадлежащие множеству одного бюллетеня, могут одновременно принадлежать и другому. Поэтому первой задачей является формирование основного множества списка станций, при условии, что в нем не будет общих элементов, то есть, повторяющихся строк информации, приуроченной к регистрирующим станциям.

2. Выборка данных для районов интереса

Учитывая отсутствие плотной сети сейсмических наблюдений практически во всей России, территории планируемого изучения могут занимать площади с условным размером 100 х 100 км. Визуализация расположения станций на таких площадях позволит дать предварительную оценку разрешения планируемого изучения глубинных структур. При необходимости районы интереса могут быть выделены путем сдвига исходных площадей после анализа визуализации расположения станций. Наборы данных для рассматриваемых районов должны иметь структурированный формат для того, чтобы файлы были пригодны для считывания оригинальными исследовательскими программами. Например, текстовые числовые значения параметров станций должны размещаться строго в пределах одного столбца, не выходя за границы определенных ячеек для считывания.

3. Описание базового алгоритма и элементов разработанной программы

Алгоритм создания набора данных о географических координатах сейсмических станций для выбранной территории включает следующие этапы.

Шаг 1. Утверждается период времени, который характеризует динамику состояния геологической среды. Например, до или после значимых эпизодов: тектонических подвижек, формирования разломов, поднятия фундамента, образования кратеров и т. д. Для тестирования алгоритма может быть установлено произвольное временное окно. После выделения бюллетеней, входящих в искомый промежуток времени осуществляется процедура разделения наборов данных, относящихся к каждому из них. Начальный формат определяет матрицу параметров всех станций. Как было сказано выше, каждая строка матрицы –это перечисление текстовых значений аббревиатуры и месторасположения отдельно взятой станции, численные значения ее координат и высоты над уровнем моря. Программный модуль, реализующий парсинг данных, используется для структурирования матрицы параметров выбранных станций.

Отметим, что следующий формат относится к набору данных о параметрах событий, зарегистрированных станциями из списка. Он имеет слож-

ную структуру. Для каждого события определены строки с перечислением числовых значений, даты, географических координат эпицентра, глубины, времени возникновения и т.д. Затем следуют строки, содержащие аббревиатуру станции, азимут, тип волны, невязки времен вступлений. Для каждой станции может быть выделена одна или несколько строк. Кроме этого, часто присутствуют строки, относящиеся к информации о поверхностных волнах, нарушающие основную структуру данных.

В рамках данной статьи мы изучаем данные, содержащие информацию о станциях, действующих на территории России и в соседних странах, прилегающих к границе для того, чтобы получить полное понимание о ГИС, связанной с распределением станций, которое, как известно, неравномерно. Приведение начального формата к мировому стандарту играет ключевую роль при использовании программ для определения внутренней структуры искомого объекта геологической среды. Значения каждой строки набора станционных данных должны быть расположены строго одно под другим без нарушения порядка считывания. Препятствием для упорядоченности является разное количество пробелов, разделяющих столбцы с разнородными данными: текстовыми и числовыми. Теоретически современные текстовые редакторы позволяют исправить этот недостаток с помощью инструментов поиска и замены. Однако на практике подобные решения приводят к появлению множества «багов» в файле с последующей рутиной для их исправления [2]. Для преодоления этой проблемы был разработан отдельный программный модуль, заменяющий разнородный разделитель на «;» с последующим сохранением файла в csv-формате.

Надо отметить, что помимо перечисленных основных сложностей при объединении совокупности бюллетеней дублируется большой процент наименований станций. Программа отсеивает лишнюю информацию, оставляя пригодный для считывания общий список. Рис. 1 иллюстрирует пример одного из сейсмических бюллетеней ЕГС РАН, содержащий данные с 1 по 10 января 2024 г [3].

Шаг 2. Выборка данных из массива информации о станциях, зарегистрировавших события в течение сегментов определенного периода времени. Из длинного списка, структурированных на

```
      Р О С С И Й С К А Я   А К А Д Е М И Я   Н А У К
                ФИЦ ЕГС (Обнинск)
       Сейсмологический бюллетень (ежедекадный)
                1-10 января    2024

При подготовке настоящего бюллетеня использованы  данные
следующих сейсмических станций:
 AAA   Alma-Ata              43.2086N   76.9150E   920 m
 AAK   Ala-Archa             42.6390N   74.4941E  1645 m
 AB31  Akbulak               49.2556N   59.9431E   243 m
 ABPO  Ambohimpanom          19.0174S   47.2275E  1528 m
 ADK   Adak                  51.8836N  176.6844W   116 m
 ADZR  Андозеро              63.9295N   38.2853E    51 m
 AFI   Afiamalu              13.9103S  171.7772W   706 m
 AIS   Amsterdam Island      37.7964S   77.5692E    36 m
 AKAR  Акташ                 50.3200N   87.6200E  1380 m
 AKASG Malin array           50.7012N   29.2242E   160 m
 AKBB  Malin Array           50.7012N   29.2242E   160 m
 AKH   Akhalkalaki           41.4001N   43.4834E     0 m
 AKT   Ахты                  41.4780N   47.7170E  1200 m
 AKTO  Aktyubinsk            50.4348N   58.0164E   379 m
 ALE   Alert                 82.5031N   62.3500W    65 m
 ALIG  Mtskhetisjvari        42.1047N   43.6490E   815 m
 ALKP  Alkanpan              29.9418S   22.2467E  1065 m
 ALN   Alexandroupolis       40.8973N   26.0456E   110 m
 ALU   Alushta               44.7000N   34.4167E    61 m
 ANDR  Анадырь               64.7340N  177.4960E    55 m
```

Рис. 1. Пример начальной части бюллетеня, содержащего данные с 1 по10 января 2024 г. Всего 700 станций [3]

Шаге 1, данных выбираются те, которые относятся к району интереса. Такими районами, как уже сказано выше, могут быть места разведки ископаемых и их добычи, а также экологических катастроф кратеров выброса газа.

Программа обработки станционных координат, сформированных на Шаге 1, была протестирована для произвольного района Восточной Сибири. Рис. 2 показывает результат обработки данных (произвольная выборка). Информация о сейсмостанциях, содержащаяся в бюллетенях ЕГС РАН была импортирована в реляционную БД с последующим решением проблемы дубликатов. Это позволило сделать выборку ряда станций по координатам для района Восточной Сибири.

Шаг 3. Применение программ GMT [4] и данных глобальной цифровой модели рельефа GLOBE [5] для визуализации распределения станций.

4. Применение разработанной технологии к области действующих месторождений Сибири

Реализуя Шаг 1, разработанное на языке программирования Python программное обеспечение было задействовано для обработки данных бюллетеней ЕГС РАН за период с 2017 по 2024 годы. Из 278 файлов бюллетеней было выделено 153832 станции, включая дубликаты. Временной сегмент был выбран не случайно. 15 апреля 2017 года в России была открыта сейсмическая станция «Бованенково» (BVNN), расположенная на полуострове Ямал, в Ямало-Ненецком автономном округе. Так как областью нашего интереса является территория действующих месторождений Западной Сибири и прилегающих к ней участков геологической среды, то близость расположения данной станции к извлечению запасов чрезвычайно важна. «Бованенково» становится элементом сети сейсмических станций, ответственных за регистрацию локальной сейсмичности, за контроль сигналов от тектонических подвижек, за активизацию сейсмических кластеров перед выбросами газов. В результате работы программного модуля, отсекающего дубликаты строк матрицы с информацией о станциях, было определено 1270 уникальных станций, добавленных в базу данных.

Район интереса включал ряд месторождений, расположенных в Ямало-Ненецком автономном

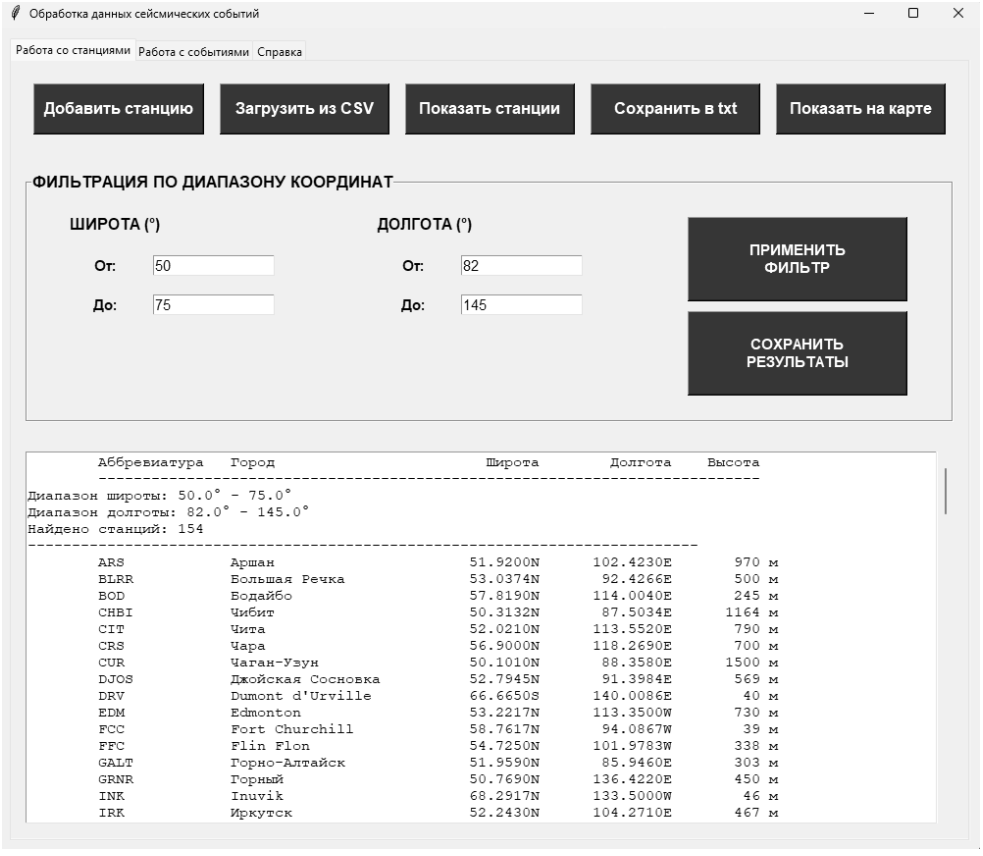


Рис. 2. Программно реализованная функция выборки данных из реляционной БД для заданной области [3]

округе, среди которых: Северо-Комсомольское нефтегазоконденсатное, Харампурское нефтяное, Русское нефтегазоконденсатное, Уренгойское газовое, Заполярное нефтегазоконденсатное, Ямбургское нефтегазоконденсатное, Новопортовское, Северо-Русское газоконденсатное, Южно-Мессояхское газоконденсатное, Пякяхинское нефтегазоконденсатное месторождение. Также в связи с близким расположением, в данный список вошло Ванкорское нефтегазовое месторождение, расположенное в Красноярском крае.

Чтобы в дальнейшем иметь представление о неоднородной структуре в окрестности добывающих скважин, мы расширили радиус поиска станций, с каждым разом увеличивая шаг до 100 км. В результате на расстоянии 400 км от исследуемого района были обнаружены еще две станции: Норильск (NRIK) и Соснино (SSN). Еще две станции: – Красноярск (KRAR) и Анжеро-Судженск (ASR1) были выявлены на расстоянии до 800 км. В диапазон до 1000 км попало 6 станций: Берчикуль (BRCR), Балахонка (BALAH), Кемерово (KEM), Железногорск (GZL и GZLN) и Железногорск-3 (SHTK).

На рис. 3, построенном с помощью программного обеспечения GMT [4] и данных глобальной цифровой модели рельефа GLOBE [5] показана визуализация распределения станций.

Заключение

Исследование, описанное в данной статье, проводится впервые для систематизации больших данных Геофизической Службы РАН с целью получить представление о неоднородностях геологической среды в районах интереса по данным локальной сейсмичности методами, которые не требуют бурения дорогостоящих поисково-разве-

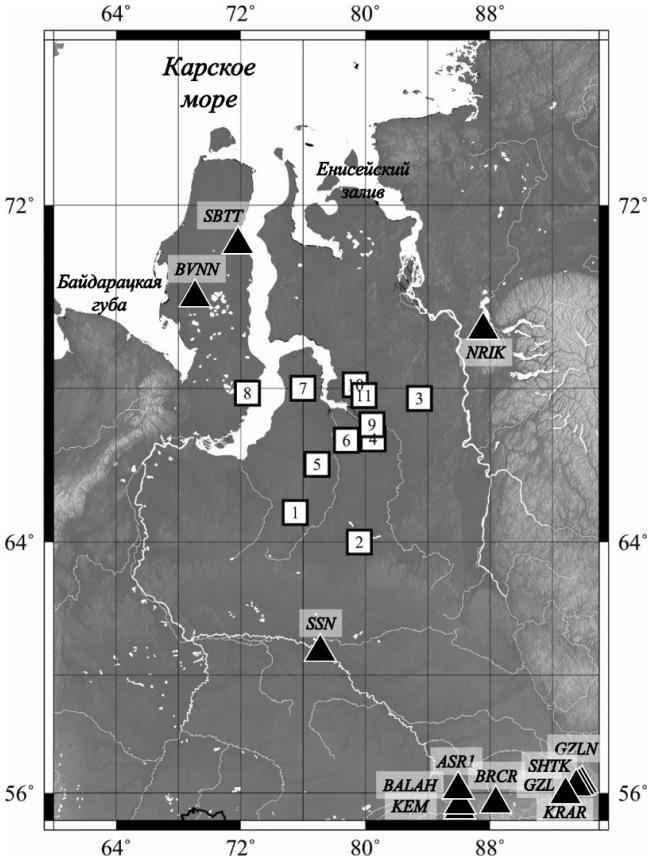


Рис. 3. Визуализация распределения станций и месторождений в Ямало-Ненецком и Красноярском округах. Обозначения месторождений: 1 – Северо-Комсомольское нефтегазоконденсатное, 2 – Харампурское нефтяное, 3 – Ванкорское нефтегазовое, 4 – Русское нефтегазоконденсатное, 5 – Уренгойское газовое, 6 – Заполярное нефтегазоконденсатное, 7 – Ямбургское нефтегазоконденсатное, 8 – Новопортовское, 9 – Северо-Русское газоконденсатное, 10 – Южно-Мессояхское газоконденсатное, 11 – Пякяхинское нефтегазоконденсатное месторождение. Обозначение станций: Бованенково (BVNN), Сабетта (SBTT), Норильск (NRIK), Соснино (SSN, Красноярск (KRAR), Анжеро-Судженск (ASR1), Берчикуль (BRCR), Balakhonka (BALAH), Кемерово (KEM), Железногорск (GZL и GZLN) и Железногорск-3 (SHTK)

дочных скважин, являясь экологически чистыми, обеспечивая комплексную оценку планирования мест добычи. Препятствием для решения этой задачи является крайне неравномерное размещение станций по территории России. В густонаселенных районах оно может быть приемлемо, в то время как в условиях жестких природных условий станции разнесены на большое расстояние или вовсе отсутствуют. Для того чтобы визуализировать присутствие станций в энергетически важных регионах мы разработали технологию обработки больших данных, которая не только определяет границы изучаемых площадей по координатам из созданного набора данных, но и в случае необходимости варьирует контуры этих границ.

Программное обеспечение разработано первым автором статьи, применяя язык программирования Python. Программные модули, ответственные за адекватность выборки данных, за парсинг, протестированы на произвольных наборах информационных данных. Новая технология применена к интересующей области, где сосредоточены месторождения Сибири, запасы которых извлекаются крупными нефтяными компаниями. В результате использования программ, как стандартных, так и авторских, была построена карта, приведенная в настоящей работе. Эта авторская карта, которая до сих пор не демонстрировалась, показывает состояние дел по распределению сейсмических станций в Западной Сибири. Вывод однозначен: рядом с территорией извлекаемых запасов их мало. Однако к югу имеется достаточно плотная сеть, что дает шанс разведки района, прилегающего к территории добычи с Юга. Это важный результат, так как он определяет область планируемой комплексной оценки структуры среды с целью возможной разработки.

Будущее направление развития системы визуализации больших данных, соответствующих информации о локальной сейсмичности России, мы видим в разработке технологии, которая позволит отобрать координаты гипоцентров событий, зарегистрированных станциями на поверхности районов добычи и проектируемых разработок. Это позволит получить изображение сейсмических источников и приемников их сигналов в объеме геологического пространства для того, чтобы оценить степень детальности просвечивания неоднородностей среды: флюидов, конфигураций разломов и т.д.

Литература

1. Сейсмологические бюллетени Единой Геофизической Службы РАН [Электронный ресурс]. URL: http://www.gsras.ru/new/wf/ (дата обращения: 27.05.2026).
2. Смагличенко А.В., Смагличенко Т.А. Построение алгоритма парсинга для структуризации сейсмических данных // Информационные технологии и вычислительные системы. 2026. №2. С.13-18. https//doi.org/10.14357/20718632260202.
3. Смагличенко А.В., Смагличенко Т.А. Задача интеграции сейсмических данных из различных узлов информационной сети // Современные информационные технологии: тенденции и перспективы развития. Материалы XXXIII научной конференции (СИТО 2026), Ростов-на-Дону, 16-18 апреля 2026 г. Таганрог: Издательство Южного федерального университета, 2026. С. 44-48.
4. Универсальные картографические инструменты (GMT) [Электронный ресурс]. URL: https://www.generic-mapping-tools.org/ (дата обращения: 27.05.2026).
5. Национальное Управление Океанических и Атмосферных Исследований. Цифровая модель рельефа GLOBE [Электронный ресурс]. URL: https://www.ngdc.noaa.gov/mgg/topo/gltiles.html (дата обращения: 27.05.2026).

Смагличенко Александр Вадимович. Институт сейсмологии и геодинамики, структурное подразделение Крымского федерального университета им. В.И. Вернадского, г. Симферополь. Инженер. Кандидат технических наук. Область научных интересов: математическое моделирование, программирование, приборостроение, сейсмическая разведка. E-mail: losaeylin@gmail.com (Ответственный за переписку).

Смагличенко Татьяна Александровна. Федеральное государственное бюджетное учреждение науки «Институт проблем нефти и газа Российской академии наук» г. Москва, Россия. Ведущий научный сотрудник. Кандидат физико-математических наук. Область научных интересов: компьютерная томография, геофизика. E-mail: tasmaglichr@gmail.com

Саянкина Мария Константиновна. Федеральное государственное бюджетное учреждение науки «Институт проблем нефти и газа Российской академии наук» г. Москва, Россия. Младший научный сотрудник. Область научных интересов: геофизика. E-mail: msayankina@gmail.com

The developed technology for processing big data from seismic stations with an application

A.V. Smaglichenko[I], T.A. Smaglichenko[II], M.K. Sayankina[II]

[I] Institute of Seismology and Geodynamics, a structural subdivision of the V.I. Vernadsky Crimean Federal University, Simferopol, Russia

[II] Oil and Gas Research Institute, Russian Academy of Sciences, Moscow, Russia

Abstract. Local seismicity data, containing information about the stations that recorded signals, are used to diagnose heterogeneities in the process of energy resource exploration. To assess the details of the result, it is necessary to visualize the distribution of stations in provinces of practical interest. This requires processing of big data, which is available in open access on the websites of the Unified Geophysical Survey of the Russian Academy of Sciences. The data includes geographic coordinate values of stations operating around the world. In this paper, a technology for selecting information related to any region of Russia has been developed. The technology was tested on a dataset from Eastern Siberia and applied to the oil-field region of Western Siberia. For the first time, a constructed map of station distribution shows that the number of stations is extremely small and stations are very far from the exploration fields. Nevertheless, the study of areas that are neighboring to the oil reserves can provide valuable information about the presence of fluids and gas accumulations.

Keywords: *eco- friendly technologies; Siberian deposits; distribution of seismic stations; big data*.
DOI: 10.14357/20790279260306 **EDN:** QRHCAQ

References

1. *Seismological bulletins of the Geophysical Survey of the Russian Academy of Sciences.* http://www.gsras.ru/new/wf/ [Accessed 27 May 2026].
2. Smaglichenko A.V., Smaglichenko T.A. Building a parsing algorithm for the structurization of seismic data // *Informatsionnyye tekhnologii i vychislitel'nyye sistemy*. 2026;(2):13-18. (In Russ). https//doi.org/10.14357/20718632260202.
3. Smaglichenko A.V., Smaglichenko T.A. Zadacha integratsii seismicheskikh dannikh iz razlichnikh uzlov informatsionnoi seti = The task of obtaining seismic data from various nodes of information networks. In: *Sovremennie informatsionnie tekhnologii: tendentsii i perspektivi razvitiya. Materiali XXXIII nauchnoi konferentsii (SITO 2026) = Modern information technologies: trends and development prospects. Proceedings of the XXXIII scientific conference (SITO 2026),* April 16-18, 2026, Rostov-on-Don. Taganrog: Publishing House of the Southern Federal University, 2026:44-48. (In Russ).
4. *The Generic Mapping Tools (GMT)*. https://www.generic-mapping-tools.org/ [Accessed 27 May 2026].
5. *National Oceanic and Atmospheric Administration. GLOBE digital elevation model.* https://www.ngdc.noaa.gov/mgg/topo/gltiles.html [Accessed 27 May 2026].

Smaglichenko Aleksandr V. PhD. Engineer. Institute of Seismology and Geodynamics, a structural subdivision of the V.I. Vernadsky Crimean Federal University, 295007, Prospekt Vernadskogo 4, Simferopol, Crimea, e-mail: losaeylin@gmail.com

Smaglichenko Tatiana A. PhD. Oil and Gas Research Institute of Russian Academy of Sciences, 119333, 3, Gubkina str., Moscow, Russia, e-mail: tasmaglichr@gmail.com

Sayankina Maria K. Oil and Gas Research Institute of Russian Academy of Sciences, 119333, 3, Gubkina str., Moscow, Russia, e-mail: msayankina@gmail.com