

# ГЕОФИЗИЧЕСКИЕ МЕТОДЫ ИЗУЧЕНИЯ НЕДР

УДК 550.8: 004.43

DOI:10.7242/echo.2022.4.11

## ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ЯЗЫКА ПРОГРАММИРОВАНИЯ R В ЗАДАЧАХ ГЕОЛОГИИ И ГЕОФИЗИКИ

И.В. Геник

*Горный институт УрО РАН, г. Пермь*

**Аннотация:** Статья является продолжением цикла работ, посвященных использованию свободного программного обеспечения для решения задач геологии и геофизики. Применение свободного программного обеспечения, включая готовые математические библиотеки и специальные пакеты для конкретной научной области, позволяет существенно снизить время и затраты на научные исследования вследствие использования готовых пакетов функций разного назначения, хорошо проработанных и протестированных, а также значительно повысить воспроизводимость результатов.

При использовании свободных программ или библиотек важную роль играет степень распространенности языка программирования, с помощью которого они созданы. На популярность языка программирования влияет много факторов, включая общее число пакетов функций, активность сообщества, документирование возможностей, наличие разнообразных областей применения. Современные языки программирования вместе со своими пакетами расширения (математическими, статистическими и прочими) играют важную роль в унификации подходов к анализу данных и моделированию, постепенно обеспечивая все более единообразные методы оперирования с данными в различных науках.

В настоящее время самыми используемыми языками программирования для решения научных задач являются Python, R, Matlab, Fortran, Julia, но наибольшую популярность имеют два первых. Python и R, относятся к свободному программному обеспечению, являются языками с динамической типизацией, во многом похожими друг на друга, но отличающимися от C/C++ более простым синтаксисом, что обеспечивает большую легкость и скорость как на начальном этапе освоения, так и впоследствии при написании сложного кода. Язык программирования R, несмотря на лидирующую роль языка Python в научных расчетах, имеет ряд важных достоинств, обеспечивающих R заслуженное второе место. Во-первых, лучшая классификация библиотек, намного сокращающая подбор необходимого пакета. Во-вторых, единообразная форма описания пакетов, что дополнительно ускоряет выбор нужных функций. В-третьих, большая распространенность тестовых наборов данных, что облегчает освоение и понимание работы алгоритмов. В-четвертых, наличие научного периодического издания («The R Journal») информирующего как о новых научных разработках, так и об их практической реализации в виде готовых к использованию пакетов функций.

В статье охарактеризованы основные группы пакетов языка, наиболее подробно рассмотрены группы, отвечающие за вычислительную математику, анализ пространственных данных, цифровую обработку сигналов, геологию и геофизику. Среди геофизических пакетов больше всего библиотек посвящено задачам сейсмологии и сейсморазведки, являющихся наиболее сложными в вычислительном отношении. Имеются пакеты для решения обратных задач геофизики, решения некоторых задач гравиметрии и магнитометрии. Интерес также представляет функциональное наполнение пакетов, не связанных с науками о Земле, но интенсивно развивающихся и дающих возможности исследовать работу новых математических методов.

**Ключевые слова:** геофизика, геология, науки о Земле, программное обеспечение, свободное программное обеспечение, язык программирования R.

### Введение

Интенсивному развитию наук о Земле всегда сопутствовали большие заимствования из смежных наук. Например, возникновение разведочной геофизики стало возможным после успехов теоретической и прикладной механики, астрономии, геодезии, теории электричества и магнетизма. Другой яркий пример заимствования связан с геотектоническими системами, которые значительно повысили эффективность работ в геологии, геофизике и в других науках о Земле.

За последние 40 лет геология и геофизика приобрели новое качество за счет широкого использования новинок информатики: математические пакеты символьных и численных вычислений, геоинформационные системы, программное обеспечение для геологического моделирования месторождений, новые языки программирования.

Можно выделить важнейшие компоненты в перечисленных системах: математические вычисления, базы данных, двумерная и трехмерная компьютерная графика, картографические системы, специализированные геологические, геофизические и иные алгоритмы, в том числе машинного обучения, искусственного интеллекта, анализа временных рядов, изображений и пространственных данных. Указанные подсистемы могут быть интегрированы в готовые пакеты программ, например, «КОСКАД 3D», СИГМА3D так и быть реализованы в виде библиотек и пакетов в математических системах численных вычислений или отдельных языков программирования.

### **Методы и подходы**

Цель работы – анализ возможностей языка программирования R для решения задач геологии и геофизики с указанием его достоинств и недостатков, описанием возможных направлений использования. Использование языка R позволяет: а) непосредственно применять свободные пакеты функций и наборы данных, что обеспечивает получение воспроизводимых результатов; б) попутно осваивать новые направления анализа и обработки данных, возникшие в последние 20-30 лет и плохо представленные в русскоязычных источниках и, тем более, почти никак не представленные в русскоязычной среде в виде алгоритмов со свободным доступом (аналогов советских книг [1, 3] первой половины 1970-х гг.); в) изучать современную англоязычную терминологию анализа данных.

### **Результаты и обсуждение**

Из всей совокупности языков программирования наилучшими возможностями (простота использования математических алгоритмов, баз данных, графики и картографирования, специализированные алгоритмы для наук о Земле) для решения возникающих задач геологии и геофизики обладают языки Python, R, Matlab, Fortran, Julia.

Языки Python, Matlab и R входят в первую двадцатку языков программирования (на основе подсчета результатов поисковых запросов, содержащих название языка). В этом рейтинге в ноябре 2022 года Python был на первом месте, R на двенадцатом, Matlab на пятнадцатом. Из следующих по популярности тридцати языков программирования Fortran на 24 месте, Julia на 32 месте.

Первая открытая версия Julia появилась в 2012 году, язык создавался как конкурент Python, R и Matlab. Язык развивается, но перспективы пока еще полностью не ясны: число пакетов Julia меньше, чем для Python и R, также ниже популярность языка. Общей проблемой многих языков, созданных, подобно Julia, «для замены устаревших» в последние четверть века (например D, Rust, Go как замена C/C++), являются большие ожидания, но намного меньшее реальное использование; очень важным для развития языка является первоначальное занятие некоторой предметной области.

Fortran также ограничен по числу пакетов и несет большой груз старых проблем, которые стали более-менее преодолеваться в последние 5-10 лет. Перспективы языка также туманны сравнительно с Python и R.

Matlab единственный из рассматриваемой пятерки языков не относится к свободному программному обеспечению. Использование в РФ Matlab в настоящее время проблематично, хотя в Matlab имеются хорошие геологические и геофизические пакеты.

В настоящее время Python и R имеют наилучшие перспективы для использования в задачах, связанных с науками о Земле. Python был разработан в 1991 г. как язык общего назначения, важнейшие его библиотеки были созданы: NumPy (1995/2005), Scipy (2001), Matplotlib (2003). Язык R начали публиковать с 1993 г., а с 1995 г. под свободной лицензией вышел первый официальный релиз, репозиторий пакетов начал функционировать с 1997 г. Язык R первоначально создавался для статистической обработки и анализа данных, работать с ним можно из многих статистических пакетов (SPSS, Statistica, Power BI, Gretl и др.).

Python и R являются языками с динамической типизацией, изначально проектируемыми как скриптовые. Это обеспечило возможность быстрого и относительно легкого начала работы для новых пользователей, в первую очередь для научных сотрудников, не являющихся профессиональными программистами и не нуждающихся большей частью в языковых особенностях для системного программирования, что позволило значительно упростить языки в отличие от C/C++. Дальнейшая разработка также ускоряется по сравнению с C/C++. В результате Python и R имеют во многом сходный синтаксис, что при необходимости обеспечивает быструю смену языка программирования.

Если сравнивать библиотеки сходной тематики в количественном плане, то Python имеет примерно в 1.5-2.0 раз больше пакетов, чем R, но важным достоинством репозитория последнего является более детальная классификация и лучшее документирование пакетов. Описание пакетов намного более упорядочено (по каждому направлению четко выделяются по значимости основные и обычные библиотеки), чем в Python, издается даже «The R Journal» – рецензируемый научный журнал с открытым доступом, impact factor 1.673 (2021). «The R Journal» публикует исследования в области статистических вычислений, а также является официальным бюллетенем с информацией о новых пакетах. Особенностью журнала является включение в статьи полного кода, согласно которому читатели могут воспроизводить результаты и примеры. Важным достоинством значительного числа ключевых пакетов R является включение в них помимо функций для обработки данных самих наборов данных. Такие возможности появились позже в больших библиотеках Python, но в R включение данных в пакеты носит более массовый и органичный характер, что связано с изначальным ориентированием на статистические выводы и особым вниманием к воспроизводимости результатов.

В настоящее время наиболее интенсивное развитие пакетной базы языков Python и R связано с решением задач анализа данных, искусственного интеллекта, биологии, химии, медицины, т.е. с научными направлениями, связанными с обработкой и интерпретацией больших наборов данных. Тем не менее, среди пакетов R имеется много библиотек, которые прямо или косвенно можно использовать в задачах наук о Земле.

Рассмотрим основные группы пакетов языка R, представленные в CRAN (The Comprehensive R Archive Network), которые можно использовать для решения задач геологии, геофизики и иных наук о Земле, по указанным выше группам: математические вычисления, базы данных, графика и картография, специальные геологические, геофизические и иные алгоритмы.

Первую, самую большую группу, естественно, составляют пакеты для статистики, относимые к разделам: байесовский вывод, причинно-следственный вывод, кластерный

анализ, распределения вероятностей, робастные статистические методы, эконометрика, метаанализ, анализ экстремальных значений, функциональный анализ данных, графовые вероятностные модели, планирование экспериментов и анализ экспериментальных данных.

При помощи байесовских выводов можно корректировать значение вероятности, например, изменение вероятности появления новых геофизических аномалий, по мере их выделения. Причинно-следственные выводы можно применять для определения различий между вариациями исходных данных, которые могут быть случайными или быть связаны с определенным причинно-следственным механизмом. Использование кластерного анализа возможно для классификации геофизических аномалий, выделения групп, имеющих одну и ту же группу признаков. В случае нефтегазопоисковых задач анализ имеющихся выборок исходных геофизических полей, их трансформант, аномалий, выделяемых по трансформантам, позволяет определить вид распределений вероятностей для этих объектов, а также спрогнозировать вероятность появления геофизических аномалий, не превышающих заданный размер и в конечном счете спрогнозировать максимально возможный объем ресурсов углеводородов на исследуемой площади.

Применение робастных методов позволяет выделять выбросы и снижать их влияние на выборку, что особенно важно при анализе геофизических данных для распределений, которые не являются нормальными. Эконометрические методы используются при построении разного рода регрессионных моделей. Метаанализ сводной статистики или уровней значимости используют для объединения результатов нескольких исследований с проверкой одной или нескольких взаимосвязанных научных гипотез.

Статистический анализ экстремальных значений используется в случае, если экстремальная часть выборки может иметь большое значение (задачи гидрологии, климата, оценки окружающей среды и др.) и продемонстрировать высокий потенциал риска. Функциональный анализ данных – раздел статистики, который анализирует данные, предоставляя информацию о кривых и поверхностях, изменяющихся непрерывно, что используется при построении функциональных моделей линейной регрессии. Графовые вероятностные модели – модели, в которых в виде графа представлены зависимости между случайными величинами, такие модели используются в задачах извлечения информации, компьютерного зрения и обнаружения объектов.

Вторая группа пакетов предназначена непосредственно для основных математических операций: решения дифференциальных уравнений (ДУ) всех типов, высокопроизводительные и параллельные вычисления, численные методы, методы оптимизации.

Пакеты для дифференциальных уравнений предполагают решение обыкновенных ДУ, уравнений в частных производных, дифференциальных уравнений с запаздыванием, дифференциально-алгебраических систем уравнений.

Пакеты высокопроизводительных и параллельных вычислений позволяют использовать интерфейс передачи сообщений (MPI) и реализовывать методы параллельного программирования с явным и неявным параллелизмом, использовать графические процессоры, менеджеры ресурсов и пакетные планировщики, инструменты профилирования, т.е. все необходимые ресурсы для организации и контроля процесса параллельных вычислений.

Группа пакетов численных методов позволяет проводить различные операции с матрицами (в том числе с большими и разреженными), решать линейные уравнения, системы линейных уравнений, задачи методом наименьших квадратов; вычислять специ-

альные математические функции; выполнять операции с полиномами и поиск корней; дифференцирование и интегрирование; решать задачи аппроксимации; решать задачи в символьном виде; предоставляет интерфейсы к языкам Python и Julia, математическим системам SAGE, MATLAB, Octave.

Оптимизационные пакеты реализуют широкий набор методов оптимизации: различные варианты метода множителей Лагранжа, алгоритм Бройдена – Флетчера – Гольдфарба – Шанно, ньютоновские и квазиньютоновские методы; методы квадратичной оптимизации; глобальную и стохастическую оптимизацию; комбинаторную оптимизацию и другие методы. Важным достоинством является наличие пакетов для генерирования тестовых функций для различных видов оптимизации.

Пакеты для работы с базами данных выполнены более единообразно, чем рассмотренные выше, предоставляя доступ и операции со всеми основными реляционными и нереляционными базами данных: Oracle, Microsoft SQL Server, PostgreSQL, MySQL, MariaDB, SQLite, Redis, Elasticsearch, MongoDB и некоторые другие.

Пакеты для графики и картографии для языка R разнообразны и многочисленны. Есть базовая графика, позволяющая отображать разные виды диаграмм (линейные, точечные, круговые, гистограммы и др.) Имеются продвинутые пакеты, например, ggplot2, позволяющие строить (и форматировать) не только диаграммы, но и выполнять различные картографические построения. Вся совокупность пакетов картографирования по классификации CRAN (The Comprehensive R Archive Network) делится на две большие группы: анализ пространственных данных и обработка и анализ пространственно-временных данных. Охарактеризуем кратко обе группы.

Группа пакетов анализа пространственных данных включает: классы для пространственных данных и метаданных; обработка векторных и растровых пространственных данных (инструменты для работы с кубами данных, пространственная выборка, визуализация и др.); интерфейсы к различным геоинформационным системам (ArcGIS, QGIS, PostGIS и др.), веб-сервисам (Earth Engine, GeoServer, стандартные веб-службы OGC) и другим источникам геопространственных данных; геостатистика (анализ вариограмм; интерполяция данных, включая kriging, сплайны, триангуляцию); анализ пространственных данных; работа с другими библиотеками, такими как PROJ, GDAL или GEOS.

Группа пакетов анализа пространственно-временных данных предусматривает: работу с различными представлениями данных; анализ данных, включая геостатистику; отслеживание движения и траекторий; модели, описывающие случайные блуждания и другие виды движения.

Отдельно можно выделить библиотеки для разработки веб-приложений на основе пакета Shiny, на основе которого создаются приложения с графическим интерфейсом пользователя (GUI).

Пакеты, специализированные по различным наукам, предназначены для решения задач: геологии, геофизики, цифровой обработки сигналов (включая анализ изображений), химии (включая хемометрику, спектроскопию и спектральный анализ), анализ сетей и графов, теории надежности и системного анализа, географии (включая гидро- и метеорологию, гидрологию, климатологию, экологию, дистанционные зондирования Земли, океанологию), биологии (включая биостатистику и филогенетический анализ) и медицины.

Геологические пакеты дают возможность: анализа данных палеонтологии (в том числе палеоэкология, палеоклиматология, анализ фоссилей), стратиграфии (в т.ч. цик-

лостратиграфия), кристаллографии, осадконакопления, построения геохронологических шкал, геологических карт, радиоизотопного датирования, решения задач морфометрии, анализа изображений микроскопии.

Геофизические пакеты включают: анализ данных сейсмологии и работа с данными сейсморазведки, геодезические расчеты и работа с картографическими проекциями, анализ топографических и батиметрических данных, 2.5D гравитационное моделирование, вычисление магнитного поля Земли по моделям IGRF и WMM, численные методы для решения обратных геофизических задач (в т.ч. оптимизационные методы), конечноэлементное моделирование двумерных механических задач, 3D визуализацию геофизических данных [2]. В целом наибольшее число и наиболее крупные пакеты связаны с задачами сейсмологии и сейсморазведки, являющимися наиболее сложными геофизическими методами.

Пакеты цифровой обработки сигналов позволяют: выполнять предварительную обработку данных, bootstrapping; анализировать стационарные и нестационарные, одномерные и многомерные временные ряды; выполнять различного рода фильтрацию (сглаживание, удаление шумов и трендов, калмановскую фильтрацию, устранение сезонности); производить различные виды спектрального анализа; выделять аномалии в данных; выполнять прогнозирование временных рядов; синтезировать сигналы; анализировать и преобразовывать изображения разными способами; выполнять сегментацию временных рядов и изображений; генерировать случайные наборы данных разной размерности.

Пакеты, не связанные прямо или косвенно с геологией и геофизикой, интересны своими математическими методами, которые в настоящее время интенсивно развиваются. Для таких наук как химия, биология, медицина представляют интерес подходы, связанные с работой с дискретными элементами, аналогами которых в геофизике могут быть сейсмические структуры и приподнятые участки, гравитационные, магнитные и электрометрические аномалии.

### **Выводы**

Показано, что интенсивное развитие геофизики за последние десятилетия связано с использованием нового программного обеспечения разного типа, в котором важное место занимают языки программирования. Рассмотрены вопросы функциональности языков программирования, которые наиболее часто применяются для решения геофизических задач: Python, R, Matlab, Fortran, Julia. Среди указанных языков наибольшее значение имеют Python и R, лидируя по популярности, имея самое большое число библиотек. Важнейшими факторами, обеспечивающими интерес к двум указанным языкам, является то, что они относятся к свободному программному обеспечению (в отличие от Matlab), имеют большие сообщества пользователей (в отличие от Fortran и Julia), существуют более четверти века (в отличие от Julia) и сумели быстро распространиться на более широкий круг задач, чем предназначалось изначально (в отличие от Fortran).

Важным достоинством языка программирования R являются: лучшее документирование функций и более детальная классификация пакетов, публикация статей о новых пакетах в специальном журнале. Благодаря этому язык R вышел из круга первоначальных задач математической статистики, применяясь в настоящее время для решения задач математики, обработки сигналов и изображений, геологии и геофизики, географии, биологии, химии, медицины. Развитие пакетной базы языка, связанное с проникновением информатики в науки, ранее слабо математизированные, с одной стороны, увеличивает число и повышает качество математических пакетов общего назначения, а с другой стороны, интенсифицирует опробование новых математических методов на новых объектах.

Охарактеризованы группы пакетов, связанных с вычислительными задачами, базами данных, графикой и картографией, а также предназначенных для конкретных наук. Наиболее полно рассмотрены пакеты, предназначенные для математических вычислений, анализа пространственных данных, цифровой обработки сигналов, геологии и геофизики.

Таким образом, язык программирования R, наряду с Python, является современным средством решения широкого круга задач в различных науках, включая науки о Земле, в том числе геология и геофизика.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Библиотека алгоритмов. Вып. 1. 1б-50б: справ. пособие. – М.: Совет. радио, 1975. – (Б-ка технич. кибернетики).
2. Геник И.В. Свободное программное обеспечение для задач гравиразведки и магниторазведки // Горное эхо. – 2020. – № 3 (80). – С. 60-66. – DOI: 10.7242/echo.2020.3.12.
3. Уилкинсон Дж.Х. Справочник алгоритмов на языке АЛГОЛ: Линейная алгебра. – М.: Машиностроение, 1976. – 390 с.

УДК 550.834

DOI:10.7242/echo.2022.4.12

## СЕЙСМОРАЗВЕДОЧНЫЙ МОНИТОРИНГ ЖЕЛЕЗНОДОРОЖНОГО УЗЛА

А.А. Жикин

*Горный институт УрО РАН, г. Пермь*

**Аннотация:** Представлены результаты сейсморазведочного мониторинга объекта транспортной инфраструктуры, расположенного на подработанной территории калийного месторождения. Рассмотрены особенности наблюдения и обработки малоглубинных сейсмических данных. Показаны результаты режимных исследований динамики изменения упругих параметров массива горных пород.

**Ключевые слова:** сейсморазведочный мониторинг, малоглубинная сейсморазведка, цифровая обработка сейсмических данных.

Риски возникновения аварийных ситуаций при разработке водорастворимых месторождений, как правило, связаны с нарушением устойчивости грунтов, что в свою очередь приводит к частичному или полному разрушению инфраструктурных объектов, попадающих в зону влияния максимальных негативных изменений прочностных свойств подработанного массива.

Серия аварийных событий на Верхнекамском месторождении калийных солей (ВКМКС), в результате которых произошли обрушения грунта, повреждения конструкций промышленного, транспортного и гражданского строительства, обусловила создание современных регламентов комплексного обеспечения безопасности ведения горных работ и контроля подработанных территорий. Особое внимание уделяется объектам жилой застройки и транспортной системе, на которых ведутся режимные исследования, включающие широкий набор геофизических, гидрогеологических, сейсмологических и газогеохимических наблюдений.