*ФИЗИКА ЗЕМЛИ, 2022, № 1, с. 3–34*

**ОБЗОР**

*УДК 550.3*

# БОЛЬШИЕ ДАННЫЕ В ГЕОФИЗИКЕ И ДРУГИХ НАУКАХ О ЗЕМЛЕ

**© 2022 г. А. Д. Гвишиани[1,2], М. Н. Добровольский[1], Б. В. Дзеранов[1,3], Б. А. Дзебоев[1,3,*]**

[1]*Геофизический центр РАН, г. Москва, Россия*

[2]*Институт физики Земли им. О.Ю. Шмидта РАН, г. Москва, Россия*

[3]*Геофизический институт Владикавказского научного центра РАН, г. Владикавказ, Россия*

******E-mail: b.dzeboev@gcras.ru*

Поступила в редакцию 24.03.2021 г.
После доработки 13.04.2021 г.
Принята к публикации 16.04.2021 г.

Термин “Большие Данные” (“Big Data”) в последнее десятилетие стал весьма популярен. Частота его использования в научных статьях, докладах и широкой прессе непрерывно растет. В статье описывается история возникновения и развития теории и практики Больших Данных (БоД), как научной дисциплины, приводятся их главные характеристики и методы обработки и анализа, представляются формализм и семейство V-характеристик БоД, перечисляются примеры источников растущих БоД, оказывающих фундаментальное воздействие на развитие геофизики и смежных наук о Земле. В качестве примеров источников БоД в науках о Земле рассмотрены: дистанционное зондирование Земли, метеорология, геоэкология на примере глобальной иерархической сети SMEAR (Stations Measuring Earth surfaces and Atmosphere Relations) и сейсмическая разведка. Также рассмотрены данные сейсмического мониторинга, которые могут стать БоД при комбинировании с другими геофизическими данными, и геомагнитные данные, не являющиеся БоД, но, тем не менее, представляющие большую научную ценность.

*Ключевые слова:* Большие Данные (БоД), науки о Земле, дистанционное зондирование Земли, метеорологические наблюдения, сейсмический мониторинг, сейсморазведка, геоэкология, геомагнитные наблюдения.

**DOI:** 10.31857/S0002333722010033

## ВВЕДЕНИЕ

В настоящее время наблюдается беспрецедентный непрерывный рост объемов данных, собираемых и анализируемых в различных областях науки и техники [Gvishiani, Soloviev, 2020]. Важнейшими источниками данных являются: интернет (социальные сети, мессенджеры, рекламная аналитика и т.д.), интернет вещей (Internet of Things – IoT) [Internet Of Things, 2020], сети сенсорных наблюдений, данные спутниковых наблюдений и многое другое. Это приводит к необходимости хранения и обработки поистине гигантских массивов данных, не имеющих аналогов в прошлом. Особенностью таких данных является то, что они часто являются неструктурированными или частично структурированными [Dedić, Stanier, 2017]. Сбор и обработка этих данных должны выполняться исключительно оперативно. При этом часто речь идет об их обработке в реальном или близком к реальному времени.

Указанные особенности современных исходных данных определяют общемировую тенденцию, а собственно информационные массивы называются Большими Данными (Big Data). Далее по тексту мы будем сокращать их как БоД.

БоД можно охарактеризовать как данные огромных объемов, представленные в разнородных форматах, создаваемые и обрабатываемые с большой скоростью. Это затрудняет, если вообще делает возможной, работу с БоД на основе традиционных технологий реляционных баз данных. Поэтому, чтобы работать с постоянно быстро растущими объемами, разнообразием и скоростями поступления и обработки БоД развиваются специальные технологии. Это в первую очередь коснулось интернет-компаний. Например, Google обрабатывает более петабайта данных в день [Mayer-Schönberger, Cukier, 2014]. Единицы измерения количества информации приведены в таблице.

В последние годы термин Большие Данные стал чрезвычайно популярным как в научной, так и в деловой литературе и средствах массовой информации [Gvishiani et al., 2021]. БоД начинают появляться в новых областях науки и техники. При этом использование этого термина стало

**Таблица 1.** Единицы измерения количества информации в байтах [ГОСТ 8.417-2002, 2013]

| Название | Обозначение | Степень |
|---|---|---|
| Байт | Б | $10^0$ |
| Килобайт | Кбайт | $10^3$ |
| Мегабайт | Мбайт | $10^6$ |
| Гигабайт | Гбайт | $10^9$ |
| Терабайт | Тбайт | $10^{12}$ |
| Петабайт | Пбайт | $10^{15}$ |
| Эксабайт | Эбайт | $10^{18}$ |
| Зеттабайт | Збайт | $10^{21}$ |
| Йоттабайт | Ибайт | $10^{24}$ |

своего рода модой и не всегда является оправданным.

Более того, классификация как БоД информационных массивов, таковыми в реальности не являющихся, может приводить к применению к ним неадекватных математических методов. Свою недавнюю книгу об этом известный американский специалист по БоД Кэти О'Нил (Cathy O'Neal) [O'Нил, 2018] посвящает ущербу для отдельно взятых людей от того, что она называет ОМП – Оружием Математического Поражения.

Когда успешно развитые сегодня методы хранения и обработки БоД неоптимально подбираются или применяются к массивам, ошибочно классифицированным как БоД, последние могут превратиться в ОМП, направленное против отдельных индивидуумов [O'Нил, 2018]. Рассматриваемый в данной статье вопрос о строгом и самодостаточном определении БоД становится со временем все более актуальным как для общества в целом, так и для отдельных личностей.

Впервые термин Большие Данные был введен в 2008 г. по аналогии с такими терминами как Большая Вода, Большие Деньги, Большая Нефть и т.д. 03.09.2008 г. редактор журнала "Nature" Клиффорд Линч (Clifford Lynch) подготовил выпуск на тему "Как технологии, открывающие возможности работы с большими объемами данных, могут повлиять на будущее науки". С этого момента термин прочно вошел в научную и деловую литературу, сначала в США, а затем и во всем мире.

В 2010–2020 гг. БоД стали академической дисциплиной в университетах многих стран мира. Причинами быстрого принятия и распространения этого раздела науки о данных стали огромные новые возможности, которые обещает использование БоД в науке и бизнесе, а также создание и развитие собственно технологий для работы с БоД. Если проследить историю возникновения термина Большие Данные еще дальше, то можно обнаружить его использование уже в 1990-х годах. Популяризация термина в это время была связана с именем Джона Маши (John Mashey) [Lohr, 2013], научного руководителя Silicon Graphics.

Когда говорят о БоД, то, прежде всего, имеют в виду, что их объем трудно хранить и обрабатывать с помощью традиционных программ, таких как реляционные базы данных [Snijders et al., 2012]. В то же время размер данных, которые считаются большими, со временем и развитием технологий постоянно меняется. Из-за этих особенностей для извлечения знаний из этих данных также требуются специальные методы и технологии [Hashem et al., 2015].

Сегодня научная дисциплина БоД располагает весьма обширной библиографией. Так в монографии [Mayer-Schönberger, Cukier, 2014] объясняется как БоД изменяют парадигму науки о данных, да и нашу жизнь в целом. Происходит смена знаний ***почему*** на ***что именно***.

Книга [Karimi, 2014] знакомит читателей с геопространственными приложениями, относящимися к категории БоД. В ней исследуются новые тенденции, такие как геокраудсорсинг и облака точек LiDAR. Рассматривается ряд тем, связанных с методами и технологиями сбора БоД в геоинформатике, включая распределенные вычисления, анализ геопространственных данных, социальные сети и географическую информацию, предоставляемую на добровольной основе.

Практическое руководство [Kleppmann, 2017] описывает плюсы и минусы различных технологий обработки и хранения данных. В работе [Hurwitz et al., 2013] дается определение БоД и объясняется их важность. Работа [Marz, Warren, 2015] посвящена созданию БоД-систем, используя архитектуру, разработанную специально для сбора и анализа данных в веб-масштабе. Монография [Marr, 2015] является введением в практику применения БоД. На многочисленных реальных примерах показаны пять этапов модели **SMART**: начните со стратегии (**S**tart with Strategy), измеряйте показатели и данные (**M**easure Metrics and Data), применяйте аналитику (**A**pply Analytics), сообщайте о результатах (**R**eport Results), трансформируйте (**T**ransform). В работе [Karau et al., 2015] представлена Apache Spark – система кластерных вычислений с открытым исходным кодом, которая делает анализ данных быстрым в написании и быстрым в выполнении. Книга [Inmon, Linstedt, 2014] рассматривает более широкую архитектурную картину того, как БоД вписываются в существующую информационную инфраструктуру.

В настоящей статье приводится описание современного понимания БоД и обзор их использования в геофизике и других науках о Земле. Излагаются оригинальные методы анализа БоД. Демонстрируется, что в науках о Земле БоД имеют

место в метеорологии, геоэкологии, дистанционном зондировании Земли и сейсмологии. Оценивается возможность применения подхода БоД в геомагнетизме.

## ФОРМАЛИЗМ БОЛЬШИХ ДАННЫХ

Возникновение БоД как самостоятельного явления было обусловлено появлением в конце нулевых годов нынешнего века новых технологических возможностей для анализа гигантских информационных массивов. Начиная с этого времени предлагались различные определения понятия Большие Данные. Одно из современных определений выглядит так [Manyika et al., 2011; Preimesberger, 2011]:

**Определение 1.** Большие Данные означают структурированные и неструктурированные массивы информации огромных объемов и существенного многообразия, эффективно обрабатываемые горизонтально масштабируемыми алгоритмическо-программными инструментами, альтернативными традиционным системам управления базами данных (СУБД) и аналогичным решениям.

Критерий “трех V” устанавливает определяющие характеристики того, что некоторый набор объектов представляет собой Большие Данные. **Три V** – это **физический объем (Volume)**, **скорость (Velocity)** и **разнообразие (Variety)**. Последнее V понимается не только как широкое разнообразие видов и типов данных, но и как возможность одновременной обработки структурированных и частично структурированных данных (рис. 1).

**Объем (Volume)** данных определяет их ценность и потенциальные результаты их анализа, а также собственно возможность вообще считать их БоД.

**Скорость (Velocity)** – суть скорость, с которой данные генерируются, передаются и обрабатываются. В случае БоД это часто происходит в режиме реального времени.

**Разнообразие (Variety)** определяется типами используемых данных. Это может быть текстовая или графическая информация, аудио, видео и т.д.

Критерий “трех V” был предложен в 2001 г., первоначально вне связи с БоД. Он появился в работах Meta Group [Laney, 2001] в связи с популяризацией концепции центрального хранилища данных. Более чем 15 лет спустя критерий “трех V” был успешно взят на вооружение теорией и практикой БоД. Вопрос о том, дает ли этот критерий необходимые и достаточные условия выполнения Определения 1 строго говоря остается открытым. Для получения ответа здесь необходимо более глубокое аксиоматическое проникновение в логику БоД.

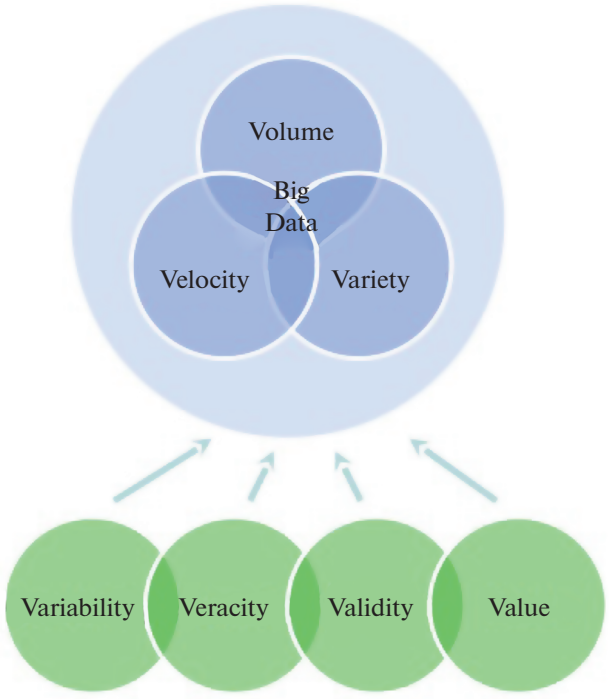


**Рис. 1.** Основные и некоторые дополнительные признаки V для БоД [Gvishiani, Soloviev, 2020].

Помимо приведенных основных “трех V”, для характеристики БоД используют также и некоторые дополнительные признаки. Их английские названия также начинаются на букву V. К ним прежде всего относятся: способность к **вариативности (Variability)**, **достоверность (Veracity)**, **обоснованность (Validity)** и **полезность (Value)** (рис. 1). Дополнительные V в критерии несут, вообще говоря, иную функциональную нагрузку по отношению к трем основным. Они характеризуют важные свойства тех или иных уже выявленных БоД.

Приведем краткую расшифровку дополнительных признаков.

**Достоверность (Veracity)** относится к качеству БоД [Onay, Öztürk, 2018], влияющему на результаты выделения знаний. Речь идет о надежности источника данных и о том, насколько осмысленным является анализ на основе данных из этого источника.

**Обоснованность (Validity)** определяет точность и корректность данных.

**Ценность (Value)** определяет возможность получения полезной информации из данных.

**Способность к вариативности (Variability)** – возможность усиления вариативности со временем. Этот дополнительный признак имеет важное значение, т.к. он расширяет используемую в опреде-

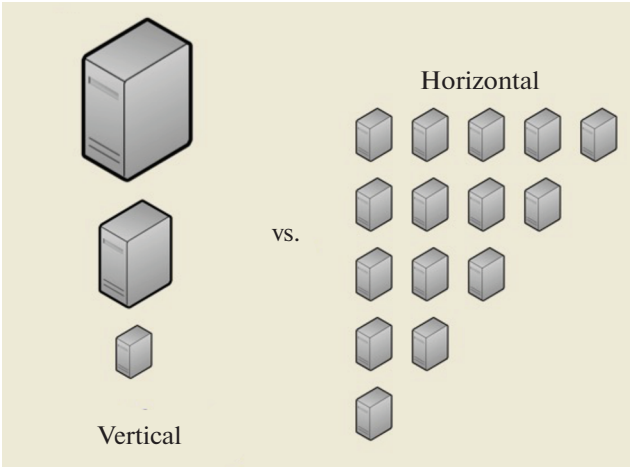


**Рис. 2.** Вертикальное и горизонтальное масштабирование [Gvishiani, Soloviev, 2020].

лении БоД основную характеристику **разнообразия (V – Variety)**.

Пусть разворачивающиеся во времени данные обладает большим объемом и высокими скоростями их приема и обработки. При этом наблюдается постоянный рост разнообразия видов и типов данных, т.е. имеет место усиление их **вариативности**. Тогда, даже если пороги **разнообразия (V – Variety)** еще не достигнуты, данные могут быть классифицированы как стремящиеся к БоД. Для этого достаточно, чтобы указанный рост вариативности имел неуклонный тренд.

Можно выделить еще несколько дополнительных характеристик БоД:

**Полнота (Exhaustive)**. Обеспечивают ли данные полный охват всей наблюдаемой системы?

**Расширяемость (Extensional)**. Имеется ли возможность легко добавлять и/или изменять поля данных.

**Масштабируемость (Scalability)**. Определяет возможность быстрого увеличения объема данных.

Основным принципом обработки Больших Данных является горизонтальная масштабируемость. Речь идет об обработке распределенных данных на большом количестве вычислительных узлов. Этот принцип заложен в следующее определение БоД, разработанное NIST (National Institute of Standards and Technology, Национальный институт стандартов и технологий США):

**Определение 2.** Большие Данные состоят из громадных наборов данных, главным образом, по характеристикам объема, разнообразия, скорости и/или изменчивости, которые требуют масштабируемой архитектуры для эффективного хранения, обработки и анализа [NIST..., 2019].

В силу отсутствия на сегодняшний день аксиоматики как для дисциплины БоД, так и для науки о данных в целом, вопрос эквивалентности определений 1 и 2 остается открытым.

**Масштабируемость** означает способность системы, сети или процесса справляться с увеличением нагрузки и/или увеличивать свою производительность путем добавления аппаратных ресурсов.

**Определение 3.** Система называется масштабируемой, если:

1) она обладает возможностью наращивания дополнительных ресурсов без структурных изменений собственно системы;

2) она способна увеличивать свою производительность пропорционально дополнительным ресурсам.

Различают два вида масштабирования: вертикальное и горизонтальное (рис. 2).

Вертикальное масштабирование определяется увеличением производительности каждого компонента системы. При этом, вообще говоря, не требуется изменения программного обеспечения, используемого для работы с системой.

Горизонтальное масштабирование представляет собой добавление к системе дополнительных компонентов. В этом случае может потребоваться изменение программного обеспечения, используемого для работы с системой.

Одной из ключевых особенностей БоД является то, что в основном это данные с низкой плотностью информации. Задачами обработки БоД являются выявление связей и зависимостей внутри этой информации и построение на их основе прогнозов различного рода. В настоящее время эти задачи решаются, в том числе, с использованием методов и алгоритмов искусственного интеллекта.

Как отмечается в работе [Фрэнкс, 2014], БоД представляют интерес не из-за своих размеров, а из-за новых мощных средств их анализа. При этом значительная часть БоД может вообще не представлять никакого интереса. Соответственно, важной задачей является фильтрация БоД для выбора нужной и полезной информации.

В работе [Фрэнкс, 2014] выделяются такие отличия БоД от традиционных источников данных, как их автоматическая генерация без участия человека, принципиально новые источники информации, частая не дружественность информации к пользователю, возможность большого содержания бесполезной информации. Помимо преимуществ, от использования БоД появляются и риски.

Согласно работе [Майер-Шенбергер, Кукьер, 2014] **БоД представляют собой три шага к новому способу анализа информации. Первый шаг** состоит в анализе огромного количества данных, что позволяет во многих случаях отказаться от выборок и обрабатывать все имеющиеся данные, связанные

изучаемым явлением. **Второй шаг** состоит в снижении претензий к точности. Он позволяет ценой некоторой неточности на микроуровне, достигать существенных результатов на макроуровне. И наконец, **третий шаг** состоит в отходе от поиска причинности. Вместо этого математические методы 21-го века ищут корреляции в огромных массивах данных.

## ИСТОЧНИКИ И ПРИМЕРЫ БОЛЬШИХ ДАННЫХ

Основным источником БоД является Интернет, где в минуту происходят десятки миллионов обращений к таким сервисам как Google, Twitter, Facebook, Yahoo, Yandex, Instagram, Netflix и др. Эти обращения генерируют огромные объемы данных, которые необходимо анализировать.

В последнее время также стремительно развивается так называемый Интернет вещей (Internet of Things, IoT) – сеть физических объектов со встроенными технологиями для связи и взаимодействия с их внутренними состояниями или внешней средой [Internet Of Things, 2020]. Организация таких сетей рассматривается как явление, способное реорганизовать научно-исследовательские, экономические и общественные процессы и до какой-то степени исключить человека из сетевых операций [Ashton, 2009].

Сама концепция появилась в 1999 г. благодаря пониманию перспектив радиочастотной идентификации для распознавания взаимодействий физических объектов друг с другом и с окружающей средой. А в 2010-е годы она стала одной из тенденций развития информационных технологий [LeHong, Fenn, 2012]. Причинами этого стали стремительное развитие беспроводных межкомпьютерных телекоммуникационных сетей, появление облачных вычислений и возможностей архивирования облачных данных, а также активный рост технологий межмашинной связи.

БоД также могут быть получены из внутренней информации крупных предприятий, научно-исследовательских институтов и других организаций [Chen et al., 2014].

Помимо перечисленных, к источникам БоД однозначно относятся:

– потоки данных о местоположении мобильных пользователей;

– потоки данных аудио- и видеозаписи в реальном времени. Например, с камер видеонаблюдения;

– потоки данных непрерывных измерений, при условии достаточного разнообразия регистрируемых характеристик;

– потоки данных фондовых, валютных и товарных бирж в режиме реального времени.

В науках о Земле к основным источникам БоД относятся:

– дистанционное зондирование Земли (ДЗЗ);

– метеорологические наблюдения;

– геоэкологический мониторинг;

– сейсмическая разведка.

В последние годы к ним добавились наблюдения беспилотных летательных аппаратов [Ofli et al., 2016], данные беспроводных сенсорных сетей [Chen et al., 2013], данные численного моделирования [MacLachlan et al., 2015], а также данные краудсорсинга [Poblet et al., 2014].

Мы живем в период бурного развития БоД. В ближайшем будущем БоД проникнут в научные исследования, коммерческий сектор и сферу государственного учета и управления.

## ТЕХНОЛОГИИ ХРАНЕНИЯ И МЕТОДЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

В науках о Земле среди технологий работы с БоД можно выделить: кубы данных [Nativi et al., 2017], онлайн-портал USGS Earth Explorer (https://earthexplorer.usgs.gov), Copernicus Sentinel Hub (https://scihub.copernicus.eu) [Jutz, Milagro-Perez, 2020], портал GEOSS (www.geoportal.org) [Giuliani et al., 2017]. Для эффективного анализа БоД используются такие решения как Google Earth Engine [Gorelick et al., 2017] и EarthServer [Baumann et al., 2016].

С целью изучения океанографии в NASA создали Платформу анализа научных данных (Science Data Analytics Platform, SDAP) [Armstrong et al., 2019]. На ее основе были созданы: Состояние океана (State of the Ocean, SOTO) [Huang et al., 2017] и Портал по измерению уровня моря (Sea Level Portal) [Huang, 2017]. Помимо этого, для использования в области наук о Земле достижений БоД можно отметить такой проект, как NASA Earthdata Cloud [McInerney, 2020].

Аналитические системы БоД в науках о Земле в основном используют облачные вычисления [Gomes et al., 2020]. При этом в некоторых случаях для обеспечения полного контроля над данными используется частное облако. Но чаще используются сторонние поставщики облачных решений, что упрощает поддержку и обслуживание систем. Используются также гибридные облака, когда конфиденциальные данные помещаются в частное облако, а для обеспечения основного доступа пользователей используется публичное облако [Islam, Reza, 2019].

Извлечение знаний из БоД в науках о Земле состоит из подготовки, преобразования и анализа больших объемов пространственно-временных и спектральных данных различных типов [Kempler, Mathews, 2017].

Подготовка и преобразование состоят в изучении и очистке исходных данных, которые могут быть избыточными и зашумленными. Это может занимать до 50–80% всего времени анализа данных [Kempler, Mathews, 2017]. Методы подготовки и преобразования данных можно разделить на методы извлечения, конвертации, оценки и сжатия данных.

К методам извлечения относятся: выделение аномалий [Corizzo et al., 2019], фильтрация [Moore, 2014] и удаление выбросов [Balta et al., 2018]. К методам преобразования – конвертация форматов [Esch et al., 2018], сглаживание [Katzfuss, Cressie, 2011], нормализация, пространственная интерполяция [Mitas, Mitasova, 1999], преобразование координат [Kang et al., 2018]. К методам оценки – коррекция смещения данных [Maraun et al., 2017], анализ чувствительности [Salimi, Hammad, 2020], оценка [Caers, 2011]. К методам сжатия данных – устранение избыточности, машинное обучение [Rodriguez-Galiano, Chica-Rivas, 2014], кластеризация [Chowdhury, Al-Zahrani, 2014], классификация [Tao et al., 2020], выделение признаков [Stromann et al., 2020], факторный анализ [Watson, 2017].

Извлечение знаний из БоД после их предварительной обработки состоит в выявлении скрытых закономерностей, поиске неизвестных корреляций и выделении другой полезной информации. Используемые для этого методы можно разделить на методы моделирования и прогнозирования, статистические методы, машинное обучение и глубокое обучение.

К методам моделирования и прогнозирования относятся: агентное моделирование [Wu et al., 2017] и моделирование систем Земли (климат, погода, поверхность суши, океан) [Guo, 2017]. К статистическим методам – определение распределения данных, регрессия, корреляция, кластеризация, классификация, уменьшение модели [Cressie, 1993]. К методам искусственного интеллекта и машинного обучения – классификация (логистическая регрессия, метод опорных векторов, деревья классификации, случайные леса, метод ближайшего соседа), кластеризация (смешанные модели, алгоритм K-средних, агломеративная кластеризация), регрессия (линейная регрессия и ее расширения, усиленные деревья регрессии, нейронные сети, функциональный анализ), уменьшение размерности (метод главных компонент) [Holloway, Mengersen, 2018]. Наиболее популярным и быстро развивающимся разделом искусственного интеллекта и машинного обучения в настоящее время является глубокое обучение – автокодировщики, сверточные нейронные сети, глубокие сети доверия, рекуррентные нейронные сети [Ball et al., 2017].

Традиционные статистические методы основаны на определенных допущениях о характере распределения исследуемых данных. Они широко используются для выявления сложных взаимоотношений между наблюдаемыми параметрами и определения географического и частотного распределения в БоД наук о Земле [Pei et al., 2020]. Для характеристики данных о Земле могут использоваться такие простые статистики, как среднее и стандартное отклонение. Для оценки точности получаемых оценок используются доверительные интервалы. Проверка гипотез позволяет сравнивать распределение разных наборов данных. Методы корреляции, интерполяции и экстраполяции позволяют строить прогнозы на основании известных данных [Stein et al., 2006]. Регрессионные методы позволяют выявлять скрытые взаимосвязи в системах Земли [Becker-Reshef et al., 2010]. Методы кластеризации – группировать схожие пространственные объекты в классы [Besic et al., 2016]. Методы сокращения статистических моделей уменьшают вычислительную сложность при численном моделировании [Fernández-Martínez, 2015]. Но для эффективного применения статистических методов необходимо правильно оценивать распределение используемых данных и их статистические свойства.

Методы искусственного интеллекта и машинного обучения позволяют строить модели, основанные на характеристиках и особенностях наблюдаемых данных. Такие модели позволяют выявлять в данных скрытые закономерности. Они обычно лучше, чем традиционные статистические методы выявляют нелинейные зависимости, что особенно важно при исследовании систем Земли [Fouedjio, Klump, 2019]. Регрессия, классификация и кластеризация, полученные из статистических методов также могут использоваться как методы искусственного интеллекта и машинного обучения, поэтому точное разделение между машинным обучением и статистическими методами не всегда понятно. Например, с помощью искусственных нейронных сетей можно строить регрессию для прогнозирования [Pyo et al., 2019].

Методы глубокого обучения, возникшие на основе машинного обучения, предлагают уникальные возможности в выделении признаков на основе БоД в науках о Земле. Они показали свою эффективность в таких областях, как компьютерное зрение [Voulodimos et al., 2018], обработка естественного языка [Lopez, Kalita, 2017], построение рекомендательных систем [Zhang et al., 2019] и т.д. В науках о Земле их используют для таких задач, как распознавание объектов на снимках дистанционного зондирования [Hoeser et al., 2020; Song et al., 2019], изучение изменения климата [Hassani et al., 2019], предсказание погоды [Schultz et al., 2021; Weyn et al., 2019] и др.

## БОЛЬШИЕ ДАННЫЕ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ (ДЗЗ)

**Дистанционным зондированием** называется метод получения информации об объекте или явлении без непосредственного физического контакта. Как научную дисциплину дистанционное зондирование, как правило, рассматривают в качестве раздела географии. В современном его понимании, термин, в основном, относится к технологиям воздушного или космического зондирования Земли при помощи распространения и регистрации сигналов с целью обнаружения и описания объектов земной поверхности, а также атмосферы и океана. Соответственно, для названия метода сегодня в основном используется термин ДЗЗ (Дистанционное Зондирование Земли).

Первые дистанционные снимки Земли связывают с именем французского воздухоплавателя Гаспара-Феликса Турнашона, также известного как Надар, который в 1859 г. сделал фотоснимки Парижа с воздушного шара. Существенное развитие аэрофотосъемка получила в результате Первой и Второй мировых войн [Jong et al., 2004].

Бурное развитие ДЗЗ связано с запуском космических аппаратов и началом исследования Земли из космоса. Как следствие, в 1960 г. Эвелин Пруитт (Evelyn Pruitt) из Управления военно-морских исследований США ввела термин **дистанционное зондирование (remote sensing)** [National Research Council, 1998]. Сильное влияние на развитие ДЗЗ оказала американская спутниковая программа Landsat [Baumann, 2009] изучения ресурсов Земли (с 1972 г. по н.в.) [National Research Council, 1998].

Среди пионеров ДЗЗ можно выделить имя Джеймса Ван Аллена (James Van Allen), открывшего радиационные пояса Ван Аллена [Van Allen et al., 1958; Van Allen, Frank, 1959]. Еще одним пионером в этой области был Гарри Векслер (Harry Wexler), продвигавший использование спутников в метеорологии. Одним из отцов-основателей спутниковой метеорологии стал и Вернер Эдвард Суоми (Verner Edward Suomi), энергично создававший Всемирную службу погоды, возникшую в 1963 г. уже после его смерти [National Research Council, 2008].

В нашей стране первым космическим экспериментом ДЗЗ был полет второго космонавта Германа Титова, который 6–7 августа 1961 г. в течение 25 ч своего орбитального полета вел систематическую фото- и киносъемку с помощью отечественной кинокамеры “Конвас” и фотокамеры “Зенит” [Батурин, Щербинин, 2011] (рис. 3). 04.10.1957 г., запущенный в СССР в рамках Международного геофизического года, Первый искусственный спутник Земли определил плотность верхних слоев атмосферы.

**Рис. 3.** Второй космонавт Герман Титов с кинокамерой “Конвас” во время подготовки к полету, 1961 г. (из личного архива Батурина Ю.М.)

Системные наблюдения Земли из космоса начались в 1975 г. на орбитальной станции “Салют-4” силами космонавтов А.А. Губарева и Г.М. Гречко. Наблюдения были продолжены в том же году вторым экипажем станции: П.И. Климук и В.И. Севастьянов. С тех пор наблюдения ДЗЗ проводились на всех орбитальных советских и российских станциях, включая “Мир” и МКС (рис. 4).

Говоря о ДЗЗ, мы вспоминаем имя Константина Эдуардовича Циолковского, первым предложившего ракету как средство покинуть земную атмосферу и заложившего этим основы всей современной космонавтики. Дальнейшие успехи ДЗЗ в нашей стране были бы невозможны без Сергея Павловича Королева, чьими усилиями, во многом, была задумана и реализована советская программа освоения космоса.

ДЗЗ разделяют на активное (сигнал сначала излучается самолетом или космическим спутником, рис. 5) и пассивное (регистрируется только сигнал от естественных источников, рис. 6) [Liu, Mason, 2013]. Отраженный солнечный свет – наиболее часто используемый источник излучения, регистрируемый пассивными сенсорами. Примерами пассивного дистанционного зондирования являются цифровая и пленочная фотография, наблюдение инфракрасных лучей, а также эмиссия радиометров.

Активные приборы излучают сигнал с целью сканирования объекта и объемлющего его пространства. После этого сенсор способен обнаружить и измерить излучение, отраженное или образованное путем обратного рассеивания объектом зондирования.

Примерами активных сенсоров ДЗЗ являются радар и лидар. Радар использует метод радиолокации, основанный на излучении радиоволн и регистрации их отражений от объектов. Лидар является активной оптической системой, основан-

**Рис. 4.** Космонавт Юрий Батурин ведет съемку Земли с борта космического корабля (из личного архива Батурина Ю.М.)

ной на поглощении и рассеянии света. В обоих случаях на основе анализа отраженного сигнала определяется размещение, скорость и направление движения объекта.

ДЗЗ позволяет получать данные об опасных, труднодоступных и быстро движущихся объектах и одновременно проводить наблюдения на обширных территориях. Среди приложений ДЗЗ – мониторинг вырубки лесов, городское планирование, управление и развитие, определение состояния ледников в Арктике и Антарктике, измерение глубины океана, наблюдение за морскими судами, глобальный и локальный мониторинг распространения заболеваний, пожаров, наводнений, нефтяных пятен, наблюдения за последствиями землетрясений и оползней, извержений вулканов, мониторинг урожаев сельскохозяйственных культур и природопользования в целом, наблюдение за развитием экосистем [Chi et al., 2016]. Рынок ДЗЗ относится к категории наиболее быстро растущих в мире [Toth, Jóźków, 2016].

Среди ведущих организаций, осуществляющих ДЗЗ, следует выделить: AARS – Asian Association on Remote Sensing, Australian Bureau of Meteorology, BNSC – British National Space Centre, CNES – Centre National d'Etudes Spatial, CSA – Canadian Space Agency, China National Space Administration, DLR – Deutches Zentrum für Luft- und Raumfahrt, ESA – European Space Agency, ISRO – Indian Space Research Organization, JAXA – Japan Aerospace Exploration Agency, NASA – National Aeronautics and Space Administration, NOAA – National Oceanographic and Atmospheric Administration, NRO – National Reconnaissance Office, USGS – United States Geological Survey.

По своей природе ДЗЗ является динамической информационной средой с постоянно меняющейся ситуацией на поверхности Земли и спутниками, находящимися в непрерывном движении. В этом динамическом состоянии ДЗЗ выделяются стационарная и нестационарная части. К стационарной части относятся изменения, вызванные вращением Земли вокруг Солнца и своей оси. К нестационарной части относятся изменения, связанные с антропогенными факторами и влиянием природных процессов и экстремальных событий. Наличие нестационарной составляющей в наблюдаемых процессах затрудняет анализ ДЗЗ статистическими методами [Liu, 2015].

Используемые в ДЗЗ космические аппараты классифицируются по способу получения изображения, по области применения и по периоду повторного пролета. По способу получения изображения спутники делятся на [Liu, 2015]:

1) оптические (SPOT, LandSat, IKONOS),

2) микроволновые (TerraSAR-X, RADARSAT, Envisat),

3) многорежимные (MODIS).

По области применения аппараты различаются на [Liu, 2015]:

1) наземные (LandSat),

2) океанические (ERS),

3) метеорологические (MODIS).

Классификация по периоду повторного пролета выглядит так [Liu, 2015]:

1) непрерывные измерения по всей поверхности Земли (GOES),

2) короткий период повторного пролета (MODIS, WorldView, RapidEye),

3) длинный период повторного пролета (EO-1, LandSat).

Такая классификация космических аппаратов говорит о большом разнообразии ДЗЗ, то есть о характерном для БоД высоком значении V – Variety. В плане разнообразия (V – Variety) ДЗЗ можно говорить и о том, что данные поступают от раз-

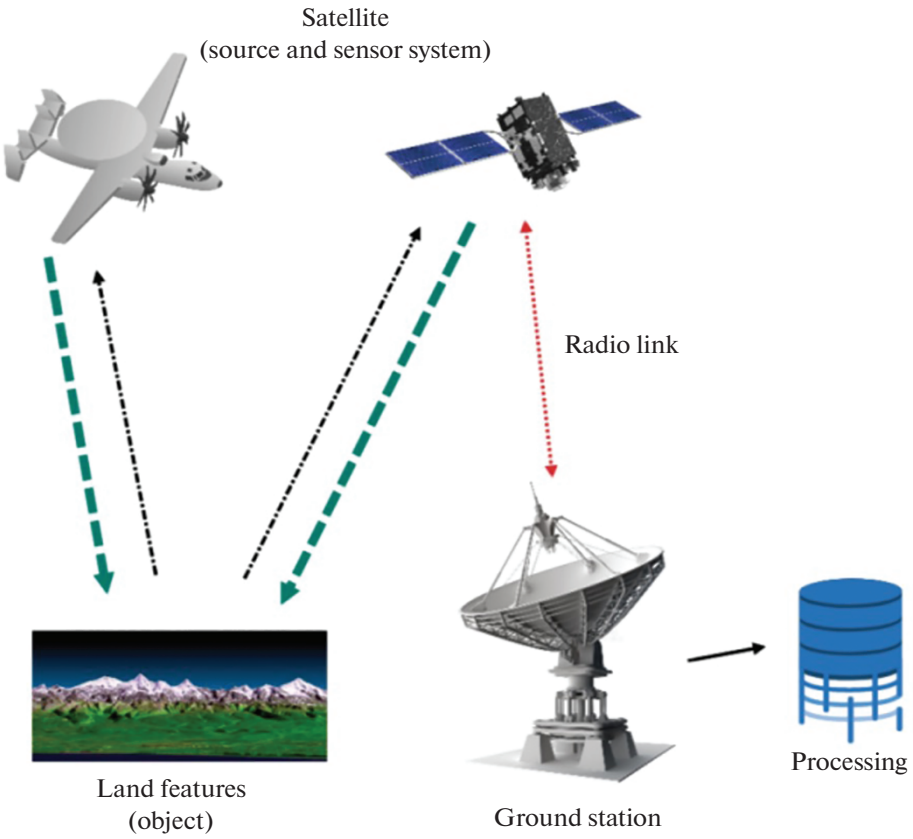


**Рис. 5.** Схема активного ДЗЗ.

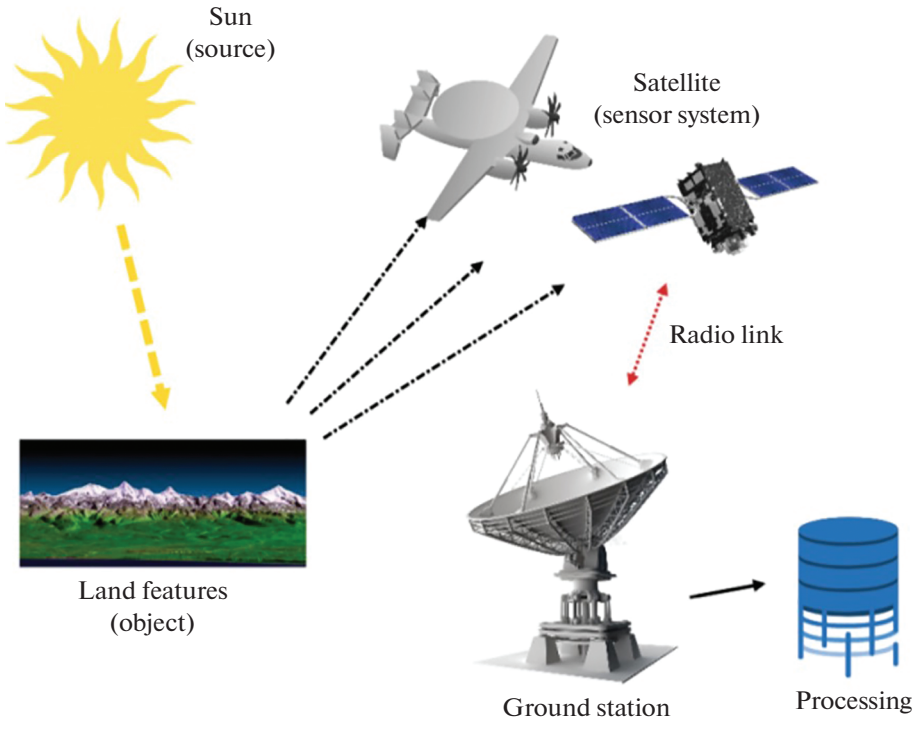


**Рис. 6.** Схема пассивного ДЗЗ.

личных приборов регистрации (лазеры, радары, оптические системы и др.). Кроме того, данные относятся к разным моментам времени и имеют различное пространственное разрешение. Информация ДЗЗ относится и к разным научным дисциплинам. Соответственно, интегрально ДЗЗ несомненно достигает характерного для БоД порога разнообразия в 1000 единовременно поступающих различных типов данных. Так, например, архивы ДЗЗ NASA включают в себя несколько тысяч различных типов данных [Chi et al., 2016].

Важной характеристикой БоД ДЗЗ является масштаб измерений. Он включает в себя разрешение, временной интервал, спектральный диапазон, телесный угол или направление поляризации [Wu, Li, 2009]. Пространственный масштаб определяет пространственное разрешение данных. Разномасштабность БоД ДЗЗ означает, что при работе с ними необходим выбор правильного масштаба и учет эффектов масштабирования [Liu, 2015].

По пространственному разрешению спутниковые данные можно также разделить на три группы [Liu, 2015]:

1) низкого разрешения (MODIS, Envisat),

2) среднего разрешения (Landsat, EO-1, Terra, RADARSAT),

3) высокого разрешения (QuickBird, IKONOS, WorldView).

Основными типами спутниковых данных с точки зрения механизма получения изображений являются оптические, микроволновые и данные в виде облака точек. Также можно выделить стереографические пары, полученные с нескольких фотографий и часто используемые для создания трехмерных или топографических карт. Использование информации, полученной с помощью различных механизмов наблюдений, позволяет анализировать информацию ДЗЗ с различных точек зрения. Это также требует выбора источника данных наиболее подходящего для решаемой задачи [Liu, 2015].

Для создания точных карт на основе данных дистанционного зондирования необходима трансформация, устраняющая геометрические искажения. Снимок поверхности Земли аппаратом, направленным точно вниз, содержит неискаженную картинку только в центре снимка. При смещении к краям расстояния между соответствующими точками на снимке и на Земле все более различаются. Такие искажения корректируются в процессе фотограмметрии. С начала 1990-х большинство коммерческих спутниковых изображений продается уже скорректированными [Konecny, 2014].

Кроме того, может требоваться радиометрическая или атмосферная коррекция ДЗЗ. Радиометрическая коррекция преобразует дискретные уровни сигнала, например от 0 до 255, в их истинные физические значения [Teillet, 1986]. Атмосферная коррекция устраняет спектральные искажения, внесенные наличием атмосферы [Hadjimitsis et al., 2004].

Частота пролета спутника имеет значение при исследовании серий изображений, например, при изучении динамики лесов [Potapov et al., 2015].

Благодаря функционированию перечисленных выше спутников и интеграции наблюдений ДЗЗ сегодня генерируются огромные объемы информации. RSGS (Remote Sensing Satellite Ground Station) – единственная большая приемная наземная станция ДЗЗ в Китае расположена в районе Миюньскоого водохранилища в округе Миюнь в 70 км от Пекина. Объемы хранения информации на подобных больших приемных станциях измеряются в терабайтах [Liu, 2015]. Исторические архивы данных отдельных стран, таких как Китай или США (NASA`s Earth Observing System Data and Information System – EODIS), измеряются в петабайтах [Chi et al., 2016; Liu, 2015], а глобальный архив данных ДЗЗ – в экзабайтах [Liu, 2015]. При этом такие объемы информации непрерывно растут и при комплексном аэрокосмическом мониторинге могут достигать ~0.5 терабайт в сутки и более [Бондур, 2014]. Экстраполируя оценки роста объемов данных ДЗЗ [Бондур, 2014] можно утверждать, что их объемы, полученные с помощью космических аппаратов, самолетов, вертолетов и беспилотников могут достигнуть к 2025 г. суммарных объемов порядка 300 экзабайт.

Таким образом, для ДЗЗ с большим запасом достигаются необходимые пороги БоД в аспекте V – Volume. Большие объемы данных ДЗЗ порождаются, главным образом, их спектральной и временной размерностями.

Наконец скорость (V – Velocity) в БоД ДЗЗ определяется как скоростью поступления самих данных, так и скоростью их обработки и анализа. Это особенно важно, например, в случае природных катастроф. Скорость поступления данных на индийскую Antarctica Ground Station for Earth Observation Satellites (AGEOS), расположенную в Антарктиде, составляет порядка 100 гигабайт в сутки, а скорости поступления данных в RSGS или EODIS измеряются уже в терабайтах в сутки [Liu, 2015].

Таким образом можно с уверенностью констатировать выполнение достаточных условий 3V (Volume, Variety, Velocity) у Больших Данных ДЗЗ. Структуру данных ДЗЗ, при этом, характеризует их разномасштабность, большая размерность и нелинейность изучаемых систем [Liu, 2015].

## БОЛЬШИЕ ДАННЫЕ МЕТЕОРОЛОГИЧЕСКИХ НАБЛЮДЕНИЙ

В целом ряде стран, включая Россию, метеорологию в ее теоретической части называют **физикой атмосферы** [Houghton, 2002]. Ряд крупных ученых метеорологов считают, что такое название точнее соответствует сегодняшнему статусу развития и задачам этой науки. Значительная часть физиков атмосферы занимается изучением погоды, климата и атмосферы.

Данные метеорологических измерений крайне разнообразны. **Большие значения (V – Variety) в этом случае не вызывают сомнений.** Их крупномасштабными **источниками являются:**

1. Информация радиолокационных и космических исследований атмосферных явлений (физическая метеорология) [Rauber, Nesbitt, 2018].

2. Данные изучения физических механизмов атмосферных процессов (динамическая метеорология) [Gordon et al., 2016].

3. Гигантские данные исследований закономерностей изменения и прогнозирования погоды (синоптическая метеорология) [Ahmad et al., 2017b].

4. Климатологическая информация [Miller, 2019].

5. Данные аэрологии (наука, изучающая верхние слои атмосферы до нескольких десятков километров от поверхности Земли).

**Дополнительные метеорологические данные** поступают от мониторинга и целевых исследовательских измерений, проводимых в рамках ряда прикладных разделов метеорологии. Среди них:

6. Авиационная метеорология [Gultepe et al., 2019].

7. Агрометеорология [Ahmad et al., 2017a].

8. Биометеорология (наука, изучающая влияние атмосферных процессов на человека и другие живые организмы) [Hondula et al., 2017].

9. Ядерная метеорология (научная дисциплина, изучающая естественную и искусственную радиоактивность, распространение в атмосфере радиоактивных примесей, влияние ядерных взрывов).

10. Радиометеорология (наука, изучающая распространение радиоволн в атмосфере) [Pavlikov et al., 2017].

11. Спутниковая метеорология [Pelton et al., 2013].

12. Лесная метеорология (акцент на изучение пожаров) [Moon et al., 2019].

Базовым инструментом, создающим метеорологические БоД, является **всемирная сеть метеостанций.** Метеостанцией называется учреждение, ведущее метеорологические наблюдения, обладающее метеоплощадкой, удовлетворяющей определенным требованиям, на которой установлены стандартные приборы для непрерывных или квазинепрерывных наблюдений за погодой и климатом по единой методике в определенной последовательности и передаче зарегистрированных данных в центр данных или их потребителям (http://meteorologist.ru/meteorologicheskaya-stantsiya.html).

**Классическая метеостанция** оснащена:

1) термометрами для измерения температуры воздуха и почвы на разных глубинах, термографом для непрерывной регистрации температуры воздуха,

2) барометром для измерения атмосферного давления, барографом для определения барометрической тенденции,

3) гигрометром для измерения влажности воздуха,

4) гигрографом для непрерывной регистрации влажности воздуха,

5) психрометром для измерения температуры и влажности воздуха,

6) флюгером для измерения скорости и направления ветра,

7) осадкомером для измерения осадков,

8) плювиографом для непрерывной регистрации жидких осадков в теплый период года,

9) гололедным станком для измерения гололедно-изморосевых отложений,

10) ледоскопом для определения измороси и инея,

11) рейкой для измерения высоты снежного покрова.

**Продвинутые метеостанции** дополнительно оснащаются:

12) трансмиссометром для измерения метеорологической оптической дальности,

13) облачным прожектором (облакомером) для измерения нижней границы облаков,

14) испаромером для измерения величины испарения с земной поверхности,

15) гелиографом для непрерывной регистрации солнечного сияния.

При расширении спектра наблюдений на метеостанциях возможно размещение и иных приборов.

Основным официальным метеостанциям мира присвоены **синоптические индексы**. В зависимости от установленного объема и качества наблюдений метеостанциям присваивается определенный разряд. Обширные потоки метеорологической информации также поступают и от метеорологических спутников.

Таким образом, информация, предоставляемая метеорологическими спутниками, тесно пересекается с данными ДЗЗ. Данные, производи-

мые спутниками этого типа, являются базовыми для прогноза погоды.

Метеорологические спутники оснащены приборами для наблюдения за температурой поверхности Земли, а также ее облачным, снеговым и ледовым покровом [Su et al., 2018; Zhao et al., 2017]. Методы получения информации с помощью метеоспутников и способы ее обработки изучает спутниковая метеорология. Добавление информации спутниковой метеорологии к гигантским объемам и разнообразию данных сетей метеостанций еще более увеличивает **V – Volume** и **V – Variety** метеорологической информации. Для примера, общий объем необработанных данных, накопленных только одной глобальной метеорологической сетью FLUXNET, давно исчисляется в петабайтах информации и состоит из более чем 200 регистрируемых параметров. Данная сеть включает микрометеорологические башни, которые используют вихревые ковариационные методы для измерения обмена углекислого газа, водяного пара и энергии между биосферой и атмосферой [Novick et al., 2018].

Метеоспутники вместе со станциями приема и обработки данных образуют международную метеорологическую космическую систему. В России эксплуатацией метеоспутников занимается организация "НИЦ "Планета", а в странах Европейского союза международная организация EUMETSAT. Высокие скорости передачи и обработки данных гарантируют наличие у метеорологической информации и третьего **V – Velocity**.

Метеорологические БоД дополняются целенаправленными потоками информации непосредственно из атмосферы, поступающих от метеозондов. Метеозонд – это беспилотный аэростат, предназначенный для изучения атмосферы. Он состоит из резиновой или пластиковой оболочки, наполненной водородом или гелием, и подвешенного к ней контейнера с аппаратурой.

Приборы метеозонда измеряют давление воздуха, влажность, температуру и другие параметры. Замеры перемещения аэростата позволяют определять скорость ветра на разных высотах. Информация, как правило, передается по радиоканалу. До внедрения радио на метеозондах устанавливали метеорографы, которые нужно было возвращать на землю. Если шар запускают только для измерения скорости ветра, то его называют "шар-пилот".

Высотные метеозонды могут достигать высот 30–40 км. Рекорд высоты для ультратонкого полиэтиленового метеозонда с диаметром 60 м составляет 53.7 км (20.09.2013 г., Япония, Тихий океан).

Для получения метеоданных в верхних слоях атмосферы используются также и **метеорологические ракеты.**

Метеорологическая информация интегрируется в мировом масштабе в целях глобального прогноза погоды [Kawasaki et al., 2017; Thorne et al., 2017]. Это с большим запасом обеспечивает необходимые для БоД 3V – Volume, Variaty и Velocity. **Центром интеграции является находящаяся в Женеве штаб-квартира Всемирной Метеорологической организации (ВМО/World Meteorological Organization/WMO).** ВМО была основана в 1950 г., а в 1951 г. стала специализированным учреждением ООН, занимающимся вопросами состояния атмосферы Земли, океанов, климата [History of WMO, 2020]. В 2019 г. членами ВМО являлись 187 государств и 6 территорий. На рис. 7 представлена схема глобальной наблюдательной системы ВМО. Для сбора, обмена и распространения данных наблюдений используется глобальная телекоммуникационная система, схема которой представлена на рис. 8.

## БОЛЬШИЕ ДАННЫЕ СЕЙСМИЧЕСКОГО МОНИТОРИНГА

Сейсмический мониторинг территории заключается в регистрации, обработке и анализе сейсмических сигналов. Он является неотъемлемой частью систем обеспечения безопасности ответственных сооружений (атомных электростанций, гидротехнических сооружений, скважин, шахт, мостов и др.). Сейсмический мониторинг включает в себя не только регистрацию и дальнейшую оперативную обработку полученной информации, но и интерпретацию сейсмологических данных с выходом на прогнозные оценки. Бурная современная активизация сейсмического мониторинга и сейсмологии в целом связана с "взрывным" развитием в 21 веке глобальных сейсмических сетей.

Первой глобальной сейсмической сетью, начавшей функционировать в 1960-х годах и включавшей в себя около 120 станций, была The World-Wide Standardized Seismograph Network (WWSSN). В результате ее работы был получен беспрецедентный набор высококачественных сейсмических данных. Эти данные позволили прояснить механизмы очагов землетрясений и структуру земной коры [Amnion et al., 2010], а также способствовали развитию теории тектоники плит [Richards, Zavales, 1996].

В конце 1970-х 13 станций WWSSN были дооборудованы цифровыми регистраторами. Эти станции получили название DWWSSN и стали частью Глобальной цифровой сейсмографической сети (GSDN). Современным преемником WWSSN является Глобальная сейсмографическая сеть (GSN), управляемая Объединенными исследовательскими институтами сейсмологии (IRIS). В то же время на территории СССР была создана Единая Система Сейсмических Наблюдений (ЕССН),

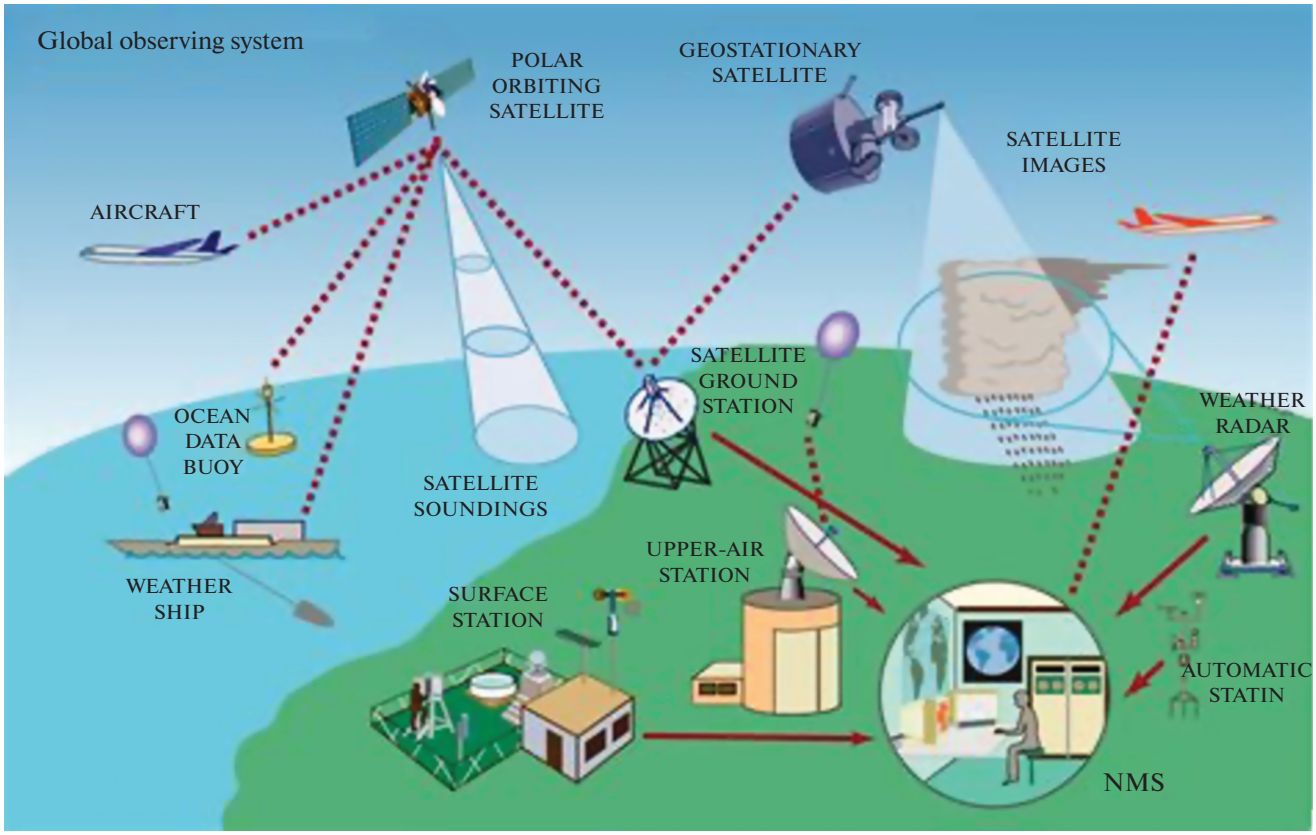


**Рис. 7.** Схема глобальной наблюдательной системы ВМО [Global..., 2016].

которая стала основой современной Геофизической службы Российской академии наук.

Среди ведущих зарубежных международных организаций, осуществляющих сейсмический мониторинг, следует отметить следующие: NIED – National Research Institute for Earth Science and Disaster Prevention, CEA – China Earthquake Administration, USGS – United States Geological Survey, ANSS – The Advanced National Seismic System, IRIS – Incorporated Research Institutions for Seismology, CTBTO – The Comprehensive Nuclear-Test-Ban Treaty Organization, CEA – Le Commissariat à l'énergie atomique et aux énergies alternatives, BGR – Bundesanstalt für Geowissenschaften und Rohstoffe, GSC – Geological Survey of Canada, EMSC – European-Mediterranean Seismological Centre, SED – Swiss Seismological Service и другие.

В России сейсмический мониторинг осуществляется Федеральным исследовательским центром "Единая геофизическая служба Российской академии наук" (ФИЦ ЕГС РАН). Ее наблюдательные сети объединяют около 450 сейсмических станций. Доступ к каталогам землетрясений предоставляется через сайт ФИЦ ЕГС РАН (www.gsras.ru/new/catalog). В состав глобальной сейсмометрической сети GSN входит 12 станций ФИЦ ЕГС РАН.

В зависимости от размеров охватываемой территории сейсмический мониторинг подразделяется на глобальный, региональный и локальный. На глобальном уровне можно выделить следующие сети: CBTBO, GSN, GEOSCOPE. Благодаря тесному международному сотрудничеству за последние десятилетия наметился определенный процесс глобализации, выражающийся в том, что одни и те же станции входят в состав разных сетей. Например, почти 30 станций GSN одновременно входят в состав сети CBTBO. На рис. 9 приведены схемы расположения станций GSN и GEOSCOPE.

По уровню организации наиболее развитой является система мониторинга CTBTO (Организация Договора о всеобъемлющем запрещении ядерных испытаний), состоящая из следующих элементов:

1. IMS – Международная система мониторинга.
2. IDC – Международный центр данных.
3. GCI – Инфраструктура глобальной связи.

IMS состоит из 50 базовых и 120 дополнительных сейсмических станций, 60 инфразвуковых станций, 11 гидроакустических, 80 радионуклидных, 79 станций мониторинга благородных газов на радионуклидных станциях и 16 радионуклид-

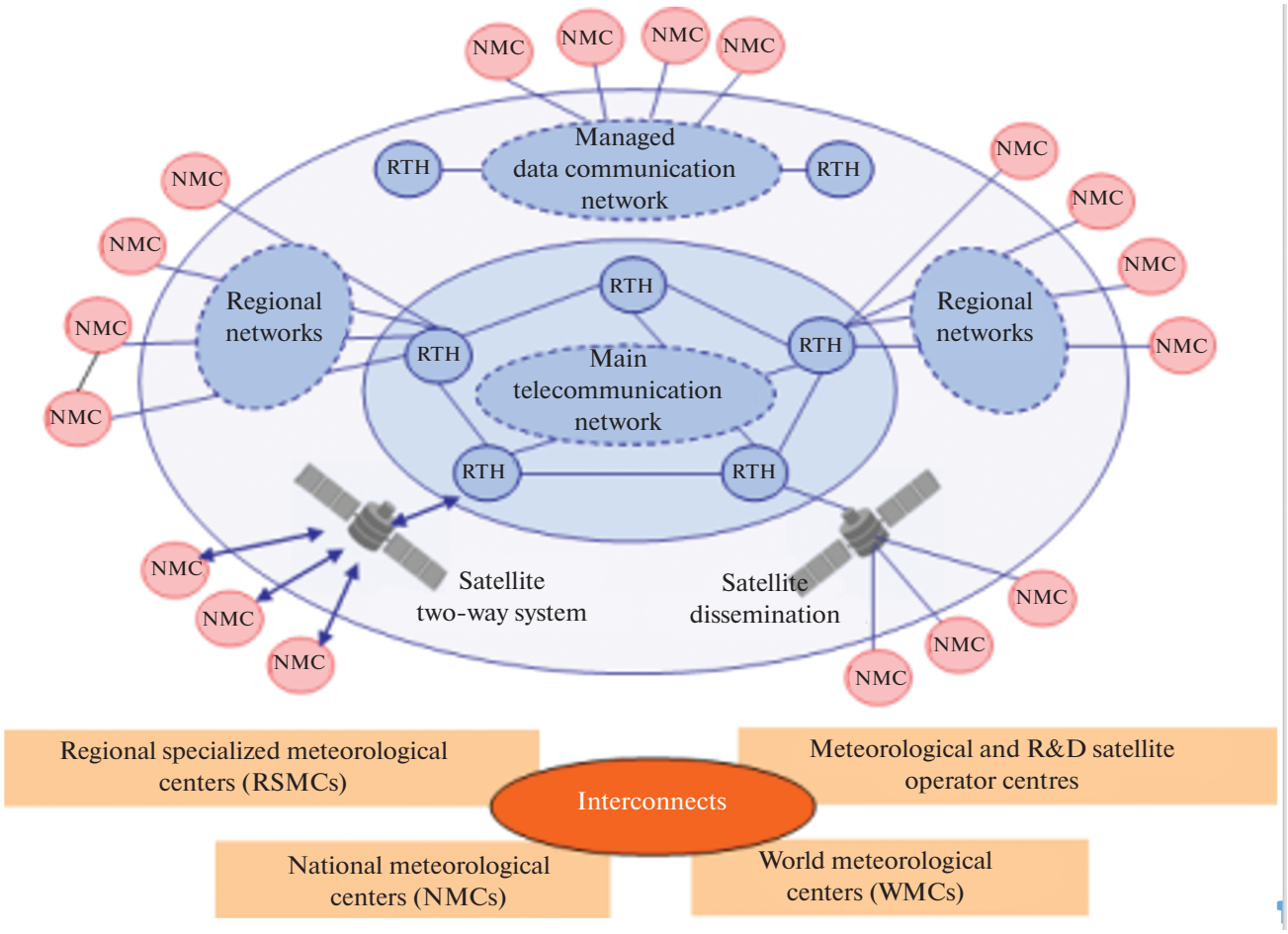


**Рис. 8.** Схема глобальной телекоммуникационной системы ВМО [GTS..., 2020].

ных лабораторий. Система предназначена для обнаружения любого ядерного взрыва, проводимого на земле, под землей, под водой или в атмосфере.

IMS связана с Международным центром данных (IDC), расположенным в штаб-квартире СТВТО в Вене, Австрия. IDC обрабатывает и анализирует данные станций мониторинга и выпускает бюллетени для государств-членов организации.

Инфраструктура глобальной связи (GCI) используется для передачи данных IMS в IDC. Необработанные данные и бюллетени от IDC передаются государствам-членам СТВТО. GCI обеспечивает прием и распространение данных через сеть из шести спутников на три наземных хаба, а затем в IDC.

Станции глобальных сейсмических сетей оснащены, в основном, широкополосными сейсмоприемниками и ведут непрерывную запись. Это создавало трудности с передачей и хранением сейсмограмм. Для оптимизации работы сейсмических сетей и обмена цифровыми данными о землетрясениях был создан формат SEED (Standard for the Exchange of Earthquake Data). Данные в формате SEED сводят к минимуму занимаемое пространство для хранения и передачи. Это дает дополнительное преимущество: можно получить доступ к данным с меньшим количеством запросов на ввод. При этом система индексов (перекрестных ссылок) логических записей обеспечивает эффективный доступ к данным [SEED Reference Manual..., 2012].

Объемы данных сейсмического мониторинга рассмотрим на примере Центра управления данными IRIS. Он аккумулирует основную часть сейсмических данных со всего мира. Данные состоят из временных рядов (сейсмограммы), информации о сейсмических событиях, метаданных (схемы расположения станций и настройки оборудования) и исторических данных (данные аналоговых источников). Станции глобальных сетей ведут непрерывную запись, остальные функционируют, большей частью, в триггерном режиме. Данные от стационарных станций поступают и архивируются в формате SEED. Отдельные данные, поступающие от временных станций, имеют форматы отличные от SEED. Сеть из 100 станций

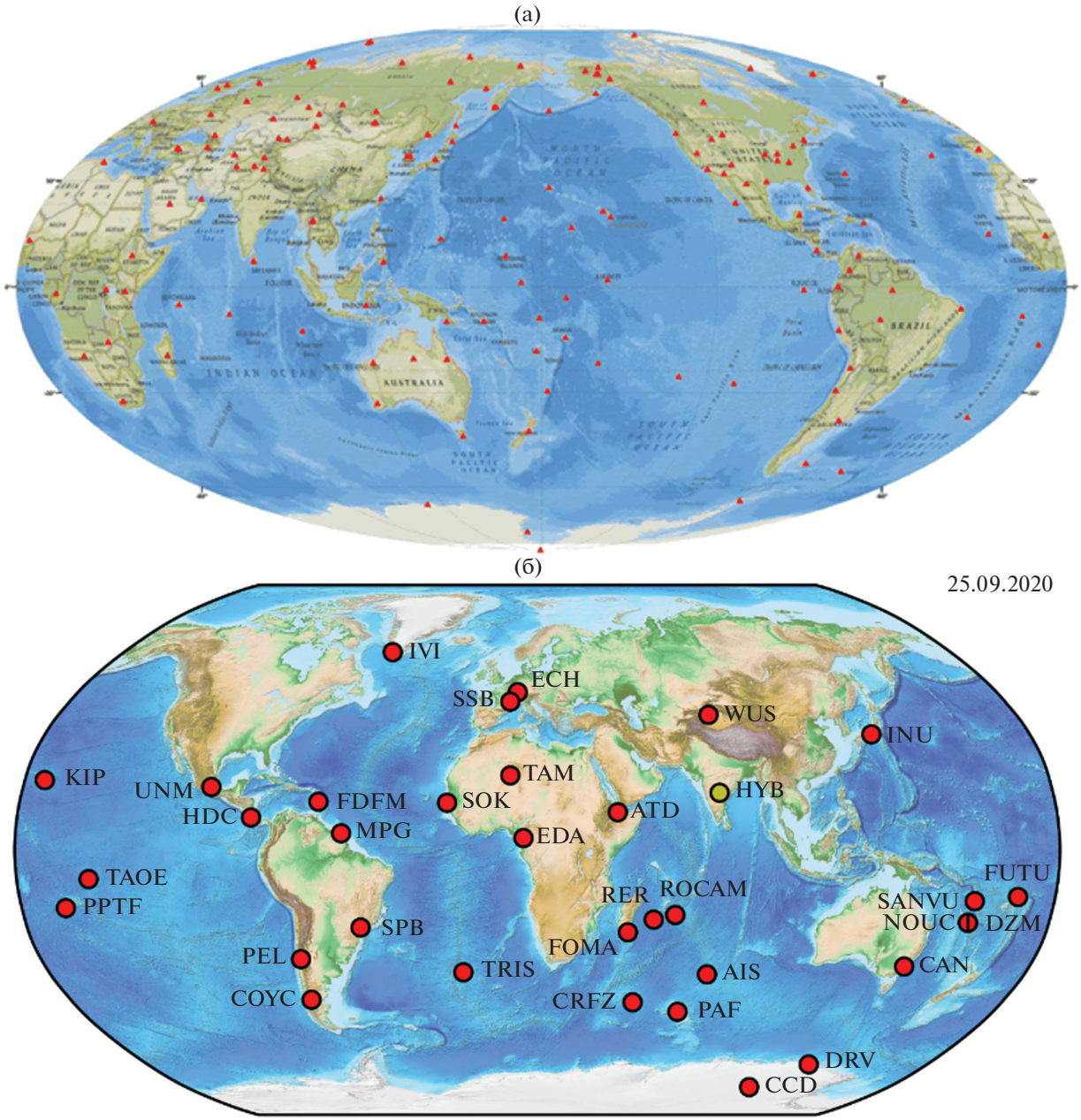


**Рис. 9.** Схемы расположения станций сетей (а) GSN (www.iris.edu) и (б) GEOSCOPE (http://geoscope.ipgp.fr/index.php/en/).

может производить до 2 гигабит данных в формате SEED в сутки [Newman, 2006]. Исходя из этого, можно сделать вывод, что средний объем данных такой сети, как GSN в сутки составит всего 375 мегабайт или 136 гигабайт в год.

Динамика роста объемов данных, поступающих в Центр, приведена на рис. 10. Общий объем накопленных данных составляет 687 тебибайт, что эквивалентно 755 терабайтам. Необходимо отметить, что это объем сейсмологических данных с более чем 8500 станций (в том числе инженерных сетей на территории США и по всему миру). Учитывая темпы роста объемов данных сейсмического мониторинга можно заключить, что достижение ими размера в петабайт вопрос ближайшего времени. Таким образом, с определенной долей условности можно сделать заключение об их соответствии характеристике БоД V – Volume.

Данные большинства сетей передаются и обрабатываются в режиме времени, близком к реальному. Это говорит о соответствии характеристике БоД V – Velocity.

Сейсмические данные обладают ограниченным набором параметров, что не соотносится с

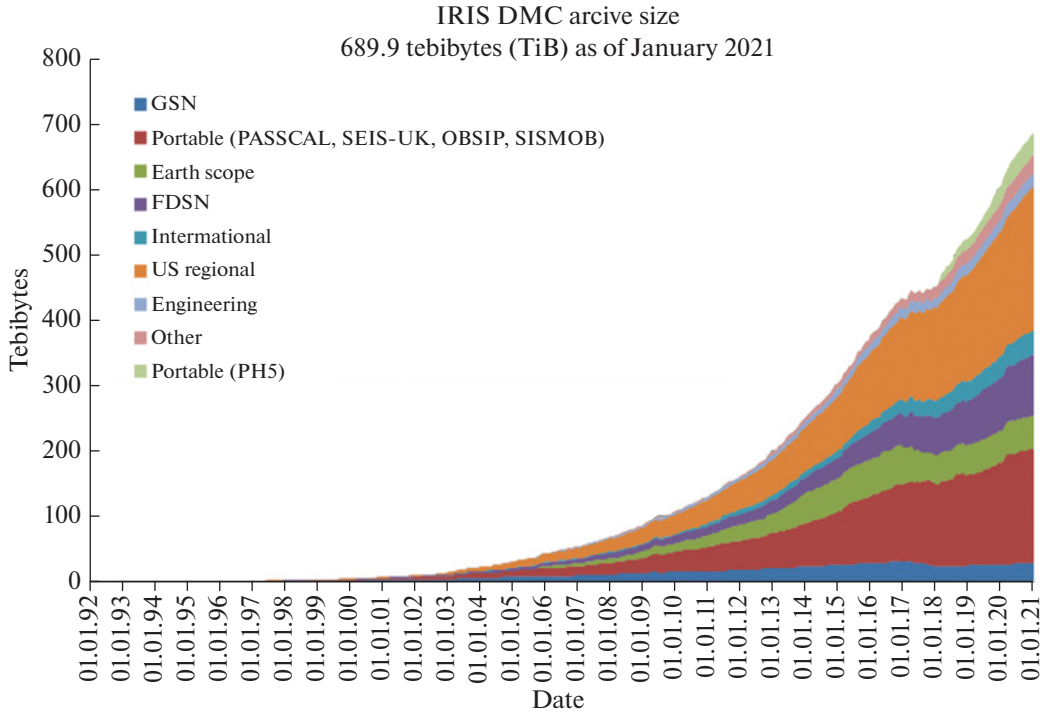


**Рис. 10.** Динамика роста объемов сейсмических данных, поступающих в Центр управления данными IRIS (www.iris.edu).

характеристикой БоД V – Variety. СТВТО использует в своей работе сейсмические данные в комплексе с данными инфразвуковых, гидроакустических и радионуклидных станций, включая станции мониторинга благородных газов. Кроме того, известно, что отдельные сейсмические станции сети GSN фактически переоборудуются в геофизические обсерватории, оснащенные микробарографами, анемометрами, магнитометрами и приемниками GPS. Это позволяет сделать вывод, что данные сейсмического мониторинга обладают потенциалом к разнообразию. Однако в данном случае уже более корректно будет говорить не о сейсмическом, а о геофизическом мониторинге.

В настоящее время сейсмические данные стремительно развиваются с точки зрения их качества и достоверности. Если рассматривать сейсмический мониторинг отдельно от других геофизических наблюдений, то можно констатировать, что его сети постоянно расширяются, увеличивается скорость передачи в центры сбора данных и частота дискретизации наблюдений. Повышается общий уровень автоматизации сейсмических сетей. Однако это не происходит такими темпами, которые позволили бы нам говорить о переходе данных сейсмического мониторинга в категорию БоД.

Таким образом, на данном этапе трудно констатировать выполнение достаточных условий 3V (Volume, Variety, Velocity) отдельно для данных сейсмического мониторинга. Но в случае их комплексного использования с другими геофизическими данными ситуация кардинально меняется, что было проиллюстрировано на примере эволюции сетей СТВТО и GSN.

## БОЛЬШИЕ ДАННЫЕ СЕЙСМИЧЕСКОЙ РАЗВЕДКИ

Сейсмическая разведка или сейсморазведка (СР) является одним из ведущих геофизических методов исследования структуры, строения и состава земных недр. Она основана на регистрации искусственно возбуждаемых упругих волн, которые отражаются или преломляются от различных пластов горных пород. СР обычно используется в качестве эффективного средства при поиске и разведке залежей нефти, газа и твердых полезных ископаемых. Она также используется в инженерных, археологических и других научных исследованиях. СР применяется для изучения упругих свойств геологической среды.

Сейсморазведка наиболее эффективна при исследованиях недр Земли на глубинах до 10 км. Для этого интервала глубин СР обладает разрешающей способностью до десятков метров. Поскольку эта область недр Земли включает почти все запасы нефти и газа, СР играет важную роль в энергетической отрасли [Talwani, Kessinger, 2003].

Развитие методов сейсморазведки в нефтегазовой отрасли началось в 1920-х годах. В последующие десятилетия произошел значительный прогресс в методологии благодаря улучшению регистрирующей аппаратуры и вычислительного оборудования. Позже двумерные сейсмограммы были заменены детализированными трех- и четырехмерными [Thomas, Hoover, 2021]. Основными форматами сейсморазведочных данных являются SEG-D и SEG-Y, которые, благодаря бинарной структуре, позволяют минимизировать объемы хранимой и передаваемой информации [Barry et al., 1975].

Данные СР в нефтегазовой отрасли представляют собой статическую информацию о структуре коллектора: горизонтальная протяженность, толщина, разломы, пористость и др. Покадровая 4D-сейсморазведка, полученная в разные моменты времени, отображает изменения в состоянии коллектора. Непрерывный сбор и обработка сейсмических данных в 4D является проблемой. Кроме того, данный тип исследований связан с компьютерной визуализацией и генерацией больших объемов данных.

Характеристика БоД V – Volume проявляется в различных секторах нефтегазовой отрасли, таких как разведка, бурение и добыча. СР генерирует большие объемы данных, используемые для создания 2D-, 3D- и 4D-изображений подземных слоев.

Рост объемов данных СР рассмотрим на пример компании CGG (Compagnie Générale de Géophysique), которая первой в 1971 г. осуществила исследования в области 3D сейсморазведки. В период с 2005 по 2009 гг. объем данных СР компании CGG вырос со 100 гигабайт до 2 терабайт. В 2009 г. в результате 3D СР Мексиканского залива методом StagSeis было собрано более 1.5 петабайт данных. Поток данных СР компании составлял от 200 до 400 мегабайт в секунду, а ее крупнейший вычислительный центр обрабатывает более 100 петабайт активных данных в сутки [Boman, 2015].

Скорость передачи и обработки данных, характеристика БоД V – Velocity, является наиболее актуальной проблемой для нефтегазовой отрасли. Обработка огромного количества данных СР даже для крупных компаний часто становится трудно разрешимой задачей. Во многих случаях в нефтегазовой отрасли критически важна обработка данных в реальном времени. Задача интерпретации сейсмических данных требует значительных вычислительных мощностей и возможностей визуализации.

Характеристика БоД V – Variety в СР относится к числу качественно различных типов данных, которые генерируются, хранятся и анализируются. Информация, регистрируемая различными датчиками, имеет разные объемы и форматы. Последние могут быть численными, текстовыми, графическими, аудио или видео.

Напомним, что БоД бывают структурированные и неструктурированные. При этом до 90% БоД неструктурированные [Ishwarappa, Anuradha, 2015]. Отметим, что большая часть данных СР в нефтегазовой отрасли является структурированной [Rezai et al., 2017]. Источниками неструктурированных данных являются журналы скважин, ежедневные письменные отчеты о бурении и цифровые чертежи [Feblowitz, 2013].

Примером реализации БоД в сейсморазведке является работа [Roden, 2016], в которой авторы разработали метод многокомпонентного сейсмического анализа, проводимого в пять этапов. На первом этапе ставится геологическая задача; на втором этапе определяются ключевые атрибуты; на третьем – происходит анализ нейронной сетью с использованием инструментов машинного обучения; на четвертом – полученные результаты дополнительно анализируются с помощью двухмерных карт для определения важных геологических особенностей; наконец, на пятом этапе производится уточнение результатов с учетом различных атрибутов и сценариев обучения.

Рассмотрим сервис для управления данными Katalyst (www.katalystdm.com), представляющий собой интегрированное решение для управления геологическими данными в нефтегазовой отрасли. Пользователям предоставляется мультиоблачное решение для хранения данных, позволяющее легко получить доступ к архивам для анализа БоД СР. Услуги включают полный цикл управления для архивов сейсмических и скважинных данных глобальным центром обработки.

Основываясь на вышеизложенном, можно с уверенностью констатировать, что уже сегодня данные СР соответствуют характеристикам “трех V” и тем самым являются БоД. Базой этого, в первую очередь, является бурное развитие приборов и программного обеспечения сейсморазведочных исследований. Сегодня мы уже имеем большое число исследований, посвященных разным решениям в области БоД сейсморазведки. На примере Katalyst мы видим, что существуют коммерческие организации, предоставляющие свои услуги для анализа БоД СР. Эти услуги включают предоставление вычислительных мощностей и облачных хранилищ для данных нефтегазовых компаний.

## БОЛЬШИЕ ДАННЫЕ ГЕОЭКОЛОГИЧЕСКИХ НАБЛЮДЕНИЙ

Ярким примером БоД в науках о Земле является информация, производимая международной сетью станций SMEAR (Station for Measuring Earth surface – Atmosphere Relations). Этот ресурс также

связан с метеорологическими данными и, в смысле V – Variety, расширяет последние. Сеть SMEAR задумывалась в финском институте INAR (Institute for Atmospheric and Earth System Research, University of Helsinki, Finland) как старт глобальной сети обсерваторий для изучения взаимодействия земной поверхности и атмосферы [Hari et al., 2016]. Сегодня план INAR успешно превращается в жизнь.

Необходимость развертывания такой **глобальной геоэкологической** сети объясняется все более возрастающим изменением климата в сторону глобального потепления, в частности, в результате деятельности человека.

Для анализа последствий применения экологических мер регулирования необходимо иметь адекватные количественные оценки соответствующих геоэкологических процессов. Данные, производимые в настоящий момент существующими международными геоэкологическими сетями наблюдений ICOS (Integrated Carbon Observation System), ANAEE (Infrastructure for Analysis and Experimentation on Ecosystems), ACTRIS (Aerosols, Clouds, and Trace gases Research Infra-Structure Network), GOOS (Global Ocean Observing System), IASOA (International Arctic Systems for Observing the Atmosphere), LifeWatch, SIOS (Svalbard Integrated Observation System), InGOS (Integrated Non-CO2 Greenhouse Gas Observing System), EXPEER (Distributed infrastructure for EXPErimentation in Ecosystem Research), оказываются недостаточными. Это относится как к полноте проводимых измерений, так и к географическому покрытию поверхности Земли.

Известные финские метеорологи Перти Хари (Pertti Hari) и Марку Кулмала (Markku Kulmala) [Hari et al., 2016] разработали принципиально новую концепцию глобальной иерархической геоэкологической сети GlobalSMEAR. Данные, производимые этой сетью, представляют собой яркий пример БоД.

Уникальное измерительное оборудование существующих станций и эффективный менеджмент сбора, хранения и предварительной обработки информации явились основой для достижения впечатляющих научных результатов. По результатам научного анализа БоД SMEAR было опубликовано 2500 статей в реферируемых журналах (из них 45 в Nature и Science). В частности, атмосферные наблюдения SMEAR показали, что повышенный уровень загрязнения влияет на погодные условия [Ding et al., 2013; Zilitinkevich et al., 2013].

Развертываемая **иерархическая сеть SMEAR** включает в себя флагманские, продвинутые и стандартные станции. На **флагманских станциях** устанавливается оборудование, позволяющее изучать большинство основных экосистем Земли. **Продвинутые станции** измеряют параметры потоков протекающих процессов. На нижнем уровне находятся **стандартные станции,** выполняющие базовый набор измерений.

Планируемое в проекте количество стандартных станций SMEAR составляет порядка 10 тысяч, продвинутых – около 1000, флагманских – порядка 50. В настоящее время количество флагманских станций измеряется единицами. Для создания полномасштабной сети планируется использовать, в том числе, станции существующих систем наблюдений с их дооснащением при необходимости новым, дополнительным оборудованием. В проектируемом виде сеть SMEAR будет обеспечивать глобальное измерение и накопление практически всех параметров, интересующих сегодня геоэкологов.

На сегодня флагманской обсерваторией сети является SMEAR II, Hyytiala, Finland, производящая более 1250 различных типов измерений (рис. 11). Основные из них будут разобраны ниже. К продвинутым станциям относятся три обсерватории в Финляндии: SMEAR I, Värriö Lapland, 1990; SMEAR III, Urban, Helsinki, Finland, 2004; SMEAR IV, Puijo, 2008; а также одна обсерватория в Эстонии и две в Китае: SMEAR-Estonia, Järviselja, 2011; SORPES в Нанкине; SMEAR-Beijing в Пекине. Последняя обсерватория по своим функциям достаточно близка к статусу флагманской.

Планируется, что GlobalSMEAR будет охватывать практически все основные экосистемы Земли. Список параметров, измеряемых на той или иной флагманской станции, определяется тем на какую конкретную экосистему эта станция нацелена. **Экологической системой** или **экосистемой** называют любое единство, включающее все организмы на данном участке и взаимодействующее с физической средой таким образом, что поток энергии создает четко определенную трофическую структуру, видовое разнообразие и круговорот веществ внутри системы [Одум, 1975].

Остановимся ниже на трех крупных направлениях наблюдений сети SMEAR. Из приведенного ниже их анализа следует, что для информационных массивов GlobalSMEAR выполняются все три необходимых V-условия БоД. В данном случае не вызывают сомнений ни гигантские объемы данных, ни их огромное разнообразие, ни оперативная скорость получения информации в штабквартире SMEAR в INAR, расположенной в университете Хельсинки.

Для **лесной экосистемы** стандартные станции описывают породы деревьев, их диаметр, высоту и количество, а также измеряют стандартные почвенные параметры. На продвинутой станции к этому добавляются: измерения потоков углекислого газа, воды и тепла между экосистемой и атмосферой, ретроспективные измерения разви-

тия лесонасаждений. Флагманские станции значительно расширяют набор регистрируемых параметров. Соответственно, быстро растут V – Volume и V – Variety. Это усиливается и тем, что информация собирается в радиусе до 1000 километров вокруг станции. Таким образом, флагманская станция собирает разнообразные данные по довольно большой площади.

Среди экологических параметров, регистрируемых на флагманских станциях: массы и концентрации протеинов, целлюлозы, лигнина, крахмала и жиров в компонентах растительности, концентрации протеинов, целлюлозы, лигнина, крахмала и жиров в почвенных слоях, содержание углерода, азота и кислорода в растительности и почвенных слоях, измерения обмена углекислого газа из почвы, запасы воды в почве, выпадение осадков, поверхностный сток воды и концентрации растворенных органических и неорганических карбонатов в стоке, перечень характеристик животных, млекопитающих, птиц, насекомых в окружающей местности.

Объем и разнообразие данных сети SMEAR также увеличиваются за счет ориентации на **изучение экосистемы морей и океанов** [Vihma et al., 2019]. Стандартные и продвинутые станции в этом случае, как правило, базируются на буях, мигрирующих на поверхности льда или в открытом море. Носителями стандартных станций также могут быть воздушные дроны или подводные глайдеры.

Стандартные станции измеряют атмосферное давление и температурные профили. Измерения ведутся, начиная с поверхности воды, включая лед и снег, и заканчивая воздухом, с вертикальным разрешением 2 см. Одновременно стандартные станции оснащаются приемниками ГНСС GPS/ГЛОНАСС/GALILEO, которые обеспечивают определение для дрейфующих буев траектории дрейфа льда и направления океанических течений.

На продвинутых станциях к этому добавляются: профили температуры и ветра в нижних метрах атмосферы, профили температуры, солености и направления течений в верхних десятках или сотнях метров морской воды, компоненты солнечной коротковолновой и тепловой длинноволновой радиации, потоки водяных паров и остаточных газов, концентрации углекислого газа в воде и воздухе.

Как стандартные, так и продвинутые станции работают в автоматическом или телеуправляемом режимах без присутствия людей. Они мигрируют в зоне контроля обслуживающего их судна. В то же время флагманские обсерватории функционируют на побережье материка или на островах Северного-Ледовитого океана. Последние требуют непрерывного присутствия персонала на станциях.

**Рис. 11.** Станции SMEAR [SMEAR, 2021].

Развитие SMEAR в сторону изучения экосистемы морей и океанов должно привести к созданию “Высшей наблюдательной системы морской Арктики” (Advanced Observation System for the Marine Arctic). Как и весь SMEAR, сегодня эта система находится в стадии планирования и одновременно быстрого развития. Ключевые роли в этом процессе играют российские и финские океанологи, географы и экологи.

Развиваясь как существенная часть SMEAR, “Высшая наблюдательная система морской Арктики” входит одновременно и в другую крупную международную программу PEEX (Pan-Eurasian Experiment) [Vihma et al., 2019] (рис. 12).

**Атмосферные наблюдения – третье направление сети SMEAR**, вносящее большой вклад в статус БоД этого информационного ресурса. Изучение атмосферы перекликается с классическим метеорологическим мониторингом, который сам по себе является источником БоД.

Метеорологический блок SMEAR имеет свою специфику. Он ориентирован на изучение аэрозолей. На продвинутой станции измеряются распределения аэрозолей по количеству и размерам, тепловые потоки, а также потоки остаточных газов и летучих органических соединений.

На флагманской станции, помимо этого, измеряются химический состав аэрозолей, характеристики их вертикального профиля, распреде-

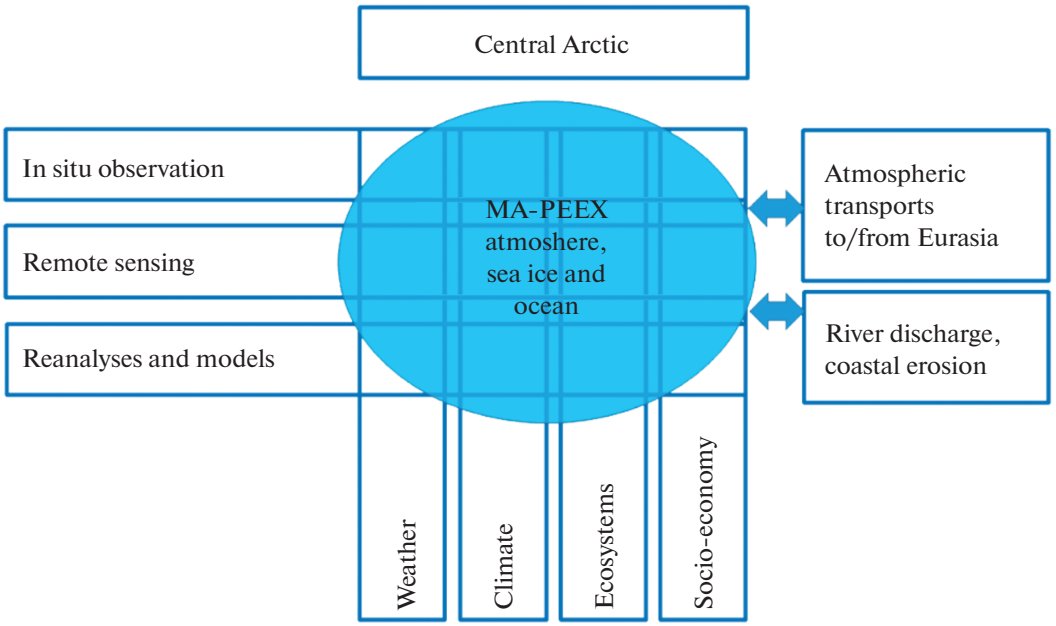


**Рис. 12.** Морская арктическая компонента PEEX [Vihma et al., 2019].

ление ионов в атмосфере, характеристики концентраций остаточных газов и оксидантов и параметры атмосферной турбулентности, а также потоков остаточных газов и аэрозолей на разных высотах. Особая зона внимания флагманских станций SMEAR – изучение характеристик облаков, спектральной зависимости солнечной радиации, поглощенной и отраженной радиации.

Иерархическое увеличение количества измеряемых параметров с ростом ранга станции сохраняется и для других наблюдаемых экосистем, например, степей. На флагманских станциях количество видов выполняемых измерений, как правило, превышает 1000.

Частота измерений определяется временной изменчивостью наблюдаемых процессов. Например, процессы, протекающие в почве, во многом определяются температурой. Последняя имеет достаточно стабильные значения. Поэтому для наблюдения за такими процессами достаточно ежечасных измерений. В то же время для процессов, протекающих в атмосфере, таких как движение облаков, характерны гораздо более высокие скорости. Здесь уже нужны ежеминутные измерения.

Система SMEAR направлена и на изучение целого ряда других экосистем и отдельных экологических явлений. Высокая частота регистрации в сочетании с большим количеством измеряемых параметров и существенным количеством станций приводит к огромным объемам данных (V – Volume). Оперативный доступ к существующим станциям SMEAR осуществляется через приложение SmartSMEAR (https://avaa.tdata.fi/web/smart). Этот инструмент обеспечивает высокую скорость передачи данных (V – Velocity) как в INAR, так и пользователям, а также их предварительной обработки.

Наконец, исключительное разнообразие (V – Variety) информации SMEAR обусловлено широким спектром научных дисциплин, данные которых регистрируются на станциях, а также иерархической структурой самой системы. Таким образом, данные экологической измерительной сети SMEAR, концентрируемые в INAR и на других флагманских станциях, имеют характерные для БоД большие значения всех 3V (Volume, Velocity, Variety), а значит являются БоД.

## ГЕОМАГНИТНАЯ ИНФОРМАЦИЯ И ФОРМАЛИЗМ БоД. ГЕОМАГНИТНЫЕ НАБЛЮДЕНИЯ: ОБСЕРВАТОРИИ, МАГНИТОВАРИАЦИОННЫЕ СТАНЦИИ, МОРСКИЕ И АЭРОМАГНИТНЫЕ ИССЛЕДОВАНИЯ

Необходимость постоянных геомагнитных наблюдений обусловлена непрерывными изменениями **Магнитного Поля Земли (МПЗ)** во времени и пространстве. Эти наблюдения проводятся с целью получения исходных данных для решения широкого круга задач солнечно-земной и космической физики, радиофизики, физики Земли, экологии, геологии, геодезии и разведки полезных ископаемых [Гвишиани, Лукьянова, 2018; Гвишиани и др., 2019а].

**Геомагнитные данные** получают путем измерений в различных точках Земли полного вектора геомагнитного поля **B** и трех его ортогональных составляющих. Наблюдения проводятся при по-

мощи магнитометрической аппаратуры наземного (полномасштабные магнитные обсерватории и магнитовариационные станции) и космического (низкоорбитальные спутники) базирования [Hulot et al., 2010].

Дополнительным источником данных являются измерения литосферного магнитного поля. В масштабах от 250 до 3000 км оно может быть картировано на глобальном уровне по данным низкоорбитальных спутников [Hulot et al., 2010]. В масштабах меньше 250 км литосферное поле становится слишком малым, чтобы его можно было обнаружить на высоте пролета спутника. В этом случае используются морские либо аэромагнитные исследования [Гвишиани, Лукьянова, 2018; Гвишиани и др., 2019а].

**Полномасштабная наземная геомагнитная обсерватория** регистрирует полную напряженность вектора геомагнитного поля и вариации его компонент, а также регулярно проводит абсолютные геомагнитные наблюдения. Продукция обсерватории – суточные файлы секундных или минутных значений вариаций геомагнитного поля, а также данные абсолютных измерений [Hulot et al., 2010].

Важнейшей задачей является достижение максимальной точности наблюдений. Также необходимо обеспечить непрерывность регистрации вариаций МПЗ. Пропуски в геомагнитных данных являются безвозвратными и могут привести к потере особо важной информации для объяснения неожиданных геофизических событий или техногенных катастроф. Еще одной характеристикой высококачественных геомагнитных наблюдений является оперативность сбора информации.

В связи с большими величинами индукции МПЗ (40000–70000 нТл) и необходимостью измерять его вариации с большой точностью (до 0.1 нТл) современные магнитные обсерватории ведут регистрацию с помощью минимум трех типов магнитометров [Rasson, 2007; Jankowsky, Sucksdorf, 1996]. **Измерения полного вектора МПЗ (абсолютные измерения)** выполняются протонным магнитометром или феррозондовым инклинометром/деклинометром. Эти приборы обеспечивают требуемую точность. При этом цикл измерений занимает 30–40 мин. и требует ручных операций. Такие измерения на обсерваториях выполняются один–два раза в неделю [Hulot et al., 2010].

**Скалярные измерения (измерения модуля)** МПЗ выполняются с помощью протонного прецессионного магнитометра, магнитометра Оверхаузера или магнитометра с оптической накачкой. Скалярные магнитометры могут быть изготовлены весьма точными (погрешность менее 0.5 нТл) и неподверженными долговременным дрейфам. Соответственно, они используются в качестве эталонных как на магнитных обсерваториях, так и на геомагнитных спутниках [Hulot et al., 2010].

**Контроль вековых вариаций** МПЗ на обсерватории осуществляется с помощью регулярных абсолютных измерений. Для этого используется сочетание магнитометра с теодолитом для измерения магнитного склонения. На полномасштабных магнитных обсерваториях, где работают наблюдатели, присутствуют все три типа магнитных измерений [Hulot et al., 2010].

В удаленных районах, где трудно обеспечить присутствие квалифицированного персонала, геомагнитные наблюдения ведутся автоматически **магнитовариационными станциями (МВС)**. Последние оборудованы векторными магнитометрами. При этом фиксируются только изменения МПЗ [Hulot et al., 2010]. МВС могут быть обслуживаемыми и необслуживаемыми, способными работать автономно до года. И обсерватории, и МВС должны располагаться в местах с малым пространственным градиентом магнитного поля.

**При проведении морских исследований** скалярный магнитометр буксируется или прикрепляется к надводному или подводному судну. Для гарантии качества данных буксировка проводится на тросе длиной не менее 200 м. Таким образом, гарантируется отсутствие зашумления регистрируемой информации магнитным полем корабля. Большой вклад в морской геомагнетизм внесли измерения советской немагнитной шхуны “Заря”, которая используется и в настоящее время [Гвишиани, Лукьянова, 2018; Гвишиани и др., 2019а].

**В случае аэромагнитных наблюдений** скалярный магнитометр (точность до 1 нТл) прикрепляется к летательному аппарату – самолету, вертолету, воздушному шару или дрону (беспилотнику). Масштабы даже в несколько сотен метров могут быть восстановлены аэромагнитными съемками на малой высоте. Еще меньшие масштабы могут быть восстановлены с помощью других средств, например, пешеходных съемок, потенциал которых весьма ограничен [Гвишиани, Лукьянова, 2018; Гвишиани и др., 2019а].

Отметим, что точность съемки существенно зависит от того, насколько хорошо определено пространственное положение магнитометра. Появление Глобальных навигационных спутниковых систем (GPS, ГЛОНАСС, Beidou, Galileo) привело к значительному увеличению точности определения положения прибора [Гвишиани, Лукьянова, 2018; Гвишиани и др., 2019а].

Для каждого из приведенных выше типов геомагнитных наблюдений их стандарты и распределение потоков данных согласованы на национальном и глобальном уровнях.

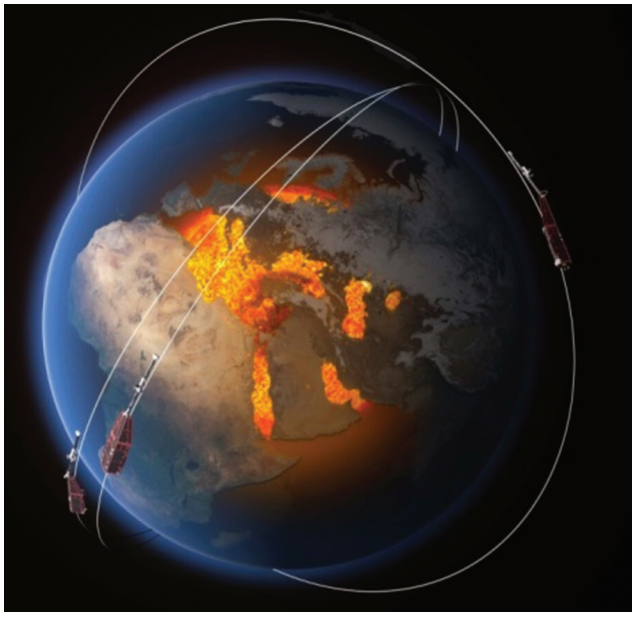

**Рис. 13.** Группировка спутников Swarm [ESA..., 2020].

### Геомагнитные наблюдения искусственных спутников Земли

Помимо описанных выше наземных наблюдений для исследований геомагнитного поля используются также искусственные спутники Земли. Они имеют низкую орбиту (400–800 км), лежащую в меридиональной плоскости и пересекающую полярные области земного шара [Olsen, Stolle, 2012]. Спутниковые наблюдения геомагнитного поля начались в СССР в 1960-х годах: “Космос-36” и “Космос-49” в 1964–1965 гг. и “Космос-321” в 1970 г. [Гвишиани, Лукьянова, 2018; Гвишиани и др., 2019а; Krasnoperov et al., 2020]. До начала 2000-х на орбите находился целый ряд, в основном американских и советских, спутников, оснащенных магнитометрами. Они проводили фрагментарные магнитные измерения вдоль отдельных траекторий пролетов.

С конца 1990-х для мониторинга магнитных вариаций были задействованы запущенные для телекоммуникационных целей космические аппараты системы спутниковой связи Iridium (США–Япония). Наблюдения осуществляются 66 спутниками. Высота орбиты 700–800 км. Параметры орбиты спутников позволяют полностью покрывать земную поверхность [Chini et al., 2009].

В 2000-х были произведены запуски европейских спутников нового поколения Oersted (Дания, 1999–2001 гг.) и CHAMP (Европейский Союз, 2000–2011 гг.) [Hulot et al., 2010; Whaler, 2007]. Работая на особо низкой орбите в 200 км, CHAMP внес исключительно важный вклад в изучение магнитного поля Земли. Небольшая высота орбиты спутника позволила провести весь цикл геомагнитных наблюдений, включая исследования главного магнитного поля, геомагнитных вариаций и неоднородностей литосферного поля. Совместная обработка спутниковых измерений, проведенных Oersted и CHAMP, помогла существенно улучшить модели главного магнитного поля Земли и магнитных аномалий [Maus et al., 2007a, 2007b; Olsen, Stolle, 2012].

Сегодня сложились уникальные условия для дальнейшего изучения МПЗ. В ноябре 2013 г. проектом Swarm Европейского космического агентства была запущена группировка из трех спутников с полярной орбитой, оснащенных высокоточными магнитометрами [Hulot et al., 2015]. Два спутника были выведены на круговую орбиту на высоте 450 км, а третий – на орбиту высотой 530 км (рис. 13). Расчетная длительность работы Swarm составляла 4.5 г. Однако это время уже существенно превышено, а группировка продолжает успешно функционировать.

Непрерывные записи МПЗ, производимые тремя спутниками Swarm, включают в себя значения трех ортогональных компонент вектора магнитного поля в системе отсчета вариометра и в системе NEC (Север–Восток–Центр), значения суммарной интенсивности вектора геомагнитного поля, время и координатную привязку. Частота измерений составляет 1 Гц. Данные сохраняются в оперативно пополняемом архиве в двоичном формате данных (CDF). Доступ к данным 1-го уровня (откалиброванные и проверенные данные приборов) и 2-го уровня (продукция) осуществляется для пользователей через FTP-сервер [Hulot et al., 2015].

Длинные ряды данных регистрации МПЗ важны для решения целого ряда научных и прикладных задач. Фундаментальной проблемой здесь является исследование свойств, собственно, главного магнитного поля Земли и его изменений во времени [Dormy, Le Mouël, 2008; Roberts, King, 2013].

Важнейшей задачей является определение значений магнитного поля при заданных географических координатах и моменте времени. Для этого строятся модели магнитного поля. Существует достаточно большое количество различных моделей, ориентированных на решение различных задач [Hulot et al., 2010; Maus et al., 2006; Olsen et al., 2006]. Модели делятся на глобальные и региональные. Наиболее распространенными глобальными моделями являются Международный стандарт геомагнитного поля (International Geomagnetic Reference Field, IGRF) [Thébault et al., 2015] и Всемирная модель геомагнитного поля (World Magnetic Model, WMM) [Chulliat et al., 2015]. Обе эти модели рассчитываются на основе разложения магнитного поля по сферическим

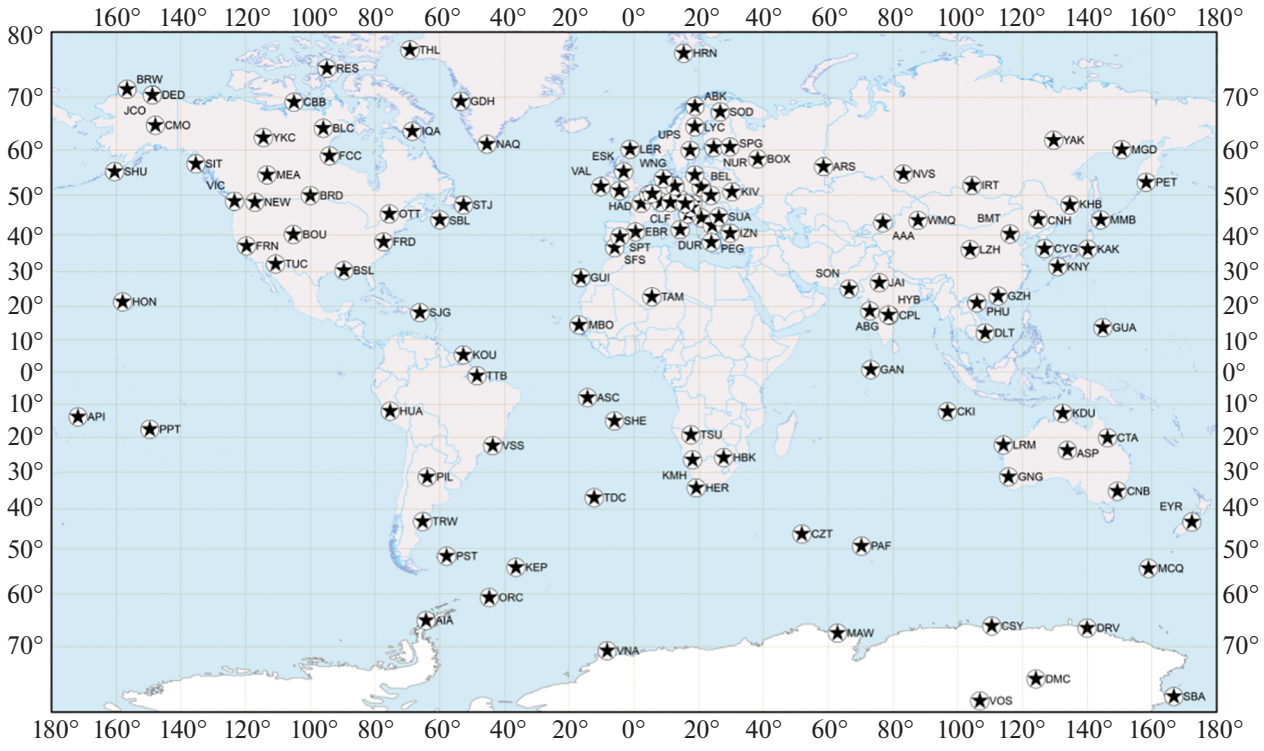


**Рис. 14.** Карта обсерваторий INTERMAGNET [Gvishiani, Soloviev, 2020].

гармоникам. Коэффициенты разложения подбираются на основе данных спутниковых и наземных магнитных наблюдений.

### Сети сбора, хранения и распространения геомагнитных данных

Основным и наиболее полным источником наземных наблюдений МПЗ являются геомагнитные обсерватории. Лучшая часть из них входит в международную сеть INTERMAGNET (International Real-time Magnetic Observatory Network) [Гвишиани и др., 2019б; St-Louis et al., 2012]. Сеть была создана для обмена данными между обсерваториями, внедрения современных стандартов геомагнитных измерений в обсерваторскую практику и подготовки информационных продуктов на основе данных геомагнитных измерений в близком к реальному времени. Карта обсерваторий, входящих в сеть INTERMAGNET в настоящее время, приведена на рис. 14.

К обсерваториям сети предъявляются строгие требования. Целью требований являются обеспечение качества и непрерывности измерений вместе с оперативностью передачи данных в узлы сети INTERMAGNET. Помимо этого, обсерватории обязаны готовить на основе исходных данных так называемые окончательные (definitive) магнитограммы. Последние представляют собой временные ряды регистрации МПЗ, очищенные от техногенных помех и скорректированные с учетом абсолютных измерений. Процедура подготовки окончательных данных выполняется ежегодно и включает в себя как алгоритмически автоматизированную, так и ручную коррекцию [Гвишиани, Лукьянова, 2018; Гвишиани и др., 2019а].

С целью автоматизации подготовки окончательных данных в Геофизическом центре РАН был создан аппаратно-программный комплекс MAGNUS [Gvishiani et al., 2016], существенно упрощающий решение этой задачи. Значительная часть работы по подготовке окончательных данных выполняется аппаратно-программным комплексом автоматизировано без участия оператора.

Всего за время существования сети INTERMAGNET в нее входило порядка 150 геомагнитных обсерваторий. Их список приведен на интернет-странице https://www.intermagnet.org/imos/imotblobs-eng.php. Обсерватории, переставшие функционировать или выполнять требования INTERMAGNET к предоставляемым данным, исключаются из сети.

На основе данных обсерваторских наблюдений рассчитываются индексы геомагнитной активности, которые характеризуют геомагнитную обстановку в целом. Основными из них являются: Kp-индекс [Menvielle, Berthelier, 1991] – обобщенный трехчасовой среднеширотный индекс; Dst-индекс [Gonzalez et al., 1994], являющийся

мерой экваториальной токовой системы в магнитосфере; AE-индекс [Davis, Sugiura, 1966], показывающий интенсивность аврорального электроджета. Индексы геомагнитной активности готовятся под эгидой ISGI (International Service for Geomagnetic Indices) (http://isgi.unistra.fr/) [Reay et al., 2011].

Магнитовариационные станции, дополняющие пространственное покрытие сети геомагнитных обсерваторий, объединены в различные национальные и региональные сети наблюдений. Среди таких сетей наблюдений можно выделить национальные сети Геофизической службы (USGS Magnetic Service) США, Японскую национальную геомагнитную сеть, канадскую CARISMA (Canadian Array for Realtime Investigations of Magnetic Activity), сеть Национального геомагнитного центра (CEA) Китая, сеть станций в Гренландии Датского национального космического института, сеть Южноафриканского национального космического агентства (SANSA). Региональные сети – IMAGE (International Monitor for Auroral Geomagnetic Effects), SAMBA (South American Meridional B-Field Array), WAMNET (West African Magnetometer NETwork), MACCS (Magnetometer Array for Cusp and Cleft Studies), AMBER.

В России к сетям геомагнитных наблюдений относятся арктическая сеть станций Арктического и антарктического научно-исследовательского института (ААНИИ) и российский сегмент сети INTERMAGNET [Гвишиани и др., 2018]. Последний развертывается при участии Геофизического центра РАН и его данные доступны на сайте Российско-украинского центра геомагнитных данных (http://geomag.gcras.ru/).

Попыткой собрать данные различных сетей наблюдений в одном месте является проект SuperMAG (http://supermag.jhuapl.edu/), реализуемый в Лаборатории прикладной физики университета Джонса Хопкинса, США (Johns Hopkins University, Applied Physics Laboratory, JHU-APL). Он включает в себя данные более чем 300 пунктов наблюдений, зарегистрированные, начиная с 1980 г., как магнитовариационными станциями, так и полномасштабными обсерваториями (рис. 15).

Для геомагнитных данных SuperMAG в автоматическом режиме выполняется приведение к единой системе координат, очистка и определение базовой линии. При этом качество выполняемой очистки данных и, соответственно, самих получаемых записей оказывается более низким, чем для окончательных данных сети INTERMAGNET. Для обсерваторий сети INTERMAGNET, передающих данные в SuperMAG, данные на сайте SuperMAG могут отличаться от окончательных данных INTERMAGNET. Согласно [Gjerloev, 2012], несмотря на более низкое качество данных МВС они могут успешно использоваться в задачах солнечно-земной физики наравне с данными полномасштабных обсерваторий.

Среди ведущих центров, предоставляющих доступ к данным геомагнитных наблюдений, необходимо выделить входящие в Мировую систему данных (ISC World Data System) Мировые центры данных по геомагнетизму в Эдинбурге и Киото [Reay et al., 2011]. Они уступают SuperMAG по количеству доступных пунктов магнитных наблюдений, однако предоставляют пользователям широкий спектр смежной информации.

### Геомагнитные данные и формализм БоД

Для оценки объемов данных наземных геомагнитных наблюдений рассмотрим сеть SuperMAG, как объединяющую наибольшее количество магнитных станций и обсерваторий. Доступный период наблюдений составляет здесь порядка 40 лет. Согласно [Gjerloev, 2012], это соответствует более чем 15000 лет минутных данных. При размере суточного файла наблюдений в текстовом формате порядка 100 килобайт это составит примерно $15000 \times 365 \times 100$ килобайт $\approx$550 гигабайт. При этом реально данные хранятся в бинарном формате и их объемы еще меньше.

Таким образом, мы видим, что даже все данные мировой сети геомагнитных наблюдений не являются большими в смысле V – Volume. Отметим, что вместо минутных в скором будущем будут использоваться в основном секундные измерения МПЗ. Переход к ним происходит, начиная с 1990-х годов. Но даже в случае секундных данных мы получим объемы порядка 30 терабайт, что тоже сложно классифицировать как БоД.

Данные геомагнитных наблюдений не являются большими и в смысле признака V – Velocity скорости их поступления и обработки. Хотя сеть INTERMAGNET и стремится к передаче данных в режиме, близком к реальному времени, но далеко не все обсерватории могут это обеспечить. Поэтому актуальное требование INTERMAGNET состоит лишь в обеспечении поступления данных в центры сбора в течение 72 ч. Скорость обработки сырых и подготовки окончательных геомагнитных данных тем более не является высокой. Окончательные данные готовятся, как правило, ежегодно, потому что процесс подготовки достаточно трудоемок и требует ручной коррекции данных. Даже скорость подготовки более оперативных квазиокончательных данных измеряется, в лучшем случае, сутками.

Основными путями увеличения скорости обработки геомагнитных данных являются совершенствование приборной базы, например, внедрение автоматизированных абсолютных измерений и совершенствование алгоритмов очистки и обработки данных. Но в ближайшее время в этой

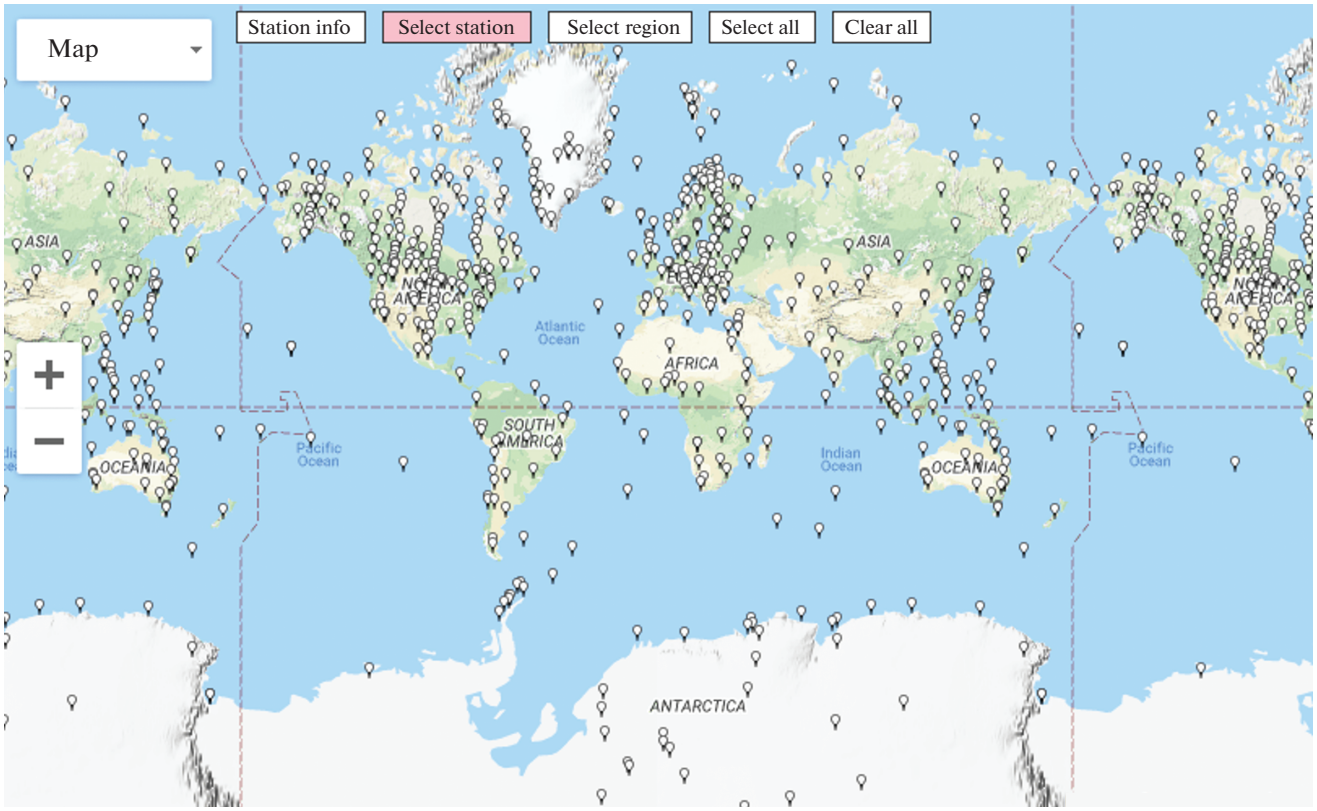


**Рис. 15.** Карта станций SuperMAG [SuperMAG..., 2021].

области сложно ожидать перехода к обработке и подготовке очищенных данных в режимах, близких к реальному времени.

Геомагнитные данные не отличаются и разнообразием в смысле признака V – Variety. Они представляют собой наблюдения трех компонент и полной напряженности вектора магнитного поля. На обсерваториях к этому добавляются результаты абсолютных измерений. На основе магнитных данных рассчитываются индексы геомагнитной активности, количество которых даже с учетом различных вариантов составляет не более пары десятков [Menvielle, Berthelier, 1991; Menvielle et al., 2011]. Этим разнообразие магнитных данных, в основном, и исчерпывается. Тем не менее, отметим, что по аналогии с данными сейсмического мониторинга магнитные данные могут иметь разнообразие в случае их рассмотрения в совокупности с данными других разделов геофизики.

Безусловно, сейчас идет постоянное расширение сетей геомагнитных наблюдений, увеличиваются скорость передачи данных в центры сбора информации и частота дискретизации наблюдений. Возрастает и уровень автоматизации геомагнитных сетей в целом. Однако это происходит не такими темпами, чтобы можно было ожидать в обозримом будущем перехода информации об МПЗ в категорию БоД. Геомагнитные данные развиваются скорее в плане их качества и точности.

Дальнейшее расширение пространственного покрытия сетей геомагнитных наблюдений, увеличение скорости передачи данных и частоты их дискретизации являются важными задачами магнитометрии. Эти направления развития подробно отражены как в планах развития Международной ассоциации геомагнетизма и аэрономии (IAGA) и сети INTERMAGNET, так и в планах национальных и региональных сетей.

В то же время существующая мировая система наблюдений МПЗ уже достаточно адекватна для изучения многих актуальных задач геомагнетизма. В числе важнейших из них построение упомянутых выше моделей МПЗ, оценки его векового хода, оценки геомагнитной активности и т.д.

Таким образом, данные магнитных наблюдений дают нам яркий пример геофизической информации, которая не входит в категорию БоД и вряд ли станет таковой в ближайшее время, но, тем не менее, представляет огромную научную и прикладную ценность, сравнимую со многими массивами данных наук о Земле, имеющими статус БоД.

## ОБСУЖДЕНИЕ

БоД имеют огромное практическое и теоретическое значение как инновационная технология. Они способны изменить наши представления и восприятие природных процессов и явлений. Со значительно расширившимися объемами и масштабами информации возникает новая научная парадигма. Происходит смена знаний **почему** на **что именно** [Mayer-Schönberger, Cukier, 2014]. При этом большинство имеющихся на данный момент знаний и фактов подвергаются сомнению. Если раньше мы пытались понять суть и природу явлений путем оптимальных выборок и формализованных процедур моделирования, то при исследованиях БоД на первое место выходит поиск корреляций.

Наборы данных в науках о Земле в основной массе имеют все атрибуты больших данных – объем, достоверность, скорость, ценность и разнообразие. Для тех наук, где широко используются спутниковые снимки и проводятся измерения большого количества параметров (ДЗЗ, СР, метеорологические и экологические наблюдения), соответствие характеристикам БоД не вызывает никаких сомнений. Это подтверждается большим количеством исследований с использованием методов БоД, ведущихся в настоящее время. В сочетании с доступностью и снижением стоимости высокопроизводительных вычислений и аналитических процедур возможности для анализа БоД в науках о Земле постоянно растут.

Помимо преимуществ, внедрение технологии БоД и обработка непрерывно растущих объемов данных создают и определенные проблемы. Перечислим некоторые из них.

Для обработки и интерпретации различных типов данных часто требуется специально разработанное для этих целей программное обеспечение. Проблема усугубляется исторически неэффективной практикой управления данными. Наборы данных редко сопоставляются после завершения исследований. Они сохраняются с недостаточным количеством метаданных для их точного описания. Большое количество данных оказывается неполным, непригодным или же недоступным. Независимо от причин, это создает огромные проблемы для современной науки.

До сих пор успешные применения аналитики БоД в основном относились к закрытым системам или анализу самых распространенных типов цифровых данных. Остаются серьезные проблемы, включая недостаток данных, с которыми можно было бы взаимодействовать в реальном времени. Существуют трудности при объединении нескольких наборов данных о взаимосвязанных явлениях. Часто отсутствуют адекватные метаданные о доступных данных для понимания их контекста и объема [Kingdon et al., 2016].

Несмотря на перечисленные проблемы, БоД дают нам инструменты для поисков взаимосвязей между данными наблюдений различных наук о Земле. Рассматривая различные типы данных в отрыве друг от друга, мы в определенном смысле уходим от самого определения БоД. Ведь мы допускаем главную ошибку, ограничиваясь выборкой данных, относящихся к той или иной науке о Земле. Скорее всего, интеграция данных из различных областей для аналитики БоД станет следующей ступенью развития. Определенные подвижки в данном направлении уже наблюдаются.

## ФИНАНСИРОВАНИЕ РАБОТЫ

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-15-50125.

## БЛАГОДАРНОСТИ

В работе использовались данные и сервисы, предоставленные ЦКП “Аналитический центр геомагнитных данных” Геофизического центра РАН (http://ckp.gcras.ru/).

Авторы выражают благодарность Герою Российской Федерации, чл.-корр. РАН Юрию Михайловичу Батурину за предоставленные фотоматериалы из личного архива.

## СПИСОК ЛИТЕРАТУРЫ

*Батурин Ю.М., Щербинин Д.Ю.* Фото- и кинотехника на борту отечественных пилотируемых космических аппаратов (1961–2000) // Вопросы истории естествознания и техники. 2011. Т. 32. № 3. С. 87–104.

*Бондур В.Г.* Современные подходы к обработке больших потоков гиперспектральной и многоспектральной аэрокосмической информации // Исслед. Земли из космоса. 2014. № 1. С. 4–16.
https://doi.org/10.7868/S0205961414010035

*Гвишиани А.Д., Лукьянова Р.Ю.* Оценка влияния геомагнитных возмущений на траекторию наклонно-направленного бурения глубоких скважин в Арктическом регионе // Физика Земли. 2018. № 4. С. 19–30.
https://doi.org/10.1134/S0002333718040051

*Гвишиани А.Д., Соловьев А.А., Сидоров Р.В., Красноперов Р.И., Груднев А.А., Кудин Д.В., Карапетян Дж.К., Симонян А.О.* Успехи организации геомагнитного мониторинга в России и ближнем зарубежье // Вестник ОНЗ РАН. 2018. Т. 10. NZ4001.
https://doi.org/10.2205/2018NZ000357

*Гвишиани А.Д., Лукьянова Р.Ю., Соловьев А.А.* Геомагнетизм: от ядра Земли до Солнца. М.: РАН. 2019а. 186 с.

*Гвишиани А.Д., Кафтан В.И., Красноперов Р.И., Татаринов В.Н., Вавилин Е.В.* Геоинформатика и системный анализ в геофизике и геодинамике // Физика Земли. 2019б. № 1. С. 42–60.
https://doi.org/10.31857/S0002-33372019142-60

ГОСТ 8.417-2002 – Государственная система обеспечения единства измерений. Единицы величин. 2013. 32 с.

*Майер-Шенбергер В.М., Кукьер К.* Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М.: Манн, Иванов и Фербер. 2014. 240 с.

*Одум Ю.* Основы экологии. М.: Мир. 1975. 741 с.

*О'Нил К.* Убийственные большие данные. Как математика превратилась в оружие массового поражения. М.: АСТ. 2018. 140 с.

*Фрэнкс Б.* Укрощение больших данных: как извлекать знания из массивов информации с помощью глубокой аналитики. М.: Манн, Иванов и Фербер. 2014. 352 с.

*Ahmad L., Habib Kanth R., Parvaze S., Sheraz Mahdi S.* Experimental Agrometeorology: A Practical Manual. Cham: Springer. 2017a.
https://doi.org/10.1007/978-3-319-69185-5

*Ahmad L., Habib Kanth R., Parvaze S., Sheraz Mahdi S.* Synoptic Meteorology. Experimental Agrometeorology: A Practical Manual. Cham: Springer, 2017b. P. 119–121.
https://doi.org/10.1007/978-3-319-69185-5_16

*Amnion C.J., Lay T., Simpson D.W.* Great Earthquakes and Global Seismic Network // Seismological Research Letters. 2010. V. 81. Is. 6. P. 965–971.
https://doi.org/10.1785/gssrl.81.6.965

*Armstrong E.M., Bourassa M.A., Cram T.A., DeBellis M., Elya J., Greguska F.R., III, Huang T., Jacob J.C., Ji Z., Jiang Y., Li Y., Quach N., McGibbney L., Smith S., Tsontos V.M., Wilson B., Worley S.J., Yang C., Yam E.* An Integrated Data Analytics Platform // Frontiers in Marine Science. 2019. V. 6. Is. JUN. 354.
https://doi.org/10.3389/fmars.2019.00354

*Ashton K.* That 'Internet of Things' Thing // RFID Journal. 2009.
https://www.rfidjournal.com/that-internet-of-things-thing-3 [дата обращения: 17.03.2021].

*Miller A.A.* Climatology. Routledge. Tylor Press, NY, USA. 2019. 326 p.

*Ball J.E., Anderson D.T., Chan Sr. C.S.* Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community // J. Applied Remote Sensing. 2017. V. 11. Is. 4. P. 042609.
https://doi.org/10.1117/1.JRS.11.042609

*Balta H., Velagic J., Bosschaerts W., De Cubber G., Siciliano B.* Fast statistical outlier removal based method for large 3D point clouds of outdoor environments // IFAC-PapersOnLine. 2018. V. 51. P. 348–353.
https://doi.org/10.1016/j.ifacol.2018.11.566

*Barry R.M., Cavers D.A., Kneale C.W.* Recommended standards for digital tape formats // Geophysics. 1975. V. 40. P. 344–352.

*Baumann P., Mazzetti P., Ungar J. et al.* Big Data Analytics for Earth Sciences: The EarthServer Approach // International Journal of Digital Earth. 2016. V. 9. Is. 1. P. 3–29.
https://doi.org/10.1080/17538947.2014.1003106

*Baumann P.R.* History of Remote Sensing, Satellite Imagery, Part II. 2009. http://employees.oneonta.edu/baumanpr/geosat2/RS%20History%20II/RS-History-Part-2.html [дата обращения: 18.03.2021].

*Becker-Reshef I., Vermote E., Lindeman M., Justice C.* A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data // Remote Sensing of Environment. 2010. V. 114. Is. 6. P. 1312–1323.
https://doi.org/10.1016/j.rse.2010.01.010

*Besic N., Figueras i Ventura J., Grazioli J., Gabella M., Germann U., Berne A.* Hydrometeor classification through statistical clustering of polarimetric radar measurements: A semi-supervised approach // Atmospheric Measurement Techniques. 2016. V. 9. Is. 9. P. 4425–4445.
https://doi.org/10.5194/amt-9-4425-2016

*Boman K.* Big data growth continues in seismic surveys. Rigzone, Sept. 2, 2015.
https://www.rigzone.com/news/oil_gas/a/140418/big_data_growth_continues_in_seismic_surveys/ [дата обращения: 18.03.2021].

*Caers J.* Modeling uncertainty in the earth sciences. John Wiley & Sons. 2011. 246 p.

*Chen D., Wang L., Dou M., Liu Z.* Natural Disaster Monitoring with Wireless Sensor Networks: A Case Study of Data-intensive Applications upon Low-Cost Scalable Systems // Mobile Networks and Applications. 2013. V. 18. P. 651–663.
https://doi.org/10.1007/s11036-013-0456-9

*Chen M., Mao S., Zhang Y., Leung V.C.M.* Big Data. Related Technologies, Challenges, and Future Prospects. Springer, 2014. 100 p.
https://doi.org/10.1007/978-3-319-06245-7

*Chi M., Plaza A., Benediktsson J.A., Sun Z., Shen J., Zhu Y.* Big Data for Remote Sensing: Challenges and Opportunities // Proceedings of the IEEE. 2016. V. 104. Is. 11. P. 2207–2219.
https://doi.org/10.1109/JPROC.2016.2598228

*Chini P., Giambene G., Kota S.* A survey on mobile satellite systems // International Journal of Satellite Communications and Networking. 2009. V. 28. Is. 1.
https://doi.org/10.1002/sat.941

*Chowdhury S., Al-Zahrani M.* Water quality change in dam reservoir and shallow aquifer: analysis on trend, seasonal variability and data reduction // Environmental Monitoring and Assessment. 2014. V. 186. P. 6127–6143.
https://doi.org/10.1007/s10661-014-3844-0

*Chulliat A., Alken P., Nair M. et al.* The US/UK World Magnetic Model for 2015–2020 // Technical Report, National Geophysical Data Center, NOAA. 2015. V. 107. Is. A2.
https://doi.org/10.7289/V5TB14V7

*Corizzo R., Ceci M., Japkowicz N.* Anomaly detection and repair for accurate predictions in geo-distributed big data // Big Data Research. 2019. V. 16. P. 18–35.
https://doi.org/10.1016/j.bdr.2019.04.001

*Cressie N.* Statistics for spatial data. John Wiley & Sons. New York. 1993. 928 p.

*Davis T.N., Sugiura M.* Auroral electrojet activity index AE and its universal time variations // J. Geophysical Research. 1966. V. 71. Is. 3. P. 785–801.
https://doi.org/10.1029/JZ071i003p00785

*Dedić N., Stanier C.* Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery / Piazolo F., Geist V., Brehm L., Schmidt R. (eds.) Innovations in Enterprise Information Systems Management and Engineering. ERP Future 2016. LNBIP. 2017. V. 285. P. 114–12. Springer, Cham.
https://doi.org/10.1007/978-3-319-58801-8_10

*Ding A.J., Fu C.B., Yang X.Q., Sun J.N., Petäjä T., Kerminen V.-M., Wang T., Xie Y., Herrmann E., Zheng L.F., Nie W., Liu Q., Wei X.L., Kulmala M.* Intense atmospheric pollution modifies weather: a case of mixed biomass burning with fossil fuel combustion pollution in eastern China // Atmospheric

Chemistry and Physics. 2013. V. 13. Is. 20. P. 10545–10554. https://doi.org/10.5194/acp-13-10545-2013

*Dormy E., Le Mouël J.-L.* Geomagnetism and the dynamo: where do we stand? // Comptes Rendus Physique. 2008. V. 9. Is. 7. P. 711–720. https://doi.org/10.1016/j.crhy.2008.07.003

ESA – Swarm probes weakening of Earth's magnetic field. 2020. http://www.esa.int/Applications/Observing_the_Earth/Swarm/Swarm_probes_weakening_of_Earth_s_magnetic_field [дата обращения: 18.03.2021].

*Esch T., Üreyen S., Zeidler J., Metz-Marconcini A., Hirner A., Asamer H., Tum M., Böttcher M., Kuchar S., Svaton V., Marconcini M.* Exploiting big earth data from space – first experiences with the timescan processing chain // Big Earth Data. 2018. V. 2. Is. 1. P. 36–55. https://doi.org/10.1080/20964471.2018.1433790

*Feblowitz J.* Insights IDCE, Analytics in Oil and Gas: the Big Deal about Big Data. Paper presented at the SPE Digital Energy Conference, The Woodlands, Texas, USA, March 2013. https://doi.org/10.2118/163717-MS

*Fernández-Martínez J.L.* Model reduction and uncertainty analysis in inverse problems // The Leading Edge. 2015. V. 9. P. 1006–10016. https://doi.org/10.1190/tle34091006.1

*Fouedjio F., Klump J.* Exploring prediction uncertainty of spatial data in geostatistical and machine learning approaches // Environmental Earth Sciences. 2019. V. 78. Is. 1. 38. https://doi.org/10.1007/s12665-018-8032-z

*Hulot G., Vigneron P., Léger J.-M. et al.* Swarm's absolute magnetometer experimental vector mode, an innovative capability for space magnetometry // Geophysical Research Letters. 2015. V. 42. Is. 5. P. 1352–1359. https://doi.org/10.1002/2014GL062700

*Giuliani G., Lacroix P., Guigoz Y., Roncella R., Bigagli L., Santoro M., Mazzetti P., Nativi S., Ray N., Lehmann A.* Bringing GEOSS services into practice: a capacity building resource on spatial data infrastructures (SDI) // Transactions in GIS. 2017. V. 21. Is. 4. P. 811–824. https://doi.org/10.1111/tgis.12209

*Gjerloev J.W.* The SuperMAG data processing technique // J. Geophysical Research: Space Physics. 2012. V. 117. Is. A9. 9213. https://doi.org/10.1029/2012JA017683

Global Observing System (GOS). 2016. https://www.wmo.int/pages/prog/www/OSY/GOS.html [дата обращения: 18.03.2021].

*Gomes V.C., Queiroz G.R., Ferreira K.R.* An overview of platforms for big earth observation data management and analysis // Remote Sensing. 2020. V. 12. 1253. https://doi.org/10.3390/rs12081253

*Gonzalez W.D., Joselyn J.A., Kamide Y., Kroehl H.W., Rostoker G., Tsurutani B.T., Vasyliunas V.M.* What is a geomagnetic storm? // J. Geophysical Research: Space Physics. 1994. V. 99. Is. A4. P. 5771–5792. https://doi.org/10.1029/93JA02867

*Gordon A., Grace W., Byron-Scott R., Schwerdtfeger P.* Dynamic Meteorology. Routledge, 2016. 271 p.

*Gorelick N., Hancher M., Dixon M., Ilyushchenko S., Thau D., Moore R.* GoogleEarth Engine: Planetary-Scale Geospatial Analysis for Everyone // Remote Sensing of Environment. 2017. V. 202 (Suppl C). P. 18–27. https://doi.org/10.1016/j.rse.2017.06.031

GTS Main. WMO. 2020. https://www.wmo.int/pages/prog/www/TEM/index_en.html [дата обращения: 18.03.2021].

*Gultepe I., Sharman R., Williams P.D. et al.* A Review of High Impact Weather for Aviation Meteorology // Pure and Applied Geophysics. 2019. V. 176. P. 1869–1921. https://doi.org/10.1007/s00024-019-02168-6

*Guo H.* Big Earth data: A new frontier in Earth and information sciences // Big Earth Data. 2017. V. 1. Is. 1–2. P. 4–20. https://doi.org/10.1080/20964471.2017.1403062

*Gvishiani A., Dobrovolsky M., Rybkina A.* Big Data and FAIR Data for Data Science / Roberts F.S., Sheremet I.A. (eds). Resilience in the Digital Age. Lecture Notes in Computer Science. 2021. V. 12660. Chapter 6. Springer, Cham. https://doi.org/10.1007/978-3-030-70370-7_6

*Gvishiani A., Soloviev A.* Observations, Modeling and Systems Analysis in Geomagnetic Data Interpretation. Springer International Publishing. 2020. https://doi.org/10.1007/978-3-030-58969-1

*Gvishiani A., Soloviev A., Krasnoperov R., Lukianova R.* Automated Hardware and Software System for Monitoring the Earth's Magnetic Environment // Data Science J. 2016. V. 15. Is. 18. https://doi.org/10.5334/dsj-2016-018

*Hadjimitsis D.G., Clayton C.R.I., Hope V.S.* An assessment of the effectiveness of atmospheric correction algorithms through the remote sensing of some reservoirs // International J. Remote Sensing. 2004. V. 25. Is. 18. P. 3651–3674. https://doi.org/10.1080/01431160310001647993

*Hari P., Petäjä T., Bäck J., Kerminen V.-M., Lappalainen H.K., Vihma T., Laurila T., Viisanen Y., Vesala T., Kulmala M.* Conceptual design of a measurement network of the global change // Atmospheric Chemistry and Physics. 2016. V. 16. Is. 2. P. 1017–1028. https://doi.org/10.5194/acp-16-1017-2016

*Hashem I.A.T., Yaqoob I., Anuar N.B., Mokhtar S., Gani A., Khan S.U.* The rise of "big data" on cloud computing: Review and open research issues // Information Systems. 2015. V. 47. P. 98–115. https://doi.org/10.1016/j.is.2014.07.006

*Hassani H., Huang X., Silva E.* Big Data and Climate Change // Big Data and Cognitive Computing. 2019. V. 3. Is. 1. 12. P. 1–17. https://doi.org/10.3390/bdcc3010012

History of WMO. 2020. https://public.wmo.int/en/about-us/who-we-are/history-of-wmo [дата обращения: 18.03.2021].

*Hoeser T., Bachofer F., Kuenzer C.* Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review – Part II: Applications // Remote Sensing. 2020. V. 12. Is. 18. 3053. https://doi.org/10.3390/rs12183053

*Holloway J., Mengersen K.* Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review // Remote Sensing. 2018. V. 10. Is. 9. 1365. https://doi.org/10.3390/rs10091365

*Hondula D.M., Balling R.C., Andrade R. et al.* Biometeorology for cities // International Journal of Biometeorology.

2017. V. 61. P. 59–69.
https://doi.org/10.1007/s00484-017-1412-3

*Houghton J.* The Physics of Atmospheres. Cambridge University Press. 2002. 320 p.

*Huang T.* NASA Sea Level Change Portal-it is not just another portal site. 2017.
https://trs.jpl.nasa.gov/bitstream/handle/2014/48852/CL%2317-6319.pdf [дата обращения: 18.03.2021].

*Huang T., Armstrong E., Jacob J. et al.* An introduction to OceanWorks ocean science platform. ESTF. 2017. 21 p.
https://trs.jpl.nasa.gov/bitstream/handle/2014/48094/CL%2317-2720.pdf [дата обращения: 18.03.2021].

*Hulot G., Finlay C.C., Constable C.G., Olsen N., Mandea M.* The Magnetic Field of Planet Earth // Space Science Reviews. 2010. V. 152. Is. 1–4 P. 159–222.
https://doi.org/10.1007/s11214-010-9644-0

*Hurwitz J.S., Nugent A., Halper F., Kaufman M.* Big Data for Dummies. Wiley. 2013. 336 p.

*Inmon W.H., Linstedt D.* Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault. Morgan Kaufmann. 2014. 378 p.

*Internet Of Things* (iot). Information Technology Gartner Glossary. 2020.

*Ishwarappa J., Anuradha J.* A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology // Procedia Computer Science. 2015. V. 48. P. 319–324.
https://doi.org/10.1016/j.procs.2015.04.188

*Islam M., Reza S.* The Rise of Big Data and Cloud Computing // Internet of Things and Cloud Computing. 2019. V. 7. Is. 2. P. 45–53.
https://doi.org/10.11648/j.iotcc.20190702.12

*Jankowsky J., Sucksdorf C.* Guide for magnetic measurements and observatory practice. Warsaw: IAGA. 1996. 238 p.

*Jong S.M., Meer F.D., Clevers J.G.* Basics of Remote Sensing. Remote Sensing Image Analysis: Including The Spatial Domain. Remote Sensing and Digital Image Processing. V. 5. / Jong S.M.D., Meer F.D.V. (eds) Dordrecht: Springer. 2004. P. 1–15.
https://doi.org/10.1007/978-1-4020-2560-0_1

*Jutz S., Milagro-Perez M.P.* Copernicus: the European Earth Observation programme // Revista de Teledetección. 2020. Is. 56. P. V–XI.
https://doi.org/10.4995/raet.2020.14346

*Kang X., Li J., Fan X.* Spatial-Temporal Visualization and Analysis of Earth Data Under Cesium Digital Earth Engine. Proceedings of the 2018 2nd International Conference on Big Data and Internet of Things (BDIOT 2018). New York: Association for Computing Machinery. 2018. P. 29–32.
https://doi.org/10.1145/3289430.3289447

*Karau H., Konwinski A., Wendell P., Zaharia M.* Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly Media. 2015. 276 p.

*Karimi H.A.* Big Data: Techniques and Technologies in Geoinformatics. CRC Press. 2014. 312 p.

*Katzfuss M., Cressie N.* Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets // J. Time Series Analysis. 2011. V. 32. Is. 4. P. 430–446.
https://doi.org/10.1111/j.1467-9892.2011.00732.x

*Kawasaki A., Yamamoto A., Koudelova P., Acierto R., Nemoto T., Kitsuregawa M., Koike T.* Data Integration and Analysis System (DIAS) Contributing to Climate Change Analysis and Disaster Risk Reduction // Data Science Journal. 2017. V. 16. 41.
https://doi.org/10.5334/dsj-2017-041

*Kempler S., Mathews T.* Earth Science Data Analytics: Definitions, Techniques and Skills // Data Science Journal. 2017. V. 16. 6.
https://doi.org/10.5334/dsj-2017-006

*Kingdon A., Nayembil M.L., Richardson A.E., Smith A.G.* A geodata warehouse: Using denormalisation techniques as a tool for delivering spatially enabled integrated geological information to geologists // Computers and Geosciences. 2016. V. 96. P. 87–97.
https://doi.org/10.1016/j.cageo.2016.07.016

*Kleppmann M.* Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. O'Reilly Media. 2017. 616 p.

*Konecny G.* Geoinformation: Remote Sensing, Photogrammetry and Geographic Information Systems, Second Edition. Taylor & Francis. 2014. 436 p.

*Krasnoperov R., Peregoudov D., Lukianova R., Soloviev A., Dzeboev B.* Early Soviet satellite magnetic field measurements in the years 1964 and 1970 // Earth System Science Data. 2020. V. 12. Is. 1. P. 555–561.
https://doi.org/10.5194/essd-12-555-2020

*Laney D.* 3D Data Management: Controlling Data Volume, Velocity, and Variety. Meta Group. 2001. https://studylib.net/doc/8647594/3d-data-management–controlling-data-volume–velocity–an [дата обращения: 18.03.2021].

*LeHong H., Fenn J.* Key Trends to Watch in Gartner 2012 Emerging Technologies Hype Cycle // Forbes. 2012.
https://www.forbes.com/sites/gartnergroup/2012/09/18/key-trends-to-watch-in-gartner-2012-emerging-technologies-hype-cycle-2 [дата обращения: 18.03.2021].

*Liu J.G., Mason P.J.* Essential Image Processing and GIS for Remote Sensing. John Wiley & Sons. 2013. 464 p.

*Liu P.* A survey of remote-sensing big data // Frontiers in Environmental Science. 2015. V. 3. 45.
https://doi.org/10.3389/fenvs.2015.00045

*Lohr S.* The Origins of 'Big Data': An Etymological Detective Story. The New York Times, 2013. https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/ [дата обращения: 18.03.2021].

*Lopez M.M., Kalita J.* Deep Learning applied to NLP. arXiv preprint. 2017. arXiv:1703.03091

*MacLachlan C., Arribas A., Peterson K.A. et al.* Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system // Quarterly J. Royal Meteorological Society. 2015. V. 141. P. 1072–1084.
https://doi.org/10.1002/qj.2396

*Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Hung Byers A.* Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. 2011. 156 p.

*Maraun D., Shepherd T.G., Widmann M., Zappa G., Walton D., Gutierrez J.M., Hagemann S., Richter I., Soares P.M.M., Hall A., Mearns L.O.* Towards processinformed bias correction of climate change simulations // Nature Climate Change. 2017. V. 7. P. 764–773.
https://doi.org/10.1038/NCLIMATE3418

*Marr B.* Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance. Wiley. 2015. 256 p.

*Marz N., Warren J.* Big Data: Principles and best practices of scalable realtime data systems. Manning Publications. 2015. 328 p.

*Maus S., Lühr H., Rother M., Hemant K., Balasis G., Ritter P., Stolle C.* Fifth-generation lithospheric magnetic field model from CHAMP satellite measurements // Geochemistry, Geophysics, Geosystems. 2007a. V. 8. Is. 5.
https://doi.org/10.1029/2006GC001521

*Maus S., Rother M., Stolle C., Mai W., Choi S., Lühr H., Cooke D., Roth C.* Third generation of the Potsdam Magnetic Model of the Earth (POMME) // Geochemistry, Geophysics, Geosystems. 2006. V. 7. Is. 7.
https://doi.org/10.1029/2006GC001269

*Maus S., Sazonova T., Hemant K., Fairhead J.D., Ravat D.* National Geophysical Data Center candidate for the World Digital Magnetic Anomaly Map // Geochemistry, Geophysics, Geosystems. 2007b. V. 8. Is. 6.
https://doi.org/10.1029/2007GC001643

*Mayer-Schönberger V., Cukier K.* Big Data: A Revolution That Will Transform How We Live, Work, and Think. Eamon Dolan/Mariner Books. 2014. 272 p.

*McInerney M.* NASA Earthdata Cloud. 2020.
https://ntrs.nasa.gov/api/citations/20200001222/downloads/20200001222.pdf?attachment=true [дата обращения: 18.03.2021].

*Menvielle M., Berthelier A.* The K-derived planetary indices: Description and availability // Reviews of Geophysics. 1991. V. 29. Is. 3. P. 415–432.
https://doi.org/10.1029/91RG00994

*Menvielle M., Iyemori T., Marchaudon A., Nose M.* Geomagnetic indices. Geomagnetic Observations and Models. IAGA Special Sopron Book Series. V. 5. / Mandea M., Korte M. (eds). 2011. P. 183–228.
https://doi.org/10.1007/978-90-481-9858-0_7.

*Mitas L., Mitasova H.* Spatial Interpolation. *Longley P., Goodchild M.F., Maguire D.J., Rhind D.W.* (eds.). Geographical Information Systems: Principles, Techniques, Management and Applications. GeoInformation International. Wiley. 1999. P. 481–492.

*Moon N.H., Shin M.Y., Moon G.H., Chun J.* Trends and Prospects of Forest Meteorological Studies Based on the Publications in Korean Journal of Agricultural and Forest Meteorology // Korean J. Agricultural and Forest Meteorology. 2019. V. 21. Is. 3. P. 121–134.
https://doi.org/10.5532/KJAFM.2019.21.3.121

*Moore J.M.* Nonlinear filtering techniques for noisy geophysical data: Using big data to predict the future. AGU Fall Meeting 2014. 2014. Abstract id. NG23A-3789.

National Research Council. Earth Observations from Space: The First 50 Years of Scientific Achievements. Washington, DC: The National Academies Press. 2008. 143 p.
https://doi.org/10.17226/11991

National Research Council. People and Pixels: Linking Remote Sensing and Social Science. Washington, DC: The National Academies Press. 1998. 267 p.
https://doi.org/10.17226/5963

*Nativi S., Mazzetti P., Craglia M.* A view-based model of data-cube to support big earth data systems interoperability // Big Earth Data. 2017. V. 1. Is. 1–2. P. 75–99.
https://doi.org/10.1080/20964471.2017.1404232

NIST Big Data Interoperability Framework: V. 1, Definitions. Version 3. NIST Special Publication 1500-1r2. 2019.
https://doi.org/10.6028/NIST.SP.1500-1r2

*Novick K.A., Biederman J.A., Desai A.R., Litvak M.E., Moore D.J.P., Scott R.L., Torn M.S.* The AmeriFlux network: A coalition of the willing // Agricultural and Forest Meteorology. 2018. V. 249. P. 444–456.
https://doi.org/10.1016/j.agrformet.2017.10.009

*Ofli F., Meier P., Imran M., Castillo C., Tuia D., Rey N., Briant J., Millet P., Reinhard F., Parkan M., Joost S.* Combining human computing and machine learning to make sense of big (aerial) data for disaster response // Big Data. 2016. V. 4. Is. 1. P. 47–59.
https://doi.org/10.1089/big.2014.0064

*Olsen N., Lühr H., Sabaka T.J., Mandea M., Rother M., Tøffner-Clausen L., Choi S.* CHAOS – a model of the Earth's magnetic field derived from CHAMP, Ørsted, and SAC-C magnetic satellite data // Geophysical J. International. 2006. V. 166. Is. 1. P. 67–75.
https://doi.org/10.1111/j.1365-246X.2006.02959.x

*Olsen N., Stolle C.* Satellite Geomagnetism // Annual Review of Earth and Planetary Sciences. 2012. V. 40. P. 441–465.
https://doi.org/10.1146/annurev-earth-042711-105540

*Onay C., Öztürk E.* A review of credit scoring research in the age of Big Data // Journal of Financial Regulation and Compliance. 2018. V. 26. Is. 3. P. 382–405.
https://doi.org/10.1108/JFRC-06-2017-0054

*Pavlikov V.V., Ruzhentsev N.V., Sobkolov A.D., Salnikov D.S., Tsopa A.I.* Ground-based radiometric complex of millimeter wave band for meteorology and telecommunications applications // Telecommunications and Radio Engineering. 2017. V. 76. Is. 16. P. 1477–1488.
https://doi.org/10.1615/TelecomRadEng.v76.i16.70

*Pei T., Song C., Guo S. et al.* Big geodata mining: Objective, connotations and research issues // J. Geographical Sciences. 2020. V. 30. Is. 2. P. 251–266.
https://doi.org/10.1007/s11442-020-1726-7

*Pelton J.N., Madry S., Camacho-Lara S.* Handbook of Satellite Applications. New York: Springer. 2013.
https://doi.org/10.1007/978-1-4419-7671-0

*Poblet M., García-Cuesta E., Casanovas P.* Crowdsourcing tools for disaster management: A review of platforms and methods. AI Approaches to the Complexity of Legal Systems / Casanovas P., Pagallo U., Palmirani M., Sartor G. (eds.) Springer: Berlin/Heidelberg. 2014. P. 261–274.

*Potapov P.V., Turubanova S.A., Tyukavina A., Krylov A.M., McCarty J.L., Radeloff V.C., Hansen M.C.* Eastern Europe's forest cover dynamics from 1985 to 2012 quantified from the full Landsat archive // Remote Sensing of Environment. 2015. V. 159. P. 28–43.
https://doi.org/10.1016/j.rse.2014.11.027

*Preimesberger C.* Hadoop, Yahoo, 'Big Data' Brighten BI Future. 2011.
https://www.webcitation.org/67j1TSO8N?url=
http://www.eweek.com/c/a/Data-Storage/TBA-Hadoop-Yahoo-Big-Data-Brightens-BI-Future-254079/ [дата обращения: 18.03.2021].

*Pyo J., Duan H., Baek S., Kim M.S., Jeon T., Kwon Y.S., Lee H., Cho K.H.* A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery //

Remote Sensing of Environment. 2019. V. 233. 111350. https://doi.org/10.1016/j.rse.2019.111350

*Rasson J.* Observatories, instrumentation. Encyclopedia of Geomagnetism and Paleomagnetism / Eds. Gubbins D., Herrero-Bervera E. New York: Springer. 2007. P. 711–713.

*Rauber R.M., Nesbitt S.W.* Radar Meteorology: A First Course. John Wiley & Sons. 2018. 488 p.

*Reay S.J., Herzog D.C., Alex S., Kharin E., McLean S., Nosé M., Sergeyeva N.* Magnetic Observatory Data and Metadata: Types and Availability. Geomagnetic Observations and Models. IAGA Special Sopron Book Series. V. 5. / Mandea M., Korte M. (eds.) 2011. P. 149–181. https://doi.org/10.1007/978-90-481-9858-0_7

*Rezai A., Keshavarzi P., Moravej Z.* Key management issue in SCADA networks: A review // Engineering Science and Technology, an International Journal. 2017. V. 20. Is. 1. P. 354–363. https://doi.org/10.1016/j.jestch.2016.08.011

*Richards P.G., Zavales J.* Seismological Methods for Monitoring a CTBT: The Technical Issues Arising in Early Negotiations / Husebye E.S., Dainty A.M. (eds.) Monitoring a Comprehensive Test Ban Treaty. NATO ASI Series. 1996. V. 303. P. 53–81.

*Roberts P.H., King E.M.* On the genesis of the Earth's magnetism // Reports on Progress in Physics. 2013. V. 76. Is. 9. 096801. https://doi.org/10.1088/0034-4885/76/9/096801

*Roden R.,* Seismic Interpretation in the Age of Big Data. Conference: SEG Technical Program Expanded Abstracts 2016. https://doi.org/10.1190/segam2016-13612308.1

*Rodriguez-Galiano V.F., Chica-Rivas M.* Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models // International J. Digital Earth. 2014. V. 7. Is. 6. P. 492–509. https://doi.org/10.1080/17538947.2012.748848

*Salimi S., Hammad A.* Sensitivity analysis of probabilistic occupancy prediction model using big data // Building and Environment. 2020. V. 172. 106729. https://doi.org/10.1016/j.buildenv.2020.106729

*Schultz M.G., Betancourt C., Gong B., Kleinert F., Langguth M., Leufen L.H., Mozaffari A., Stadtler S.* Can deep learning beat numerical weather prediction? // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2021. V. 379. Is. 2194. 20200097. https://doi.org/10.1098/rsta.2020.0097

SEED Reference Manual. Standard for the Exchange of Earthquake Data. SEED Format Version 2.4. August, 2012. International Federation of Digital Seismograph Networks, Incorporated Research Institutions for Seismology, United States Geological Survey. 224 p.

*Newman S.-F.* Seismographic Data Compression – Applying Modified Tunstall Coding. Institute of Technology, University of Washington. Tacoma. 2006. 46 p.

SMEAR, 2021. https://www.atm.helsinki.fi/SMEAR/ [дата обращения: 18.03.2021].

*Snijders C., Matzat U., Reips U.-D.* "Big Data": Big Gaps of Knowledge in the Field of Internet Science // International Journal of Internet Science. 2012. V. 7. Is. 1. P 1–5.

*Song J., Gao S., Zhu Y., Ma C.* A survey of remote sensing image classification based on CNNs // Big Earth Data. 2019. V. 3. Is. 3. P. 232–254. https://doi.org/10.1080/20964471.2019.1657720

*Stein A., van der Meer F.D., Gorte B.* Spatial statistics for remote sensing. Springer Science & Business Media. 2002. 284 p. https://doi.org/10.1007/0-306-47647-9

*St-Louis B.J., Sauter E.A., Trigg D.F. et al.* INTERMAGNET Technical Reference Manual. Version 4.6. Edinburgh: INTERMAGNET. 2012. 100 p.

*Stromann O., Nascetti A., Yousif O., Ban Y.* Dimensionality Reduction and Feature Selection for Object-Based Land Cover Classification based on Sentinel-1 and Sentinel-2 Time Series Using Google Earth Engine // Remote Sensing. 2020. V. 12. Is. 1. 76. https://doi.org/10.3390/rs12010076

*Su Z., Timmermans W., Zeng Y. et al.* An Overview of European Efforts in Generating Climate Data Records // Bulletin of the American Meteorological Society. 2018. V. 99. P. 349–359. https://doi.org/10.1175/BAMS-D-16-0074.1

SuperMAG: Download Data. 2021. https://supermag.jhuapl.edu/mag/?fidelity=low&tab=customdownload [дата обращения: 18.03.2021].

*Talwani M., Kessinger W.* Exploration Geophysics. Encyclopedia of Physical Science and Technology (Third Edition). 2003. P. 709–726. https://doi.org/10.1016/B0-12-227410-5/00238-6

*Tao R., Gong Z., Ma Q., Thill J.-C.* Boosting Computational Effectiveness in Big Spatial Flow Data Analysis with Intelligent Data Reduction // ISPRS International Journal of Geo-Information. 2020. V. 9. Is. 5. 299. https://doi.org/10.3390/ijgi9050299

*Teillet P.M.* Image correction for radiometric effects in remote sensing // International J. Remote Sensing. 1986. V. 7. Is. 12. P. 1637–1651. https://doi.org/10.1080/01431168608948958

*Thébault E., Finlay C.C., Beggan C.D. et al.* International Geomagnetic Reference Field: the 12th generation // Earth, Planets and Space. 2015. V. 67. 79. https://doi.org/10.1186/s40623-015-0228-9

*Thomas J.W., Hoover G.M.* Exploration Seismology. Encyclopedia of Geology (Second Edition). 2021. https://doi.org/10.1016/B978-0-12-409548-9.12538-2

*Thorne P.W., Allan R.J., Ashcroft L. et al.* Toward an Integrated Set of Surface Meteorological Observations for Climate Science and Applications // Bulletin of the American Meteorological Society. 2017. V. 98. P. 2689–2702. https://doi.org/10.1175/BAMS-D-16-0165.1

*Toth C., Jóźków G.* Remote sensing platforms and sensors: A survey // ISPRS J. Photogrammetry and Remote Sensing. 2016. V. 115. P. 22–36. https://doi.org/10.1016/j.isprsjprs.2015.10.004

*Van Allen J.A., Frank L.A.* Radiation around the Earth to a radial distance of 107400 km // Nature. 1959. V. 183. P. 430–434.

*Van Allen J.A., Ludwig G.H., Ray E.C., McIlwain C.E.* Observation of high intensity radiation by satellites 1958 Alpha and Gamma. U.S. National Academy of Sciences. I.G.Y. Satellite Report Series 3. 1958. P. 73–92. https://doi.org/10.2514/8.7396

*Vihma T., Uotila P., Sandven S. et al.* Towards an advanced observation system for the marine Arctic in the framework

of the Pan-Eurasian Experiment (PEEX) // Atmospheric Chemistry and Physics. 2019. V. 19. Is. 3. P. 1941–1970. https://doi.org/10.5194/acp-19-1941-2019

*Voulodimos A., Doulamis N., Doulamis A., Protopapadakis E.* Deep learning for computer vision: A brief review // Computational intelligence and neuroscience. 2018. V. 2018. 7068349.
https://doi.org/10.1155/2018/7068349

*Watson J.C.* Establishing Evidence for Internal Structure Using Exploratory Factor Analysis // Measurement and Evaluation in Counseling and Development. 2017. V. 50. Is. 4. P. 232–238.
https://doi.org/10.1080/07481756.2017.1336931

*Weyn J.A., Durran D.R., Caruana R.* Can machines learn to predict weather? Using deep learning to predict gridded 500-hPa geopotential height from historical weather data // J. Advances in Modeling Earth Systems. 2019. V. 11. Is. 8. P. 2680–2693.
https://doi.org/10.1029/2019MS001705

*Whaler K.* Geomagnetism in the satellite era // Astronomy & Geophysics. 2007. V. 48. Is. 2. P. 23–29.

*Wu H., Li Z.-L.* Scale issues in remote sensing: a review on analysis, processing and modeling // Sensors. 2009. V. 9. P. 1768–1793.
https://doi.org/10.3390/s90301768

*Wu S.R., Li X., Apul D., Breeze V., Tang Y., Fan Y., Chen J.* Agent-based modeling of temporal and spatial dynamics in life cycle sustainability assessment // J. Industrial Ecology. 2017. V. 21. Is. 6. P. 1507–21.
https://doi.org/10.1111/jiec.12666

*Zhang S., Yao L., Sun A., Tay Y.* Deep learning based recommender system: A survey and new perspective // ACM Computing Surveys (CSUR). 2019. V. 52. Is. 1. P. 1–38.
https://doi.org/10.1145/3285029

*Zhao S., Wang Q., Li Y., Liu S., Wang Z., Zhu L., Wang Z.* An overview of satellite remote sensing technology used in China's environmental protection // Earth Science Informatics. 2017. V. 10. P. 137–148.
https://doi.org/10.1007/s12145-017-0286-6

*Zilitinkevich S.S., Elperin T., Kleeorin N., Rogachevskii I., Esau I.* A Hierarchy of Energy- and Flux-Budget (EFB) Turbulence Closure Models for Stably-Stratified Geophysical Flows // Boundary-Layer Meteorology. 2013. V. 146. Is. 3. P. 341–373.
https://doi.org/10.1007/s10546-012-9768-8

# Big Data in Geophysics and Other Earth Sciences

**A. D. Gvishiani[a,b], M. N. Dobrovolskii[a], B. V. Dzeranov[a,c], and B. A. Dzeboev[a,c,*]**

[a] *Geophysical Center, Russian Academy of Sciences, Moscow, Russia*

[b] *Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, Russia*

[c] *Geophysical Institute, Vladikavkaz Scientific Center, Russian Academy of Sciences, Vladikavkaz, Russia*

**e-mail: b.dzeboev@gcras.ru*

The term "Big Data" has become very popular over the past decade. The frequency of its use in the scholarly papers, reports, and broad press has been steadily increasing. This work describes the origin and development of the theory and practice of Big Data as a scientific discipline, outlines the main characteristics of and methods for Big Data processing and analysis, discusses the formalism and family of Big Data V-characteristics, and presents the examples of the sources of the growing Big Data which have fundamental effect on the development of geophysics and related Earth sciences. The examples of the sources of Big Data in the Earth sciences are remote sensing, meteorology, geoecology (in terms of the global hierarchical network SMEAR (Stations Measuring Earth surfaces and Atmosphere Relations)), and seismic prospecting. Besides, we discuss seismic monitoring data which can become Big Data when combined with other geophysical information and consider geomagnetic data which are not Big Data but nevertheless have a great scientific value.

*Keywords:* Big Data, Earth sciences, remote sensing of the Earth, meteorological observations, seismic monitoring, seismic prospecting, geoecology, geomagnetic observations