
ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ

INFORMATICS AND INFORMATION PROCESSES

УДК 004.67: 537.86.029
DOI: 10.17586/0021-3454-2023-66-12-993-1001

ПРИМЕНЕНИЕ МЕТОДОВ BIG DATA ДЛЯ СРАВНЕНИЯ ДАННЫХ ГЕОМАГНИТНЫХ ОБСЕРВАТОРИЙ СЕТИ INTERMAGNET

А. Г. КОРОБЕЙНИКОВ

*Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн
им. Н. В. Пушкова Российской академии наук, Санкт-Петербург, Россия
Korobeynikov_A_G@mail.ru*

Аннотация. Методы обработки больших данных применяют для решения различных задач, например, сбора, хранения, анализа, визуализации и интерпретации больших массивов информации, получаемой из различных источников: сети Интернет, мобильных приложений и социальных сетей. Применение специальных технологий и инструментария, например MapReduce, Hadoop, Spark, ускоряет процесс за счет параллельной и распределенной обработки данных. Проведено сравнение данных пяти геомагнитных обсерваторий, входящих в международную сеть INTERMAGNET, при помощи визуализации, являющейся одной из составляющих технологии Big Data. В каждой обсерватории сети INTERMAGNET информация о текущем состоянии магнитного поля Земли собирается при помощи специально аттестованной магнитометрической аппаратуры. Достаточно часто вызывает научный и практический интерес анализ этой информации, полученной на большом интервале времени. В этом случае информация представляет собой большие данные, т.е. данные, которые не помещаются в оперативную память используемого компьютера. Представлены графики исходных данных наблюдений за период с 1 января 2018 г. по 31 июля 2023 г. В качестве инструментария использована система MatLab, где реализованы методы Big Data.

Ключевые слова: *Big Data, INTERMAGNET, MatLab, хранилище данных, магнитное поле Земли, геомагнитная обсерватория*

Благодарности: работа выполнена при поддержке гранта РФФИ № 23-27-00011.

Ссылка для цитирования: *Коробейников А. Г.* Применение методов Big Data для сравнения данных геомагнитных обсерваторий сети INTERMAGNET // Изв. вузов. Приборостроение. 2023. Т. 66, № 12. С. 993—1001. DOI: 10.17586/0021-3454-2023-66-12-993-1001.

APPLICATION OF BIG DATA METHODS FOR COMPARING DATA OF GEOMAGNETIC OBSERVATORIES IN THE INTERMAGNET NETWORK

A. G. Korobeynikov

*Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the RAS,
St. Petersburg, Russia
Korobeynikov_A_G@mail.ru*

Abstract. Big data processing methods are used to solve various problems, for example, collecting, storing, analyzing, visualizing and interpreting large amounts of information received from various sources: the Internet, mobile applications and social networks. The use of special technologies and tools, such as MapReduce, Hadoop, Spark, speeds up the process due to parallel and distributed data processing. A comparison of data from five geomagnetic observatories included in the international INTERMAGNET network is carried out using visualization, which is one of the components of Big Data technology. In each observatory of the INTERMAGNET network, information about the current state of the Earth's magnetic field is collected using specially certified magnetometric equipment. Quite often the analysis of this information obtained over a long period is of scientific and practical interest. In this case, the information is big data, that

is, data that does not fit into the RAM of the computer being used. Graphs of initial observation data for the period from January 1, 2018 to July 31, 2023 are presented. The MatLab system with Big Data methods implemented in it, is used as a toolkit.

Keywords: Big Data, INTERMAGNET, MatLab, datastore, Earth's magnetic field, geomagnetic observatory

Acknowledgments: This work was supported by the Russian Science Foundation, grant No. 23-27-00011.

For citation: Korobeynikov A. G. Application of Big Data methods for comparing data of geomagnetic observatories in the INTERMAGNET network. *Journal of Instrument Engineering*. 2023. Vol. 66, N 12. P. 993—1001 (in Russian). DOI: 10.17586/0021-3454-2023-66-12-993-1001.

Введение. Данные большого объема применяют при решении многих научных и практических задач различной тематики, например, связанных с геомагнетизмом [1]. Необходимо отметить, что в российской науке принято под термином Big Data подразумевать не только объект исследования (сами большие массивы данных), но также технологии их обработки. Под такими технологиями понимается программное обеспечение, основное назначение которого состоит в анализе, обработке, визуализации и извлечении информации из больших наборов данных, которые нельзя обработать традиционными технологиями. Таким образом, ключевым объектом исследования специалистов становится не объем информации, а возможности по ее обработке.

Основные отличия подходов Big Data от традиционной обработки больших данных следующие: обработка сразу всего объема данных вместо последовательного анализа относительно небольших массивов; обработка данных в их исходном виде вместо необходимости их предварительного редактирования [2].

Обработку больших объемов данных производят при помощи различных инструментов, таких как: MapReduce — модули сегментирования данных с последующей параллельной обработкой; Hadoop — распределенная файловая система для хранения и обработки больших объемов данных; Spark — фреймворк, обрабатывающий данные на основе распределенного хранилища.

В работе использован программный инструмент MatLab, при помощи которого можно решать разные задачи [3—11], используя технологии и методы работы с большими данными, например MapReduce и Hadoop.

Постановка задачи. Отметим, что под „большими“ в работе понимаются данные, не помещающиеся в оперативную память компьютера [1, 2].

В процессе написания статьи использованы исходные данные, полученные в международной сети геомагнитных обсерваторий INTERMAGNET (International Real-Time Magnetic Observatory Network) [12]*. К этим обсерваториям предъявляются строгие требования, диктуемые необходимостью обеспечения точности измерений. INTERMAGNET обеспечивает доступ к данным 150 геомагнитных обсерваторий, расположенных по всему миру. Эти данные используют для различных научных исследований, например, при изучении геомагнитных бурь, полярных сияний, геомагнитных возмущений.

Необходимо отметить, что с появлением технологии Big Data у исследователей геомагнитных явлений появились надежные инструменты для решения различных научных задач, где необходимы данные сети INTERMAGNET за достаточно длительный период. Постановка и решение таких задач помогает ученым лучше понимать процессы, происходящие, например, в магнитосфере Земли, а также прогнозировать влияние этих процессов на работоспособность современной техники.

* <https://www.intermagnet.org>.

В работе представлены результаты сравнения данных о геомагнитном поле, получаемых с пяти геомагнитных обсерваторий: входящая в состав СПбФ ИЗМИРАН геомагнитная обсерватория „Санкт-Петербург“ (международный IAGA-код SPG), геомагнитная обсерватория „Nurmijarvi“ (Финляндия, Финский метеорологический институт, международный IAGA-код NUR), геомагнитная обсерватория „Sodankyla“ (Финляндия, авроральная обсерватория, международный IAGA-код SOD), геомагнитная обсерватория „Lycksele“ (Швеция, Геологическая служба Швеции, международный IAGA-код LYC), геомагнитная обсерватория „Abisko“ (Швеция, Геологическая служба Швеции, международный IAGA-код АВК). Выбор этих пунктов сбора информации обусловлен расположением их на относительно небольших расстояниях друг от друга. Сами исходные данные представляют собой результаты измерения компонентов состояния магнитного поля Земли (МПЗ) без антропогенных возмущений и скорректированные с учетом абсолютных измерений в реальном масштабе времени с частотой дискретизации $f_d = 1 \text{ мин}^{-1}$ за период с 1 января 2018 по 31 июля 2023 года (объем этих данных превышает объем оперативной памяти компьютера, поэтому их обработка возможна при помощи методов работы с большими данными).

Сравнение данных геомагнитных обсерваторий проведено с использованием одной из технологий Big Data — Data Visualization, с целью наглядного представления различий и сходств.

Обработка данных. В настоящее время технологии обработки больших данных обеспечивают: хранение; интеллектуальный анализ данных; визуализацию.

Последовательность действий, выполняемых при решении поставленной задачи, начинается с переноса данных из базы INTERMAGNET любым удобным для пользователя способом на компьютер пользователя в отдельный раздел, например Intermagnet Data, который будет являться во время работы с MatLab подразделом рабочей директории (Current Folder). Для работы с такими данными необходимо применять специальные средства, например, Datastore (хранилище данных) или tall (высокие массивы) [1, 2]. В системе MatLab, используя „живой“ редактор (Live Editor), это можно сделать, например, так:

```
clc clear
fileName = 'spg20*.*';
f_Datastores
=fileDatastore(fileName, 'ReadFcn', @read_file, 'FileExtensions', '.min');
My_Datastore = readall(f_Datastores);
Name_Station=[];
for j=1 : size(My_Datastore,1)
    work_Datastore =My_Datastore{j};
    work_Datastore.DATETIME=datetime(work_Datastore.DATETIME, ...
        'InputFormat', 'yyyy-MM-dd HH:mm:ss.SSS');
    work_Datastore =table2timetable(work_Datastore);
    work_Datastore = removevars(work_Datastore, "DOY");
    Name_Station=[Name_Station; work_Datastore];
end
Name_Station = sortrows(Name_Station, "DATETIME");
clearvars work_Datastore;
stackedplot(Name_Station);
function My_table = read_file(filename, in1, in2)
    if nargin<=2
        in1 = 29; in2 = inf;
    end
    My_s = '%24s%3f%13f%10f%10f%f%[\n\r]';
    file_ID = fopen(filename, 'r');
    My_d_A = textscan(file_ID, My_s, in2(1)-in1(1)+1, ...
        'Delimiter', ',', 'WhiteSpace', '', 'TextType', 'string', 'HeaderLines', ...
        in1(1)-1, 'ReturnOnError', false, 'EndOfLine', '\r\n');
    for block=2:length(in1)
        frewind(fileID);
        My_d_ABlock = textscan(fileID, My_s, ...
```

```

in2(block)-in1(block)+1, 'Delimiter', ',', 'WhiteSpace', ' ', ...
'TextType', 'string', 'HeaderLines', in1(block)-1, 'ReturnOnError', ...
false, 'EndOfLine', '\r\n');
for col=1:length(My_d_A)
    My_d_A{col} = [My_d_A{col};My_d_ABlock{col}];
end
end
My_d_A{1} = strtrim(My_d_A{1});
fclose(file_ID);
My_table = table(My_d_A{1:end-1}, 'VariableNames', ...
    {'DATETIME', 'DOY', 'SPGX', 'SPGY', 'SPGZ', 'SPGF'});
end

```

В результате выполнения этого кода для визуализации данных на экран компьютера будет выведен график (рис. 1, исходные данные с SPG за период 01.01.2018—31.07.2023); X , Y , Z — компоненты данных; F — их модуль (корень квадратный суммы квадратов компонентов).

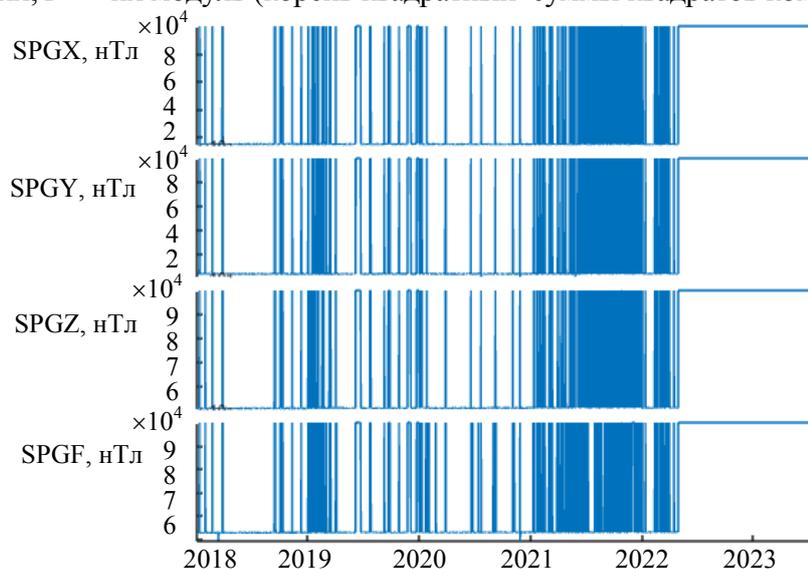


Рис. 1

Для построения графика другой обсерватории следует в исходном коде заменить кодовое наименование станции, например, SPG на LYC (на рис. 2—5 представлены графики исходных данных со станций NUR, LYC, SOD и АВК).

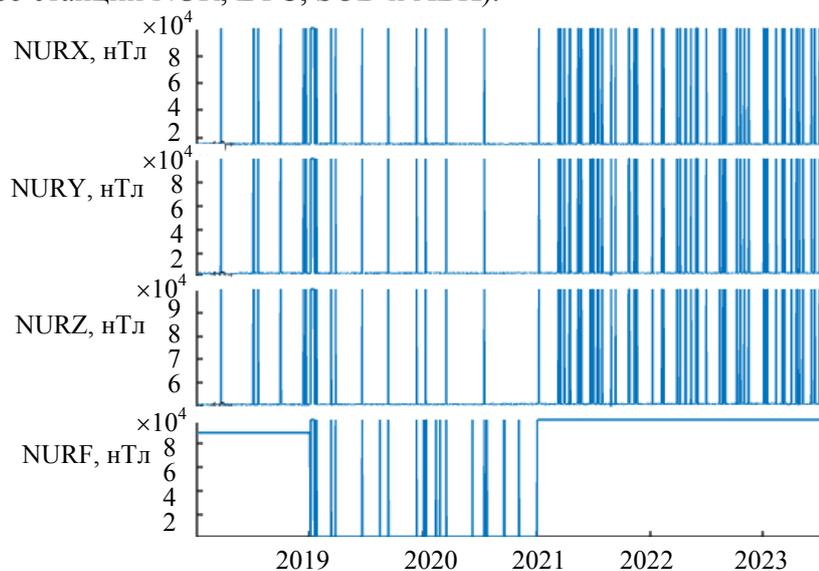


Рис. 2

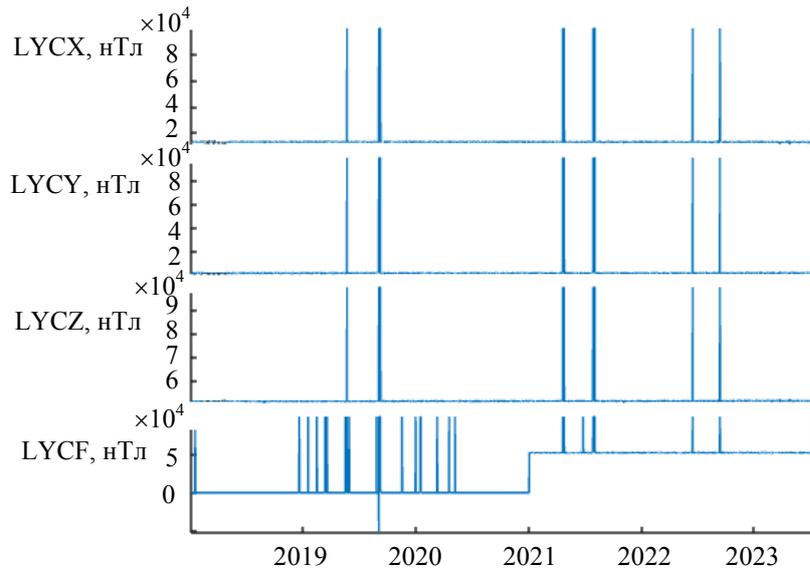


Рис. 3

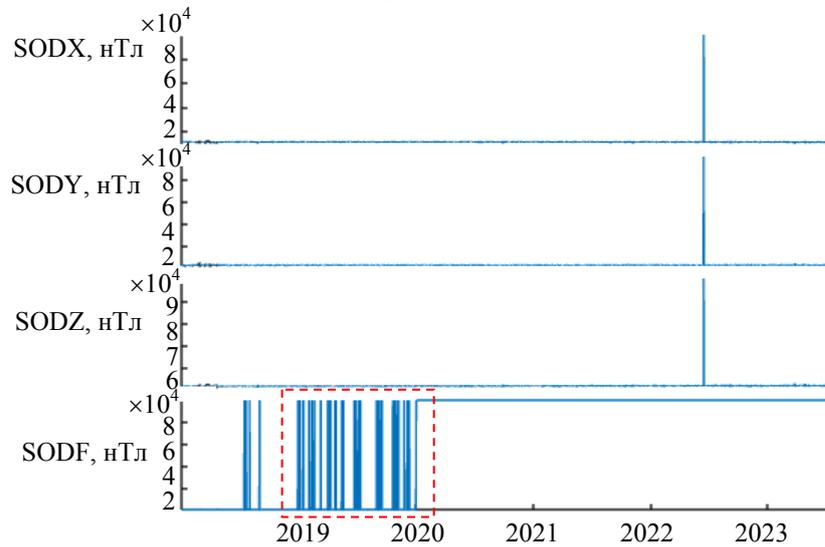


Рис. 4

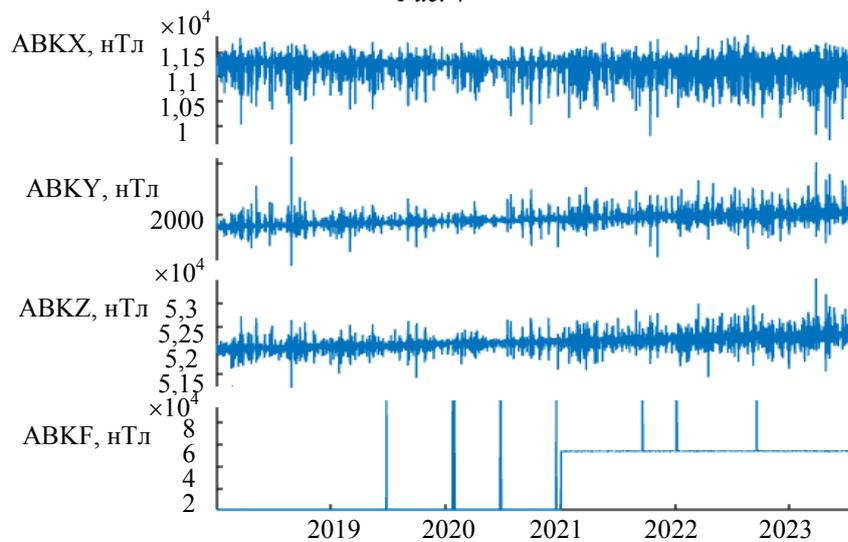


Рис. 5

Для анализа выбран интервал времени 01.03—14.03.2020, когда все станции работали без сбоев. Анализируемый интервал задается при помощи следующего кода:

```
Interval_begin=find(Name_Station.DATETIME=="2020-03-01 19:39:00.000");
```

```
Interval_end=find(Name_Station.DATETIME=="2020-03-14 07:08:00.000");
stackedplot(Name_Station (Interval_begin: Interval_end,:));
```

На рис. 6—10 приведены в графической форме исходные данные исследуемых обсерваторий.

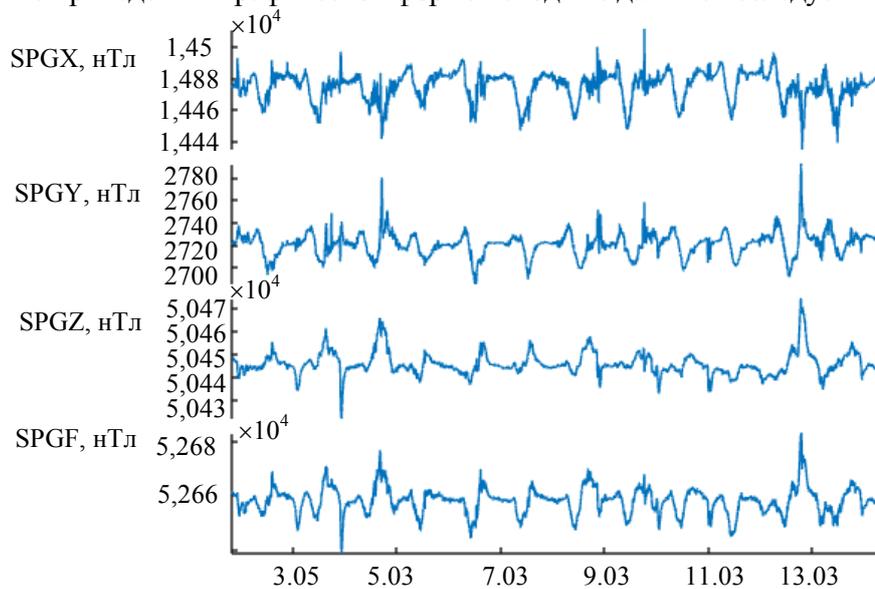


Рис. 6

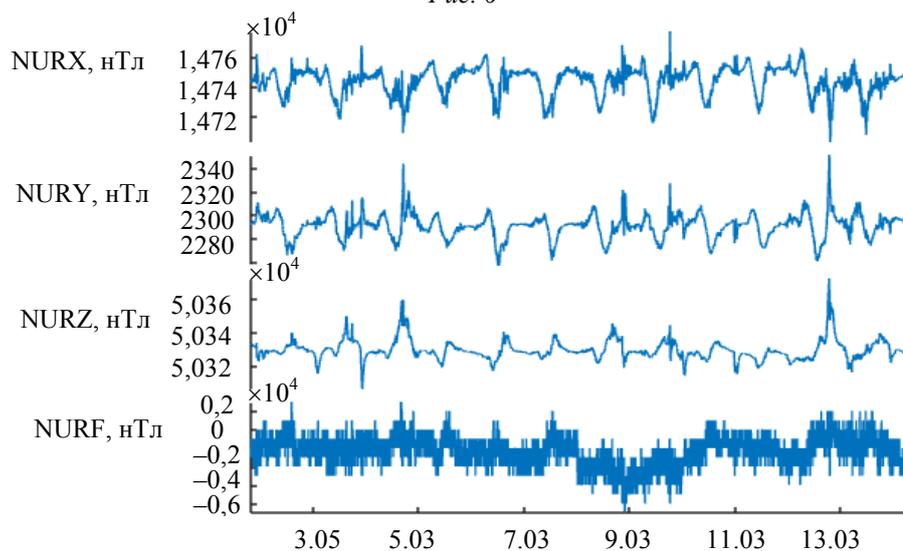


Рис. 7

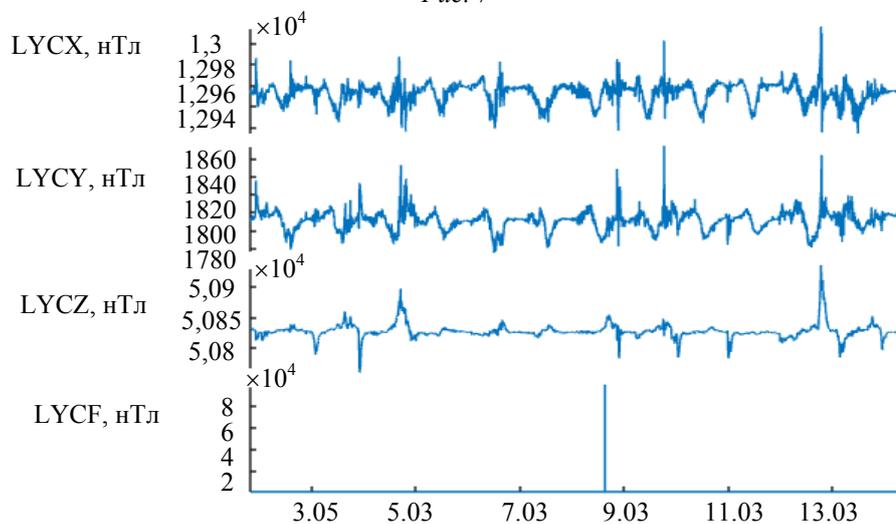


Рис. 8

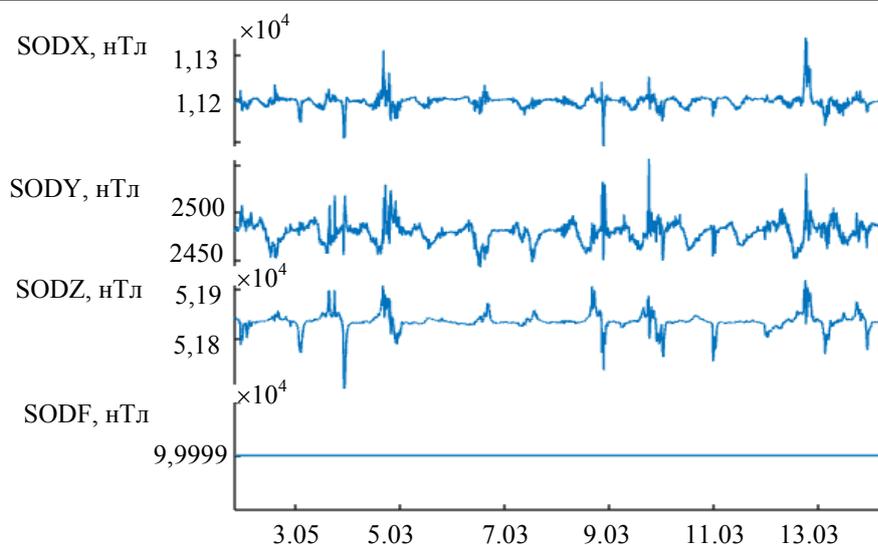


Рис. 9

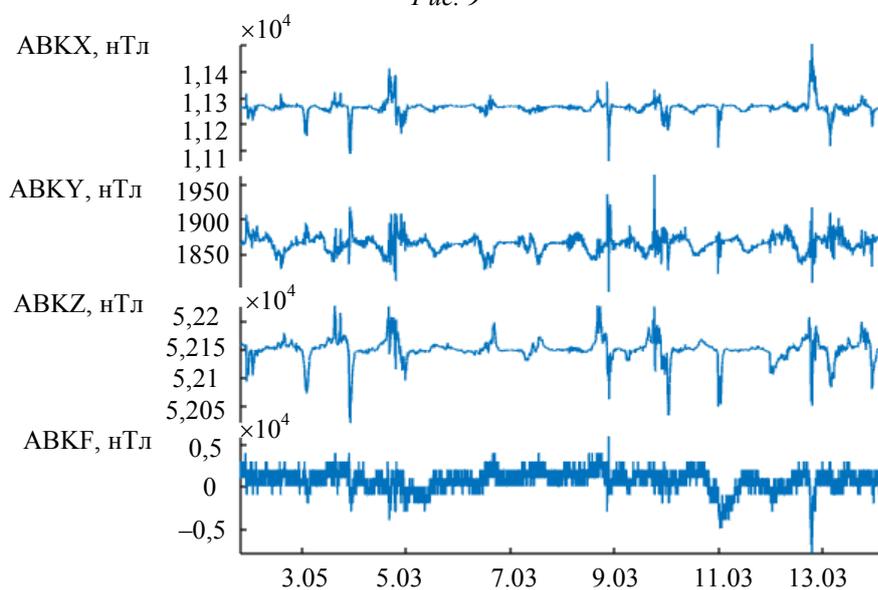


Рис. 10

Обсуждение результатов. Анализ представленных на рис. 1—5 результатов визуализации показывает, что передаваемые в INTERMAGNET данные геомагнитной обсерватории „Abisko“ в период 01.01.2018—31.07.2020 содержали меньше всего пропусков, по сравнению с другими обсерваториями. На втором месте, за исключением параметра F , обсерватория „Sodankyla“. Замыкает список обсерватория „Санкт-Петербург“.

А вот с 1 по 14 марта 2020 г. обсерватория „Санкт-Петербург“ передавала в INTERMAGNET данные с наименьшим числом пропусков. Обсерватория „Abisko“ стоит на втором месте, „Lycksele“ замыкает тройку. На обсерватории „Sodankyla“ не работал датчик полного поля, согласно рис. 4 (выделено пунктиром).

Кроме того, визуальный анализ позволяет делать и другие выводы. Например, анализ данных полного поля с обсерватории „Lycksel“ показывает, что датчик не работал до конца 2020 г., но с начала 2021 г. пошла достоверная информация. Это видно из рис. 4, где $F \approx 52\,000$ нТл, т.е. можно сделать вывод, что сенсор полного поля установили в самом конце 2020 года.

Заключение. На основании представленных результатов можно утверждать, что MatLab позволяет эффективно работать с большими массивами данных при решении задач в различных предметных областях, в частности, связанных с геофизикой. Следует также отметить, что применение графических процессоров позволяет существенно увеличить скорость

обработки данных. Отсюда следует, что применение технологий Big Data, реализованных в MatLab, окажет существенную поддержку исследователям при решении фундаментальных и прикладных задач.

СПИСОК ЛИТЕРАТУРЫ

1. Коробейников А. Г. Применение методов BIG DATA для предобработки данных сети INTERMAGNET // Изв. вузов. Приборостроение. 2023. Т. 66, № 7. С. 533—538. DOI: 10.17586/0021-3454-2023-66-7-533-538.
2. Макишанов А. В., Журавлев А. Е., Тындыкарь Л. Н. Большие данные. Big Data. СПб: Лань, 2022. 188 с. ISBN 978-5-8114-9834-5.
3. Korobeynikov A. G., Grishentsev A. Y., Velichko E. N., Aleksanin S. A., Fedosovskii M. E., Bondarenko I. B., Korikov C. C. Calculation of Regularization Parameter in the Problem of Blur Removal in Digital Image // Optical Memory & Neural Networks (Information Optics). 2016. Vol. 25, N 3. P. 184—191.
4. Дьяконов В. П. MATLAB и SIMULINK для радиоинженеров. М.: ДМК Пресс, 2016. 976 с.
5. Новгородцев А. Б. Расчет электрических цепей в MATLAB: Учебный курс. СПб: Питер, 2004. 250 с.
6. Матюшкин И. В. Моделирование и визуализация средствами MATLAB физики наноструктур. М.: Техносфера, 2011. 168 с.
7. Korobeynikov A. G., Fedosovsky M. E., Zharinov I. O., Shukalov A. V., Gurjanov A. V. Development of conceptual modeling method to solve the tasks of computer-aided design of difficult technical complexes on the basis of category theory // Intern. J. of Applied Engineering Research. 2017. Vol. 12, N 6. P. 1114—1122.
8. Гайдук А. Р., Беляев В. Е., Пьявченко Т. А. Теория автоматического управления в примерах и задачах с решениями в MATLAB: Учеб. пособие. СПб: Лань, 2016. 464 с.
9. Поршнев С. В. Компьютерное моделирование физических процессов в пакете MATLAB: Учеб. пособие. СПб: Лань, 2011. 736 с.
10. Фриск В. В., Ганин В. И., Степанова А. Г. Компьютерный анализ и моделирование электрических цепей постоянного тока в среде MATLAB: учеб.-метод. пособие. М.: СОЛОН-ПРЕСС, 2021. 32 с.
11. Коробейников А. Г. Применение искусственных нейронных сетей в системах автоматического управления магнитной левитацией // Программные продукты и системы. 2022. Т. 35, № 3. С. 452—457. DOI: 10.15827/0236-235X.139.452-457.
12. Коробейников А. Г. Обработка и анализ данных с российского сегмента мировой сети магнитных обсерваторий ИНТЕРМАГNET // Междунар. журн. гуманитарных и естественных наук. 2018. № 8. С. 91—98.

Сведения об авторе

Анатолий Григорьевич Коробейников

— д-р техн. наук, профессор; Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн им. Н. В. Пушкова Российской академии наук; дирекция; зам. директора по науке;
E-mail: Korobeynikov_A_G@mail.ru

Поступила в редакцию 21.08.2023; одобрена после рецензирования 30.08.2023; принята к публикации 27.10.2023.

REFERENCES

1. Korobeynikov A.G. *Journal of Instrument Engineering*, 2023, no. 7(66), pp. 533–538, DOI: 10.17586/0021-3454-2023-66-7-533-538. (in Russ.)
2. Makshanov A.V., Zhuravlev A.E., Tyndykar L.N. *Bol'shiye dannyye. Big data (Big Data. Big Data)*, St. Petersburg, 2022, 188 p., ISBN 978-5-8114-9834-5. (in Russ.)
3. Korobeynikov A.G., Grishentsev A.Y., Velichko E.N., Aleksanin S.A., Fedosovskii M.E., Bondarenko I.B., Korikov C.C. *Optical Memory & Neural Networks (Information Optics)*, 2016, no. 3(25), pp. 184–191.
4. Dyakonov V.P. *MATLAB i SIMULINK dlya radioinzhenerov (MATLAB and SIMULINK for Radio Engineers)*, Moscow, 2016, 976 p. (in Russ.)
5. Novgorodtsev A.B. *Raschet elektricheskikh tsepey v MATLAB (Calculation of Electrical Circuits in MATLAB)*, St. Petersburg, 2004, 250 p. (in Russ.)

6. Matyushkin I.V. *Modelirovaniye i vizualizatsiya sredstvami MATLAB fiziki nanostruktur* (Modeling and Visualization by Means of MATLAB of the Physics of Nanostructures), Moscow, 2011, 168p. (in Russ.)
7. Korobeynikov A.G., Fedosovsky M.E., Zharinov I.O., Shukalov A.V., Gurjanov A.V. *International Journal of Applied Engineering Research*, 2017, no. 6(12), pp. 1114–1122.
8. Gaiduk A.R., Belyaev V.E., Pyavchenko T.A. *Teoriya avtomaticheskogo upravleniya v primerakh i zadachakh s resheniyami v MATLAB* (Theory of Automatic Control in Examples and Problems with Solutions in MATLAB), St. Petersburg, 2016, 464 p. (in Russ.)
9. Porshnev S.V. *Komp'yuternoye modelirovaniye fizicheskikh protsessov v pakete MATLAB* (Computer Simulation of Physical Processes in the MATLAB Package), St. Petersburg, 2011, 736 p. (in Russ.)
10. Frisk V.V., Ganin V.I., Stepanova A.G. *Komp'yuternyy analiz i modelirovaniye elektricheskikh tsepey postoyannogo toka v srede MATLAB* (Computer Analysis and Modeling of DC Electrical Circuits in the MATLAB Environment), Moscow, 2021, 32 p. (in Russ.)
11. Korobeynikov A.G. *Software & Systems*, 2022, no. 3(35), pp. 452–457, DOI: 10.15827/0236-235X.139.452-457. (in Russ.)
12. Korobeynikov A.G. *International Journal of Humanities and Natural Sciences*, 2018, no. 8, pp. 91–98. (in Russ.)

Data on author

Anatoly G. Korobeynikov

— Dr. Sci., Professor; Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the RAS, St. Petersburg Branch, Directorate; Deputy Director of Science; E-mail: Korobeynikov_A_G@mail.ru

Received 21.08.2023; approved after reviewing 30.08.2023; accepted for publication 27.10.2023.