

---

## ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ

### INFORMATICS AND INFORMATION PROCESSES

---

УДК 004.67: 537.86.029  
DOI: 10.17586/0021-3454-2023-66-7-533-538

#### ПРИМЕНЕНИЕ МЕТОДОВ BIG DATA ДЛЯ ПРЕДОБРАБОТКИ ДАННЫХ ИЗ СЕТИ INTERMAGNET

А. Г. КОРОБЕЙНИКОВ

*Санкт-Петербургский филиал  
Института земного магнетизма, ионосферы и распространения радиоволн им. Н. В. Пушкова РАН,  
Санкт-Петербург, Россия  
Korobeynikov\_A\_G@mail.ru*

**Аннотация.** Достаточно часто при решении геофизических задач, связанных с магнитным полем Земли, возникает необходимость обработки данных большого объема. Полученная в результате анализа информация может служить основой для решения различных фундаментальных задач, например изучения движения магнитных полюсов, или прикладных задач, например навигации по магнитному полю Земли. Кроме того, во многих случаях представляет интерес анализ данных за продолжительный промежуток времени, что значительно увеличивает объем данных, которые необходимо обработать. Применение современных технологий для работы с большими данными, такой как, например, BIG DATA, позволяет решать достаточной широкий класс геофизических задач. Рассмотрена задача предобработки данных измерений состояния магнитного поля Земли, предоставленных геомагнитной обсерваторией Lycksele (Швеция), входящей в международную сеть INTERMAGNET. Предобработка произведена с помощью методов BIG DATA, реализованных в системе MatLab.

**Ключевые слова:** BIG DATA, INTERMAGNET, MatLab, высокие массивы, геомагнитная обсерватория, магнитное поле Земли, хранилище данных

**Благодарности:** работа выполнена при поддержке Российского научного фонда, грант № 23-27-00011.

**Ссылка для цитирования:** Коробейников А. Г. Применение методов BIG DATA для предобработки данных из сети INTERMAGNET // Изв. вузов. Приборостроение. 2023. Т. 66, № 7. С. 533—538. DOI: 10.17586/0021-3454-2023-66-7-533-538.

#### APPLICATION OF BIG DATA METHODS FOR INTERMAGNET DATA PREPROCESSING

A. G. Korobeynikov

*Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the RAS,  
St. Petersburg Branch, St. Petersburg Russia  
Korobeynikov\_A\_G@mail.ru*

**Abstract.** When solving geophysical problems related to the Earth's magnetic field, quite often there is a need to process large volume data. The information obtained as a result of the analysis can serve as the basis for solving various fundamental problems, for example, studying the movement of the magnetic poles, or applied problems, such as navigating the Earth's magnetic field. In addition, in many cases it is of interest to analyze data over a long period of time, which significantly increases the amount of data that needs to be processed. The use of modern technologies for working with large amounts of data, such as, for example, BIG DATA, allows solving a fairly wide class of geophysical problems. The problem of preprocessing the data of measurements of the state of the Earth's magnetic field provided by geomagnetic observatory Lycksele (Sweden) included in the international network INTERMAGNET. Preprocessing was performed using the BIG DATA methods implemented in the MatLab system.

**Keywords:** BIG DATA, INTERMAGNET, MatLab, tall array, geomagnetic observatory, Earth's magnetic field, datastore

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

**For citation:** Korobeynikov A. G. Application of BIG DATA methods for INTERMAGNET data preprocessing. *Journal of Instrument Engineering*. 2023. Vol. 66, N 7. P. 533—538 (in Russian). DOI: 10.17586/0021-3454-2023-66-7-533-538.

**Введение.** Система MatLab в настоящее время применяется для решения разнообразных задач в различных областях [1—10]. В настоящей статье представлены результаты, относящиеся к предметной области, связанной с применением технологий работы с большими данными (BIG DATA) для решения геофизических задач. Все возрастающий интерес к BIG DATA фактически обусловлен тем, что достаточно часто в процессе решения геофизических задач возникает необходимость обработки данных большого объема. Полученная в результате предобработки информация может служить основой для решения различных фундаментальных и прикладных задач. В случаях, связанных с данными о магнитном поле Земли (МПЗ), к множеству фундаментальных задач относятся, например, изучение движения магнитных полюсов, анализ геомагнитных пульсаций, являющихся ультранизкочастотными колебаниями МПЗ, и т.д. К множеству прикладных задач относятся, например, навигация по МПЗ, поиск полезных ископаемых и т.д. Для решения указанных задач требуются данные за большой период времени, что является причиной экспоненциального увеличения объема обрабатываемых данных. Отсюда следует необходимость применения современных технологий для работы с большими данными (BIG DATA). Реализация методов, базирующихся на применении технологии BIG DATA, осуществляется с помощью различного инструментария [11]. В данной работе в качестве такого инструментария используется MatLab.

Исходные данные измерений состояния МПЗ в реальном масштабе времени с частотой дискретизации  $f_d = 1$  мин были взяты на сайте <http://www.intermagnet.org> — официальном сайте международной сети INTERMAGNET (International Real-Time Magnetic Observatory Network) [12], где в настоящее время имеются данные о более чем 150 геомагнитных обсерваториях, включенных в эту сеть. Для проверки предлагаемых методов были использованы данные, предоставленные геомагнитной обсерваторией Lycksele (Швеция, Геологическая служба Швеции, международный IAGA-код LYC)\*, с 1 января 2018 г. по 31 декабря 2022 г.

Вычислительные эксперименты были проведены с помощью системы MatLab под управлением операционной системы Windows [13].

**Подготовка данных.** В работе под большими данными понимаются данные, которые не помещаются в оперативную память используемого компьютера.

Данные с сайта INTERMAGNET необходимо перенести любым удобным для пользователя способом в раздел, например LYC\Data, который во время работы с MatLab будет подразделом рабочей директории (Current Folder). Так как эти файлы из-за большого объема не поместятся в память компьютера, для их обработки применим методы работы с большими данными, а именно механизмы Datastore (хранилище данных) и tall (высокие массивы) [11]. В MatLab это можно сделать следующим образом.

Заведем переменную с именами файлов и занесем их (имена файлов) в хранилище:

```
fileName_LYC = 'LYC\Data\lyc20*.min';  
files_Datastores_LYC = ...  
fileDatastore(fileName_LYC, 'ReadFcn', @read_LYC_file, 'FileExtensions', '.min');  
LYC_Datastore = readall(files_Datastores_LYC);
```

\* Необходимую информацию об этой обсерватории можно найти по адресу: <https://www.intermagnet.org>.

Заведем высокий массив:

```
my_tall_array_LYC = tall(LYC_Datastore);
```

Данная операция включает механизм автоматической оптимизации вычислений в очереди за счет минимизации количества проходов и отложенного выполнения. Кроме того, включается механизм распараллеливания, т.е. в случае многоядерного процессора к обработке подключаются все ядра (workers). Далее выполняем следующие команды:

```
%  
% Производим необходимые действия над высокими массивами  
%  
my_array_cell_LYC = gather(my_tall_array_LYC);
```

**Визуализация данных.** Для визуализации исходных данных целесообразно воспользоваться специальными таблицами — timetable. Это можно сделать, например, с помощью следующих команд:

```
LYC=[];  
for k=1 : size(my_array_cell_LYC,1)  
    TT=my_array_cell_LYC{k};  
    TT.DATETIME=datetime(TT.DATETIME, ...  
        'InputFormat','yyyy-MM-dd HH:mm:ss.SSS');  
    TT=table2timetable(TT);  
    LYC=[LYC;TT];  
end  
LYC = sortrows(LYC,"DATETIME");  
clearvars TT;
```

Далее с помощью команды

```
stackedplot(LYC);
```

выводим на экран графики исходных данных за 2018—2022 гг. (рис. 1).

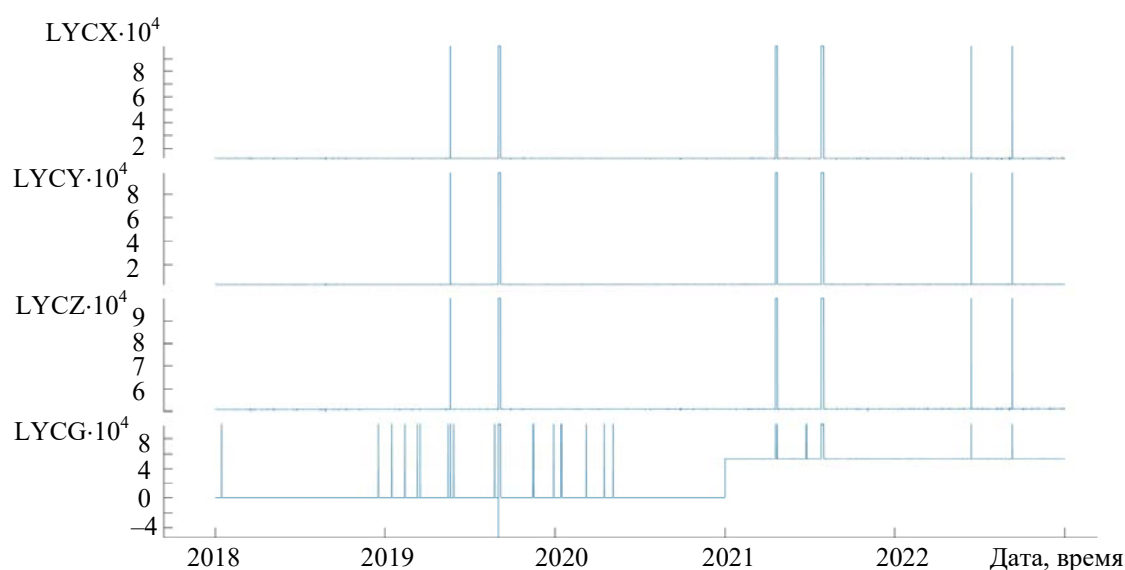


Рис. 1

На рисунке явно видны „выбросы“ по всем компонентам. Поэтому необходимо произвести предобработку данных (убрать „выбросы“), для чего можно использовать Live Editor („живой“ редактор) в разделе Task или команду

```
[newLYC,outlierIndices,thresholdLow,thresholdHigh] = filloutliers(LYC,...  
    "linear", "movmedian", days(18));
```

После этого выводим на дисплей „подчищенные“ данные (рис. 2):

```
stackedplot(newLYC);
```

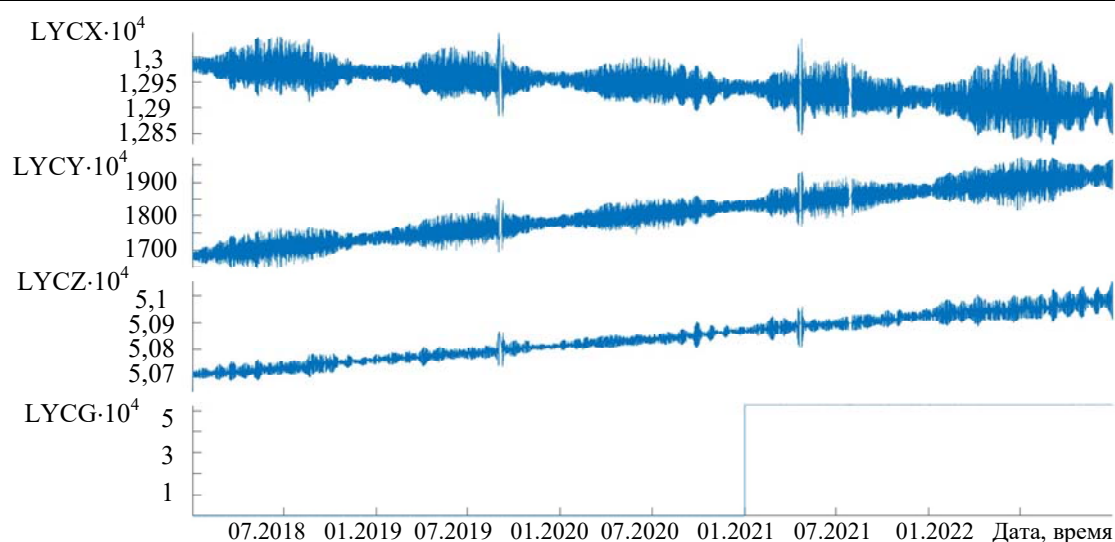


Рис. 2

Для подтверждения результатов анализа, представленного ниже, приведем график очищенных данных начиная с 01.01.2021 г. (рис. 3).

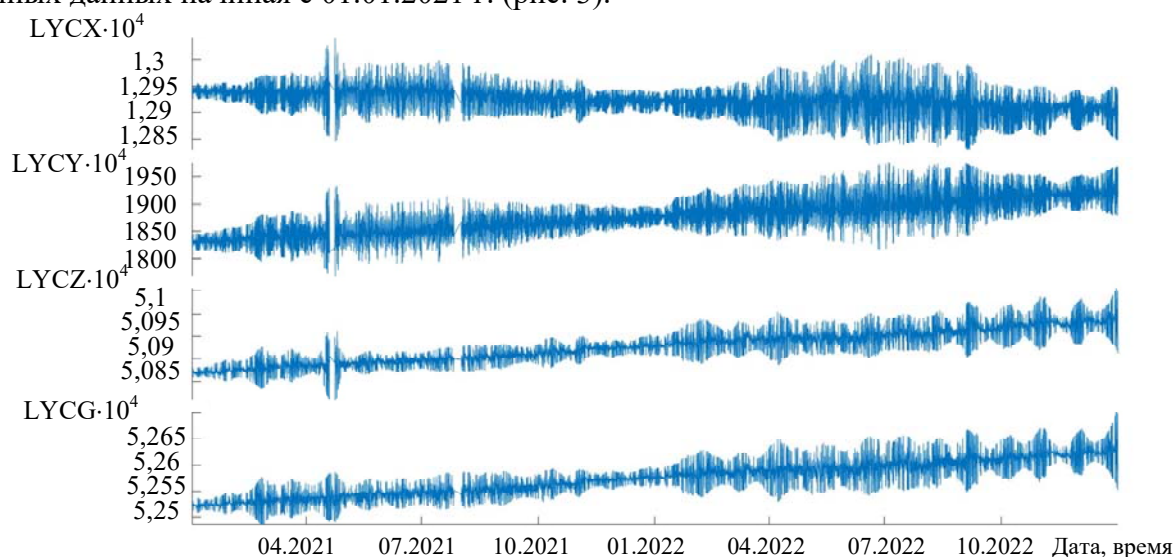


Рис. 3

**Обсуждение результатов.** Анализ рис. 1 показывает, что геомагнитная обсерватория Lyscksele поставляет в INTERMAGNET достаточно качественные данные: это видно по „выбросам“ — 21.05.2019, 02.09.2019 — 03.09.2019, 19.04.2021 — 22.04.2021, 13.06.2022 и 10.09.2022, причем в эти даты работали не все датчики.

Анализ данных модульного компонента (LYCG) на рис. 2 показывает, что по 31.12.2020 г. датчик выдавал неверные данные (значения возле нуля), а начиная с 01.01.2021 г. датчик заработал правильно. Это видно на рис. 3, где данные составляют  $\approx 52\,000$  нТл. Кроме того, явно видна корреляция изменения данных по всем датчикам. Таким образом, можно сделать вывод, что датчик измерения модульного компонента на этой станции был установлен только в конце 2020 г.

**Заключение.** Представленные результаты показывают, что применение системы MatLab позволяет достаточно эффективно работать с большими данными, необходимыми при решении задач, связанных геофизикой.

Кроме того, полученный результат предобработки может служить в качестве входных данных для решения фундаментальных и прикладных задач. Например, отношение горизон-

тальных компонентов (LYCX и LYCY) позволяет отслеживать движение магнитного полюса в горизонтальной плоскости.

В заключение следует отметить также, что применение графических процессоров (GPU) позволит повысить эффективность работы с BIG DATA.

#### СПИСОК ЛИТЕРАТУРЫ

1. 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—91.
2. Дьяконов В. П. MATLAB и Simulink для радиоинженеров. М.: ДМК Пресс, 2016. 976 с.
3. Новгородцев А. Б. Расчет электрических цепей в MATLAB: Учеб. курс. СПб: Питер, 2004. 250 с.
4. Матюшкин И. В. Моделирование и визуализация средствами MATLAB физики наноструктур. М.: Техносфера, 2011. 168 с.
5. 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. Journal of Applied Engineering Research. 2017. Vol. 12, N 6. P. 1114—1122.
6. Гайдук А. Р., Беляев В. Е., Пьявченко Т. А. Теория автоматического управления в примерах и задачах с решениями в MATLAB: Учеб. пособие. СПб: Изд-во „Лань“, 2016. 464 с.
7. Поршнев С. В. Компьютерное моделирование физических процессов в пакете MATLAB: Учеб. пособие. СПб: Изд-во „Лань“, 2011. 736 с.
8. Фриск В. В., Ганин В. И., Степанова А. Г. Компьютерный анализ и моделирование электрических цепей постоянного тока в среде MATLAB. М.: Солон-пресс, 2021. 32 с.
9. 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.
10. Николаева С. Г. Нейронные сети. Реализация в MatLab: Учеб. пособие. Казань: Казан. гос. энерг. ун-т, 2015. 92 с.
11. Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н. Большие данные. Big Data. СПб: Изд-во „Лань“, 2022. 188 с.
12. Коробейников А. Г. Обработка и анализ данных с российского сегмента мировой сети магнитных обсерваторий ИНТЕРМАГНЕТ // Междунар. журн. гуманитарных и естественных наук. 2018. № 8. С. 91—98.
13. Смолин А. А., Жданов Д. Д., Потемин И. С., Меженев А. В., Богатырев В. А. Системы виртуальной, дополненной и смешанной реальности: Учеб. пособие. СПб: Ун-т ИТМО, 2018. 59 с. [Электронный ресурс]: <<https://www.elibrary.ru/item.asp?id=46453270>>.

#### 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 РАН, зам. директора по науке; E-mail: Korobeynikov\_A\_G@mail.ru

Поступила в редакцию 06.02.2023; одобрена после рецензирования 21.02.2023; принята к публикации 31.05.2023.

#### REFERENCES

1. 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.
2. Dyakonov V.P. *MATLAB i SIMULINK dlya radioinzhenerov* (MATLAB and SIMULINK for radio engineers), Moscow, 2016, 976 p. (in Russ.)
3. Novgorodtsev A.B. *Raschet elektricheskikh tsepey v MATLAB* (Calculation of Electrical Circuits in MATLAB), St. Petersburg, 2004, 250 p. (in Russ.)
4. Matyushkin I.V. *Modelirovaniye i vizualizatsiya sredstvami MATLAB fiziki nanostruktur* (Modeling and Visualization by Means of MATLAB of the Physics of Nanostructures), Moscow, 2011, 168 p. (in Russ.)

5. 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.
6. 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.)
7. 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.)
8. 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.)
9. Korobeynikov A.G. *Software & Systems*, 2022, no. 3(35), pp. 452–457, DOI: 10.15827/0236-235X.139.452-457. (in Russ.)
10. Nikolayeva S.G. *Neyronnyye seti. Realizatsiya v Matlab* (Neural networks. Implementation in Matlab), Kazan', 2015, 92 p. (in Russ.)
11. 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.)
12. Korobeynikov A.G. *International Journal of Humanities and Natural Sciences*, 2018, no. 8, pp. 91–98. (in Russ.)
13. Smolin A.A., Zhdanov D.D., Potemin I.S., Mezhenin A.V., Bogatyrev V.A. *Sistemy virtual'noy, dopolnennoy i smeshannoy real'nosti* (Virtual, Augmented and Mixed Reality Systems), St. Petersburg, 2018, 59 p., <https://www.elibrary.ru/item.asp?id=46453270>. (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; Deputy Director for Science; E-mail: Korobeynikov\_A\_G@mail.ru

Received 06.02.2023; approved after reviewing 21.02.2023; accepted for publication 31.05.2023.