Prospects for the Application of Big Data Analysis Technologies in the Task of Searching for Harbingers of Seismic Activity by Processing the Results of Measurements Carried Out Using Space Infrastructure

Kuanysh Alipbaev[1*], Timur Suleyev[1], Zhanna Suimenbayeva[1], Yenglik Mellatova[1]
[1]*Almaty University of Power Engineering and Telecommunications named after Gumarbek Daukeev, Kazakhstan*
k.alipbayev@aues.kz

Annotation. This article analyzes studies devoted to the search for precursors of seismic activity by searching for anomalies in the measurement results of various parameters of the Earth's atmosphere and magnetosphere obtained using space infrastructure. The main limitations of studies devoted to the analysis of single facts of seismic activity are highlighted, as well as the advantages of data analysis applied to many facts of earthquakes at once. The description of the proposed method of applying technologies and methods characteristic of big data analysis to solve the problem of searching for precursors of seismic activity is given. The optimal structure of the information system proposed for solving the problem under consideration is described.

Keywords: Precursors of seismic activity, remote sensing of the Earth, processing of measurement results, big data

Introduction.

Earthquakes are still natural disasters that have a random nature of occurrence. This is primarily due to the lack of understanding of the mechanism of interaction of processes occurring shortly before the onset of seismic activity. Many researchers are trying to find earthquake precursors, on the basis of which it would be possible to create a way to predict them. Currently, one of the most promising ways to solve the problem of searching for precursors of seismic activity is to search for anomalous deviations in the measured parameters of the atmosphere, ionosphere, and Earth's magnetic field. This approach consists in analyzing the results of measurements carried out with the help of scientific equipment in the area of occurrence of the epicenter of seismic activity to compare the values recorded immediately before and after the earthquake with the values recorded earlier. At the same time, the number of facts of seismic activity to be analyzed and the time period differ.

One approach is to study in detail the changes in several physical processes for single facts of seismic activity. Examples of such studies are the analysis of data collected by a group of three spacecraft of the Swarm scientific project in relation to earthquakes that occurred in 2015 in Nepal [1, 2], in 2016-2017 in Italy [3, 4] and in 2019 in Papua, New Guinea [5]. In the course of these studies, it was possible to detect anomalies that can hypothetically be considered as harbingers of seismic activity. Such work is the optimal way to identify potential precursors of earthquakes, but it does not allow us to fully assess their effectiveness and applicability in the task of predicting seismic activity.

An alternative way to identify the precursors of seismic activity is a broader analysis devoted to a small number of parameters, but including a large number of earthquakes over a long time period. In the case of using space infrastructure as a data source for research, the choice often falls on such an ionosphere parameter as the total electronic content (TEC). Thanks to the long period of operation of global navigation systems, the history of measurements of this parameter for several decades has become available to scientists. Statistical analysis of TEC measurements allows researchers to identify similar patterns of occurrence of anomalies shortly before the start of earthquakes for samples including hundreds of earthquakes with a magnitude of more than 5 points over a period of more than 10 years [6-10]. The analysis of data on the propagation of low-frequency and ultra-low-frequency waves using ground infrastructure also demonstrates the presence of a possible connection of ionospheric disturbances with seismic activity [11-13]. With this approach, researchers are able to identify common patterns in the occurrence of anomalies preceding earthquakes, but a limited set of parameters allows us to judge only the presence of certain trends and does not allow us to uniquely identify markers preceding seismic activity. Indicative in this case is a study devoted to the analysis of changes in the temperature of the Earth's surface as an indicator of the beginning of the formation of seismic activity [14]. As a result of the work carried out, the authors concluded that it is impossible to use the selected parameter as a harbinger of earthquakes, however, the results of the study may be useful for further research.

None of the described approaches is currently able to effectively solve the problem of searching for earthquake precursors. Because in order to develop an algorithm capable of predicting the occurrence of seismic activity, it is necessary not only to identify anomalies potentially suitable for use as markers, but also to make sure that they are applicable for any earthquakes. After determining such markers, it will be necessary to determine the gradation of emerging anomalies to form a scale on the basis of which it will be possible to judge the likely parameters of a future earthquake.

Methods

In our opinion, the most promising way to solve the problem of searching for precursors of seismic activity is not to identify individual anomalies, but to determine the general pattern of their occurrence. It is the sequence of changes in the measured parameters that repeats for a large number of earthquakes that will allow us to realistically assess the probability of their occurrence, the probable magnitude and depth of the epicenter. Works devoted to the simultaneous analysis of many parameters of various physical processes demonstrate the presence of a probable effect of the interrelation of processes occurring in the lithosphere, atmosphere, and ionosphere of the Earth (LAIC effect). Understanding the mechanism underlying the LAIC effect will probably allow us to come close to developing a workable algorithm for predicting the likelihood of seismic activity. Studies on the LAIC effect are usually interdisciplinary and include the study of processes of various physical nature [15-17]. The differing physical nature of the processes under study usually implies the presence of various measuring devices, and hence data sources. This circumstance usually leads to a reduction in the number of earthquakes being analyzed, which means to some extent reduces the reliability of the identified earthquake precursors. The method proposed in this paper is to combine the advantages of the approaches used in the study of the LAIC effect and performing statistical analysis of parameters for a large sample of earthquakes. When solving the problem of searching for precursors of seismic activity in this way, there is a need to introduce special approaches and tools designed for big data analysis.

The proposed approach entails a significant complication of the data preparation process for the study. This is primarily due to a significant increase in the amount of data to be collected, as well as the number of data source systems. When consolidating data from disparate sources, they will need to be cleaned and normalized, taking into account the features of the systems that supply them. Large amounts of data being analyzed simultaneously make conventional approaches ineffective. To carry out an effective analysis of the collected data, a special infrastructure will be required, including both a large amount of computing power, as well as specialized data storage systems and software that will allow for a significant share of automation of the analysis of newly incoming data.

The optimal way to implement the proposed approach is to develop a specialized information system. Such an information system should have the following functionality:

− Automated data collection from open sources on the Internet, as well as available measuring equipment;

− Functionality for describing the collected data with the ability to display all transformations and transformations applied to them;

− Storing the received information in an optimal form for further analysis and ensuring its safety;

– Providing interfaces for uploading data from the system to third-party tools designed for data analysis and visualization;

– To provide an environment for training and launching machine learning models in order to automate the process of processing huge amounts of data in the shortest possible time and identify non-obvious patterns.

Based on the tasks set, the information system should have several mandatory components:

– a software interface designed to receive and send data in various formats;

– a database designed for data collection and storage, as well as providing the functionality of primary data processing;

– a tool for describing uploaded raw data, datasets obtained by converting them, as well as metadata;

– a tool for interacting with data stored in the system;

– an environment for running machine learning models.

Based on the described components, it is possible to determine the overall architecture of the system and designate the expected information flows between its components. The proposed architecture is shown in Figure 1.

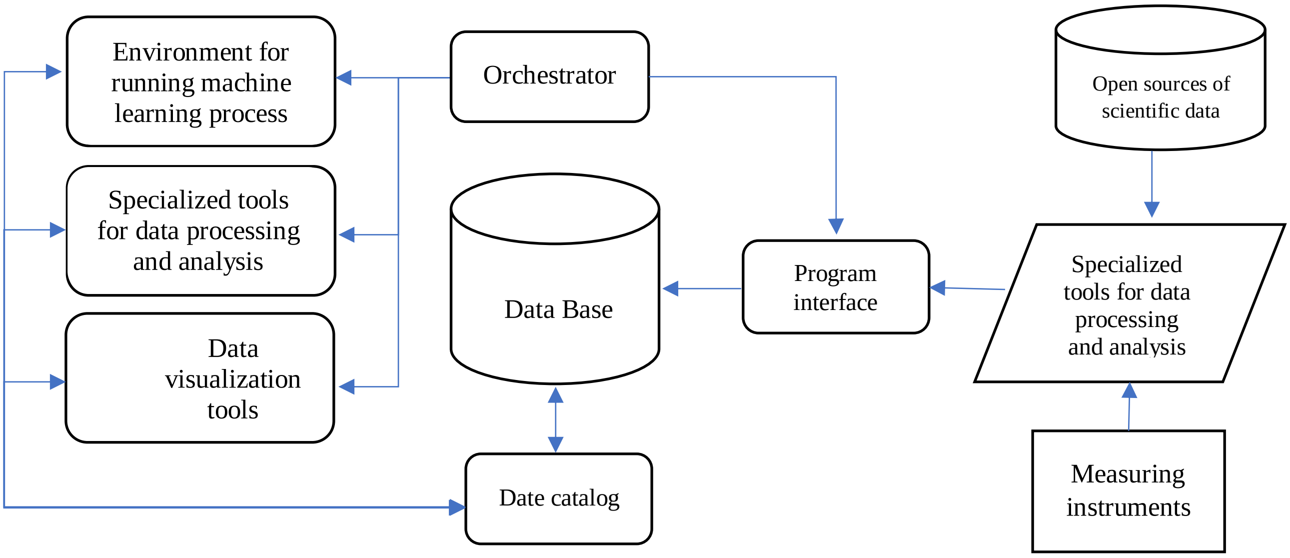

Figure 1 – General IS architecture

Results

Of course, the main expected result of the implementation of the proposed approach should be a better understanding of the mechanism underlying the LAIC effect. This will allow us to reliably assess the role of each of the studied processes in the process of earthquake formation.

The marked-up data set obtained during the study will allow to quickly check the correctness of various patterns of occurrence of anomalies accompanying the formation of seismic activity. The desired pattern, in turn, should display only the correct sequence in which anomalies should be fixed and should not impose strict restrictions on the time intervals between them and other parameters characterizing individual markers. Such differences should be brought to a scale based on which it will be possible to judge the probability of an earthquake. One of the studies devoted to the LAIC effect can be cited as an example of the work on finding a common pattern of the occurrence of anomalies [16]. In it, the authors visualized the recorded anomalies on the timeline, obtaining a similar sequence for two different earthquakes.

A side effect of the availability of a specialized information system will be the possibility of experiments on the use of machine learning models in the construction of a seismic activity forecasting system. To do this, it is necessary to have a sufficiently large set of marked data and a correct pattern (fingerprint) of the process of formation of seismic activity verified with its help.

In the case of using machine learning models, it becomes possible not only to automatically check incoming data to search for a specific pattern determined manually, but also to identify the relationships of various parameters that are not obvious to humans. This will probably allow researchers to learn more details about the processes accompanying the formation of an earthquake. Currently, attempts are already being made to use machine learning technologies in order to build a system for predicting the likelihood of seismic activity [18-19].

Conclusion

In this paper, the issue of predicting seismic activity based on the analysis of changes in the characteristics of the Earth's magnetosphere and ionosphere is considered. Examples of foreign and domestic studies are given, the results of which have been published over the past five years. Examples of measurement data sources on the basis of which these studies were conducted, as well as other available systems, both foreign and domestic, are considered. Based on the results of the analysis, conclusions are drawn about the current state of the problem and the actual problem of collecting and preparing data for analysis is determined.

In order to solve this problem, an information system is proposed designed to automate the collection and storage of measurement data from various sources, as well as providing functionality for data analysis and visualization of its results. The possible architecture of the system is given, the software capable of solving the tasks set for each component of the proposed system is listed.

The proposed system is designed to optimize the research process in the considered direction by simplifying or completely eliminating the process of data collection and preparation. In addition to optimizing the work of researchers, the system will automate the data processing procedure and thereby move from a retrospective analysis of past earthquakes to attempts to predict future seismic activity.

References

[1] Sharma S., Singh R.P., Pundhir D., Singh B. A multi-experiment approach to ascertain electromagnetic precursors of Nepal earthquakes // Journal of Atmospheric and Solar-Terrestrial Physics, Vol. 197, 2020

[2] De Santisa A., Balasisc G., Pavón-Carrasco F.J., Cianchinia G., Mandead M. Potential earthquake precursory pattern from space: the 2015 Nepal event as seen by magnetic Swarm satellites // Earth Planet. Sci. Lett. 461, 2017, p. 119–126

[3] Dedalo Marchetti, Angelo De Santis, Serena D'Arcangelo, Federica Poggio, Alessandro Piscini, Saioa A. Campuzano, Werneck V.J.O. De Carvalho, Pre-earthquake chain processes detected from ground to satellite altitude in preparation of the 2016–2017 seismic sequence in Central Italy, Remote Sensing of Environment, Volume 229, 2019, Pages 93-99, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2019.04.033.

[4] Marchetti, D., De Santis, A., D’Arcangelo, S. *et al.* Magnetic Field and Electron Density Anomalies from Swarm Satellites Preceding the Major Earthquakes of the 2016–2017 Amatrice Norcia (Central Italy) Seismic Sequence. *Pure Appl. Geophys.* **177,** 305–319 (2020). https://doi.org/10.1007/s00024-019-02138-y

[5] Akhoondzadeh M, De Santis A, Marchetti D and Shen X (2022) SwarmTEC Satellite Measurements as a Potential Earthquake Precursor Together With Other Swarm and CSES Data: The Case of Mw7.6 2019 Papua New Guinea Seismic Event. Front. Earth Sci. 10:820189. doi: 10.3389/feart.2022.820189

[6] Kon, S.; Nishihashi, M.; Hattori, K. Ionospheric anomalies possibly associated with M 6 earthquakes in Japan during 1998–2011: Case studies and statistical study. J. Asian Earth Sci. 2011, 41, 410–420

[7] Le, H.; Liu, J.Y.; Liu, L. A statistical analysis of ionospheric anomalies before 736 M6.0+ earthquakes during 2002–2010. J. Geophys. Res. Atmos. 2011, 116, A02303.

[8] Liu, J.Y.; Chen, C.H.; Tsai, H.F. A statistical study on seismo-ionospheric precursors of the total electron content associated with 146 M 6.0 earthquakes in Japan during 1998–2011; Earthquake Prediction Studies: Seismo Electromagnetics; Hayakawa, M., Ed.; TERRAPUB: Tokyo, Japan, 2013; pp. 1–13.

[9] Mustafa Ulukavak, Mualla Yalçınkaya, Emine Tanır Kayıkçı, Serkan Öztürk, Raif Kandemir, Hakan Karslı, Analysis of ionospheric TEC anomalies for global earthquakes during 2000-2019 with respect to earthquake magnitude (Mw≥6.0), Journal of Geodynamics, Volume 135, 2020, 101721, ISSN 0264-3707, https://doi.org/10.1016/j.jog.2020.101721.

[10] Munawar Shah, Arslan Ahmed, Muhsan Ehsan, Majid Khan, Muhammad Arslan Tariq, Andres Calabia, Zahid ur Rahman, Total electron content anomalies associated with earthquakes occurred during 1998–2019, Acta Astronautica, Volume 175, 2020, Pages 268-276, ISSN 0094-5765, https://doi.org/10.1016/j.actaastro.2020.06.005.

[11] Liu, J.Y.; Chen, Y.I.; Chuo, Y.J.; Chen, C.S. A statistical investigation of pre-earthquake ionospheric anomaly. J. Geophys. Res. 2006, 111, A05304.

[12] Maekawa, S.; Horie, T.; Yamauchi, T.; Sawaya, T.; Ishikawa, M.; Hayakawa, M.; Sasaki, H. A statistical study on the effect of earthquakes on the ionosphere, as based on the subionospheric LF of propagation data in Japan. Ann. Geophys. 2006, 24, 2219–2225.

[13] Hayakawa, M.; Kasahara, Y.; Nakamura, T.; Muto, F.; Horie, T.; Maekawa, S.; Hobara, Y.; Rozhnoi, A.A.; Solovieva, M.; Molchanov, O.A. A statistical study on the correlation between lower ionospheric perturbations as seen by subionospheric VLF/LF propagation and earthquakes. J. Geophys. Res. 2010, 115, A09305.

[14] Zhong-Hu Jiao, Xinjian Shan, Statistical framework for the evaluation of earthquake forecasting: A case study based on satellite surface temperature anomalies, Journal of Asian Earth Sciences, Volume 211, 2021, 104710, ISSN 1367-9120, https://doi.org/10.1016/j.jseaes.2021.104710.

[15] S. Pulinets, D. Ouzounov, Lithosphere–Atmosphere–Ionosphere Coupling (LAIC) model – An unified concept for earthquake precursors validation, Journal of Asian Earth Sciences, Volume 41, Issues 4–5, 2011, Pages 371-382, ISSN 1367-9120, https://doi.org/10.1016/j.jseaes.2010.03.005.

[16] Hayakawa, M.; Izutsu, J.; Schekotov, A.; Yang, S.-S.; Solovieva, M.; Budilova, E. Lithosphere–Atmosphere–Ionosphere Coupling Effects Based on Multiparameter Precursor Observations for February–March 2021 Earthquakes (M~7) in the Offshore of Tohoku Area of Japan. Geosciences 2021, 11, 481. https://doi.org/10.3390/geosciences11110481

[17] Dedalo Marchetti, Angelo De Santis, Xuhui Shen, Saioa A. Campuzano, Loredana Perrone, Alessandro Piscini, Rita Di Giovambattista, Shuanggen Jin, Alessandro Ippolito, Gianfranco Cianchini, Claudio Cesaroni, Dario Sabbagh, Luca Spogli, Zeren Zhima, Jianping Huang, Possible Lithosphere-Atmosphere-Ionosphere Coupling effects prior to the 2018 Mw = 7.5 Indonesia earthquake from seismic, atmospheric and ionospheric data, Journal of Asian Earth Sciences, Volume 188, 2020, 104097, ISSN 1367-9120, https://doi.org/10.1016/j.jseaes.2019.104097.

[18] Xiong, P.; Marchetti, D.; De Santis, A.; Zhang, X.; Shen, X. SafeNet: SwArm for Earthquake Perturbations Identification Using Deep Learning Networks. Remote Sens. 2021, 13, 5033. https://doi.org/10.3390/rs13245033

[19] Akhoondzadeh, M.; De Santis, A.; Marchetti, D.; Wang, T. Developing a Deep Learning-Based Detector of Magnetic, Ne, Te and TEC Anomalies from Swarm Satellites: The Case of Mw 7.1 2021 Japan Earthquake. *Remote Sens.* **2022**, *14*, 1582. https://doi.org/10.3390/rs14071582