Journal of the Georgian Geophysical Society, e-ISSN: 2667-9973, p-ISSN: 1512-1127
Physics of Solid Earth, Atmosphere, Ocean and Space Plasma, v. 25(1), 2022, pp.45-51

The Use of Structured Data for Drought Evaluation in Georgia

[1]Marika R. Tatishvili, [1]Ana M. Palavandishvili, [2]Mariam B. Tsitsagi, [2]Nikoloz E. Suknidze

[1]Institute of Hydrometeorology of Georgian Technical University, Tbilisi, Georgia
[2]Vakhushti Bagrationi Institute of Geography of Tbilisi State University, Tbilisi, Georgia

ABSTRACT

The drought is frequent phenomena in Georgia. The SPI and SPEI drought indices were calculated to analyse drought frequency and intensity on the territory of Georgia in 1991-2020 period. The structured data of hydrometeorological observation net have been used to calculate following statistical parameters: Pearson correlation, mean deviation, and absolute deviation, both for the entire period and for months. The programs R and R-instat are used to calculate and visualize these parameters. The correlation coefficient is in good agreement for all cases, and the absolute deviation shows data scattering, which should be related to the complex relief of Georgia, as well as the heterogeneity of data series. The study is important for climate change assessment, hydrometeorological disaster early warning system, as the territory of Georgia is under the risk of this phenomenon.

Keywords: *Structural data, statistical analysis, drought indices, early warning system.*

Introduction

The term Big Data has spread due to new technologies and innovations that have emerged over the past decade given the demand for the analysis of large amounts of and rapidly generated diverse data, therefore, collection and processing takes place at a high speed, which is difficult to implement with calcareous analytical tools. Big Data is a rapidly generated amount of information from a variety of sources and in a different format. Data analysis is the examination and transformation of raw data into interpretable information, while data science is a multidisciplinary field of various analyses, programming tools, and algorithms, forecasting analysis statistics as well as machine learning that aim to recognize and extract patterns in raw data. The applicability of Big Data techniques is also significantly enhanced by the novel tools that support data collection and integration. The interoperability of the systems can be improved by data warehouses and the related **ETL** (extract, transform, and load) functionalities that can also be used to gather information from multiple models and data sources. Artificial intelligence (**AI**) and machine learning (**ML**) are also the key enabler technologies of big data analysis [1]. Analysis of Big Data combines traditional methods of statistical analysis with computational approaches. The analysis of Big Data is a synthesis of quantitative and qualitative analyses. Climate computing combines multidisciplinary researches in regard to climatic, data and system sciences to efficiently capture and analyze climate-related Big Data as well as to support socio-environmental efforts.

The significance of Big Data in climate-related studies is greatly recognized and its techniques are widely used to observe and monitor changes on a global scale. It facilitates understanding and forecasting to support adaptive decision-making as well as optimize models and structures. The Big Data classifies of three types: **Structured**; Semi-Structured and Unstructured Data. In presented paper Structured Data is used. The structure of the data is the key to not only how to go about working with it, but also what insights it can produce. It is successfully used in weather, climate and environmental issues.

Study area

The complex orographic conditions and influence of the Black Sea preconditioned the formation of great variety of climate and landscapes in Georgia. Here exist most of Earths climatic types, from marine wet subtropical climate of west Georgia and steppe continental climate of east Georgia up to eternal snow and glaciers of high mountain zone of Great Caucasus, and also approximately 40% of observed landscapes. Territory of country lies between of the Major Caucasian Ridge and the Lesser Caucasus mountains. About 85 percent of the total land area occupies complex mountain ranges divided with river's valleys and ravines of different exposition. For most of these rivers spring is high water period thus ¾ of Georgia's territory is especially vulnerable to foods, flash floods and mudflows. Georgia's orography and its interaction with airflows are the basic spotting factors of synoptic processes spread in the country. Peculiarities of locally developed weather phenomena at any time a year are often characterized with diversity and extremity. Convective storms, with attendant phenomena; fog and low clouds; locally forced precipitation events; wintertime weather (snow, ice, glazed frost, avalanches) this is a short list of synoptic processes nowcasting (**NWC**) and very short range forecast (**VSRF**) of which has a great importance for Georgia [9]. The natural disasters in Georgia have to be considered as the standing negative factor for the sustainable development of the state. The importance of aroused problems from listed hazards stimulates the active investigation of reasons and physical processes involved in [5].

In the analysis of hazard and risk geo-information science and earth observation plays an increasingly important role. Remote Sensing is nowadays an essential tool in monitoring changes in the earth's surface, oceans and atmosphere, and is increasingly used as the basis for early warning for hazardous events [6].

Drought is a frequent phenomenon in eastern Georgia. Its frequency in some areas exceeded 40% in the 80-ies of the last century by certain early estimates. As a result of frequent droughts accompanying the global warming in past decades transformation of many types of natural landscapes has been observed. The desertification probability of steppe and semi-desert landscape of eastern Georgia by the end of the twentieth century has reached 25-30%. According to official figures, by the result of intense droughts area of over than 200 000 ha is strongly affected for present [4]. Property damage caused by drought is very significant.

The main meteorological factors for drought formation are dry weather, high temperature and lack of soil producing moisture. The average time of rainless period with precipitation less than 5 mm most important for agriculture is not more than 10-15 days. Besides, the mean rainfall is not more than 200-300 mm during vegetation period on the lowlands. Nevertheless, producing moisture supply is 50-200 mm per one meter of soil that corresponds to the zone of capillary agro-hydrological humidification and full spring rainfall penetration. At the same time active air temperatures sum exceeds 4000° over 10° times, and the mean duration of continuous high temperatures more than 30°C is longer than 4 hours [2].

Temperature change velocity is greatly depended on region physical-geographical conditions and seasons. Due to complex relief, mainly orographic and landscape-climatic conditions temperature change has heterogenic –mosaic character in Georgia The strong warming centers were located in east and also on west parts of Georgia. Strong warming centers were spread in moderately dry and dry subtropical semi-arid and dry step regions of eastern Georgia. During intense warming period (1975-2010) temperature annual trend has been significantly increased and on Black Sea region amounts 0.034^{0}C (Poti), Colchis Lowland -0.025^{0}C and on Dedoplistskaro-0.09-^{0}C, Gori-0.05^{0}C, in Khashuri and Telavi-0.06^{0}C.

Drought genesis in Georgia is depending on cyclonic and anticyclone motions. In first case rainy days are frequent and in second dry periods, with high temperature and low humidity of different durations have been taken place. If air masses directed from Arctic are dry and cold. They spread over long territories and stable anticyclone system is established on east-south parts of Europe. During such situation dry period happens in Georgia. If air masses are invading from east high temperature and low humidity dry weather is standing. Such periods are more brutal and dangerous [**3**].

The observation analysis shows that various degree drought may take place all over the Georgian territory. The event frequency is expected mainly on spring, summer and fall seasons. During winter due to frequent cyclonic and frontal periods dry day duration is less. The drought day number and dry period frequency increase from the Black Sea regions through east or in direction of continental climate [10].

Based on historical records Georgian territory is under drought 60% repeatability. The most drought regions are Kvemo Kartli, Shiraki and Eldari lowlands and other low parts of eastern parts of Georgia. Those regions are characterized by productive humidity shortage in soils. Two types of productive humidity stocks are common for those places: capillary moisturizing and complete spring wetting. In the first case the productive humidity stock in 1m. soil layer composes 100-200mm, and in other- 50-150mm, while in western Georgia the humidity stock doesn't exceed 400mm. Except natural factors (windy erosion and precipitation decreasing) the anthropogenic loading has significant effect on desertification process too. Namely: unmanaged use of soil, forest and water resources, soil salting.

Method

Understanding of the natural environment is increasingly important to respond to the climate change negative impacts and anthropogenic pressures on finite natural resources, and their impacts on water, energy and food security, infrastructure, human health, natural hazards. This is also a major cross-disciplinary challenge involving almost all scientific fields

In 2013, the UK government announced large-scale investment in Big Data infrastructure for science, particularly in the environmental sector starting funding for a program called CEMS (Climate and Environmental Monitoring from Space). This allowed for the creation of larger databases to cope with the upcoming Big Data revolution and to allow research partner organizations to work with more data and produce more results. With a specific focus on climate change and planetary monitoring, CEMS storage removed the need to download enormous data sets while reducing the cost of access. Along with Cloud data, this is now the standard globally for some of the world's top research institutes.

Environmental data comes from a wide variety of sources and this is increasingly rapidly with new innovations in data capture:

1. Large volumes of data are collected via remote sensing, typically from satellite sensing or aircraft-borne sensing devices, including an increasing use of drones. This includes passive sensing, such as photography or infrared imagery, and active sensing, e.g., RADAR/LIDAR. The increasing availability of open satellite data is a major trend in earth and environmental sciences. For example, the EU Copernicus program and the associated Sentinel missions, or NASA's Earth Observing System satellites, LandSat archive are regularly mined for data for a variety of applications [6,10].
2. Other data are collected via earth monitoring systems, which consist of a range of sensor technologies measuring various physical entities. Namely weather stations and monitoring systems
3. Model output is also a significant generator of environmental data with results from previous model runs often stored for subsequent analysis

There are a number of issues deserve special attention:

1. Scarcity of in-situ observations.

2. Data heterogeneity: neighboring pixels in the dataset may have been acquired at different times of day and thus are not directly comparable. In addition, there are no observations during periods of cloud cover, and it is not possible to assume these data are simply "missing for statistical modeling purposes

After data receiving we conduct inventory, which means its visualization, in order to better estimate the transmitting break, for all this we use the program R-studio and R-Instat, R-Instat is a free, open source statistical software that is easy to use, even with low computer literacy. It encourages good statistical practices and learning, by opening the door to training to emphasize concepts rather than the theory.

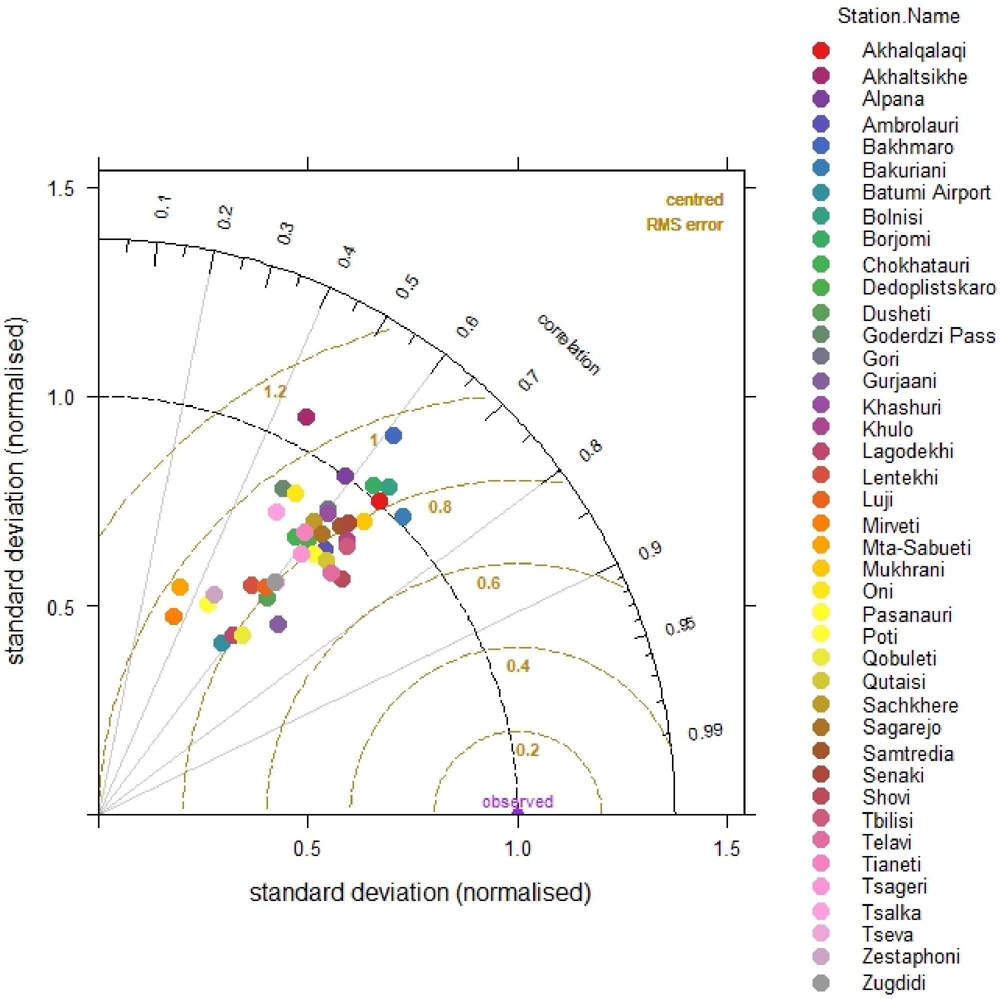

Fig.1. Taylor diagram of standard deviation (normalized) of processed stations

The standardized indices **SPI** and **SPE**I classify the precipitation and water balance anomalies with respect to the long term records. The index values directly indicate how frequent the current situation is expected to occur at the location and season of interest given the long term observations. The **SPI** (standardized precipitation index) classifies the precipitation sums on a particular date with respect to the sums of the same month in all years of the measurement record. For this purpose, the precipitation sums of the whole record within one month around the respective date are transformed into a standard normal distribution around zero. The SPI is nothing else than these transformed precipitation sums. The **SPI** value hence directly indicates the frequency of the observed precipitation amount in the corresponding month as estimated from the whole observation record. The **SPEI** (standardized precipitation evapotranspiration index) is calculated in analogy to the SPI, using the cumulative water balance instead of precipitation sums. The SPEI hence represents the standard-normal distributed water balance[8].

From above discussed stations several have been chosen for 1991-2020 year period drought research. Those stations satisfy all requirements and also are most vulnerable to draught phenomenon to calculate draught indices and conduct statistical analysis. Those stations are the following Dedoplistskaro, Gori, Khashuri, Telavi. By using the calculated indices it became possible to evaluate drought degree and intensity.

The evolution of SPI and SPEI drought indices in 1991-2020 year period are presented on the plots (Fig.2,3).

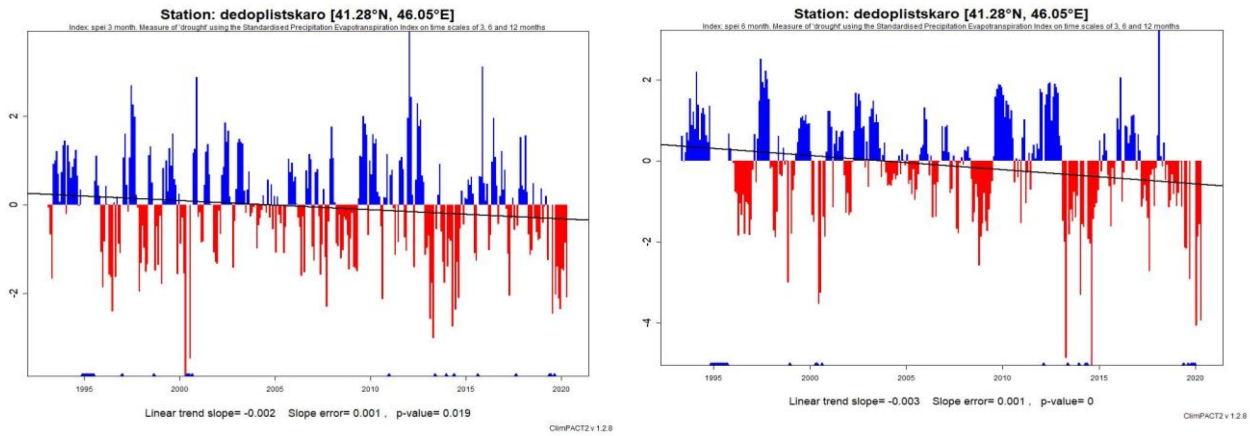

Fig.2. SPEI 3and 6 month Dedoplistskaro station

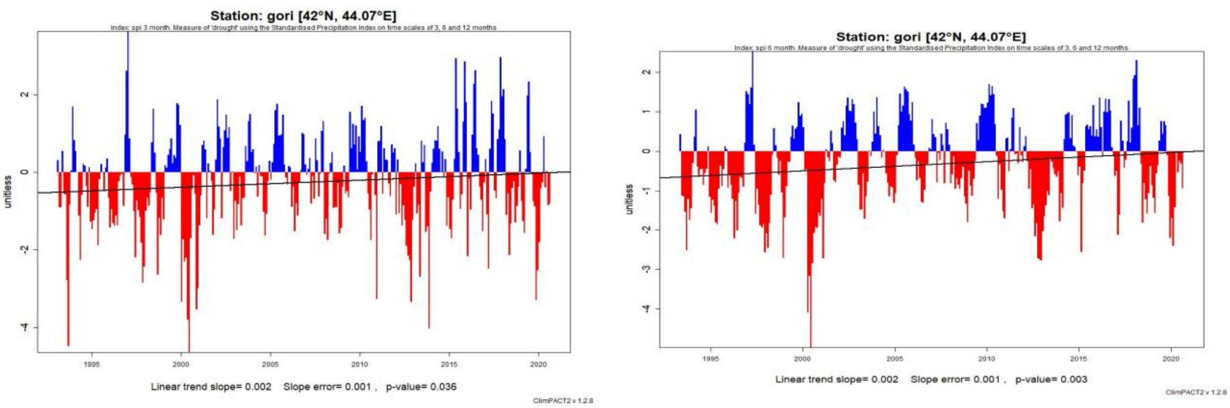

Fig.3. SPI 3and 6 month Gori station

In study Pearson correlation coefficient (PCC), determination coefficient (R^2), and root mean square error (RMSE) criteria, which are among the strong statistical criteria, were used. R^2 ranges from 0 to 1, with higher values indicating less error variance. The **RMSE** is the square root of the variance of the residuals. It indicates the absolute fit of two data set and lower the RMSE the better performance is.

In order to compare drought indices, scatter diagrams of indices were drawn and statistically evaluated. For this, R^2 and the RSME were used. Correlation between data sets is a measure of how well they are related. The most common measure of correlation in stats is the Pearson Correlation. It shows the linear relationship between two sets of data (3month) PCC, which shows linear relationship between SPI-SPEI is quite high, RMSE (SPI-SPEI) is low especially for Khashuri and Telavi; RMSE (SPI-EDI), (Tab.1)

Table1.Statistical parameters of SPI3-SPEI3 for selected stations

Station	Pearson	Pearson	Pearson	R^2-SPI	R^2-SPEI	RMSE(SPI_SPE)
Dedoplistskaro	0.935369	0.056465	0.068697	0.0233	0.0785	0.017053
Gori	0.916488	0.049104	0.114851	0.0034	0.0315	0.010956
Khashuri	0.918664	0.111081	0.130188	0.0064	0.0918	0.005407
Telavi	0.949467	0.051841	0.078509	0.0013	0.0263	0.005885

(12month) PCC for SPI-SPEI is high. R^2 is low for all stations. RMSE (SPI-SPEI) is low which means perfect fitting, (Tab.2).

The strongest relationship was observed among the indices in the same time periods. As time lag increases, the relationship between variables has weakened. Among the indices, the strongest correlation coefficient (0.94) was observed between SPI-12.

Table 2. Statistical parameters of SPI12-SPEI12 for selected stations

Station	Pearson	Pearson	Pearson	R^2-SPI	R^2-SPEI	RMSE(SPI_SPE)
Dedoplistaskaro	0.924953	0.065094	0.102553	0.0705	0.2187	0.016993
Gori	0.84161	0.055709	0.14224	0.0046	0.1264	0.010384
Khashuri	0.898084	0.13671	0.084106	0.0465	0.2861	0.005174
Telavi	0.945912	0.116814	0.148417	0.0023	0.0383	0.018411

DISCUSSION

In Europe the Global Monitoring for Environment and Security (GMES) initiative of the European Commission and the European Space Agency (ESA) is actively supporting the use of satellite technology in disaster management, with projects such as PREVIEW (Prevention, Information and Early Warning pre-operational services to support the management of risks), LIMES (Land and Sea Integrated Monitoring for Environment and Security), GMOSS (Global Monitoring for Security and Stability), SAFER (Services and Applications For Emergency Response), and GMOSAIC

(GMES services for Management of Operations, Situation Awareness and Intelligence for regional Crises). The United Nations Platform for Space-based Information for Disaster Management and Emergency Response (UN-SPIDER, 2010) has been established by the UN to ensure that all countries have access to and develop the capacity to use space- based information to support the disaster management cycle. They are working on a space application matrix that will provide the satellite-based approaches for each type of hazard and each phase of the disaster management cycle [7].

WMO has signed the Emergency Alerting Call to Action as part of its ongoing Global Multi-hazard Alert System (GMAS) development and its collaboration with governmental, non-governmental, and commercial organizations to achieve the broadest adoption of CAP worldwide. The Call to Action was launched at a special event during the Humanitarian Networks and Partnerships Weeks 2021.
For further researches application of satellite imaginary with ground observation data will be useful to construct drought hazard risk map for entire territory of Georgia.

REFERENCES

[1] Blair G.S., Henrys P., Leeson A., Watkins J., Eastoe E., Jarvis S., Young P.J. Data Science of the Natural Environment: A Research Roadmap. Frontiers in Environmental Science. www.frontiersin.org. Volume 7, 2019. doi: 10.3389/fenvs.2019.00121
[2] Chirakadze G.I. Drought, Dry Period over Caucasus. Hydrometizdat, Leningrad, 1980.
[3] Papinashvili L. Drought in Georgia. Transactions of the Institute of Hydrometeorology, v. 107, Tbilisi, Georgia, 2002, pp. 28-33.
[4] Janelidze Z. On some Reasons of Desertification and Climate Aridization of Iori Valley. Transactions of the Institute of Hydrometeorology, v. 115, Tbilisi, Georgia, 2008, pp. 51-56.
[5] Tatishvili M., Elizbarashvili E., Meskhia R., Elizbarashvili Sh. Natural Hydrometeorological Disasters, Their Causes and Prevention Measures. The Macrotheme Review. A Multidisciplinary Journal of Global Macro Trends, 148-154, A Macrotheme Capital Management, LLC Academic Publication, ISSN 1848-4735, v. 2, Iss. 1, Winter (January), 2013.
[6] Tatishvili M., Bolashvili N., Palavandishvili A. Impact of Short-Term Geomagnetic Activity on Meteorological Parameters Variability on the Middle Latituge Region. Georgian Geographical Journal, v.2 (1), 2022, pp. 45-50, https://journals.4science.ge/index.php/GGJ/article/view/756/791
[7] Copernicus European Drought Observatory (EDO): https://edo.jrc.ec.europa.eu/

[8] Vicente-Serrano S. M., Beguería S., Lorenzo-Lacruz J., Camarero J. J., López-Moreno J. I., Azorin-Molina C., Revuelto J., Morán-Tejeda E., Arturo Sanchez-Lorenzo A. Performance of Drought Indices for Ecological, Agricultural, and Hydrological Applications. Earth Interactions, 16(10), 2012. https://doi.org/10.1175/2012EI000434.1
[9] Tatishvili M. Developing Weather Forecasting System in Georgia. Ecology & Environmental Sciences 2 (7) DOI:10.15406/mojes. 2017.02.00046.
[10] Tatishvili M.R., Mkurnalidze I.P., Samkharadze I.G., Chinchaladze L.N. Application of Satellite Imaginary in Forestry for Georgia. Journal of the Georgian Geophysical Society, ISSN: 1512-1127, Physics of Solid Earth, Atmosphere, Ocean and Space Plasma, v. 21(1), 2018, pp. 18-25, https://openjournals.ge/index.php/GGS/article/view/2484

სტრუქტურირებული მონაცემების გამოყენება გვალვის შეფასებისთვის საქართველოში

მ. ტატიშვილი, ა. ფალავანდიშვილი, მ. ციცაგი, ნ. სუქნიძე

რეზიუმე

საქართველოში გვალვა ხშირი მოვლენაა. SPI და SPEI გვალვის ინდექსები გამოთვლილია გვალვის სიხშირისა და ინტენსივობის გასაანალიზებლად საქართველოს ტერიტორიაზე 1991-2020 წლებში. ჰიდრომეტეოროლოგიური დაკვირვების ქსელის სტრუქტურირებული მონაცემები გამოყენებულია შემდეგი სტატისტიკური პარამეტრების გამოსათვლელად: პირსონის კორელაცია, საშუალო გადახრა და აბსოლუტური გადახრა, როგორც მთელი პერიოდისთვის, ასევე თვეების განმავლობაში. პროგრამები R და R-instat გამოიყენება ამ პარამეტრების გამოსათვლელად და ვიზუალიზაციისთვის. კორელაციის კოეფიციენტი კარგად შეესაბამება ყველა შემთხვევისთვის, ხოლო აბსოლუტური გადახრა აჩვენებს მონაცემთა გაფანტვას, რაც დაკავშირებული უნდა იყოს როგორც საქართველოს კომპლექსურ რელიეფთან, ასევე მონაცემთა სერიების ჰეტეროგენურობასთან. კვლევა მნიშვნელოვანია კლიმატის ცვლილების შეფასებისთვის, ჰიდრომეტეოროლოგიური კატასტროფების ადრეული გაფრთხილების სისტემისთვის, რადგან საქართველოს ტერიტორია ამ მოვლენის საფრთხის ქვეშ იმყოფება.

Использование структурированных данных для оценки засухи в Грузии

М.Р. Татишвили, А.М. Палавандишвили, М.Б. Цицаги, Н.Е. Сукнидзе

Резюме

Засуха частое явление в Грузии. Индексы засухи SPI и SPEI рассчитаны для анализа повторяемости и интенсивности засух на территории Грузии в период 1991-2020 гг. По структурированным данным сети гидрометеорологических наблюдений рассчитаны следующие статистические параметры: корреляция Пирсона, среднее отклонение и абсолютное отклонение как за весь период, так и по месяцам. Для расчета и визуализации этих параметров используются программы R и R-instat. Коэффициент корреляции хорошо согласуется для всех случаев, а абсолютное отклонение показывает разброс данных, что должно быть связано со сложным рельефом Грузии, а также с неоднородностью рядов данных. Исследование важно для оценки изменения климата, системы раннего предупреждения о гидрометеорологических катастрофах, так как территория Грузии находится под угрозой этого явления.