

The Big Data for Drought Monitoring in Georgia



**Marika Tatishvili, Ana Palavandishvili, Mariam Tsitsagi,
and Nikoloz Suknidze**

Abstract The dangerous hydrometeorological phenomenon—drought is frequent in Georgia. The SPI and SPEI 3, 6 and 12 month drought indices were used to analyze drought frequency and intensity on the territory of Georgia for 1991–2020 year period. The structured data that is the part of big data, of ground hydrometeorological observation network of Georgia have been used to conduct research. The following statistical parameters were calculated: Pearson correlation coefficient (PCC), determination coefficient (R^2), and root mean square error (RMSE) both for the entire period and for months. The correlation coefficient is in a 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 precipitation data distribution. The calculated Standardized Precipitation Index (SPI) for 3 months of Kakheti region was subjected to Machine Learning. A Support Vector Machine (SVM) was selected in the Matlab space—the algorithm for Supervised Machine Learning method. The tenth model showed the best result; using of mentioned model it became possible to determine the drought probability by months at each point. Despite of obtained good parameters, it was necessary to add additional stations, because there was not enough information in the Kakheti region for the correct analysis of Machine Learning avoiding overfitting. The study is important for climate change assessment and hydrometeorological disaster early warning system implementation.

Keywords Big Data · Machine learning · Climate change · Drought indices · Natural hazard

M. Tatishvili (✉) · A. Palavandishvili · M. Tsitsagi · N. Suknidze
Institute of Hydrometeorology, Georgian Technical University, Tbilisi, Georgia
e-mail: m.tatishvili@gtu.ge

A. Palavandishvili · M. Tsitsagi · N. Suknidze
Vakhushti Bagrationi Institute of Geography of Tbilisi State University, Tbilisi, Georgia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
R. C. Geibel and S. Machavariani (eds.), *Chances and Challenges of Digital Management*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-45601-5_13

1 Introduction

The hydrometeorological data are collected from a great range of sources and this process increases rapidly with new technologies measuring various environmental data. The first basic environmental data are collected by ground monitoring systems, which consist of wide range of sensor technologies measuring various physical parameters. The weather stations and monitoring systems measure almost all hydrometeorological parameters used in weather prediction and climate change assessment topics. Those include temperature, precipitation, wind velocity and direction, solar radiation, relative humidity and etc.

In Georgia, there exist most of Earths climatic types, from marine wet subtropical climate of west part and steppe continental climate of east part up to eternal snow and glaciers of high mountain zone of Great Caucasus, and also approximately 40% of observed landscapes.

The complexity of the orographic structure of Georgian territory, along with other physical -geographical factors is the cause of wide variety of climates and landscapes. There are almost all types of climates observed on the Globe, from the climate of eternal snows of high mountains and glaciers to steppe continental climate of eastern Georgia and humid climate of the Black Sea coast subtropical zone (Tatishvili et al., 2013).

Such complicated relief has definite influence on air masses motion in atmosphere lower layers. Mainly west and eastern atmospheric processes prevailed over Georgian territory. Thus those climatic zones uphold the formation of different dangerous hydrometeorological phenomena, namely: hailstone, heavy showers, flooding, thunderstorm, draughts, and sea storms. The existing geodynamical and orographic properties of Georgia play major role in the formation of various weather patterns. Such complex relief is one of the main reason of the formation and evolution of various scaled circulation systems and heterogeneous spatial distribution of meteorological elements. This is verified by the fact, that precipitation annual distribution has diverse type, with sharply expressed spatial inhomogeneities (Tatishvili, 2017).

The hydrometeorological ground station observation in Georgia begun on the end of 19-th century gradually increased up to 200 stations at the end of 20-th century. For present there are about 80 ground based automatic weather stations that transmit weather information per 10 min (Fig. 1).

The dangerous phenomenon—drought is a frequent in eastern Georgia. Its frequency in some areas exceeded 40% in the 80-th of the last century by assured estimations. The significant transformation of many types of natural landscapes has been observed resulted from the frequent droughts accompanying the global warming in past decades. The desertification probability of steppe and semi-desert landscape of eastern Georgia by the end of the twentieth century has reached 25–30% (Tatishvili et al., 2021). According to official information, the area of over than 200 000 ha is strongly affected by the intense droughts for present. Property damage caused by drought is also very significant.

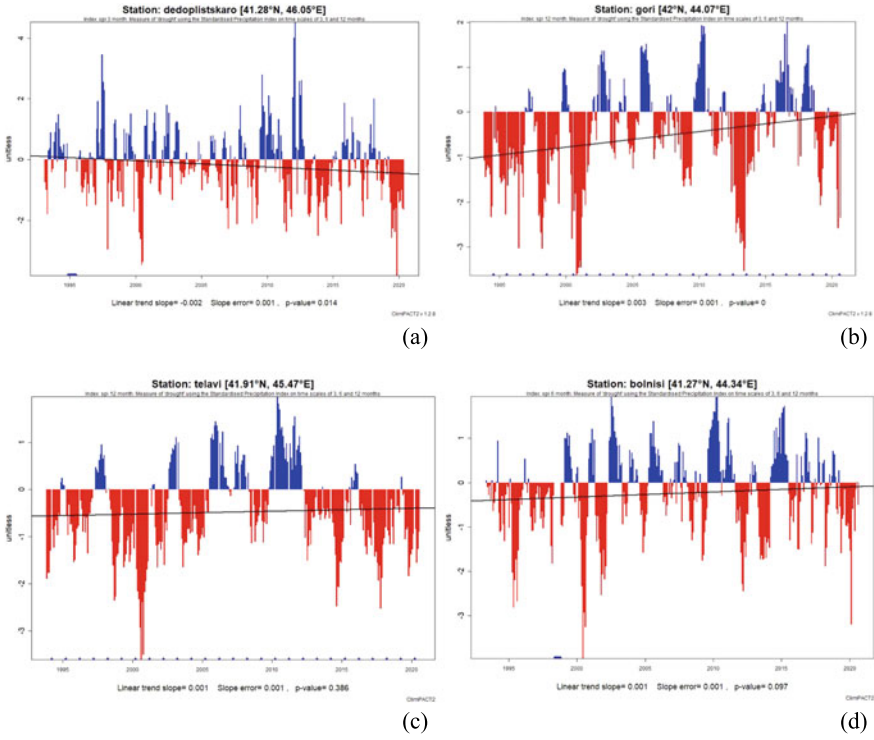


Fig. 1 Drought indices for Dedoplistskaro a SPI12, Gori b SPI12, Telavi c SPI12, Bolnisi d SPI12

It is well established, that 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 °C by 10 times and the mean duration of continuous high temperatures more than 30 °C is longer by 4 h.

2 Data and Method

Station data was recovered from the CLIDATA database of the National Environmental Agency (NEA), which has been operating since 2014. Stations were selected based on data continuity and accuracy. After station data validation where data interruption has been detected or measuring sensor transmitted incorrect information

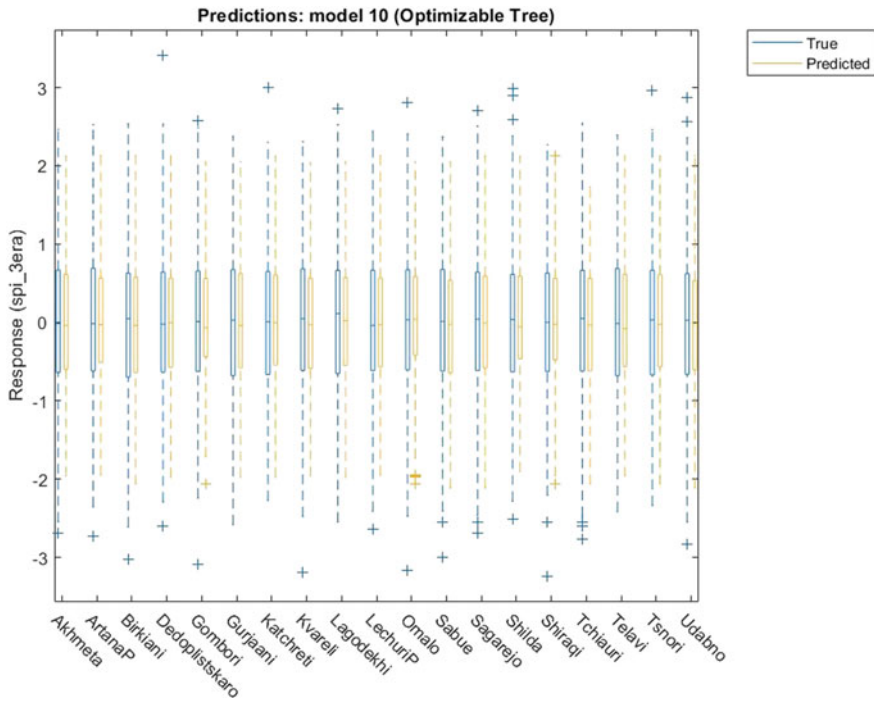


Fig. 2 The trained model 10 for Kakheti region

due to its malfunction they were removed and not analyzed. At the 21 stations the observers monitor data and except human factor the unreliability of the data are minimal, and the rest of them were operated by rain gauge produced by VAISALA (Publish House of Iv. Javakhishvili Tbilisi State University, 2022), which by its design does not measure residual precipitation. The VAISALA weather gauges are the new generation of weighing precipitation gauges. They combine techniques, the latest high-accuracy load cell technology and advanced measurement control algorithms to ensure high performance, both in liquid and solid precipitation used for almost all weather conditions (Publish House of Iv. Javakhishvili Tbilisi State University, 2022).

Drought is one of the most dangerous and widespread natural disasters of all over the globe. Drought in the historical past has become a provocateur for the disappearance of many civilizations (Evans et al., 2018; Kaniewski et al., 2015) and robust migration processes (DeMenocal, 2001). Droughts are characterized by decreased natural water availability in the form of precipitation, river runoff, or groundwater (Babre et al., 2022; Tatishvili et al., 2022a). Difficulties related to the study of droughts are determined by their diverse nature (meteorological, hydrological, ecological, and economic). Accordingly, there are different types of indexes, the complexity of which depends on the availability of the necessary data. The most widely used indices in modern studies are SPI (Standard Precipitation Index) and SPEI (Standard Precipitation Evapotranspiration Index). The SPI calculation uses the

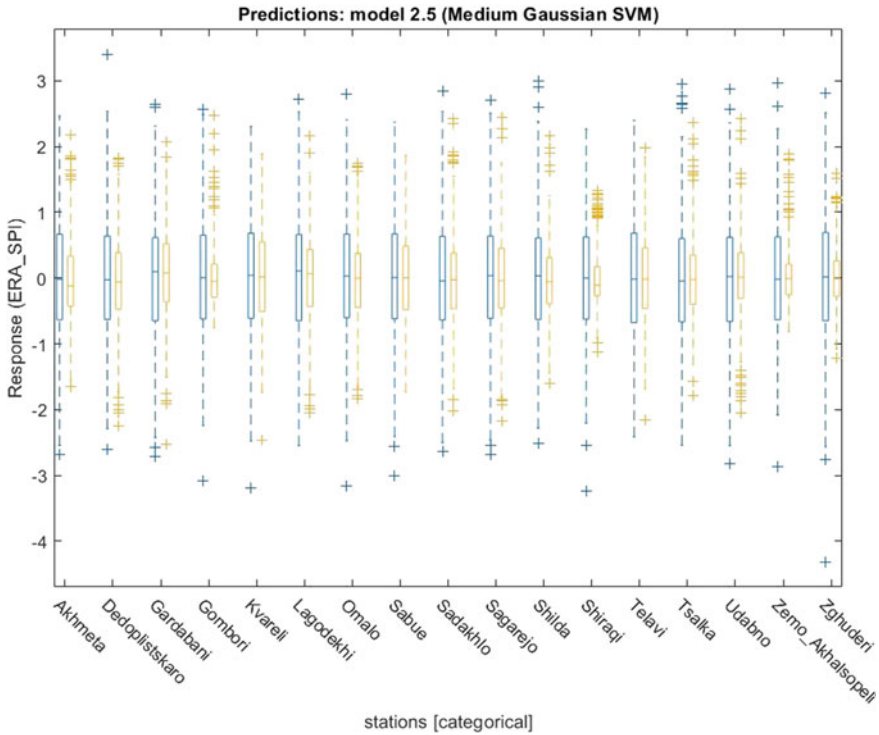


Fig. 3 Advanced SVM prediction Kakheta region model including additional seventeen stations of Kartli region

precipitation data series (McKee et al., 1993), and the SPEI is based on the cumulative water balance instead of precipitation sums (Vicente-Serrano et al., 2010). The SPEI hence represents the standard-normal distributed water balance (Vicente-Serrano et al., 2010), where the temperature is considered along with precipitation (Tatishvili et al., 2022a). To calculate above mentioned indices the ground meteorological station data of Georgian National Environmental Agency was used of 1991–2020 year period. 10 stations have been chosen from data archive that has valid data to calculate indices.

The drought indices are classified according months: 3, 6 and 12. R-stat soft was used for plots. Below are presented several draught SPEI, SPI drought indices plots from selected stations. They represent Kakheta and Kartli regions notable for agriculture: Dedoplistskaro, Gori, Telavi and Bolnisi.

3 Discussion

The correlation between data sets is a measure of how well they are related. In the presented study, the statistical parameters: Pearson correlation coefficient (PCC)-the most common measure of correlation, Determination coefficient (R^2), and Root Mean Square Error (RMSE) were used as the criteria, which are among the strong statistical measures. Generally, 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 sets and lower the RMSE the better performance is (Observatory and EDO, 2021).

The PCC (3 month) which shows linear relationship between SPI-SPEI is quite high and the RMSE (SPI-SPEI) is low for all stations, especially for Khashuri and Telavi. (Table 1).

The PCC (12 month) for SPI-SPEI (12 month) is high. The R^2 is low for all stations. The RMSE (SPI (12)-SPEI (12)) is low which means perfect fitting, (Table 2). The strongest relationship was observed among the indices using the equal time periods. By the increasing of time lag, the relationship between variables has been weakening.

The PCC for SPI-SPEI relations inside stations is very high, while between stations is relatively low. This can be explained by the facts that meteorological parameters-temperature and especially precipitation spatial-temporal distribution have diverse nature for different locations at different elevations and stations are placed in various climatic zones. The distance between stations is also the important factor.

The values of PCC are shown on Table 3.

The calculated values of indices give possibility to split negative and positive number or dry and wet day number. It's interesting to count drought and wet day ratio

Table 1 Statistical parameters PCC, R^2 , RMSE for most vulnerable regions according SPI3 and SPEI3

Station/Stat. parameter	Pearson (3)	R^2 -SPI	R^2 -SPEI	RMSE(SPI-SPE)
Dedoplistskaro	0.94	0.02	0.07	0.01
Gori	0.92	0.01	0.03	0.02
Khashuri	0.91	0.01	0.09	0.01
Telavi	0.96	0.001	0.02	0.01

Table 2 Statistical parameters PCC, R^2 , RMSE for most vulnerable regions according SPI (12) and SPEI (12)

Station/Stat. parameter	Pearson	R^2 -SPI	R^2 -SPEI	RMSE(SPI_SPE)
Dedoplistskaro	0.92	0.07	0.22	0.02
Gori	0.84	0.01	0.12	0.01
Khashuri	0.89	0.05	0.28	0.01
Telavi	0.93	0.01	0.04	0.02

Table 3 Pearson correlation of draught indices inside and between stations

Stations	Pearson SPI(3)-SPEI(3)	Pearson SPI(6)-SPEI(6)	Pearson SPI (12)-SPEI(12)
Telavi	0.96	0.94	0.93
Akhalsikhe	0.93	0.91	0.86
Ambrolauri	0.96	0.96	0.96
Gori	0.92	0.86	0.84
Mta Sabueti	0.98	0.98	0.97
Pasanauri	0.98	0.96	0.94
Poti	0.99	0.98	0.99
Kutaisi	0.98	0.97	0.98
Tbilisi	0.93	0.91	0.88
Telavi-Gori	0.47	0.48	0.54
Axalkhalakhi- Mta Sabueti	0.5	0.50	0.50
Tbilisi -Kutaisi	0.32	0.32	0.45
Tbilisi-Mta Sabueti	0.31	0.31	0.44
Bolnisi-Pasanauri	0.24	0.16	0.26

at each stations. For Akhalsikhe point wet day number exceeds drought one, severe drought day is approximately 3 and moderate- 50. At Gori station wet day number slightly exceeds drought day number, severe drought day equals 1 and moderate-60. At Telavi point drought day number greatly exceeds wet day number, severe drought day equals 5 and moderate-58. At Tbilisi point drought day number exceeds wet day one, severe drought day is 4 and moderate-62. At Kutaisi station wet day number exceeds drought one, severe drought day is 5 and moderate-55. At Mta-Sabueti both day types are approximately equal, severe drought day is 5 and moderate-54 (Tatishvili et al., 2022b). In Table 4 dry and wet day number for 3 places are presented.

The index values categorize drought intensity in given period. On Tables 5, 6, 7 and 8 SPEI and SPI categories were presented for Telavi, Tbilisi and Gori stations. All SPI, SPEI values are differing from each other and this fact has explanations.

Table 4 Negative and positive day number in Telavi, Tbilisi and Gori for 1991–2020 year period by SPEI3

Telavi	Negat	175	179	178	162	182	190
	Pos	171	164	159	183	160	146
Tbilisi	Negat	177	180	177	166	181	194
	Pos	181	175	172	191	173	153
Gori	Negat	171	171	164	168	170	163
	Pos	175	172	173	176	169	171

Table 5 SPI (3,6,12) and SPEI (3,6,12) categories for Telavi station

Drought indices		SPEI (3)	SPEI (6)	SPEI (12)	SPI (3)	SPI (6)	SPI (12)
Exceptional drought	<-2	5	7	3	15	11	5
Moderate drought	<-1	58	51	53	52	46	58
Exceptional wet	> 2	3	5	2	2	12	9
Moderately wet	> 1	62	58	66	56	49	62

Table 6 SPI (3,6,12) and SPEI (3,6,12) categories for Tbilisi station

Drought indices		SPEI (3)	SPEI (6)	SPEI (12)	SPI (3)	SPI (6)	SPI (12)
Exceptional draught	< -2	4	3	2	10	9	2
Moderate drought	< -1	62	64	62	56	48	44
Exceptional wet	> 2	3	0	2	9	7	14
Moderately wet	> 1	66	68	66	55	62	58

Table 7 SPI3(3,6,12) and SPEI (3,6,12) categories for Gori station

Drought indices		SPEI (3)	SPEI (6)	SPEI (12)	SPI (3)	SPI (6)	SPI (12)
Exceptional draught	< -2	1	3	5	8	7	9
Moderate drought	< -1	63	58	48	54	54	50
Exceptional wet	> 2	4	2	4	5	4	5
Moderately wet	> 1	58	60	55	58	58	51

According to the definition of SPEI the results of surface water deficit may be interpreted as dryness that is the indicator of drought conditions. For Telavi that represents dry region irrigation is major concern. The region is significant agriculture producer and intensively uses Alazani river water for irrigation. According to SPEI index the region is most subjected to moderate dryness than by SPI.

Tbilisi the capital of Georgia has its own climatic regime known as Tbilisi Cavern (Tsitsagi et al., 2022). It is more subjected to moderate drought conditions.

Gori represents Shida Kartli agricultural region that known by its fruit production. It is undoubtedly under moderate drought conditions.

The obtained results indicate that there is a difference between the droughts depicted by the precipitation based SPI and the evaporation influenced SPEI caused by inter-annual and seasonal variability of temperature and precipitation trends, land cover, vegetation, irrigation and etc.

Table 8 Statistical parameters of stations/models located on the Kakheti and Kartli regions

Stations	Standard deviation (station/ model)	Pearson correlation (station/model)	MAE (station/model)
Shilda	49.53	0.65	59.49
Shiraqi	41.79	0.61	47.16
Surami	35.67	0.53	60.11
Tbilisi	25.257	0.79	13.24
Tchiauri	36.88	0.54	45.30
Telavi	48.51	0.47	64.66
Tetri-tskaro	45.53	0.50	63.08
Tsalka	44.20	0.41	58.92
Tsiteli_Khidi	28.24	0.59	32.20
Tskhinvali	32.622	0.57	54.08
Tskneti	41.40	0.49	48.80
Tsnori	38.92	0.53	49.69
Udabno	31.58	0.543	37.73
Vanati	39.25	0.663	50.25
Vaneli	55.935	0.583	79.235
Zemo Akhalsopeli	41.68	0.693	28.395
Zghuderi	41.698	0.60	39.44

4 Conclusion

The Machine Learning technique is the important module of the growing field of big data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them (Publish House, 2022).

Overfitting is a concept in data science, which occurs when a statistical model fits exactly against its training data (Publish House, 2022). When this happens, the algorithm unfortunately cannot perform accurately against unseen data, defeating its purpose. Generalization of a model to new data is ultimately what allows us to use machine learning algorithms every day to make predictions and classify data.

The first region that was subjected to ML analysis is Kakheti, which was revealed as the most vulnerable to drought risk. The existed stations data of 1961–1990 year period were subjected to QC on heterogeneity and continuity test and only 19 stations passed it. Based on these data, the Standardized Precipitation Index (SPI) for 3 months was calculated and then subjected to Machine Learning. In this case, the most optimal

result was shown by the “optimized tree”, where the minimum number of leaves is equal to 45, and the training time is 30,296 s, the prediction speed_ \sim 43,000 obs/sec (Publish House, 2022).

The Support Vector Machine (SVM) was selected in the Matlab space, it is the algorithm for Supervised Machine Learning, which allows us to optimize it, for example, we can control the number of divisions in the “machine tree”, which help to achieve the higher accuracy of the model. As it was revealed the tenth model showed the best result, with the using of this model it became possible to determine the drought probability by months at each point. The model was trained based on historical records and for forecasting by moth the new 1990–2020 year period was added.

Despite obtained good parameters, it became needed to add additional stations from adjacent territories, because there was not enough information in the Kakheti region to conduct the correct analysis of Machine Learning. While comparing the model and station data the overfit was detected, this is not valid for the analysis of new data. It became necessary to add the neighbor Kartli region also. In this case it is remarkable that stations that are displayed closely to each other give better results to the model for this two regions. The stations that are far away increase the probability of overfit. Therefore, Machine Learning needs to increase the observation network density- the number of stations, the observation period, as well as the use of satellite data for the region where there is not enough observation is the better choice. Especially for mountainous country where exist places that aren't suitable to conduct ground observations.

The additional region contains 17 meteorological stations. The SPI (3) drought index uses precipitation data of those stations. The new prediction model has improved parameters: RMSE-0.77, R^2 -0.39, MSE-0.58, and MAE-0.60.

The calculated statistical parameters of this two regions stations data and prediction model are enough fitted. Therefore the following is necessary to avoid overfit: increase number of dataset; increase of time study period, more nearest supporting points of observations.

Consequently, to apply the Machine Learning technique it is essential to increase the observation network- the number of stations and also increase the observation period too. As well as use satellite data for the region where there is not enough observation data and station set up is impossible is essential. Such research is conducted firstly for Georgian territory and it'll be continued using more station data of large period to cover the whole territory. The research results are important for early warning system.

Acknowledgements The authors are grateful to the Georgian National Environment Agency (NEA) for providing the meteorological data.

Funding This work was supported by the Shota Rustaveli National Science Foundation of Georgia (SRNSFG) [FR-21–13962].

References

- 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. A Macrotheme Capital Management. LLC Academic Publication, ISSN 1848–4735, v. 2, iss.1, Winter 2013 (January), France, pp. 148–154.
- Tatishvili M. Developing Weather Forecasting System in Georgia. Ecology & Environmental Sciences, 2 (7), 2017, <https://doi.org/10.15406/mojes.2017.02.00046>.
- Tatishvili, M. R., Megrelidze, L. D., & Palavandishvili, A. M. (2021). Study of the Mean and Extreme Values, Intensity and Recurrence Variability of Meteorological Elements Based on the 1956–2015 Observation Data. Journals of Georgian Geophysical Society, 24(2). <https://doi.org/10.48614/ggs2420213325>
- Ana Palavandishvili Evaluation of natural disasters using machine learning and big data for Georgia. International Conference of Young Scientists “Modern Problems of Earth Sciences”. Proceedings, ISBN 978–9941–36–044–2, Publish House of Iv. Javakhishvili Tbilisi State University, Tbilisi, November 21–22, 2022, pp. 111–115. <http://openlibrary.ge/handle/123456789/10263>
- Evans, N. P., Bauska, T. K., Gázquez-Sánchez, F., Brenner, M., Curtis, J. H., & Hodell, D. A. (2018). Quantification of drought during the collapse of the classic Maya civilization. Science, 361(6401). <https://doi.org/10.1126/science.aas9871>
- Kaniewski, D., Guiot, J., & Van Campo, E. (2015). Drought and societal collapse 3200years ago in the Eastern Mediterranean: A review. In Wiley Interdisciplinary Reviews: Climate Change (Vol. 6, Issue 4). <https://doi.org/10.1002/wcc.345>
- DeMenocal, P. B. (2001). Cultural responses to climate change during the late holocene. In Science (Vol. 292, Issue 5517). <https://doi.org/10.1126/science.1059827>
- Babre, A., Kalvāns, A., Avotniece, Z., Retiķe, I., Bikše, J., Jemljanova, K. P. M., Zelenkevičs, A., & Dēliņa, A. (2022). The use of predefined drought indices for the assessment of groundwater drought episodes in the Baltic States over the period 1989–2018. Journal of Hydrology: Regional Studies, 40. <https://doi.org/10.1016/j.ejrh.2022.101049>
- McKee, T. B., Nolan, J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. Preprints, Eighth Conf. on Applied Climatology, Amer. Meteor. Soc., January;
- Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. Journal of Climate, 23(7). <https://doi.org/10.1175/2009JCLI2909.1>
- Tatishvili M., Palavandishvili A., Tsitsagi M., Gulashvili Z., Suknidze N. Drought Evaluation Based on Spei, Spi Indices for Georgian Territory. International Conference of Young Scientists “Modern Problems of Earth Sciences”, Proceedings, ISBN 978–9941–36–044–2, Publish House of Iv. Javakhishvili Tbilisi State University, Tbilisi, November 21–22, 2022a, pp. 119–121.
- Copernicus European Drought Observatory (EDO): <https://edo.jrc.ec.europa.eu/>
- Tatishvili, M. R., Palavandishvili, A. M., Tsitsagi, M. B., & Suknidze, N. E. (2022b). The Use of Structured Data for Drought Evaluation in Georgia. Journals of Georgian Geophysical Society, 25(No.1). <https://doi.org/10.48614/ggs2520224806>
- Mariam Tsitsagi, Zaza Gulashvili, Nana Bolashvili, Marika Tatishvili, Nikoloz Suknidze RELATIONSHIP BETWEEN NORMALISED DIFFERENCE VEGETATION INDEX, PRECIPITATION AND DROUGHT INDICES (CASE OF KAKHETI, GEORGIA) 22nd International Multidisciplinary Scientific GeoConference SGEM 2022 <https://doi.org/10.5593/sgem2022/4.1/s19.46>
- Marika Tatishvili, Zurab Khvedelidze, Inga Samkharadze, Ana Palavandishvili. Atmosphere processes and climate parameters variation in River Mtkvari basin Int. Sc. Conf. “Natural Disasters in Georgia: Monitoring, Prevention, Mitigation”. Proc., Tbilisi, 2019, pp. 117–121. <http://dspace.gela.org.ge/handle/123456789/8648>