Bulgarian Journal of Meteorology and Hydrology

Национален институт по метеорология и хидрология

National Institute of Meteorology and Hydrology

Bul. J. Meteo & Hydro 28 /2 (2024)

Automated geostatistical analysis and mapping of SPI using open-source software

Krastina Malcheva*, Tanya Marinova, Radoslav Evgeniev

National Institute of Meteorology and Hydrology,
Tsarigradsko shose 66, 1784 Sofia, Bulgaria

Abstract: Accurate analysis of droughts at regional and local scales requires high-resolution drought indices data. The rapid advancement of open-source GIS software in recent decades has greatly simplified geoprocessing tasks by implementing user-friendly tools for workflow automation that have made conducting operational spatial analyses possible. The approach presented in this study utilizes the analytical power of the R programming language in a QGIS environment to automate the calculation, spatial modeling and mapping of the Standardized Precipitation Index (SPI) using data from the operational meteorological network of the National Institute of Meteorology and Hydrology. The results show a significant reduction in processing time and improved expert control over the spatial interpolation of SPI. The generated summarized information, data, and maps facilitate the rapid preparation of reports and expert assessments of the precipitation deficit or excess.

Keywords: SPI, open-source GIS, spatial modeling, automated mapping

1. INTRODUCTION

Drought is a complex, climate-related phenomenon with significant socioeconomic impacts. Drought events vary widely regarding their severity, duration, and affected area. The National Institute of Meteorology and Hydrology (NIMH) supports the Ministry of Environment and Water (MoEW) and Basin Directorates (BDs) in using and implementing several key drought indicators, providing GIS-based operational monthly assessments of drought conditions. The Standardized Runoff Index (SRI) is a

* krastina.malcheva@meteo.bg

hydrological index used to determine the drought presence and extent in a particular river basin (Shukla&Wood, 2008). The Soil Moisture Index (SMI) estimates the available soil moisture content by measured data (Hunt et al., 2009). The Standardized Precipitation Index (SPI) is one of the most commonly used indices to evaluate meteorological drought and was recommended by the World Meteorological Organization (WMO) as a key drought indicator (WMO, 2012).

SPI was developed to determine the precipitation deficit or excess at different time scales. It represents a standardized departure of the modeled by selected probability distribution precipitation from its long-term mean (McKee et al., 1993).

Among the open-source GIS projects, QGIS (QGIS, 2024) is probably the most popular. Combining R (R Core Team, 2024) and QGIS software produces a novel, enhanced environment for geostatistical computing. QGIS provides a suitable interface, letting users integrate R scripts as “user-defined” tools. Processing Modeler in QGIS utilizes its own visual programming language to automate complex geoprocessing tasks, which describes the workflow in the form of detailed diagrams rather than as a generalized processing model (Dobesova, 2020). Automating manual tasks (such as input, transformation and pre-processing data) significantly reduces the risk of human error. Additionally, automation allows for much faster processing of large datasets or performing complex calculations by creating scripts, ultimately improving the efficiency of data analysis (QGIS, 2024).

The present study aims to demonstrate the advantages of integrating own-developed R scripts for the calculation and spatial interpolation of SPI with geoprocessing algorithms and tools in the QGIS environment for automated GIS-based monitoring and long-term drought assessment using data from the operational meteorological network of NIMH for the period 1961–2020.

2. DATA AND METHODOLOGY

The previously exploited workflow for automated calculation of the monthly SPI values at three time scales (1, 3 and 6 months) in the points of the operational meteorological network utilized the specialized software of the US National Drought Mitigation Center (NDMC, https://drought.unl.edu/monitoring/SPI/SPIProgram.aspx), recommended in WMO (2012). The computation of SPI is performed on daily precipitation data from 162 meteorological stations representative of precipitation conditions on the territory of the country (Figure 1). The main historical period with available precipitation data is from 1951 onwards. The calibration period used is 1971–2000. The file of input data, in a standard format, is automatically generated after expert control of the regularly collected information from the operational network. The developed FORTRAN program provided: 1) pre-processing of the input file, which includes re-formatting of data by stations as time series and adding them to the corresponding multi-year time series; 2) calling of the external module for calculating SPI, and 3) extracting the obtained values by stations and populating the information into a standard output file. The calculated SPI values at a 3-month time scale (SPI-3) are interpolated by the Ordinary Kriging method (with a pixel of ~25 km^2) to produce raster layers and contour maps. The segmentation by BDs clearly delineates the regional differences in drought extent. The maps of SPI-3 for each calendar month, along with the archived maps for the past 11 months, are available on the website www.hydro.bg.

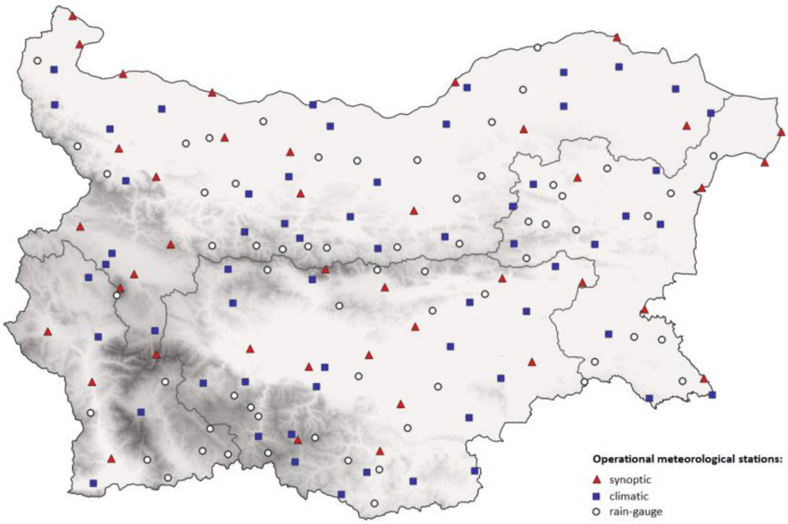

Fig. 1. Location of the NIMH operational stations on the country's territory as of 2020.

McKee et al. (1993) defined drought as a period when SPI is consistently negative and reaches an intensity of -1.0 or less. Classifying dryness and wetness conditions into more SPI categories provides a more detailed insight into the evolution of the precipitation deficit or excess. The scale used by NIMH, which includes categories *mild drought* and *slightly wet*, is presented in Table 1. Drought events occur at an intensity of -0.5 or lower, while wet events have an intensity of 0.5 or higher.

Table 1. SPI scale in nine categories, including the corresponding intensity threshold values, cumulative probabilities and the categories' labels

SPI value	Cumulative Probability	Classification	
-3.0	0.0014	Extreme drought	SPI ≤ -2.0
-2.5	0.0062		
-2.0	0.0228		
-1.5	0.0668	Severe drought	-2.0 < SPI ≤ -1.5
-1.0	0.1587	Moderate drought	-1.5 < SPI ≤ -1.0
-0.5	0.3085	Mild drought	-1.0 < SPI ≤ -0.5
0.0	0.5000	Around normal	-0.5 < SPI < 0.5
0.5	0.6915	Slightly wet	0.5 ≤ SPI < 1.0
1.0	0.8413	Moderately wet	1.0 ≤ SPI < 1.5

1.5	0.9332	Very wet	1.5 ≤ SPI < 2.0
2.0	0.9772	Extremely wet	2.0 ≤ SPI
2.5	0.9938		
3.0	0.9986		

The proposed new approach for automated calculation and mapping of SPI in the QGIS environment follows the methodology of the existing workflow of precipitation data processing and generating SPI maps, but differs significantly in its realization. It integrates the own-developed scripts for SPI calculation (based on the R-package "SPEI"; Beguería&Vicente-Serrano, 2023) and kriging interpolation (based on the R-package "automap"; Hiemstra et al., 2008) with other appropriate geoprocessing algorithms and tools using the QGIS Processing Modeler. The visual interface not only allows for the algorithmization of spatial analysis but also provides a convenient way to edit and adjust the settings at each workflow step.

The Ordinary Kriging method is widely used for spatial analysis of drought intensity and extent. This method is based on evaluating the spatial autocorrelation of data using appropriate correlation functions (Hengl, 2007). The SPI values calculated on observational data are interpolated in a regular grid of approximately 5×5 km. Choosing a model and corresponding geostatistical parameters (sill, nugget and range) to fit the empirical semivariogram is crucial in applying kriging interpolation. The *autoKrige* function in the R package "automap" automatically fits the variogram model to the data. The initial sill value is estimated as the mean of the maximum and median of the semivariance. The initial range is 0.1 times the diagonal of the bounding box of the spatial data, and the initial nugget is defined as the minimum of the semivariance (Hiemstra et al., 2008). However, the automatically estimated model parameters are sensitive to outliers and tend to the semivariogram overfitting, resulting in maps with extreme gradients occurring over short distances surrounding individual locations. These errors require expert control and are typically easy to identify at larger time scales, where maps should have a smoother transition between wet and dry areas (Lucas et al., 2022).

The *Zonal Statistics* tool in QGIS calculates one or multiple statistics from a given raster layer based on zones defined in a specified vector layer. In this study, boundary polygon layers at the country and BD levels are used. The areal statistics are extracted from ordered by month raster layers, thus forming respective time series for the period 1961–2020.

3. RESULTS AND DISCUSSION

3.1. Development of an automated workflow for calculating and mapping SPI

Both the generalized (Figure 2) and detailed (Figure 3) flowcharts of automated processing of precipitation data from the NIMH operational network depict the processing stages, the algorithms and tools used in the QGIS environment to create the final product – SPI maps. The input data are organized in tables with a suitable structure. The calculation module generates SPI data, which is supplied to the interpolation module and the point layer creation algorithm. The processing of the raster layer obtained after interpolation includes cutting along the contours of the individual BD and preparing SPI choropleth maps at a given scale and color scheme (Figure 4).

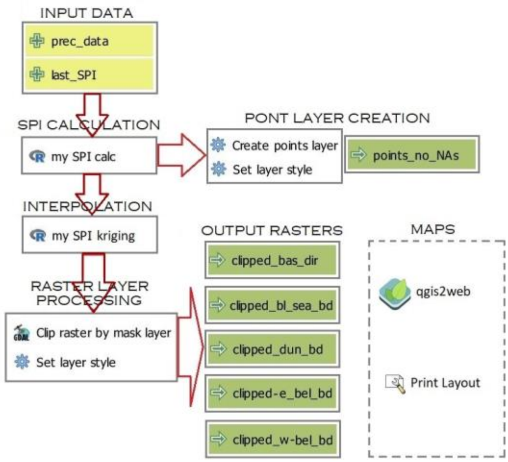

Fig. 2. Generalized flowchart of the automated processing, including the processing steps, algorithms and tools used in the QGIS environment.

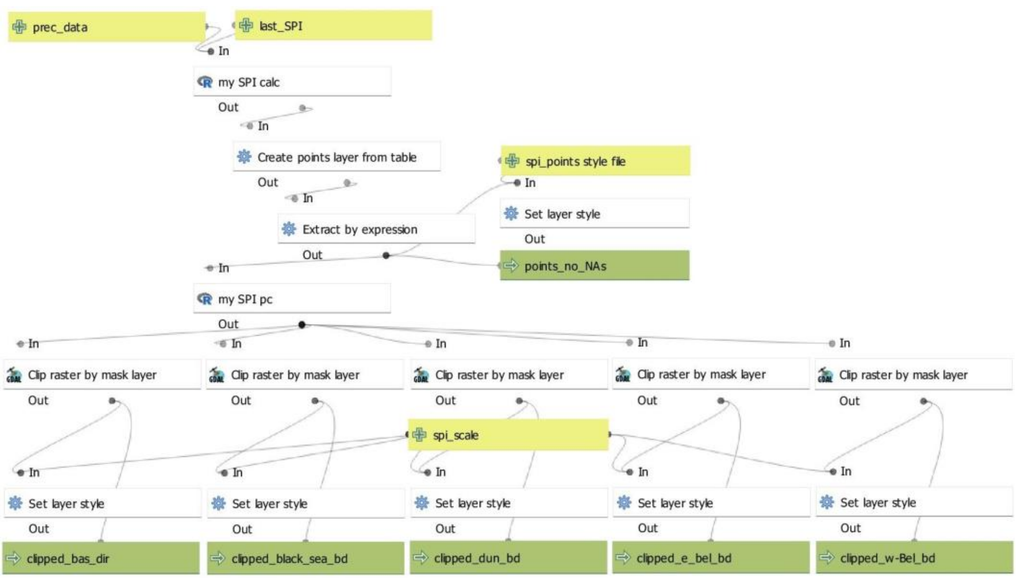

Fig. 3. Detailed flowchart of the automated processing in the QGIS environment.

The calculation module represents an R script for computing SPI at different time scales using the package “SPEI.” Pre-processing includes extracting daily precipitation data from the input file and aggregating it to monthly values by station. The calculated SPI values at the defined time scales of 1, 3 and 6 months (SPI-1, SPI-3, and SPI-6) are organized in three matrices, covering the period from 1961 onwards. For the study period (1961–2020), the matrices have dimensions of 720×162 each.

The second R script generates raster layers from the computed SPI values by kriging interpolation using the packages “automap” and “raster” (Hijmans, 2024). Under default settings, the *autoKrige* function automatically selects the most appropriate model amongst a few predefined variogram models. However, both the model type and parameters (range, sill, and nugget) can be tuned manually. This option enables flexible expert control to achieve more accurate spatial estimates of SPI, considering the impact of model selection and parameterization on the kriging efficiency.

In operational mode, the interpolation is applied only to the last row (i.e., the latest month) of the matrices created by the calculation module. However, the total number of generated rasters for the study period 1961–2020 is 10800. All rasters are cut along the contours of the country and individual BDs and input to the relevant raster stack, ordered by months. The main features, zonal statistics, and other important information are extracted from raster layers and summarized in tables for further analysis.

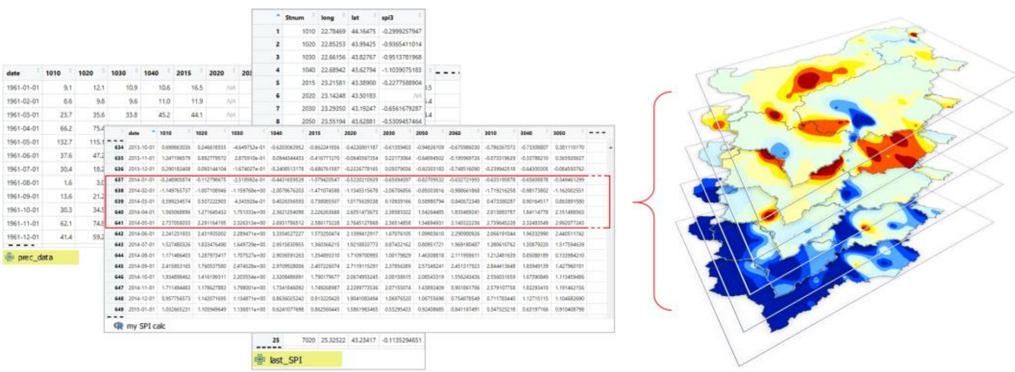

Fig. 4. Data structure at the individual processing steps and part of the generated raster stack.

The automated workflow described here allows for the fast generation of maps and expert analyses under developing droughts and easy transition to the latest WMO recommendations (WMO, 2024) on drought indicators, monitoring and early warnings. Additionally, it enables the flexible use of various spatial statistics for long-term analyses of precipitation deficit. However, expert control is necessary at this development stage to avoid errors and discrepancies between the input data, the calculated SPI values, and the automatically generated output rasters.

Along with eliminating the expert control in the next stage, the batch of five monthly SPI maps shall be automatically generated from the clipped rasters using the

QGIS Atlas tool on a standard template prepared in the QGIS Print Layout (Fig. 5, right). Including QGIS Atlas in the workflow significantly improves quality and speeds map production. In addition, the QGIS2Web plugin could be used to create interactive web maps (Figure 5, left). It is a flexible and user-friendly tool that uses the Java libraries OpenLayers or Leaflet to create web maps of the same quality as those made in the QGIS Print Layout (QGIS, 2024). The exported SPI maps could be uploaded to a web server, providing the opportunity to be updated automatically.

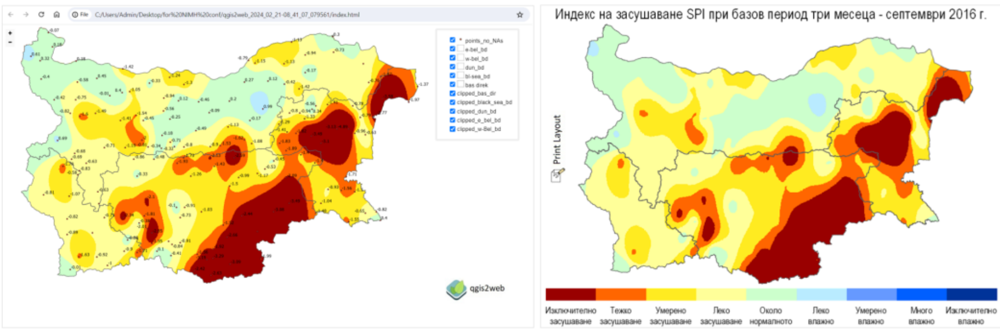

Fig. 5. Interactive map of SPI-3 for September 2016 (left) and the same map intended for the website www.hydro.bg (right) as an example of the final mapping product.

It is essential to evaluate the influence of software change on the SPI estimates. Because the SPI archive has been available since 2014, we used the period 2014–2023 to compare the calculated SPI-3 values by the "SPEI" package with the NDMC program output. As seen in Figure 6, right, *spei* reveals less wetter but drier conditions; however, nearly 88% of differences fall within the range ±0.5, and 52% do not surpass ±0.2 (Figure 6, left).

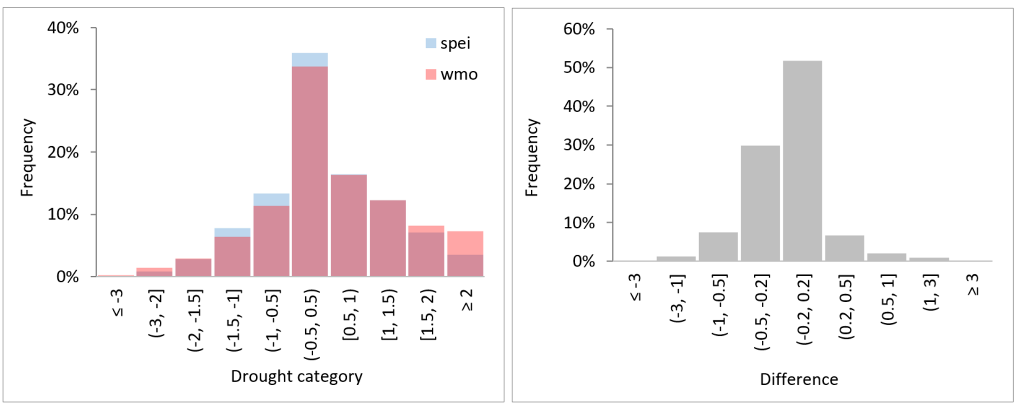

Fig. 6. Comparison of the calculated SPI-3 values by the R-package "SPEI" (noted as *spei*) with the NDMC program output (noted as *wmo*) for the period 2014–2023.

3.2. Geostatistical analysis of SPI for the period 1961–2020

The raster analysis tools in QGIS were used to process the raster stack, extract the average values at the country and BD levels and form time series.

At a country level (Figure 7, left), the period is characterized by balanced wet and dry conditions, but the intra-annual distribution shows a relative preponderance of very dry over very wet conditions (Figure 7, right).

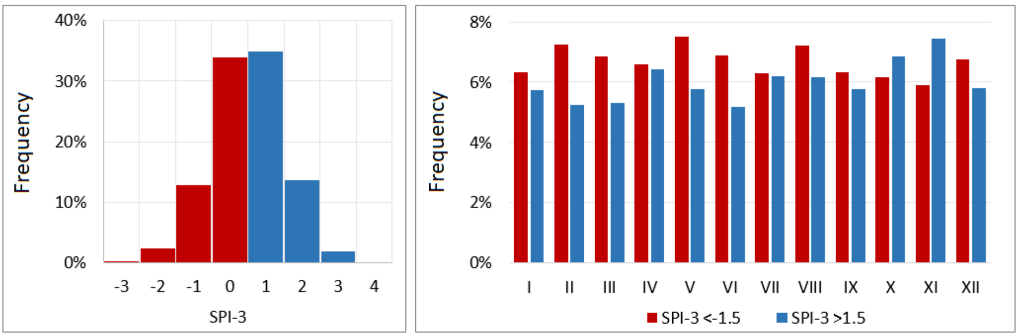

Fig. 7. Results of the SPI-3 statistical analysis for the period 1961–2020 on the generated raster stack. Left: overall frequency of SPI values; Right: frequency of the very dry and very wet conditions by month.

The analysis reveals a long period of prevailing drought, from 1982 to 1994, with over 80% of the country's territory affected in individual months. The most intense droughts were in 1968, 1990 and 2000 (in some stations, the SPI-3 values reached minus 4.5-4.7), while 1966, 2005 and 2014 stand out as very wet years (Figure 8).

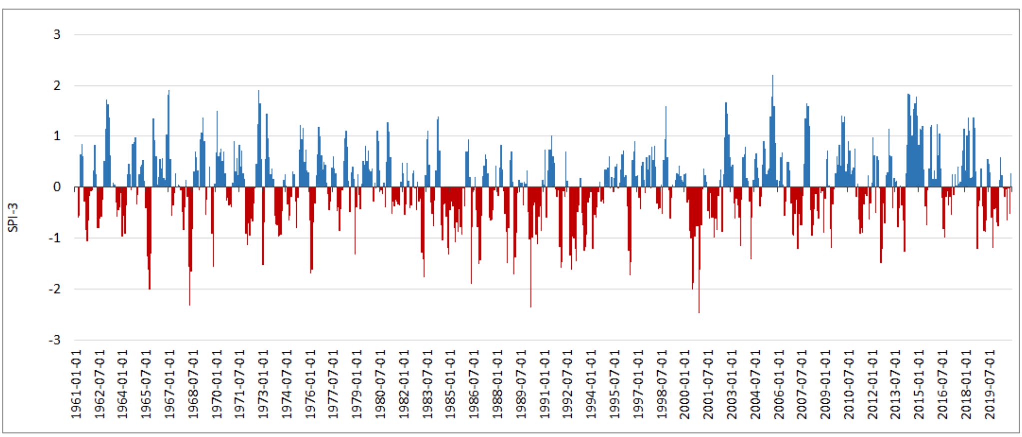

Fig. 8. Long-term variation of SPI-3 in the period 1961–2020 at a country level, extracted by the generated rasters.

4. CONCLUDING REMARKS

The proposed methodology and developed workflow speed up the overall processing of precipitation data and preparing operational drought analyses. The results show a significant reduction in total processing time and an improvement in the expert control capabilities of SPI modeling. Automating geoprocessing tasks in QGIS by incorporating R scripts in processing models allows for generating customized maps with just a few clicks, which significantly facilitates geostatistical analyses and preparing multiple maps, as well as reducing the chances of mistakes. A QGIS plugin is currently being developed, including the QGIS Atlas tool and web mapping in the workflow. It can be seen as a step towards developing an advanced monitoring system.

REFERENCES

Beguería, S., & Vicente-Serrano, S.M. (2023). SPEI: Calculation of the Standardized Precipitation-Evapotranspiration Index, https://spei.csic.es, https://github.com/sbegueria/SPEI

Dobesova, Z. (2020). Evaluation of Effective Cognition for the QGIS Processing Modeler. Appl. Sci. 10, 1446. https://doi.org/10.3390/app10041446

Hengl, T. (2007). A Practical Guide to Geostatistical Mapping of Environmental Variables, EUR 22904 EN. Luxembourg: Office for Official Publications of the European Communities, JRC38153

Hiemstra et al. (2008). Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network. Computers & Geosciences, http://dx.doi.org/10.1016/j.cageo.2008.10.011

Hijmans, R. (2024). raster: Geographic Data Analysis and Modeling. R package version 3.6-28, https://rspatial.org/raster

Hunt, E.D., Hubbard, K.G., Wilhite, D.A., Arkebauer, T.J., & Dutcher, A.L. (2009). The development and evaluation of a soil moisture index. Int. J. Climatol., 29, pp. 747-759, https://doi.org/10.1002/joc.1749

Lucas, M.P., Longman, R. J., Giambelluca, T.W., Frazier, A.G., Mclean, J., Cleveland, S.B., Huang, Y., & Lee, J. (2022). Optimizing Automated Kriging to Improve Spatial Interpolation of Monthly Rainfall over Complex Terrain. Journal of Hydrometeorology, 23(4), 561-572. https://doi.org/10.1175/JHM-D-21-0171.1

McKee, T.B., Doesken, N.J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales, In Proceedings of the 8th Conference on Applied Climatology, http://ccc.atmos.colostate.edu/relationshipofdroughtfrequency.pdf

QGIS.org (2024). QGIS 3.28. Geographic Information System User Guide. https://docs.qgis.org/3.28/en/docs/

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Shukla, S., & Wood, A.W. (2008). Use of a standardized runoff index for characterizing hydrologic drought. Geophysical Research Letters, 35, L02405, https://doi.org/10.1029/2007GL032487

WMO (2012). Standardized Precipitation Index User Guide. [M. Svoboda, M. Hayes and D. Wood], WMO-No. 1090. World Meteorological Organization, ISBN 978-92-63-11091-6, 16 p., https://library.wmo.int/index.php?lvl=notice_display&id=13682

WMO (2024). EC-78, Executive Council Abridged Final Report of the Seventy-eighth Session, Geneva, 10–14 June 2024. WMO-No. 1358. World Meteorological Organization, ISBN 978-92-63-11358-0. https://library.wmo.int/idurl/4/69109