A Data Analysis and Interactive Visualization Approach With SPI for Initial Drought Assessment in Guanajuato, Mexico

José De Jesús Sandoval-Palomares, CIATEC, Mexico
Oscar Manuel Rodríguez-Narváez, CIATEC, Mexico

The Asian Conference on the Social Sciences 2024
Official Conference Proceedings

Abstract

Droughts present a series of significant problems that affect both the environment and society. Problems arise in diferents intensity, scope, type of impact by region, for example water scarcity affects supply of water for human consumption, agriculture, industry and the environment. To address these problems, it is crucial to implement water management strategies and a continuous monitoring are essential to predict and mitigate the effects of droughts. The Standardized Precipitation Index (SPI), tells us if it has rained normally in a region during a certain time range. The SPEI allows assessing the severity of drought on different time scales. Guanajuato state, nestled in central Mexico, have a semi-arid climate, with an average annual rainfall of 650mm. This work uses data analysis techniques to conduct an initial drought assessment for state of Guanajuato. The Python informatic language was used to generate a web plataform, data visualization, generating interactive maps, informative graphs, and comprehensive data tables, we uded the information of 29 climate monitoring stations, and we analyzed data spanning from January 2000 to November 2023, resulting in a final dataset of 8,294 entries. Preliminary results from our ongoing work presents time series of SPI records helped visualize key characteristics such as trend, seasonality, and cyclicity. Each station's SPI is visualized by a heat map, where vibrant colors paint the picture of their values over time, across different years.

Keywords: SPI, Data Analysis, Mexico, GTO

Introduction

Guanajuato (Gto), nestled in central Mexico, basks in a semi-arid climate. With an average annual rainfall of 650mm and 30°C peaks in May-June, it's no stranger to warm weather.[1] But recent years have seen temperatures intensify, casting a shadow of drought across the region.

To track the dryness, scientists have turned to various tools, with the Standardized Precipitation Index (SPI) emerging as a useful tool. This handy metric, used in climate studies, quantifies precipitation anomalies in a specific area over time. It essentially weighs current rainfall against historical averages, offering a clear picture of drought, excess rain, or normalcy. Positive values signify rain abundance, negative ones of drought, and values hovering near zero shows a balanced precipitation (Table 1).[2,3]

Table 1. Range and interpretation of the "Standardized Precipitation Index" (SPI) standard.

SPI Value	Interpretation
-2.0 or ‹	Extreme drought
-1.5 to -1.99	Severe drought
-1.0 to -1.49	Moderate drought
-0.99 to 0.99	Normal precipitation
1.0 to 1.49	Moderate excess precipitation
1.5 to 1.99	Severe excess precipitation
2.0 or ›	Extreme excess precipitation

Though the SPI index unveils drought patterns, its extensive historical data can be unwieldy. Data analysis, encompassing collection, cleaning, exploration, and interpretation, offers a solution. This approach not only streamlines data management but also unlocks valuable insights to inform decisions.[4]

The PSI have some strengths like:

Data-efficient: Relies solely on precipitation data, making it applicable in regions with limited climate data.

Multi-scale analysis: Measures drought/wetness severity at different "time windows" relevant to various water resources, like soil moisture or groundwater.

Climate-independent: Performs better than PDSI (Palmer Severity Drought Index) when comparing across regions with diverse climates.

Simple calculation: Relatively straightforward to compute compared to PDSI.

And limitations like:

Ignores evapotranspiration: Doesn't consider water loss through evaporation and plant transpiration, potentially underestimating drought impact in warmer climates.

Data sensitivity: Requires reliable and long-term precipitation data (30-50 years preferred) for accurate calculations.

Precipitation intensity: Focuses solely on precipitation amount, missing the impact of intense rainfall patterns on runoff and water availability.

One of the SPI's most valuable features is its ability to adapt to different timescales 1, 3, 6, 12, 18, 24, 36, 48, and 60 months. While it can pinpoint drought intensity for specific periods like a month or five months, it also excels at capturing the bigger picture by analyzing drought patterns across various timescales simultaneously. Researchers often use precipitation data for durations ranging from one to 60 months to compute the SPI for each timeframe, providing a multi-faceted view of drought conditions. Furthermore, spatial and temporal data visualization creates a crucial reference framework.[4,5,6,7]

Rain doesn't fall evenly, It tends to come in bursts, with many small showers and fewer, heavier downpours. This uneven pattern is called an asymmetric frequency distribution. The reception distribution is usually skewed, with low values and high. To obtain a more precise result, it is advisable to fit the distribution to a mathematical function, such as the gamma distribution or the Pearson III.

The SPI, takes into account the average rainfall in a region and how much it varies over time, the bell curve representing the typical rainfall pattern. The SPI tells us how far a specific rainfall amount falls from the "sweet spot" in the middle of that curve.

The transformed precipitation data are then used to compute the dimensionless SPI value, defined as the standardized of the precipitation [4], The summary calculation of PSI is:

$SPI = (P-P^*) / \sigma p$

where P = precipitation, p* = mean precipitation.

To visualize the data, time series were used, which are sequences of data that are collected or recorded at regular time intervals. In this case, the data collected was daily, but grouped in months, from the year 2000 to 2023. Time series help visualize key characteristics such as trend (a long-term movement in the data), seasonality (patterns that occur repeat at regular intervals), and cyclicity (fluctuations that do not have a fixed pattern or that repeat at irregular intervals). A computational interface was developed with a graphical view where the user can, on the same graph, view all the data, modify the time interval or select the graph to see a set of data of interest in a zoom view, without requiring re-generation. a new query.[8]

Recognizing the limitations and insights discussed previously, this project seeks to utilize the SPI index as a synergistic indicator of drought, alongside data analysis techniques, to establish a comprehensive spatial and temporal database of drought occurrences within Guanajuato.

Methodology

The initial phase of this project involved developing the data analysis model, outlined in the following section.

Data collection: Precipitation data was retrieved from Conagua's drought monitor repository by querying historical values.[1]

Data Cleansing, Data Exploration and Data Preparation: Building the analysis model for this project began with data acquisition. We accessed historical precipitation data for all Mexican stations through CONAGUA's repository. Using Python, we then cleaned and explored the data, ultimately preparing it by filtering for stations located in Guanajuato state (prefix "11"). This step ensured our analysis focused on the relevant data for our project.

The initial data extraction from the relevant stations yielded a data structure not conducive to direct analysis, as shown in Table 2. To overcome this obstacle, we meticulously reorganized the data into a more user-friendly format, presented in Table 3. This involved several steps: firstly, the extraction of the first two rows containing geographical position data, which were subsequently placed in a dedicated file named "GPS_pos.csv". Next, we embarked labeling process, renaming the first and second columns to "year" and "month", respectively. Additionally, a new "Record" column was introduced to provide a consecutive index for each entry. Finally, the data was filtered to retain only columns beginning with "11", signifying GTO status. Importantly, other columns pertaining to the monitoring stations remained untouched throughout this process.

Table 2. Original CONAGUA precipitation data.

Stn	**1004**	**1005**	**…**	**32055**
Lat	22.01	21.78	…	22.1
Lon	-102.2	-102.14	…	-102.87
1951	1	-999.99	…	-999.99
…	…	…	…	…
2023	9	-2.71	…	-2.35
2023	10	0.4	…	-0.94

Table 3. Modifications to allow data analysis.

Item	**year**	**month**	**1004**	**1005**
2	1951	1		
3	1951	2		
6	1951	5		
…	…	…	…	…

Data analysis uncovered a data type mismatch for the " month " variable. Initially defined as "object," it was later assigned "int64" (Table 4). To ensure data consistency and accurate analysis, this discrepancy needs to be addressed.

Table 4. Conversion from "Object" data type to "Int".

Item	Int64	Item	Int64
year	Objet	year	Int64
month	float64	month	float64
11001	float64	11001	float64
11002	float64	11002	float64
11003	float64	11003	float64

We analyzed data spanning from January 2000 to November 2023, resulting in a final dataset of 8,294 entries. This equates to 286 rows and 29 columns. Notably, some station columns contained missing values represented by NaN, 0, and -999.99. These missing values constituted 0%, 0.482%, and 7.73% of the total data, respectively (Figure 1).

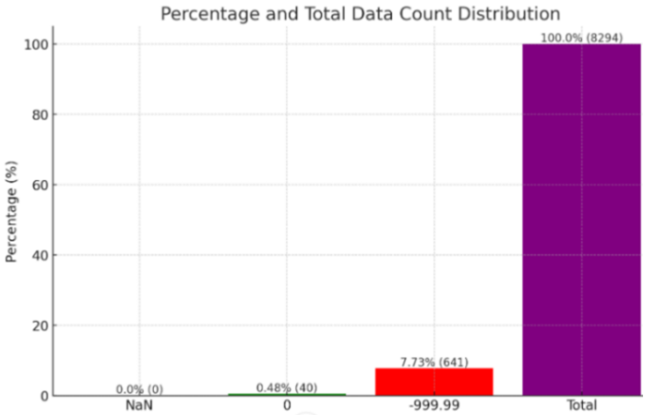


Figure 1. Representative table of missing records Vs total data from the 29 stations in the state of GTO.

Consequently, these records were filled with values, using an average between the previous and next year and the same month, to assign a referenced numerical value.

Analysis techniques: A temporal view in maps by year and a time series view of all records were used.

Data analysis, Interpretation of results, Communication of results and Decision making: These proposed stages, including a potential web interface, would be implemented in the project's second phase.

Results

Figure 2 presents time series of SPI records, while Figure 3 displays the geographical locations of the stations. These initial findings suggest localizations for mitigation actions.

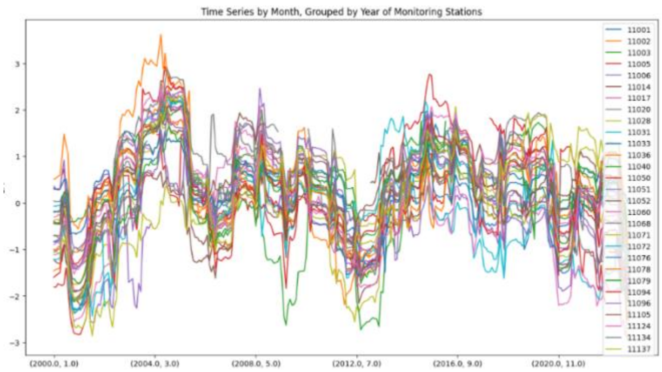
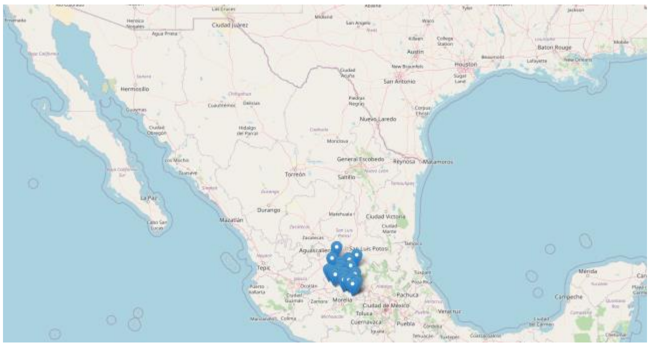
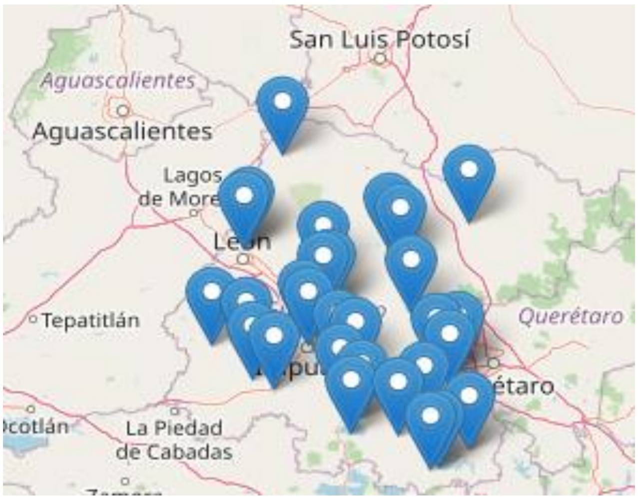


Figure 2. Time series of GTO status SPI records 2000-2023



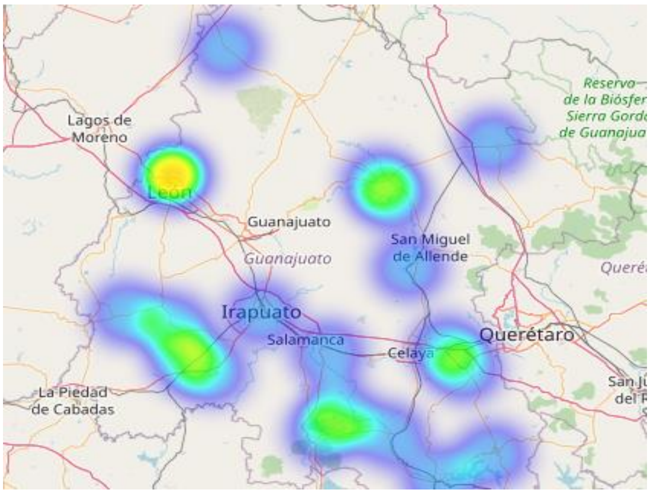
a)



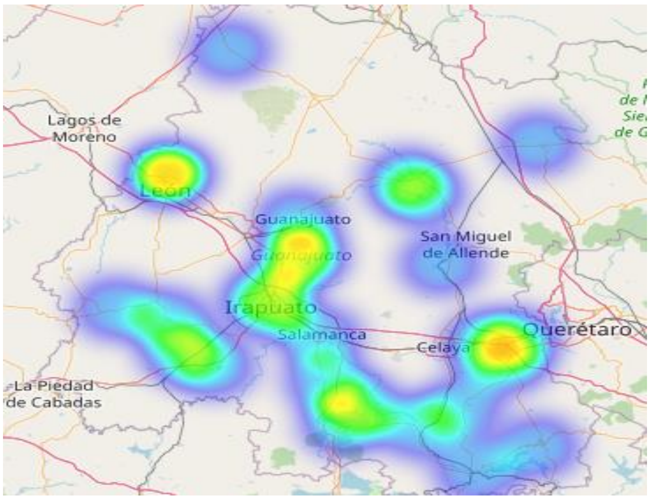
b)

Figure 3. a) Geographic location view of stations within Mexican territory, b) detailed view of stations within the state of GTO.

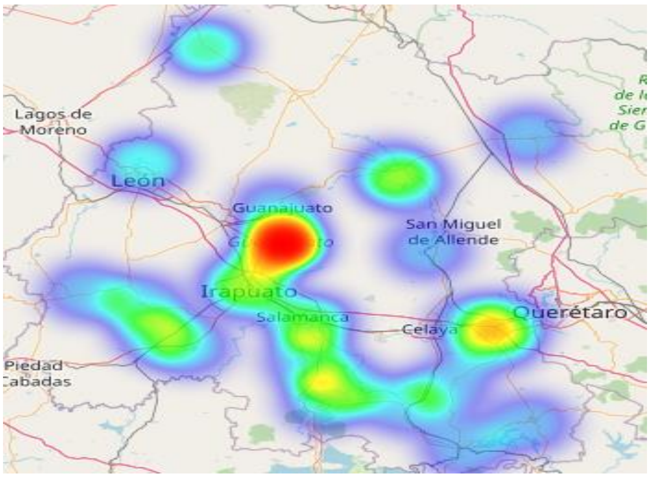
In Figure 4, a map comes alive, revealing a network of stations pulsating with data. Each station's health is visualized by a heat map, where vibrant colors paint the picture of their SPI values over time, across different years. GPS pinpoint's location right on the map, and the station numbers act as guides. a zoom in for a closer look and get a deeper understanding of its data.



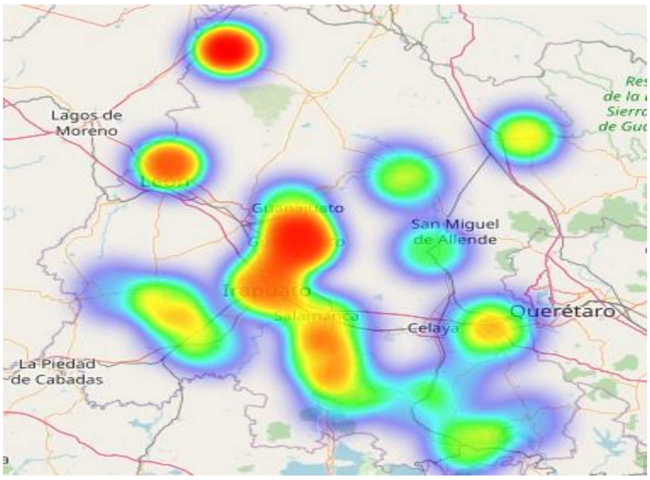
a) Year 2000



b) Year 2010



c) Year 2015



d) Year 2020

Figure 4. GPS position and heat map, of different years and color according to their SPI value.

Conclusions

The Standardized Precipitation Index (SPEI) allows assessing the severity of drought on different time scales, which is important to understand the different impacts of drought on hydrological, environmental and socioeconomic systems. Droughts are a challenging phenomenon to monitor and predict. This is because they can occur quickly, they can depend on processes that are not well resolved in forecast models, and the challenges associated with droughts vary depending on meteorological variables, regional distributions, seasonality and climatic risk factors and of the earth's surface. This suggests that there are different forms of drought development and therefore there is no single definition of drought that is universally accepted.

This study paves the way for applying the SPI methodology in a more comprehensive manner, specifically within the context of the ongoing drought affecting Mexico, particularly the state of GTO. This can serve as a valuable model for addressing similar challenges in the future.

The final interface design is evolving with input from drought researchers. Their focus on real-time data integration will equip the GTO state government with the tools to make informed decisions and address drought challenges proactively.

Acknowledgements

We thank the IDEA GTO fund [9], which is providing economic resources to integrate a project of Data Science applied to water problems, for the generation of national and international scenarios in solutions to regional droughts.

References

[1] Comisión Nacional del Agua (CONAGUA). (2024). Servicio Meteorológico Nacional. Visualizador Tabular de Modelos de Pronóstico Numérico. https://smn.conagua.gob.mx/es/modelos-de-pronostico-numerico/visualizador-tabular-de-modelos-numericos

[2] Hadley Wickham. (2016). SpringerLink: Data Analysis. https://link.springer.com/chapter/10.1007/978-3-319-24277-4_9

[3] Ian H. Witten, Eibe Frank y Christopher J. Pal. (2011). Morgan Kaufmann

[4] National Center for Atmospheric Research (NCAR). (2023). Standardized Precipitation Index (SPI). https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-index-spi

[5] Nazire Goksu Soydan Oksal. (2019). Manfred Mudelsee, Trend analysis of climate time series: A review of methods. Earth-Science Reviews, Volume 190, Pages 310-322, ISSN:0012-8252, https://doi.org/10.1016/j.earscirev.2018.12.005

[6] Oliver Schabenberger, Carol A. Gotway.Statistical Methods for Spatial Data Analysis, 1st Edition, ISBN 9781032477916

[7] Manfred Mudelsee. (2019). Trend analysis of climate time series: A review of methods, Earth-Science Reviews,Volume 190,Pages 310-322,ISSN 0012-8252, https://doi.org/10.1016/j.earscirev.2018.12.005

[8] Comparative analysis of the influence of temperature and precipitation on drought assessment in the Marmara region of Turkey: an examination of SPI and SPEI indices. Journal of Water and Climate Change 1 September 2023; 14 (9): 3096–3111. doi:https://doi.org/10.2166/wcc.2023.179

[9] Gobierno de estado de Guanajuato. (2023). Conexión internacional de la ciencia. https://idea.guanajuato.gob.mx/convocatorias

Contact emails: jsando@ciatec.mx
orodriguez@ciatec.mx