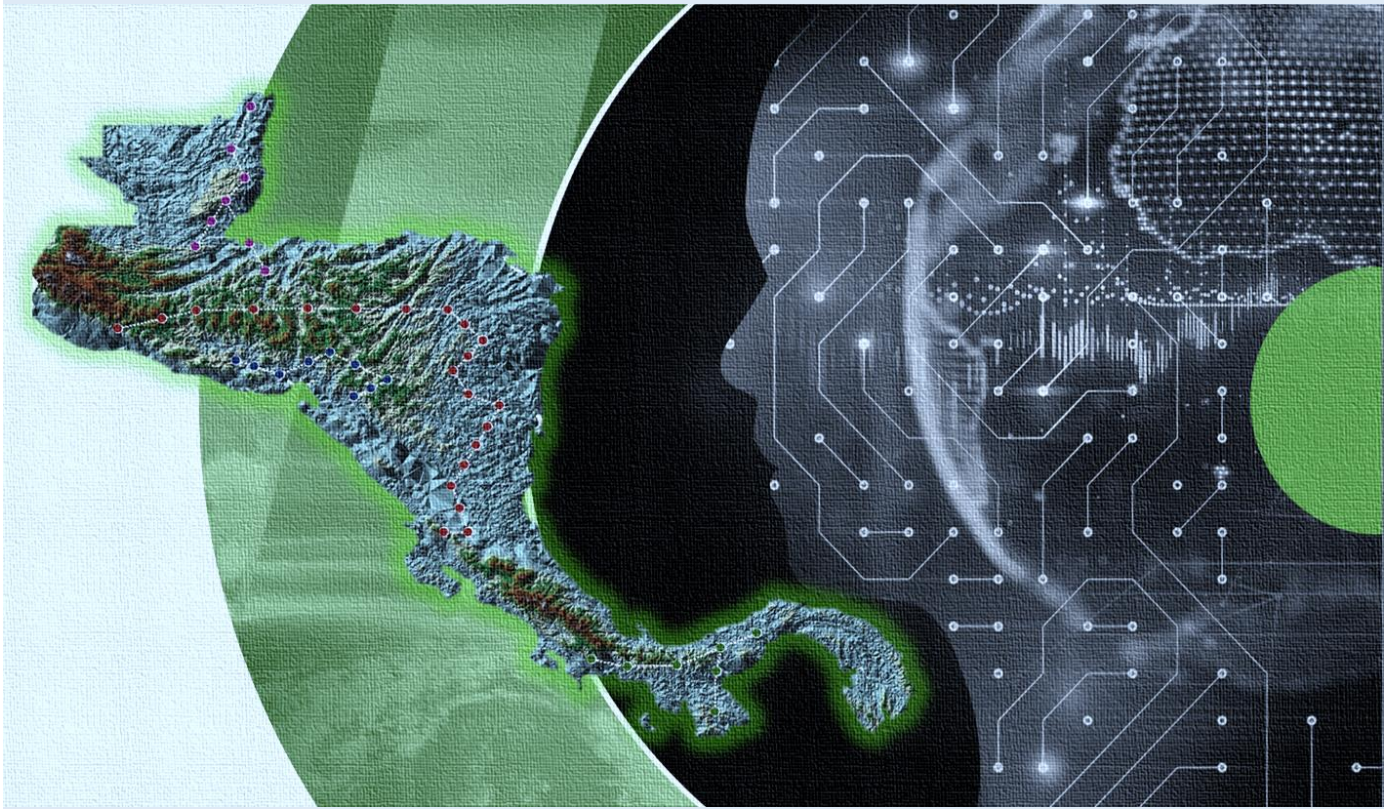
Machine Learning Methods for Characterising and Tracking Spatiotemporal Drought Events

Case Study: Central America Dry Corridor

Karel Aldrin Sánchez Hernández

MSc Thesis Identifier WSE-HI.21-22
April 2021

Updated version, April 2021



ESCUELA
COLOMBIANA
DE INGENIERÍA
JULIO GARAVITO

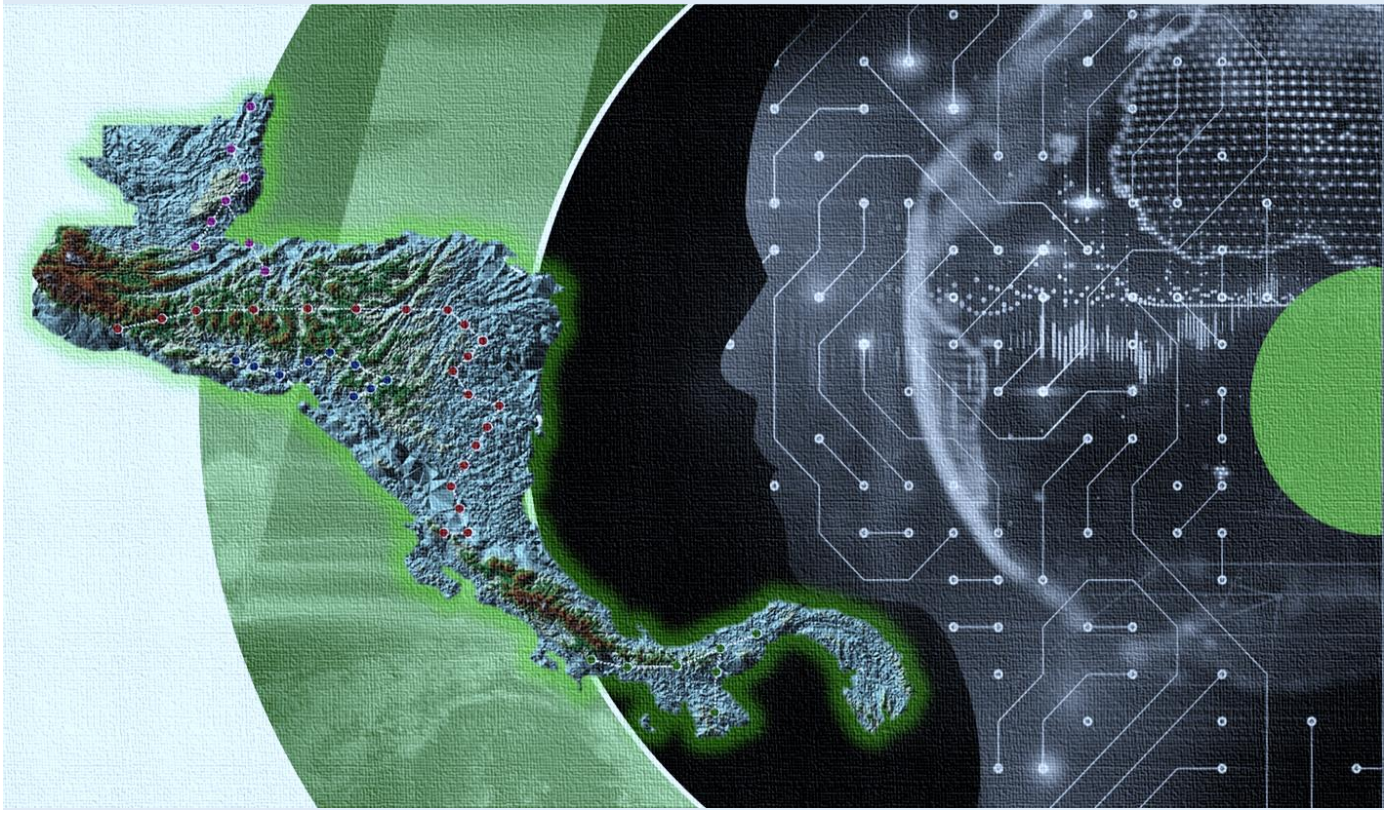
Machine Learning Methods for Characterising and Tracking Spatiotemporal Drought Events

Case Study: Central America Dry Corridor

Karel Aldrin Sánchez Hernández

MSc Thesis Identifier WSE-HI.21-22
April 2021

Updated version, April 2021



Machine Learning Methods for Characterising and Tracking Spatiotemporal Drought events
Case Study: Central America Dry Corridor

Master of Science Thesis
by
Karel Aldrin Sánchez Hernández

Supervisors
Associate Professor Schalk Van Andel
PhD. German Santos Granados

Mentors
Associate Professor Gerald A. Corzo Perez

Examination Committee
Associate Professor Gerald A. Corzo Perez
Associate Professor Schalk Van Andel
PhD. German Santos Granados
PhD. Francisco Muñoz Arriola

This research is done for the partial fulfilment of requirements for the Master of Science degree at the IHE Delft Institute for Water Education, Delft, the Netherlands.

Delft
30/03/2021

Abstract

Drought is a natural, erratic phenomenon that has a widespread and significant impact on socioeconomic and environmental development. The early monitoring and evaluation of drought through forecasting models would allow the articulation of early control and mitigation strategies, thus achieving an optimal development in planning and preparation for climate change. Therefore, this research developed a methodology for spatiotemporal analysis of drought patterns using automatic learning tools in the dry corridor of Central America. To this end, some specific milestones were defined. These include: (i) To assess temporal and spatial meteorological and agricultural droughts events, (ii) Identifying and validate results of the spatiotemporal events using computer vision techniques and finally (iii) Implementing machine learning drought forecasting models.

ERA 5 monthly land average dataset was used as input for index estimation, spatiotemporal analysis and forecasting models. The frequency of drought events was calculated using standardized SPI and SPEI indices for accumulation periods of 1,3,6,9. However, 3,6 allowed a more realistic analysis of the seasonal change conditions in the hydrological regime of the area and the identification of the existing teleconnection between drought events and scale propagation. Regarding the spatiotemporal dynamics, 97 drought events of greater extension were identified, which are generally originated in countries such as Guatemala, Nicaragua, and El Salvador between seasonal periods not longer than 7 months. Additionally, the suitability of automatic learning models such as SVR, ANN and deep learning such as LSTM for index forecasting (r2=0.80) and drought dynamics in a temporal window of 1 to 6 months ahead was verified with considerable performance.

The presented methodology provides an important basis for drought characterization and forecasting through the integration of spatiotemporal tracking models and machine learning techniques. Therefore, the methodological development can be adapted as an instrument for monitoring and forecasting, articulated to management and early mitigation policies. Finally, we suggest adapting variables related to the orographic context, relief, land use and land cover change, for instance, to improve the forecasting performance of the exposed forecasting models.

Acknowledgements

Before all, I would like to thank my family because with their unconditional support, love and sacrifice, they have pushed me to be a better person, and without their support, collaboration and inspiration, it would have been impossible to complete this research.

I would also like to express my deepest gratitude to my mentor and supervisors: Dr Gerald Corzo, Dr Schalk van Andel and Dr German Santos, since their guidance and technical and scientific advice allowed me to fully develop both the research topic and my expectations in the research environment.

Special thanks to my grandmother for her support and example of strength in difficult situations. And my classmates for their faithful support and companionship. Thank you for your love, patience, and faithful support. Thank you.

Table of Contents

List of Figures

List of Tables

Chapter 1 General Introduction

1.1 Background

Regarding the WMO (World Meteorological Organization, 2006), droughts relate to water deficiency in a particular region, and their severity has potential repercussions in diverse contexts. Recent studies indicate that droughts' frequency and severity appear to be increasing in some zones due to climate variability and change (Singh, 2010). Also, its occurrence is cyclical and is related to the El Niño Southern Oscillation ENSO (Zee Arias et al., 2012), which influences both the irregularity of the precipitation cycle and the spatiotemporal magnitude of impact at scales. Regional and local, limiting the natural resilience capacity of the territory.

Relate to (Diaz et al., 2020b), spatiotemporal methodologies applied to droughts have been developed. These have managed to generate a better understanding of spatiotemporal dynamics, strengthening their monitoring and impact reduction.

Several approaches have been proposed to describe the spatiotemporal development of drought. According to Bacanli et al., 2009; Dracup et al., 1980, they propose methodologies which apply to force a significant number of meteorological variables in the drought indexes calculation based on time series disaggregation of relating it to the spatiality of all zone, establishing a unique phenomenon in the zone. Therefore, they did not consider the spatial network's interconnection; that is to say, this approach is multivariable and not contiguous.

On the other hand, it is enough to understand past events' dynamics and generate predictions that serve to improve the contingency systems and local control. Sheffield et al., 2009. The first contextualisation refers to the predictive models in droughts that separated space and time to establish empirical relations between climatic variables and the indexes derived from the observations, in other words, statistical models, which due to their simplicity and low computational costs, have been traditionally used.(Belayneh et al., 2014)

With the advance of these techniques, today, we have more robust models, such as machine learning, neural networks (deep learning), that allow us to manage different types of information and algorithms to facilitate, optimise modelling and forecasting. (Zhang et al., 2016). Several hydro-climatology applications, such as Visual Search, Face Recognition, and Image Pattern Recognition, are being used to imitate variables interactions between predictors and predicting. That has been widely used and exceeds classical modelling and prediction performance, such as machine learning and deep learning techniques, that can incorporate space and time data for imagery collection analysis and dynamic measures.(Morid et al., 2007)

1.1.1 Droughts

Droughts are natural anomalies in rainfall patterns associated with extreme climatic events of significant impact (Belayneh et al., 2014; Dai, 2011). These occur throughout the planet and are generally associated with variations in precipitation and temperature, so which causes a water deficit in the region, such as decreased soil moisture, runoff and groundwater and unfavourable socioeconomic implications. (Mishra et al, 2010; Van Loon, 2015). The level and depth of monitoring can vary spatiotemporally and, in turn, depending on the monitoring system's robustness.

1.1.2 Droughts in Central America

The Central America Dry Corridor is made up of Mexico, Honduras, El Salvador, Guatemala, Puerto Rico, Nicaragua, and Panama, as is shown in figure 1. Due to its geographical location, the dry corridor is the most susceptible to natural disasters and to receive direct effects of climate variability. Meteorological phenomena such as hurricanes, intense waves of drought and floods, and volcanic activity are some of the events that make this region vulnerable. (ECLAC, 2015).

Figure 1 Area Delimitation of Central America Dry Corridor

Source: PISCS-UCREA,2018

This area has two main precipitation patterns, one for each ocean current. Towards the Atlantic, this zone presents a bimodal rain regime and a low extension of the dry season due to the intertropical dynamics and wind currents that make its hydrologic distribution favourable. In contrast, the Pacific borders represent the aggravating problem since they are the ones that experience intense droughts and hydric deficit. (Dorsch and Fischer, 2015).

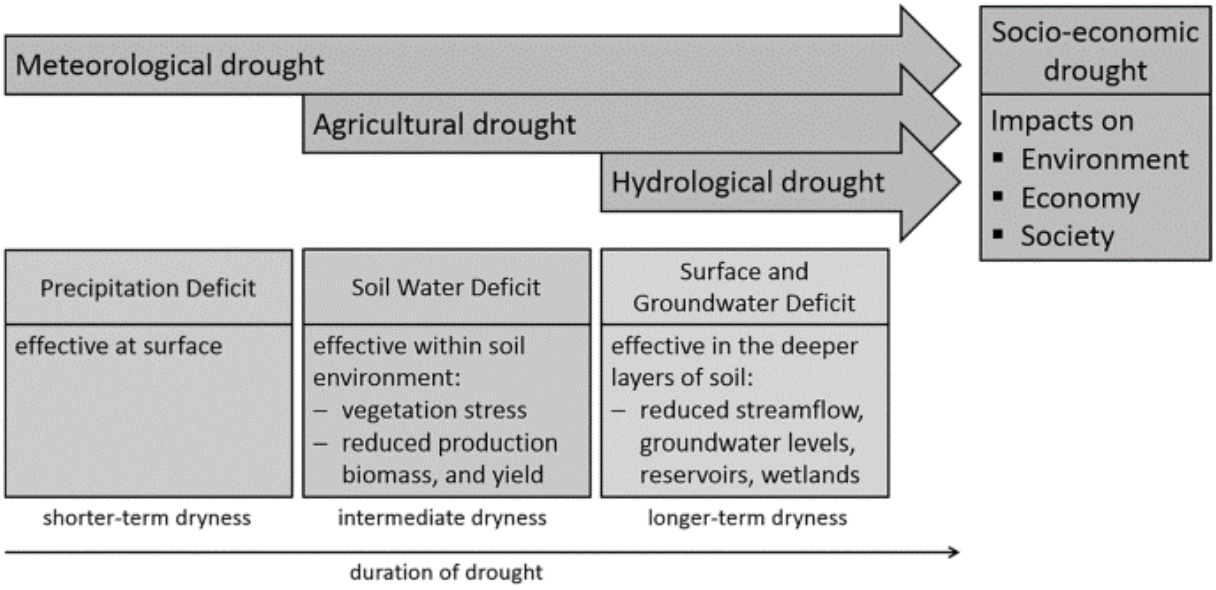
According to the Intergovernmental Panel on Climate Change (IPCC), extreme events related to climate change can magnify its impact on this area, environmental resilience, water availability would be one of the catastrophic consequences that may occur mainly due to reduced rainfall and evapotranspiration in semi-arid areas, expanding its impact on the socioeconomic sectors, energy (MAGRIN ET AL., 2014) and given the social dependence on agriculture, these would generate food insecurity (van der Zee Arias,2012)

Historically, the frequency between high-impact events allowed for the gradual recovery of ecosystems. In recent years meteorological patterns and records have focused attention on the formulation of monitoring strategies since there has been a percentage increase of more than 23% in temperature and a 12% decrease in Precipitation since 2015. (FAO, 2019)

1.1.3 Drought Assessment

Drought is a prolonged period of below-normal water availability. It is a globally recurrent phenomenon with spatiotemporal characteristics that vary significantly from one region to another. In general, droughts are clustered into five types of meteorological, agricultural, hydrologic, groundwater and socioeconomic (Van Loon, 2015). Authors such as Wilhite,1985, Van Loon,2016 have considered that droughts present propagation between groups (Fig 2), exerted according to the processes that drive their persistence and severity, such as climate, socioeconomic activities, generating multiple stays both temporally and spatially. (Weedon et al., 2011)

Figure 2Drought Types and Impacts



Source: (Crocetti et al., 2020)

Due to their complexity, drought forecasting has presented a significant challenge to scientists and decision and policymakers because of their diverse origins and occurrence. In general, three principal types of methods have been currently used for drought prediction: statistical, dynamic and hybrid methods. (Podestá et al., 2016). However, before forecasting, it is necessary to group factors into a single variable representing the study's phenomenon, which translates into drought indices. (Hao et al., 2018)

1.1.4 Drought Indices

As mentioned in the previous section, before making drought forecasts, a quantitative variable or indicator must be defined to estimate drought conditions of a specific type (meteorological, agricultural, among others), and each one varies according to its scale and factors that govern its dynamics. However, there is no universally accepted drought indicator to characterise the drought condition at a global level. (Okal et al., 2020; Van Loon, 2015)

1.1.5 Drought Forecasting

According to the Drought Monitoring framework for properly drought knowledge, it is enough to understand past events' dynamics and generate predictions that serve to improve the contingency systems and local control. The first contextualisation refers to the predictive models in droughts that separated space and time to establish empirical relations between climatic variables and the indexes derived from the observations, in other words, statistical models, which due to their simplicity and low computational costs, have been traditionally used. (Anshuka et al., 2019)

However, in recent years models have been structured that articulate the space-time of events and prediction, called hybrid models by Frieler et al., 2017; Keyantash & Dracup, 2002; Strazzo et al., 2019. Nevertheless, the predictive module has a low capacity to relate the high density of variables with time due to its complex interaction (linear/non-linear), performing low-quality forecasts regarding precision and insurability.(Heye et al., 2017; Murakami et al., 2016)

In the last decade, authors like (Belayneh et al., 2014; Poornima & Pushpalatha, 2019) have started to explore the automatic learning techniques in this field, from self-regression models like ARIMA, SARIMA and models under the Bayesian philosophy. Additionally, some authors have suggested incorporating machine learning techniques such as ANN´s, RNN´s, Hybrid Models to improve the time prediction paradigm, where currently they are employed for short and medium-term forecasting. (Djerbouai & Souag-Gamane, 2016)

1.2 Motivation

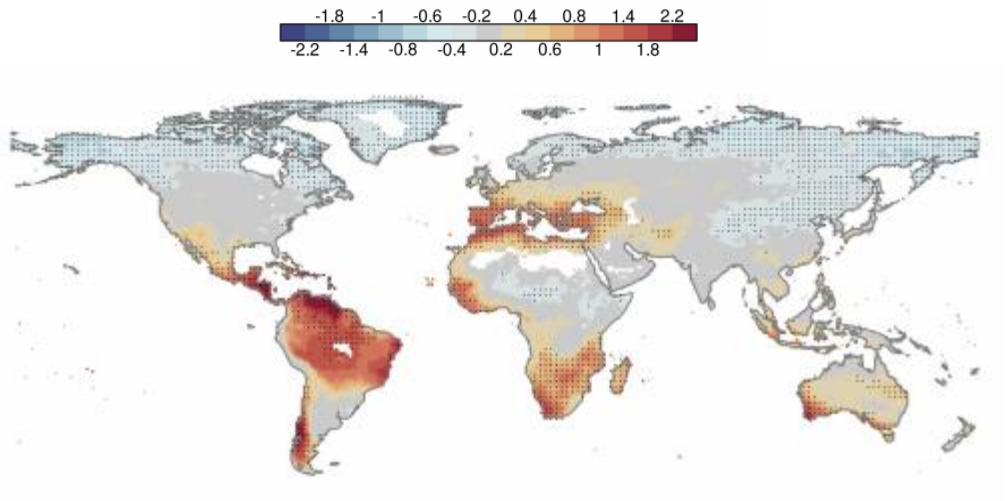
As mentioned above, drought is one of the most complex water phenomena in tracking, monitoring, and forecasting due to the density of variables influencing and modifying any place's conditions. Such studies have traditionally separated the space-time relationship in the qualification and quantification of drought events, simplifying the models.

Under probabilistic adjustments, most countries use such empirical methodologies to project and plan the territory; however, such simplicity generates a paradigm of uncertainty and precision. Current models have a limited capacity in articulating the synchronicity with spatiotemporal drought monitoring.

Changes in future drought, as well (AghaKouchak et al., 2015)exposes may be due to changes in the average and the variability of Precipitation. Central America mainly in hydric deficit projections made in 2100, is one of the regions along the Mediterranean and Africa, their increase in water deficit is a latent problem (Fig 3). Thus, offering considerable potential for

generating strategies based on the control and monitoring of drought to mitigate the underlying risk.

Figure 3 Potential Drought Projections to 2100



Based on: (AghaKouchak et al., 2015)

Despite several drought indices, the suitability of these with a temporal analysis of more than 30 years had not been tested for the dry corridor. The fact that there is a lack of an assessment of drought spatiotemporal appropriate can potentially be used to define conditions forecasting, control, and monitoring as a tool for government and non-government support.

1.2.1 Research Statements

- How can I use **spatiotemporal drought analysis** in forecasting?
- Is it possible **to develop** this type of methodology in Central America?
- Can pattern analysis and **propagation dynamics** allow for spatial relationships between drought-related synchronous events?
- How far does the **drought prediction** capability of a machine learning algorithm ensure high reliability?

1.2.2 Research Hypothesis

- Drought Indexes can identify different impacts associated with drought at different time scales
- The Spatiotemporal characterisation schemes can be modelled using data-based models.
- Machine learning (ML) algorithms can learn about hydro-climatic variable interactions and drought events.
- The implementation of machine learning models based on spatiotemporal conceptualisation can improve conventional drought modelling techniques.

1.2.3 Research Objectives

This research aims to implement machine learning methods for monitoring and forecasting the spatiotemporal characteristics of drought events in the Central American Dry Corridor.

Specific objectives are:

- To assess temporal and spatial meteorological and agricultural droughts events.
- Identify and validate results of the spatiotemporal events.
- Implement machine learning drought forecasting models.

This research promotes a robust framework for drought analysis and evaluation from a social, technical, and scientific perspective. The technical foundation and required science to support the spatiotemporal dynamics of drought can promote alternatives in innovation and aspects related to management efficiency and control measures in social, economic, and environmental areas. In addition to being a proposed methodology of easy use and application, it can face implementation and application challenges.

1.3 Innovation

Understanding climate variability and spatiotemporal drought dynamics allow improving and strengthening monitoring and forecasting systems. This research presents and develops a methodology that allows improving the spatiotemporal drought analysis from a dynamic scope to integrate to the monitoring, teleconnection, or synchronicity concepts together with automatic learning that allows improving the control and forecasting systems.

Among the critical aspects of innovation, the following are highlighted:

- Potential past events assessment using reanalysis data, thus improving conventional drought indices' computation.
- The results of the quantitative evaluation of the drought indexes used have provided information to consider the possible climatological variables responsible for the drought in Central America, which allows not only the development of new models of monitoring and evaluation of drought impacts with multiple approaches.
- Articulation of computational vision, such as the technique of connected components, has demonstrated its success in capturing drought patterns, allowing for the analysis of spatiotemporal characteristics such as location and the study time area's percentage variation.
- The developed drought forecasting models using machine learning algorithms developed contribute as a forecasting tool with applicability in early warning systems to future drought conditions under dynamic context.
- The framework for developing more robust forecasting systems using deep learning allows exploring new panoramas of spatiotemporal analysis, thus improving the

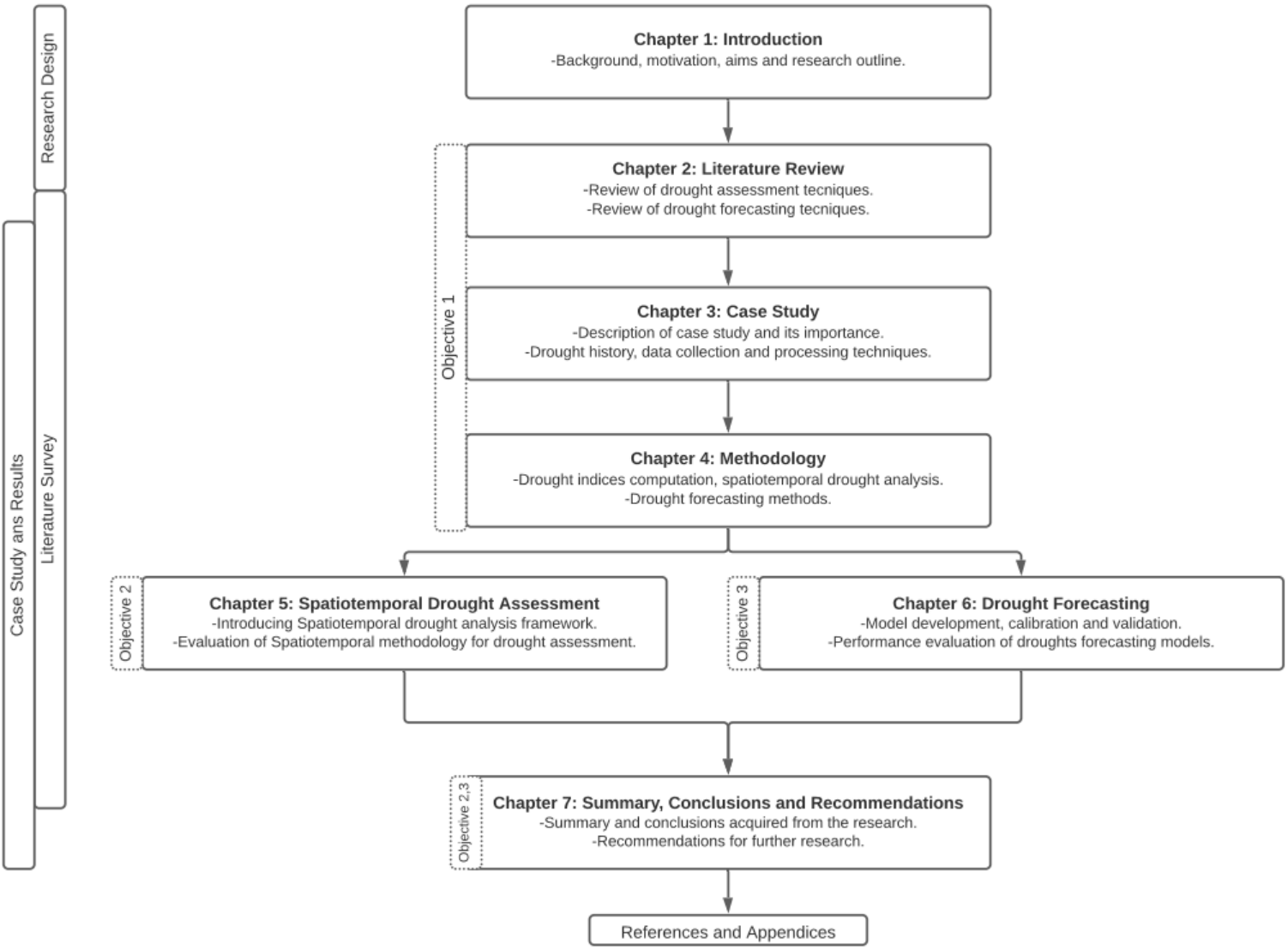
understanding of the causes, effects and timely mitigation, thus limiting potential socioeconomic, environmental and sustainable impacts.

1.4 Research Outline

This thesis is structured in seven chapters. Additionally, this thesis contains three other chapters that follow the specific objectives proposed above of the general introduction, methodology, and summary chapters (Fig 4). These will be contributing to the research framework of spatiotemporal drought assessment and forecasting.

- Chapter 2 reviews literature basics related to drought assessment, spatiotemporal drought analysis and drought forecasting frameworks.
- Chapter 3 contains the general research methodology, methods, technical requirements that are implemented along with the study
- Chapter 4 and 5 contains hydro climatological trends and relations with drought in the study zone, drought past events evaluation by indices (meteorological and agricultural), and spatiotemporal characteristics and propagation patterns analysis using computer vision techniques.
- Chapter 6 explore some approaches in the forecasting framework applied to drought, efficiency, and accuracy of machine learning algorithms to estimate drought indices and spatiotemporal characteristics.

Figure 4 Research Structure



1.4.1 General Outcomes

- The direct relationship between meteorological and agricultural conditions was explored by estimating drought indices, which analysed how the drought's magnitude is related to the hydro climatological parameters that influence its development intensity.

- The pattern tracking techniques by artificial vision, introduced in this study for the spatiotemporal analysis, capture the movements and magnitudes of drought events through their grouping and proximity analysis criteria. Synchronous relationships between events were possible by establishing propagation dynamics based on the studied indices.

- Drought event forecasting techniques allowed to evaluate the variation between the accuracy and complexity of machine learning algorithms and their promising results in consolidating multitemporal forecasting systems applicable to any area.

Chapter 2 Literature Review

This chapter aims to summarise the existing research literature on drought monitoring and forecasting. Establishing the current literature related to drought assessment and forecasting seeks to support the technical and scientific basis required in developing the research.

Different sections of this chapter also provide relevant context for the issues discussed in subsequent chapters. Being a starting point that reflects how the exposed methodology contributes to the investigation on the evaluation and spatiotemporal forecast of the drought. The objectives of this chapter are:

- Contextualise, classify, and analyse drought events and their impact on the various economic, social and political systems.
- To present a general context related to the framework of drought monitoring and forecasting.
- Generate an analysis of the methodologies and development of modelling and forecasting science.

2.1 Climate-Driven Factors and Drought Classification

Changes in land use and hydrological cycle significantly impact carbon flows and greenhouse gas (GHG) emissions, altering the atmospheric and soil composition. Hence, degradation and drought accentuate climate change because of the lower potential of geographical areas in carbon dioxide ($CO_{2)}$ capping. The most common factors directly related to drought (Sen, 2015) are described below.

- **Precipitation**

It is the most relevant climate factor for determining areas threatened by drought, as it plays a vital role in the development and distribution of flora, where the variability of rains and extreme rains can lead to soil erosion, which over a long period can lead to desertification (World Meteorological Organization, 2006).

- **Temperature**

Temperature is the primary determinant of climate and, consequently, vegetation distribution, soil formation, and precipitation. (World Meteorological Organization, 2006) However, high air temperatures can stop photosynthesis, prevent plant eggs fertilisation and induce dehydration, affecting agricultural production and impacting ecosystem balance. (Fernandez, 2013)

- **Soil Moisture**

Moisture affects organisms due to its relationship to oxygen supply, limiting soil fixating organisms (Guerrero, 2020). Therefore, soil moisture is defined as the presence or absence of water available or usable for the plant in the soil or on one of its soil layers during specific periods of the year.
For this purpose, the soil is considered dry when the water is retained with a voltage greater than 1500 kPa and is considered to have water available to the plant when the soil moisture is retained at less than 1500 kPa above zero kPa voltage. (Jaramillo, 2002)

According to the multiplicity of drought meanings, four (4) types of drought have been categorised according to scientific discipline, where (Blain, 2012; Van Loon, 2015) derive four (4) types of drought, which can be seen in Table 1 and Fig 5.

Figure 5 Spatiotemporal Drought Classification

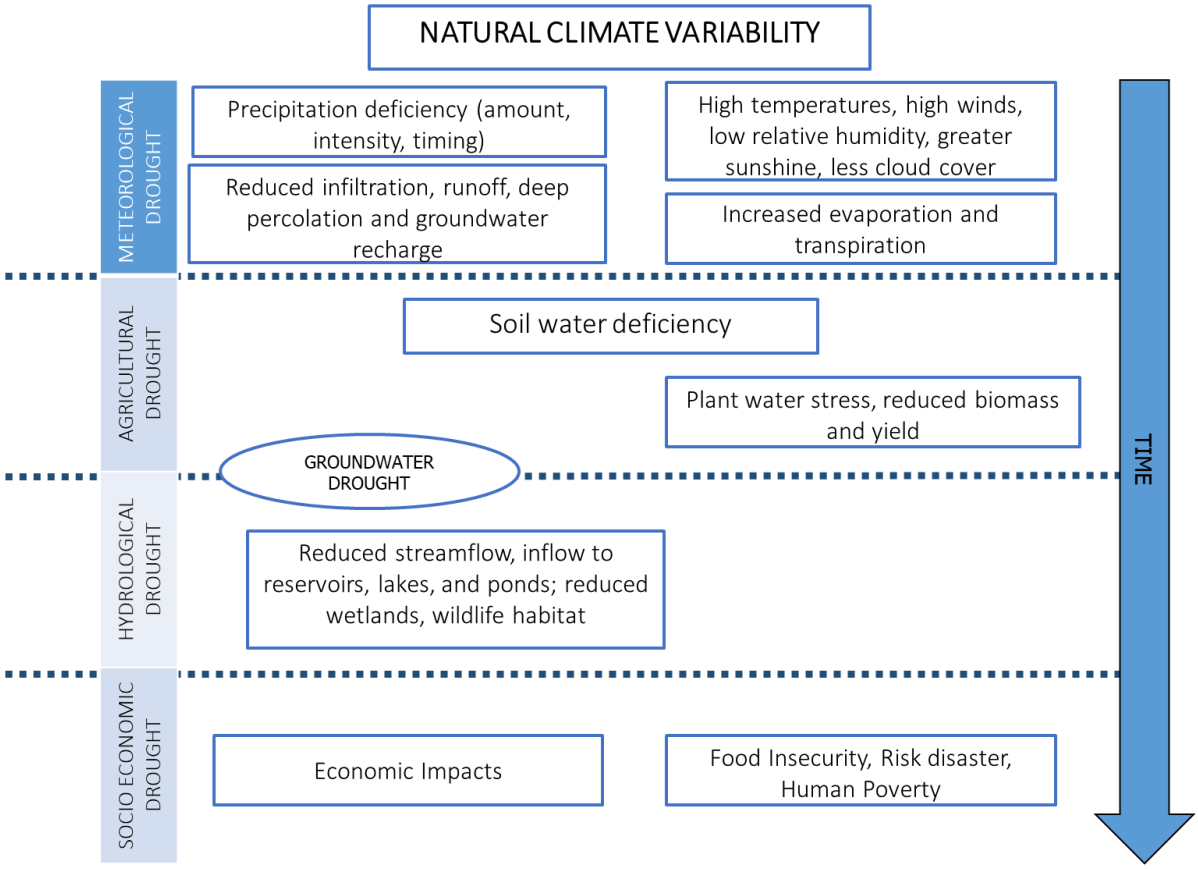


Table 1 Classification of drought description

Type	Description
Hydrological Drought	It is the deficiency of the flow or volumetry of surface or groundwater. The water resource availability in an aquatic system is lower than necessary, generating a temporary lag of various hydrological cycle levels.
Meteorological Drought	It is based on climate data and defines the relationship between precipitation deviation relative to mean over a given time,

	considering that this time may have an order of months or years. The rainfall contribution is much less than expected or appropriate over a given geographical area.
Agricultural Drought	Agricultural drought refers to the scarcity of water or moisture in the soil that prevents the development of crops in all their growth phases, severely affecting the first economic sector. It emphasises that no valid fixed thresholds can be established in this type of drought because each crop is different.
Socioeconomic Drought	This type of drought causes the highest economic and social impact because water availability decreases to the point of causing damage to the population due to water scarcity, with significant socioeconomic losses.

2.2 Drought Assessment

2.2.1 Drought Indices

This section will describe the central drought measurement rates, which are intended to quantitatively measure the impacts of drought, as Jaramillo proposed (2002).

In this way, the rainfall regime is a factor of significant influence in calculating a region's drought rates, considering that this is the difference between a quantity provided with rainwater and consumed or lost by evapotranspiration. On the other hand, the precipitated water correlates with the temperature, which affects evaporation, soil type, topography (slopes of the terrain), richness, and vegetation variety.

However, the mathematical model for calculating water balance (Equation 1) is not global to accurately explain water movements, uptake, and losses. It is a biological system in which other variables influence, but the equation is standardised for study purposes.

Equation 1General Water Balance

$$\sum INPUTS - \sum OUTPUTS = \Delta\, STORAGE\ VARIATION$$

$$\sum RAINFALL - \sum EVAPOTRANSPIRATION = \Delta\, STORAGE\ VARIATION$$

In this way, the flow of the waters through the horizons of the soil is to calculate the drought is given by different geomorphological characteristics of the soil, where they predominate: texture, plasticity, porosity, expansion, and compaction; in addition to climatic and physical variables such as temperature, wind speed, relative humidity, evaporation and radiation. (Novoa, 1998)

Consequently, Montaud (2019) states that drought indices are the values that allow to classify by intervals the degree of affectation of a territory by a drought event and also make decisions. Thus, Table 2 describes them.

Table 2Drought Indices

<table>
<tr><th>Index</th><th>Description</th><th>MATHEMATICAL FOUNDATION</th></tr>
<tr><td colspan="3">INDICES BASED ON WEATHER DATA</td></tr>
<tr><td>Standardised Precipitation Index (SPI)</td><td>It was raised to measure the rainfall deficit on multiple timescales to assess water resources' impacts based on soil moisture, snow accumulation, runoff, groundwater, and reservoir reserves. Therefore, this is calculated on timescales of 1 to 48 months, where it is precipitation dependent. (Mishra & Singh, 2010). This data is adjusted based on the Gamma distribution and standardised with the normal distribution to obtain deviations from each precipitation record.

The classification of SPI values are as follows:
<table><tr><th>SPI values</th><th>Classification</th></tr><tr><td>≥ 2,00</td><td>Extremely humid</td></tr><tr><td>Between 1.50 and 1.99</td><td>Very humid</td></tr><tr><td>Between 1.00 and 1.49</td><td>Moderately humid</td></tr><tr><td>Between -0.99 and 0.99</td><td>Approximately normal</td></tr><tr><td>Between -1.00 and -1.49</td><td>Moderately dry</td></tr><tr><td>Between -1.50 and -1.99</td><td>Severely dry</td></tr><tr><td>≤2,00</td><td>Extremely dry</td></tr></table></td><td>The SPI calculation is performed with the long-term precipitation record for the desired period. Such a long-term record conforms to a distribution of probabilities.(Blain, 2012). Positive SPI values indicate rainfall more significant than the median value, and negative values indicate precipitation below the median value.
• According to the SPI, Drought begins when the SPI value is equal to or less than -1.0 and ends when the value becomes positive.</td></tr>
</table>

Precipitation deciles index

It uses cumulative frequency curves that determine rain deciles for each series. The above is to sort the data and divide the precipitation occurrence distribution by a long period into tenths of distribution. It can apply to series that are not adjusted to a distribution
normal.

The classification of SPI values are as follows:

Range	**Probability of being lower**	**Classification**
1	Less than 20%	Well below average
2	20 – 30%	Well below average
3	30 – 40%	Below average
4	40 – 60%	Media
5	60 – 70%	Above average
6	70 – 80%	Quite above average
7	Greater than 80%	Well above average

The limits of each decile are calculated from the frequency curve; thus, the first decile is that rain is not exceeded by 10% of the lowest totals. The second decile is that rain that is not exceeded by 20% of the lowest totals and so on, until the amount of rain identified by the tenth decile is the maximum amount of precipitation within the long-term records. (Blain, 2012)
The fifth decile would be equivalent to the median, i.e. the decile that does not exceed 50% of occurrences during the registration period. This method has the advantage of being applicable to normally un distributed series.

The Standardised Evapotranspiration Precipitation Index (SPEI)

SpeI is calculated from the SPI, which are based on precipitation records, but also evapotranspiration, combining the sensitivity of changes in evaporation demand with the multitemporal of the SPI Index,
The values that SPEI takes for the determination of dry episodes are the same as those of the SPI.

SpeI values are obtained from the Global SPEI Database of the Pyrenean Institute of Experimental Ecology of Aula Dei of the Higher Council for Scientific Research (CSIC).

WATER BALANCE-BASED INDICES

Palmer Drought Severity Index (PDSI)	It is based on an original study for the Central Iowa and Western Kansas regions of the United States. It is commonly used to assess agricultural drought and is based on water balance. This index is a calibrated soil moisture algorithm for relatively homogeneous regions, so mountainous and heterogeneous regions with microclimates should be supplemented with other indices (Verdon-Kidd & Kiem, 2014). It was developed under the concept of water supply-demand, taking into account the deficit between actual precipitation and precipitation necessary to maintain standard humidity conditions in the soil for any time. His input data is the potential or reference evapotranspiration and the amount of usable water in the soil. One of the main advantages is that it measures the impact of soil moisture conditions, so it has particular application in agriculture and shows a current condition based on historical records and clearly shows dry events of significant magnitudes, such as ENSO phenomena.	Starting with an ideological water balance (BHE), made each month, which uses historical records of precipitation (P) and average temperature (Tt). The storage of moisture in the soil (S) is carried out in two layers: the first, shallow, can store (Ssmáx) up to 25 mm of water sheet; the second, more profound, has a capacity (Sumáx) that depends on the physical characteristics of soil and the depth of vegetation roots, varying from 127 to 229mm. Moisture cannot be extracted into the deep layer (Su) until all available moisture in the surface layer (Ss) is taken (or recharged). Potential evapotranspiration (PE) is estimated with thornthwaite's method, which is then exposed, and soil losses (L), due to it, occur when PE > P, with P being precipitation. Evapotranspiration losses from the surface layer (Ls) occur at the potential level; instead, those in the layer (Lu) depend on its initial moisture content, PE and storage able moisture in both layers of soil. So, when PE >P, having: $L_s = mínimize\ [S_s, (PE - P)]$ (1) $L_u = \frac{S_u[PE-P)]-L_s}{S_{smax}+SU_{máx}}$ (2)
Humidityanomaly index (z)	It is an index derived from the Palmer procedure (PDSI), which expresses a relative deviation from the current time in a given month concerning the average time conditions (Okal et al., 2020). However, for the derivation of a drought index, it is essential to consider the time factor, which is generally not the first period of low rainfall that causes severe damage to agriculture but prolonged periods. The difference in the duration of these periods is what determines the severity of the drought.	This equation that defines Z's value represents the actual precipitation deviation from the characteristic precipitation, and **K is** the d deviations product. Z = d * K

Water Availability Index (HDI)	It is similar to the method used by Thornthwaite, with variations in the value of the coefficients and the form of calculation of the water balance.	$I = \frac{ESC}{2 * ET_0} - \frac{DEF}{ET_0}$ Being ESC, decadal or monthly runoff and DEF, a deficit in the same Period.
INDICES BASED ON REMOTE SENSING SYSTEMS		
NDVI (Normalized Difference Vegetation Index)	This index is defined as the ratio between the sum and the Red and Infrared Band differences. The relationship between this index and the earth's surface temperature (NDVI – LST) provides information on vegetation stress and soil moisture conditions.	$NDVI = \frac{IRC - R}{IRC + R}$ Where: IRC: Reflectance in the near-infrared band. A: Reflectance in the Red Band

In the need to quantify drought, some parameters have been developed to classify this phenomenon into more or less critical events. So seven can be described, according to (Sen, 2015).

- Magnitude: Accumulated precipitation or flow deficit throughout the dry period. . (Verdon-Kidd & Kiem, 2014)
- Duration: Total time that precipitation or flow rate remains below the average level, often defined as the variable's historical mean in the matter. . (Verdon-Kidd & Kiem, 2014)
- Frequency: Number of events that occur during a fixed period. (Verdon-Kidd & Kiem, 2014)
- Implementation speed: Time elapsed between the beginning of the drought and its maximum deficit. . (Verdon-Kidd & Kiem, 2014)
- Temporary spacing: Average time elapsed between the start of two drought events. . (Verdon-Kidd & Kiem, 2014)
- Extension: Represents the total area affected by drought. . (Verdon-Kidd & Kiem, 2014)
- Spatial dispersion: Degree of concentration of flow or precipitation anomalies. . (Verdon-Kidd & Kiem, 2014)

2.2.2 Space and Time Drought Analysis

Spatiotemporal drought analyses are under the precept of finding and reviewing how drought evolves in a specific area about a unit of time.(Diaz et al., 2020, Podestá et al., 2016)

In the last few years, methodologies of spatiotemporal analysis applied to droughts have been developed. These have managed to generate a better understanding of spatiotemporal dynamics, strengthening their monitoring and impact reduction.

Several approaches have been proposed to describe the spatiotemporal development of drought. According to Bacanli et al., 2009, Dracup et al., 1980, they propose methodologies, which apply to force a great number of meteorological variables in the drought indexes calculation based on time series disaggregation of relating it to the spatiality of all zone, establishing a unique phenomenon in the territorial extension. Therefore, they did not consider the spatial network's interconnection; that is to say, this approach is multivariable and not contiguous.

On the other hand, Sheffield et al., 2009 propose one in which the area will be divided according to the most extensive grouping of phenomena weighted by the index using an applied grouping technique allowing not only to observe spatiality but also to attribute duration and magnitude in each area but even the spatiotemporal tracking was a limitation for these methodologies.

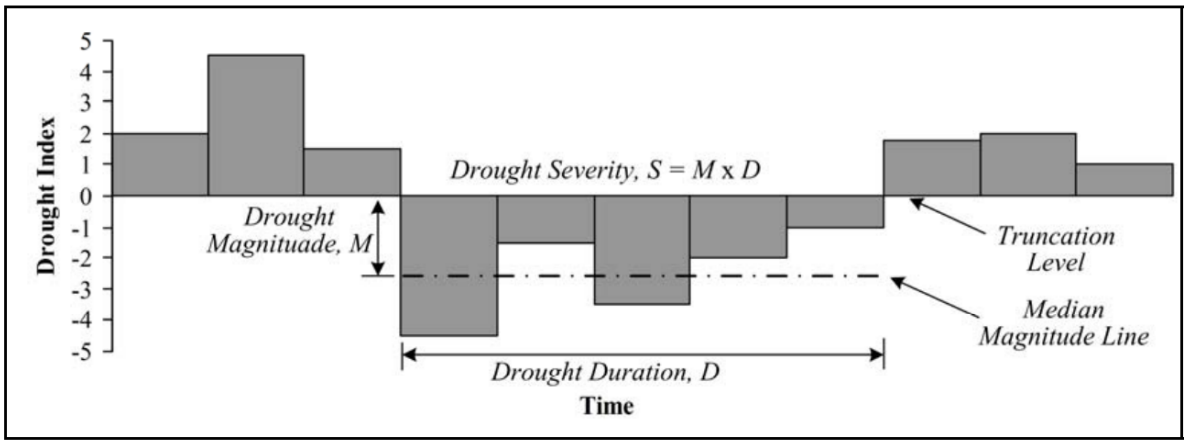
Later, Podestá et al., 2016 generates the first conceptualisation of the spatial droughts monitoring, adapting the approaches of the previous methodologies and calculating the displacement between the contiguous areas in time-based on probabilistic monitoring at a multidimensional level (magnitude, duration, time), generating a new change in the conventional paradigm of spatiotemporal modelling of droughts. Advances regarding the monitoring of these events were strengthening the characterisation methodologies, and consequently, Diaz et al., 2020a presented a methodology to build the spatial trajectory based on drought contiguous areas computation and dynamic drought characterisation. However, due to the computational cost, analysis of large volumes of data and the complexity of the model, only 1D approximations of the drought dynamic patterns were generated.

In general, within the space-time analysis, Authors like (Diaz et al., 2020a) propose certain stages to achieve this study, according to drought events, about time-series events space only and space – temporary. In this way, four (4) stages of development described below would be established:

A. Time analysis: calculating the aggregate drought indicator by regions, such as calculating time-series events, their duration and deficit.
B. Space-time analysis: Integration of time series events into drought area aggregates and calculating the percentage (PDA) each time.
C. Characterisation of space-time events.
D. Visualisation and analysis of results.

Consequently, each of the factors and the importance of highlighting the four stages for their execution within the spatiotemporal analyses are described below, following (Corzo, 2019; Diaz et al., 2020a; Van Loon, 2015)

Figure 6 Drought Index Data Analysis



A. Time Scan

A time-series event starts at the *ts moment* when the drought indicator (DI) value is below a set threshold (T) and ends at the time t_e when the drought indicator value is above the threshold. The duration *(d)* and deficit *(df)* of each i-th time-series event are calculated using the following equations:

$$d_i = t_e - t_s$$

$$df_i = \sum_{t=ts}^{te} (DI(t) - T)$$

The water deficit is standardised through a percentage expression, as shown below:

$$df_{s_i} = 100 * \frac{df_i}{\bar{x}}$$

Where:

df_{si}: standardised deficit of the i-th event of the time series

$\bar{x}$ Average of the deficit values of the time series analysed

From the above, it is possible to obtain the maps of the spatial distribution.

B. Space-Temporary Analysis

Within this stage, space-time analysis allows to evaluate changes in the coverage in the study area over time, so the drought index values must be converted into events and binary representations, based on a possible threshold.

In this way, the implementation of the threshold method is carried out, which is under the equation shown below:

$$D_s(t) = \begin{cases} 1 \text{ sí } DI(t) \leq T \\ 0 \text{ sí } DI(t) > T \end{cases}$$

After the drought index values become a binary representation of spatial coverage, the PDA is calculated, using the following expression:

$$PDA(t) = {}^{100}/_{A_{tot}} * \sum_{c=1}^{N} (D_s(t) * A)$$

Where:

A: It's the cell area
A_{tot} : Region area
PDA: Magnitude of drought
PDA time series allows the identification of extensive droughts in a region. Small areas are neglected by applying a second threshold.

C. Characterisation of space-time events.

The drought duration (DD) calculation begins in the time interval when the PDA is below the defined threshold. In a series of ADD, this threshold is estimated using the 90th percentile of the low PDA (Diaz et al., 2020a). Therefore, the space-time event starts when the PDA is below the set threshold (TPDA) and ends when it is above the setthreshold. This threshold is calculated from:

$$DD_j = t_E - t_s$$

where j is the j-th space-time event and *DD*, t_s and t_E are the duration of the drought, the start and end of the j event. The region under the PDA curve is the severity of the drought (S, expressed as a percentage).
This value is a measure of the magnitude of the drought. *S* is calculated for each j *event* with:

$$S_j = \sum_{t=t_s}^{t_E} PDA(t)$$

Intensity (I) is calculated as the ratio between S and DD, as detailed below:

$$I_j = \frac{S_j}{DD_j}$$

This Ij relationship can be interpreted as the average PDA value during DDj time. The calculation of DD, S and I, is performed on the entire time series of the PDA.

D. Visualisation and analysis of results.

As a result of spatiotemporal analyses, implementation is performed through graphs and matrices. These are the line charts (it is represented in temporal series of drought indicators (TA)), scatter charts (the duration vss deficit (TA), duration versus severity, duration and intensity (STA)), area chart (the time series is represented drought area percentage (STA), colour matrix (time-series displays drought area percentage (STA)) and image graph (a spatial indicator of drought, duration, deficit (TA) and area indicator (STA) are represented).

2.3 Drought Forecasting

Forecasting of key drought parameters has been extensively studied and implemented using physical or data-driven models(Albergel et al., 2019). Physical models are based on the interaction of the physical conditions of the environment in which the phenomenon develops, such as ocean-atmospheric conditions, which determine climatic parameters' trends. (Müller Schmied et al., 2014)

As mentioned above, although the forecasting capabilities of physical models are accurate for atmospheric factors such as land temperature, they are less accurate for parameters essential for drought modelling, such as precipitation forms.(Hudson et al., 2011; Okal et al., 2020). Physical models are also challenging to implement as they require various types of data involving very complex models. (Dikshit et al., 2020)

On the other hand, data-driven models use various machine learning (ML) algorithms to determine the relationships between predictors (inputs) and variables (outputs) (Abbot & Marohasy, 2012; Ukkola et al., 2020). Machine learning , supervised algorithms are a set of techniques in which the system learns from the data's characteristics or patterns and makes a decision. (Anandhi et al., 2008)

Conventional techniques for weather event prediction have included regression trees, support vector machines, artificial neural networks (ANNs) (Anshuka et al., 2019; Poornima & Pushpalatha, 2019; Sachindra & Kanae, 2019)since they have a low computational expense, short run time and in terms of performance they present comparable results with physical models (Anandhi et al., 2008)

2.3.1 Machine Learning and Probabilistic Framework

A summary of the described methods is presented below:

Table 3 Forecasting Models Foundation

MODEL	DESCRIPTION	MATHEMATICAL FOUNDATION

Logistic regression model	Logistic regression works much like linear regression, but with a binomial response variable, using continuous explanatory variables. Looking at multiple explanatory variables independently ignores the covariance between variables and is subject to a logistic regression that will model the probability of a result based on individual characteristics. Because probability is a reason, what will be modelled is the probability logarithm.	$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$ Where, π indicates the probability of an event, and i are the regression coefficients associated with the group reference, and the xi is the explanatory variable.
K-NN classifier-based method	The classification method based on k-NN or k-nearest neighbours is it is based on the fact that the properties of an input x data are similar to those of your neighbourhood's data, so it belongs to the same class as the most frequent of its nearest k neighbours. The nearest k-neighbours method's general algorithm assumes that all examples correspond to points in a p-dimensional Rp space with a C class set.	The first step in the K-NN model is to find the distance between training and test data. Choosing the distance measurement is an important consideration. Commonly, the Euclidean distance measurement is used: $$d(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$ X refers to parameter training data (x_1 to $x_{n)}$, and Y refers to test data with specific parameters (y_1 and y_n).
Random Forest	The model starts by building a tree forest, using the bootstrap technique, in which each tree is created independently, based on a random subset of the predictor variables. Trees grow to maximum size without pruning and, therefore, the average output of all multi-decision trees is the result, achieving practical regression	The RF model's performance was performed using the determination coefficient method ($R^{2)}$, and the error means square root (RMSE) method. R^2 determines the adequacy between predicted and original values, while RMSE measures the variance of errors between actual and predicted values. Formulas are: $$R^2 = \frac{\sum_i^N = 1(\hat{y}\imath - \bar{y}_\imath)}{\sum_i^N = 1(\hat{y}\imath - \bar{y}_\imath)^2}$$ Where, yi is the mean value; yi and yˆi are observed and predicted values, and N is the number of data points.

Support Vector Machines Model	This method is based on statistical learning theory, with non-linear classifiers through kernel functions. In general, SVR models are classified into SVR classification models and SVR regression models. The SVR classification model is used to classify data-related issues placed in different classes, and the SVR regression model is used to solve prediction problems. A hyperplane is achieved by regression over adjusted data. The least-squares method is the best-recommended method for linear regression. However, for regression problems, the use of the least-squares estimator in the presence of outliers may not be entirely reasonable: as a result, regression is poorly performing. Therefore, a robust estimator should be developed to avoid poor performance, which is not sensitive to small changes in the model. As stated, the SVR is based on minimising the risk structure derived from statistical training theory.	$L(y, f(x,a)) = \lvert y - f(x,a)\rvert\varepsilon$ The $\begin{cases} \lvert y - f(x,a\rvert \le \varepsilon \\ \lvert y - f(x,a)\rvert - \varepsilon \; si \; \lvert y - f(x,a)\rvert > \varepsilon \end{cases}$
ARIMA model	According to the well-known box-Jenkins methodology, the ARIMA model is the stochastic model and has been widely used in hydrological forecasts in recent decades. The ARIMA model is divided into a self-regulating model (AR), moving average model (MA). The self-regulating moving average model (ARMA) is classified as an "ARIMA (p, d, q)" model and is a traditional time series forecast model. P is the number of self-regulating terms, d is the number of differences required for stationarity, and "q" is the number of forecast errors delayed in the prediction equation. The selected ARIMA time segments (q, d, p) are used as training sets and test equipment for predictive analysis.	$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \delta + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q}$ $\phi(L)x_t = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)x_t = \delta + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)u_t = \delta + \Theta(L)u_t$ $\phi(L)x_1 = \delta + \Theta(L)u_t.$

SARIMA

A SARIMA model is generally best achieved through a three-stage iterative procedure based on diagnostic identification, estimation, and verification.

- Identification means the use of data and any information about how the series is generated.
- Estimation means the efficient use of data to make inferences about parameters conditioned on the entertaining model's adequacy and diagnosis.
- It is verifying means verifying the fitted model in its relationship to the data to reveal any model inadequacies and perform a model improvement.

- Identification stage:

$$\rho_k = \frac{E\left[(z_t-\mu)(z_{t+k}-\mu)\right]}{\sqrt{E\left[(z_t-\mu)^2\right]E\left[(z_{t+k}-\mu)^2\right]}} = \frac{E\left[(z_t-\mu)(z_{t+k}-\mu)\right]}{\sigma_z^2}$$

- Estimation stage:

$$AIC = -2\ln\left(L\left(\hat{\beta}\right)\right) + 2\omega$$

- Verification stage:

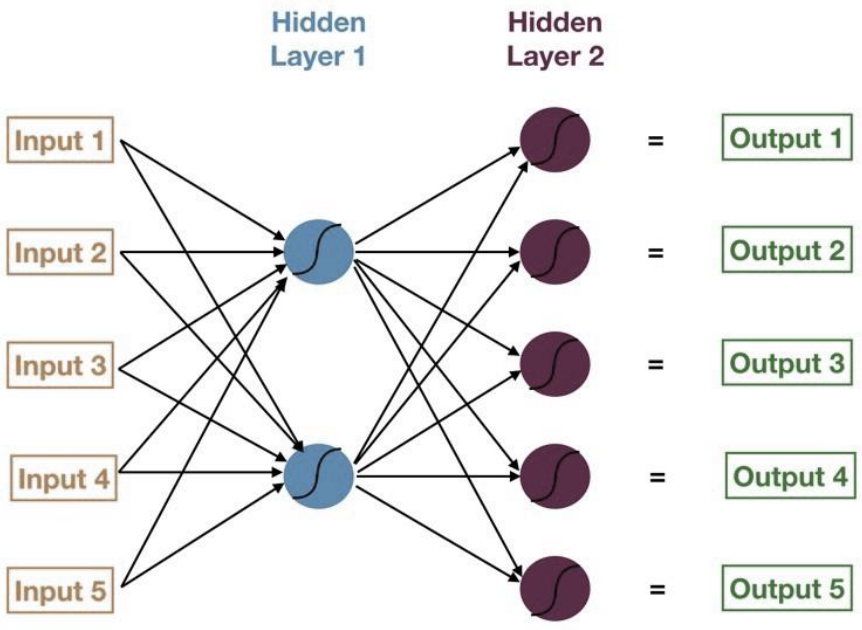
$$\hat{\alpha}_t = \hat{\theta}_q^{-1}(B)\,\hat{\Theta}_Q^{-1}(B^s)\,\hat{\phi}_p(B)\,\hat{\Phi}_P(B^s)\,\nabla^d\nabla_s^D z_t$$

$$Q = n\sum_{k=1}^{m}\hat{\rho}_k^2 \sim \chi^2(m)$$

2.3.2 Neural Networks and Deep Learning Brief Framework

The description of neural models will be described and detailed in this section. Goldberg (2016) states that neural networks are inspired by the brain's computational mechanism, consisting of computational units called neurons. (Deo & Şahin, 2015) A neuron is a computational unit with scalar inputs and outputs. Each input has an associated weight and the neuron multiplies each input by its weight and then adds them together, applies a non-linear function to the result, and passes it to its output. (Hassan & Mahmood, 2017). Neurons are connected, forming a network: one neuron's output can feed one or more other neurons' inputs. (Fig7)

Figure 7 Neural Network Scheme

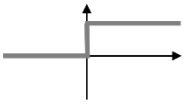
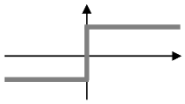
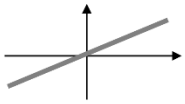
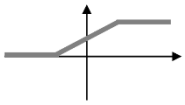
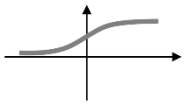
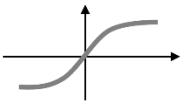
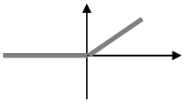
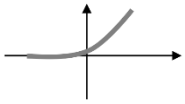


ANNs have been employed extensively to model complex, non-linear processes and systems. ANNs are self-adaptive systems driven by non-linear data that can identify and learn correlated patterns between sets of input data and corresponding output values, even when the underlying data relationship is unknown. ANN resembles a human brain in two respects; the network acquires knowledge through a learning process, and interconnection forces known as synaptic weights are used to store knowledge (Illeperuma & Sonnadara, 2009).

The ANN can be explicitly programmed to perform a task manually by setting the weights and thresholds for each Link. The process of determining weights and biases is called training. The dataset used to train the ANN is called the training dataset. The training dataset consists of input signals mapped with the corresponding target output.

- **Activation Functions**

Figure 8 Activation Functions used in Neural Networks

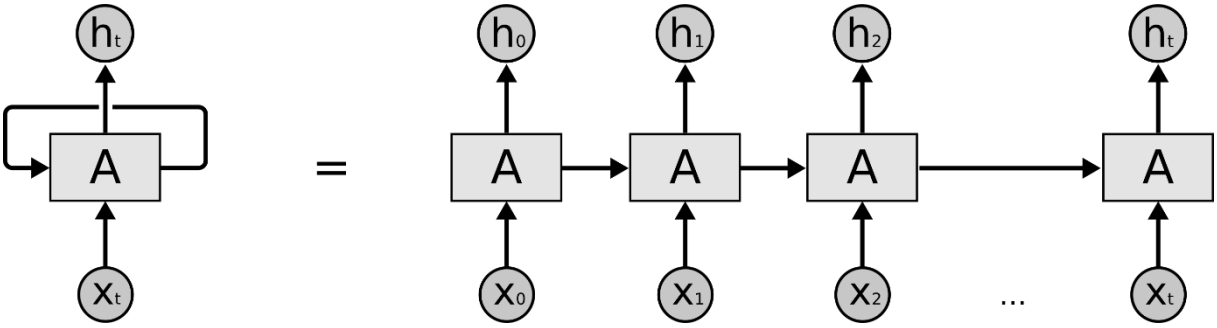
Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = max(0, z)$	Multi-layer Neural Networks	
Rectifier, softplus	$\phi(z) = \ln(1 + e^{z})$	Multi-layer Neural Networks	

2.3.2.1 Recurrent Neural Network

The recurrent neural network was characterised for early detection (prognostication) of heart failure related cases and controls using recurrent controlled units (RUs). Performance measures were compared with the neural network, K-nearest neighbour, logistic regression and support vector machine (Poornima & Pushpalatha, 2019).

The RNN is a class of ANN that allows nodes to connect to directed loops Fig9. The RNN formulation ensures that it can exhibit dynamic temporal behaviour, generating a memory of past data and processing sequential data and establishing dependencies between data from different points in time (Yingyi Chen, 2018).

Figure 9 RNN loop cell

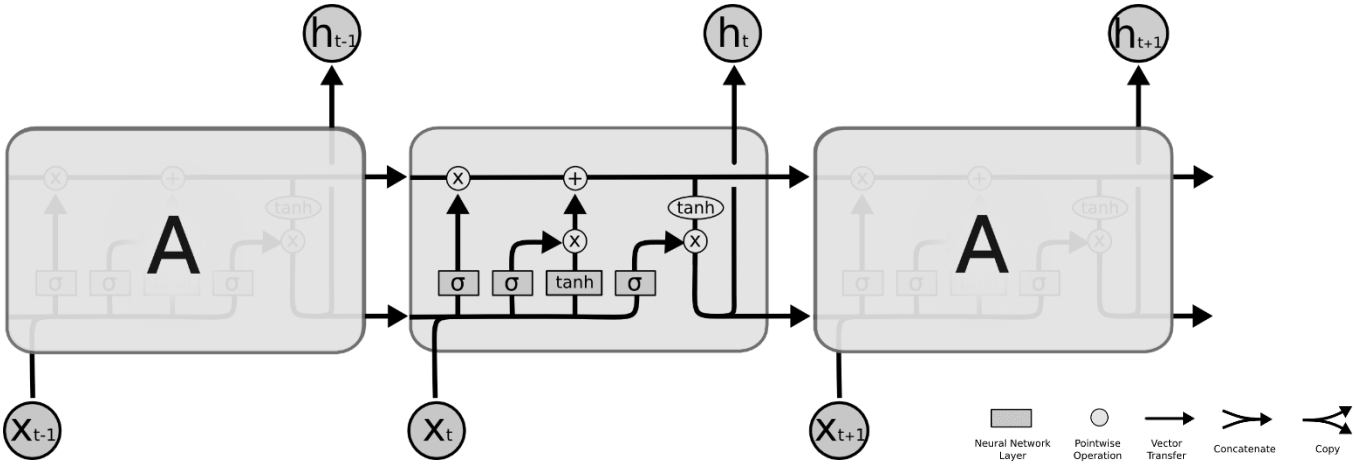


About what was raised in Yingyi Chen's (2018) theory, RNNs can handle arbitrary input sequences. The shared weight is adopted recursively, achieving that this model learns temporal dependencies and simulates a universal Turing machine, performing almost any computation.

RNNs have different variations such as the bidirectional recurrent neural network (BRNNN), which at time t is related not only to the state of the past but also to the state of the future, in addition to the long-term, short-term memory network (LSTM) is an improved RNN variant proposed by Hochreiter and Schmidhuber in the year 1997 (Dikshit, Pradhan, & Alamri, 2020).

The LSTM is characterised by its basic unit, a memory cell that can store a time state and is protected while storing, writing, and reading information (Figure 10).

Figure 10 LSTM cell



2.4 Related Works

Products related to drought monitoring and forecasting worldwide are presented below. The main proposals and projects that have been taken up in the study as a basis for work are presented.

Table 4 Related Works

Authors	Contribution	Application	Data acquisition technology	Input variables	Input dataset format	Decision-making method
Khan et al. (2015)	Wireless Sensor Network based Flood/Drought Forecasting System	Forecasting of Drought and Flood conditions	Wireless Sensor Network	Wind pressure, Humidity, Water Pressure	Numeric	Predefined threshold values
Mokhtarzad et al. (2017)	Drought forecasting by ANN, ANFIS, and SVR and comparison of the models	Drought Forecasting	NA	Precipitation, Temperature, Humidity	Numeric	ANN, ANFIS, SVR
Heydari et al. (2018)	An investigation of drought prediction using various remote sensing vegetation indices for different periods	Drought Prediction	Remote Sensing	NDVI, TCI, VCI	Satellite Images	SVR
Zou et al. (2018)	MapReduce functions to remote sensing distributed data processing-Global vegetation drought monitoring as an example	Cloud framework for drought monitoring	Remote Sensing	Vegetation	Satellite Images	MapReduce Paradigm
van Hoek et al. (2019)	A prototype web-based analysis platform for drought monitoring and early warning	Drought Monitoring and Early Warning	Remote Sensing	Precipitation, Soil Moisture, EDI, NDAI, VCI, TCI	Satellite Images Numeric	Open Source Softwares

Adede et al. (2019)	A Mixed Model Approach to Vegetation Condition Prediction NAUsing Artificial Neural Networks (ANN): Case of Kenyaâ€™s Operational Drought Monitoring	Drought Monitoring	NA	NDVI, VCI, SPI, RCI	Numeric	ANN
Poornima and Pushpalatha (2019)	Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network	Drought Prediction	NA	SPI, SPEI	Numeric	LSTM (Long ShortTerm Memory)
Proposed System	Deep Learning-based Drought Assessment and Prediction	Drought evaluation and prediction	Fog Computing, IoT	Groundwater, Streamflow, Soil moisture and Soil temperature, Temperature, Humidity, Precipitation, Evapotranspiration, Season	Numeric	DNN, SVR
Diaz, Corzo, Van, & Huijgevoort (2010)	Temporal analysis was performed to identify regions of critical variations of temporal events through their duration and deficit	Drought evaluation and prediction	Fog Computing, IoT	SPEI	Numeric	SPI
Abdullayeva (2019)	Application and Evaluation of LSTM Architectures for Energy Time-Series Forecasting	Drought evaluation and prediction	Wireless Sensor Network	humidity and temperature inside a house and temperature, pressure, humidity, wind speed, dew point outside	Images Numeric	Neural Networks, ARIMA, Standard LSTM, Stack LSTM, Sequence to Sequence LSTM

Chapter 3 Study Case: Central America Dry Corridor

Central America is one of the world's regions most exposed to the risks of natural disasters and climate variability. Due to the geographic conditions that make it more susceptible to strong heat waves, the influence of global climatic phenomena and hurricane seasons prevent a progressive recovery of its hydric condition in terms of food and environmental security. (Zee Arias et al., 2012)

Additionally, this has led to a great interest in understanding these interactions between the erratic drought behaviour and the factors that generate it. This chapter provides a detailed description of the general context of the dry corridor of the Americas. Thus, generating a starting point concerning the antecedent conditions and their potential effects that drought, as an extreme event, has historically caused.

3.1 Description

Central American Dry Corridor, which extends through seven Central American countries: Guatemala, El Salvador, Honduras, Belize, Panama, constitutes a region of dry tropical forest extended along the Pacific border of Central America (Fig 8)

The Dry Corridor is known for its irregular precipitation patterns. It has become one of the most susceptible regions globally to climate change and variability, making climate hazards the most prevalent, represented mainly by frequent droughts, severe spontaneous floods, which affect agricultural production, with greater intensity in degraded areas. (Eckstein et al., 2017)

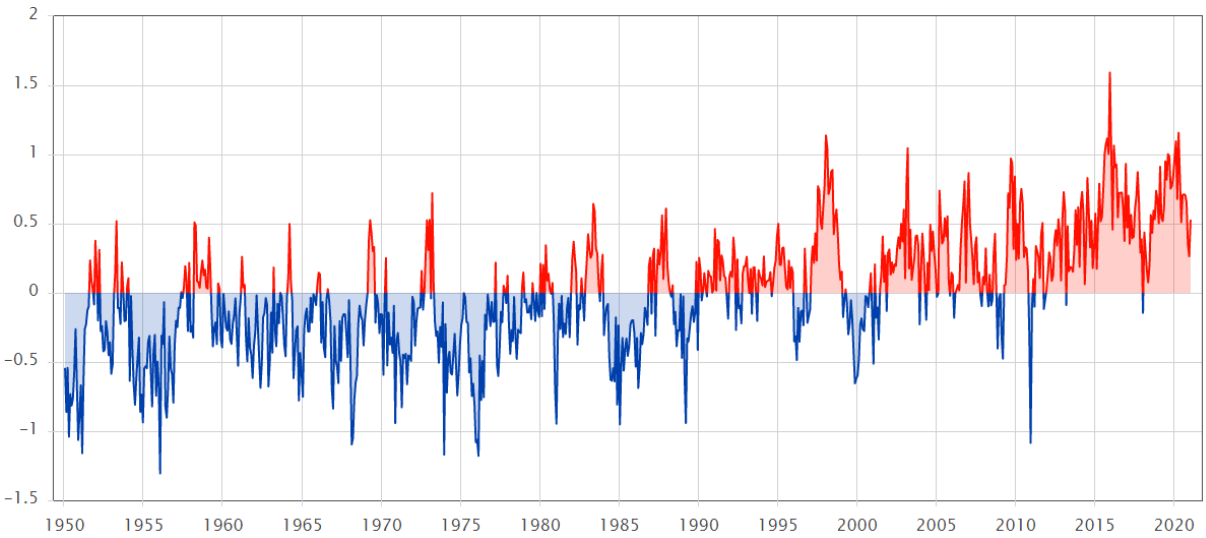
Figure 11 Central América Dry Corridor

Source: (FAO, 2019)

3.2 Climate Conditions (Climate change and variability)

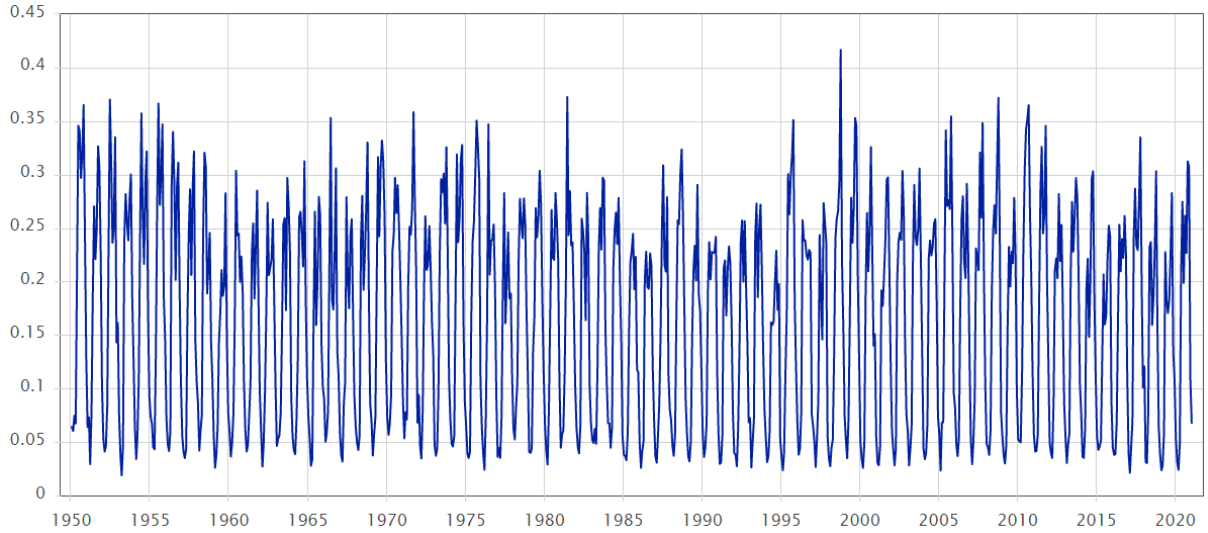
The hydro-meteorological background showed that Central America had experienced an average temperature increase close to 0.5 to 0.6 ° C during the last 40 years (Fig 9). In terms of Precipitation, these have tended to decrease at a rate of 12 mm/y during the same period (Fig 10). According to the IPCC, 2018, Central America has been classified as one of the Latin American regions with the most significant sensitivity to climate change. This report is supported by a constant increase in extreme events, including storms, floods, and droughts.

Figure 12 Surface Temperature Anomalies 1950-present



Source: Author

Figure 13Total Precipitation in meters 1950-present



Source: Author

Water availability is related to both intra-annual precipitation patterns and geographic differences. During El Niño years, Precipitation can fall between 38 and 42 percent, with prolonged periods of temperature rise and many more prolonged periods without rain periods.

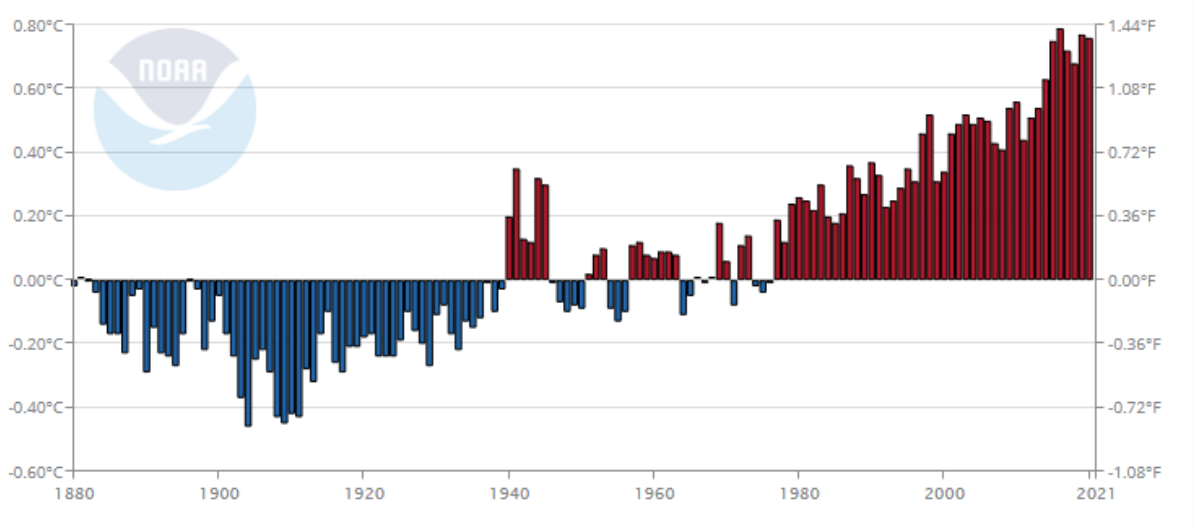
The increase in temperatures and the dry climate has disastrous consequences on the cultivation scheme and agricultural production yield and affecting food security. Added to the tropical storm season having devastating effects and preventing the recovery of the water system. (FAO, 2019). These threats generated by these erratic events (droughts and floods) have increased in recent years. Due to unsustainability and growing food insecurity, increasing climate trends, long and short variability terms, and socioeconomic stresses in the region have begun to displace the population.

3.3 Drought History in Central America

3.3.1 Drought Impacts

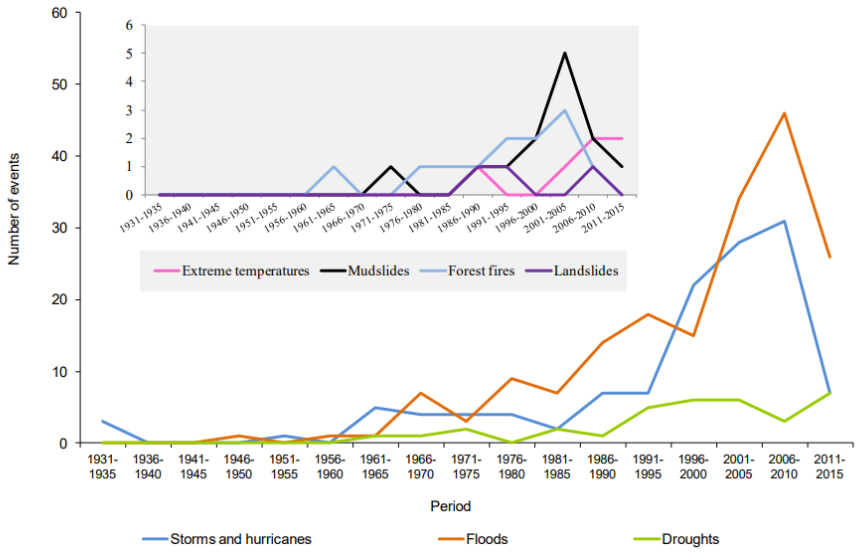
Central American Dry Corridor is a group of ecosystems existing in the dry eco tropical forest of Central America, covering the lower parts of the Pacific slope and much of the premontane central region of Guatemala, El Salvador, Honduras and Nicaragua, where periodic dry conditions are predominating, causing environmental stress at the regional, national and local levels (IDB, 2019).

Figure 14 ENSO 3.4 Anomalies



Especially these drought conditions are also associated with phenomena that can have direct implications such as ENSO, Pacific Decadal Oscillation, among others. Fig X, represents how the anomalous variation of ocean temperatures has been temporally over 30 years, revealing a change of about 0.02 degrees of variation, increasing of great magnitude in the last three decades.(Sheffield et al., 2009). According with CEPAL (Unidas, 2008) between 1931 and 2015, the Central American countries experienced 375 extreme events associated with hydro-meteorological phenomena Fig12.

Figure 15 Trends in Climatological events in Central America



Honduras experienced the greatest number, with 66, and Belize the least, with 19. The most recurrent events are floods, storms, landslides and mudslides, which account for 82% of total events, and droughts, which account for 9%. (Unidas, 2008). During the last three decades, the number of natural disasters in the region grew by 6% per year with respect to the 1970s. Table 4 specifies some of the more representative drought impacts on the Dry Corridor of Central America, where ENSO is definitely affected.

Table 5 Principal Drought Impacts reported on the Dry Corridor of Central America

Country	Impact	Source
El Salvador	• Effects on corn, bean, and sorghum crops. • Affectation of low-income families, who depend on the production of these crops for their livelihood. • Affectations in the crops of coffee, sugar cane, vegetables and pastures, concentrated in the east of the country, mainly in La Unión, San Miguel and Usulután. • For the drought of 1997 – 1998, there is talk of an impact on 25% of agricultural producers; and losses estimated at $200 million. • For the 2001 drought, 61% of producers were reported to be affected. Percentages of losses were also established in the three main crops: 77.5% corn losses, 87.6%bean losses, and 80.3% sorghum.	Inter-American Development Bank – IDB (2019)
Guatemala	• In 2009, 25 children were reported dead due to lack of food and 54,000 families affected by malnutrition. • Other crops that are also mentioned as affected, although to a lesser extent, are sorghum, rice, wheat, coffee, bananas, vegetables and sugar cane. • In the 1997–1998 drought, total losses equated to 0.5% of GDP. • In the 2001 drought, bean loss was estimated at approximately 80% of the crop. • In the 2009 drought, an impact of 450,000 families and 60 per cent of Guatemalan children were estimated.	Inter-American Development Bank – IDB (2019); Bald, Quesada, Hidalgo, & Gotlieb, (2018)
Honduras	• In2015, 75% of losses were placed in bean and maise crops. • Crops of rice, palm, sugar cane, melon, cattle, and shrimp production. • In 2010, economic losses reached $25 million, in addition to the $30 million earmarked for emergency care.	Inter-American Development Bank – IDB (2019); Saguí, Madrigal, & Estigarribia, (2017)
Nicaragua and Costa Rica	• The drought resulted in losses of 30% in the last harvests for the economic damage of USD 6.25 million, affecting 1200 ha of rice in 2009. However, there were also conditions in maise crops.	Hidalgo (2020)

3.4 Hydrometeorological Data Sources

For the assessment and tracking of past drought events at the meteorological and agricultural scales and their subsequent forecasting, it was necessary to use complete and detailed satellite information whose observations of the terrestrial system were traced back as far as possible in time.

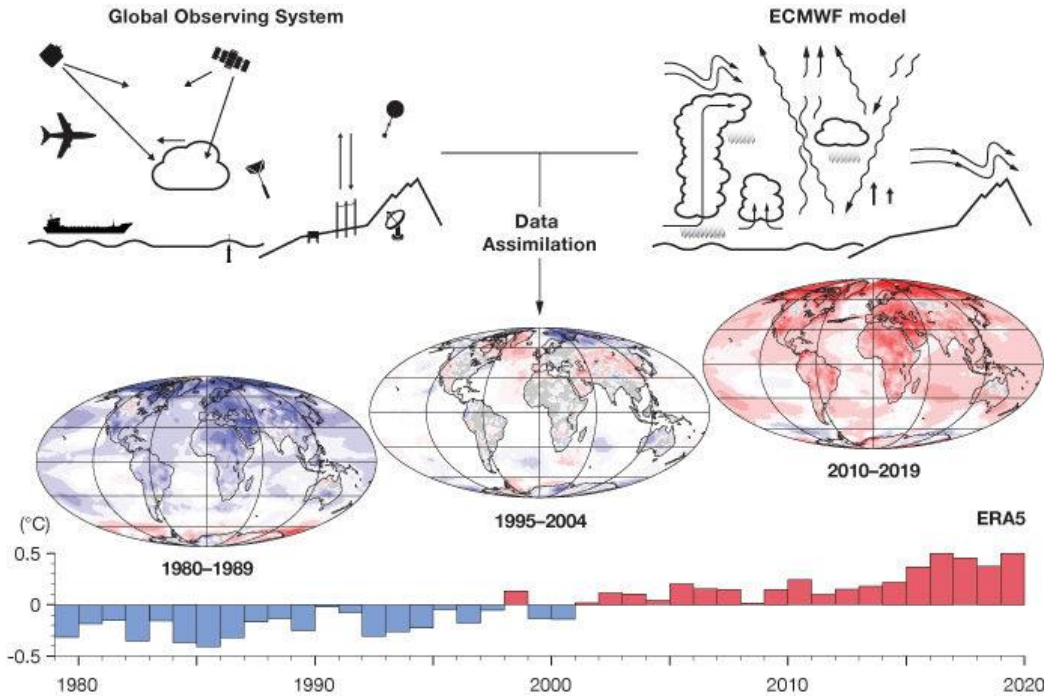
According to that, some data sources are available for climatological modelling, as shown below.

Table 6 Satellite Data Sources

Dataset	Centre/Sponsor	Variable	Time Data	Spatial Resolution
ERA-5	European Centre for Medium-Range Weather Forecasting	Precipitation Runoff Temperature Humidity Wind … ….	1950-present	0.25 °
ERA-Interim GPCC	European Centre for Medium-Range Weather Forecasting	Precipitation Runoff Climate Variables	1979-present	0.5 °
CHIRPS	NASA/ ProjectRainfall Estimates from Rain Gauge and Satellite Observations	Meteorological Variables	1981-present	0.05°

However, to better conceptualise and understand the drought history and increase the possibilities of accuracy in the forecasts, satellite reanalysis data were used. Satellite reanalysis data provide a complete picture possible of past weather and climate (Fig 13). It presents a combination of observations with past short-term weather forecasts repeated with modern weather prediction models.

Figure 16 Data Reanalysis Scheme, Based on ECMWF



ERA5 is a new climate reanalysis dataset (fifth generation) from ECMWF (the European Centre for Medium-Range Weather Forecasts). Based on data availability, atmospheric reanalysis data from 1981 to 2019 were selected, with the main variables being total Precipitation and near-surface temperature adopted for this study.

3.4.1 Other Data Sources - Local Data

Each country that makes up the dry corridor has its meteorological network, more than 286 weather stations are distributed throughout the study area, but only 22% have public access information. That information obtained will support and validate the satellite information used (CICESE,2019).

Chapter 4 Methodology

The chapter aims to present an approach to identify, characterise, track and forecast drought events from a spatiotemporal approach. This is based on the review of existing methodologies and metrics used to define past drought events' dynamic characteristics (intensity, frequency and duration) and the forecasting framework employed.

Three main stages are formulated, articulated between them, ensuring an efficient and directional workflow. Each proposed stage is structured in three phases, 1) Literature Review as mentioned in chapter 2, 2)Drought Assessment, which will be implemented and discussed in chapter 5 and finally, 3) Drought Forecasting, implemented and discussed in chapter 6.

After that, a comprehensive analysis and analysis criteria, such as data analysis techniques and methodologies, selection of accumulation periods, thresholds, tracking methodologies, the tracking and forecasting models configurations, will be presented in this chapter.

4.1 Data

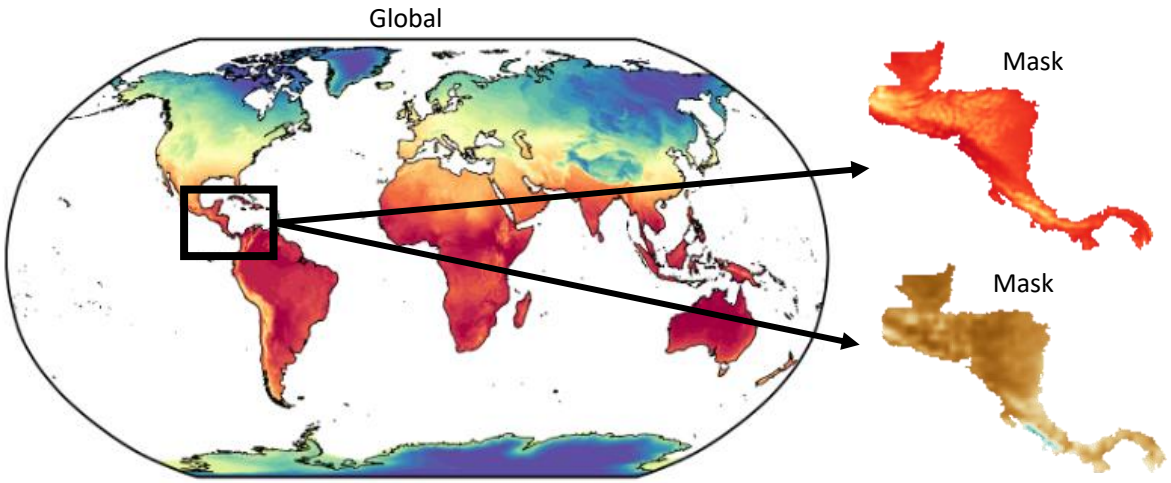
As mentioned in the previous chapter, the hydro-meteorological information used is ERA 5. The parameters used to identify meteorological and agricultural variability accompanied by a drought event included the monthly average temperatures at 2 m, wind speed at 10 m, and annual average precipitation.

Data corresponding to a period of 39 years (January 1981 to June 2020) was used. Climatological data were pre-analysed not to generate errors in the drought index computation, so metric units, spatial domain, were adjusted before use.

Table 7 Data Settings

ERA 5 LAND (10 m)	Data Raw Units	Data Units Needed
Total precipitation	m/d	mm/d
Surface temperature	Kelvin	Centigrade
Wind u	ms^{-1}	ms^{-1}
Wind v	ms^{-1}	ms^{-1}
Runoff	m/d	Mm/d
Pressure	Pa	Pa

Figure 17 Data Mask Extractions



4.1.1 Exploratory Data Analysis

Exploratory data analysis is a critical process of making initial investigations or deductions about the data to discover patterns, detect anomalies, test hypotheses, and test assumptions with summary statistics and graphical representations.

For this research, exploratory data analysis allowed us to analyse the acquired data's quality for further use. It also identifies data types, variable composition (null/missing values), descriptive measures such as Mean, Standard Deviation, Extreme Values, Percentiles.

```
precipitation_raw['tp'].to_series().describe()
```

```
count    2.681375e+06
mean     5.534115e-03
std      4.835752e-03
min      8.568168e-07
25%      1.745824e-03
50%      4.366707e-03
75%      8.119987e-03
max      7.503555e-02
Name: tp, dtype: float64
```

```
temperature_raw['t2m'].to_series().describe()
```

```
count    2.681375e+06
mean     3.010381e+02
std      5.248044e+00
min      2.799392e+02
25%      2.952777e+02
50%      2.973497e+02
75%      2.988092e+02
max      3.066421e+02
Name: t2m, dtype: float64
```

```
windu_raw['u10'].to_series().describe()
```

```
count    2.681375e+06
mean    -7.589039e-01
std      1.133135e+00
min     -1.058274e+01
25%     -1.233659e+00
50%     -5.033382e-01
75%     -7.803679e-02
max      7.090868e+00
Name: u10, dtype: float64
```

```
windv_raw['v10'].to_series().describe()
```

```
count    2.681375e+06
mean    -4.764625e-01
std      8.242971e-01
min     -7.564543e+00
25%     -8.406948e-01
50%     -3.085651e-01
75%      5.044198e-02
max      4.556436e+00
Name: v10, dtype: float64
```

The consolidated dataset consists of 9 main variables, including position and time. After cleaning, the dataset is generated with the spatial extent of mainly Central America and with time records of 39 years. These will be used for subsequent steps.

	Unnamed: 0	latitude	longitude	time	u10	v10	t2m	sro	tp	tp_mm
0	9044	19.0	-91.099998	1981-01-01	-1.406261	-1.277755	294.82864	6.351620e-07	0.000254	0.253897
1	9045	19.0	-91.099998	1981-02-01	-1.860150	-1.258704	296.89980	3.177673e-06	0.000947	0.946619
2	9046	19.0	-91.099998	1981-03-01	-1.079127	-0.782802	298.96735	2.542511e-06	0.000798	0.797771
3	9047	19.0	-91.099998	1981-04-01	-1.932966	-0.573613	301.50998	6.351620e-07	0.000250	0.250462
4	9048	19.0	-91.099998	1981-05-01	-0.900862	-0.422685	303.05142	8.262694e-06	0.001990	1.989704

4.2 Drought Assessment

4.2.1 Drought Index (DI) Computation

According to chapter 2, different index models have been developed to estimate drought according to its nature. The development of this research focuses on the analysis of meteorological and agricultural drought conditions using standardised variable indices such as SPI and SPEI.(Diaz et al., 2020b)

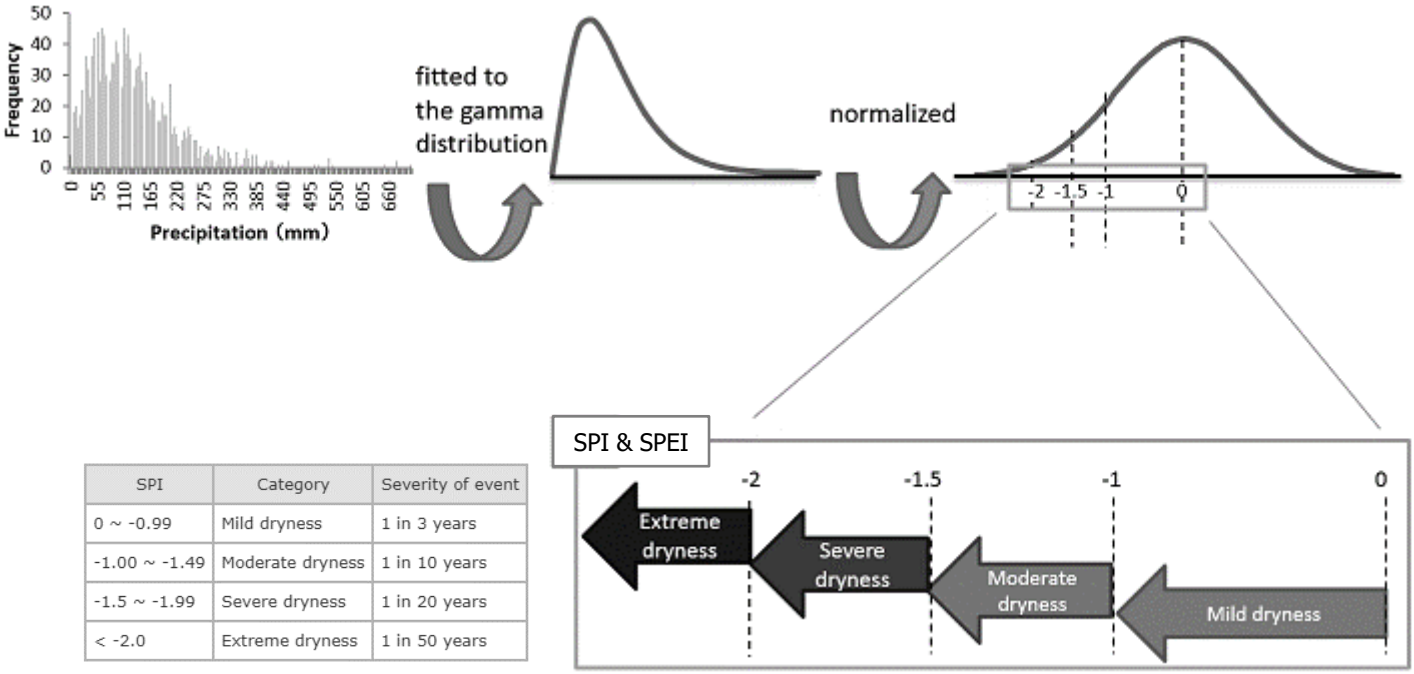
Despite the large variability of distributions in the spatiotemporal domain, SPI was designed to be on space and time-invariant with the potential capability of representing drought and flood events in a similar probabilistic way, even when different precipitation regimes are being applied. (Okal et al., 2020)

As Guttman (1998), the SPI calculation starts by determining a probability density function (pdf) capable of adequately describing the observed long-term precipitation. However, the 2-parameter gamma distribution is often employed. However, Pearson type III distribution (a 3-parameter function) is the best universal model for SPI calculations. (Blain, 2012).

Considering the above, drought indices were estimated for different time accumulations (1,3,6,9 months) to evaluate the period of concentration that can relate to the reported past drought patterns.

- SPx-1 to SPx-3: an indicator of immediate impacts, such as reducing soil moisture, snowpack, and smaller streamflow.
- SPx-3 to SPx-12: an indicator of reduced streamflow and reservoir storage.
- SPx-12 : an indicator ofthe reduced reservoir and water recharge.

Figure 18 Standarised Drought Index Class, Scheme Internal Computation



SPI	Category	Severity of event
0 ~ -0.99	Mild dryness	1 in 3 years
-1.00 ~ -1.49	Moderate dryness	1 in 10 years
-1.5 ~ -1.99	Severe dryness	1 in 20 years
< -2.0	Extreme dryness	1 in 50 years

4.2.1.1 Thresholding

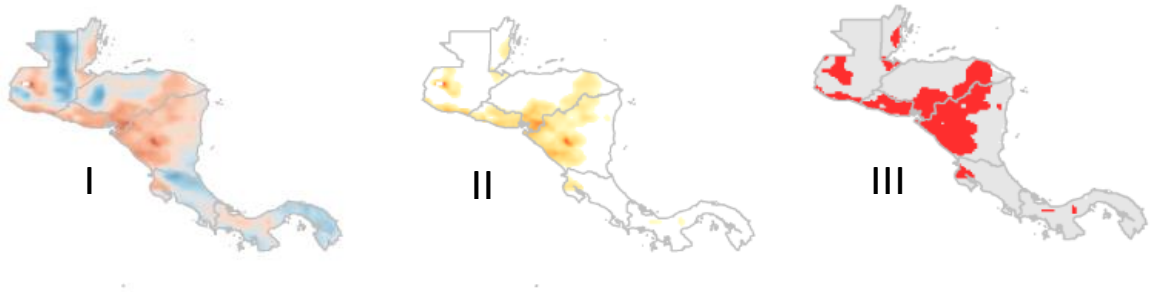
Once the drought indices have been computed, and their calculation and physical representation have been verified. The analysis threshold is established. At the level of time series and grid, this threshold allows defining a limit at which the drought will be defined for its integrated analysis (in space and time).

According to the drought classification described above, there are three main drought categories: moderate, severe and extreme.

Figure 19 Drought Classification Scale

Extremely wet	2.00 or more
Very wet	1.50 to 1.99
Moderately wet	1.00 to 1.49
Near normal	0.99 to –0.99
Moderately dry	–1.00 to –1.49
Severely dry	–1.50 to –1.99
Extremely dry	–2.00 and less

In this context, the threshold of values below the normal condition (-1.0) was established in order to extract the drought zones in the period 1981-2019. As shown below:



I indicate the typical scale values of the standardised index, II, establishes the threshold maintaining the drought categories and III represents the drought areas in a grouping.

4.2.2 Spatiotemporal Method

The spatiotemporal analysis of this research is based on the methodology proposed by Corzo and Vitali, 2018 called Contiguous Drought Area (CDA) as the basis for the development of the tracking method. The Contiguous Drought Area CDA method is applied to identify neighbouring cells that form drought clusters. From this perspective, a drought event is defined by a starting location, trajectory over time, and a final location based on the reconstructed tracks. This methodology is mainly composed of 2 phases. 1) spatial analysis or clustering and 2) clusters tracking.

4.2.2.1 Step 1. Spatial drought Units

A Contiguous Drought Area, CDA is composed of neighbouring cells in drought. These cells in drought are identified at each time step. (Corzo, 2019)When the drought indicator is less than or equal to the selected threshold, the value of 1 indicates that the cell is in drought. Otherwise, the value of 0 is used, indicating no drought. (Diaz et al., 2020b)Several approaches suggest the use of visual analysis tools for the development of this analysis. Among these, the following are highlighted:

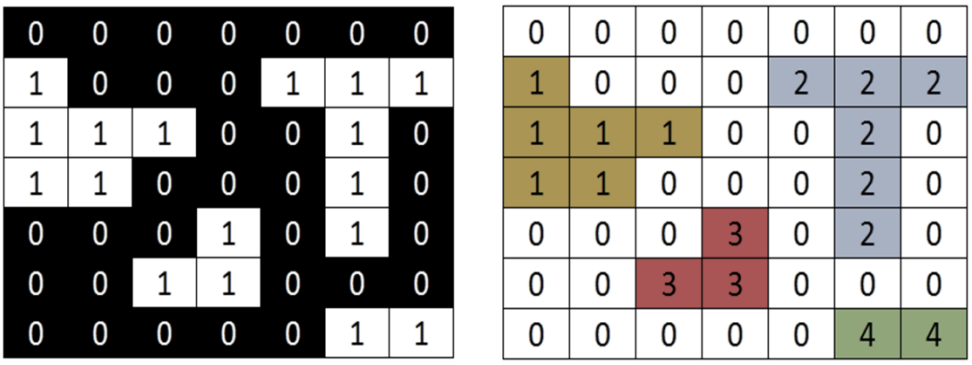
1. Graph Theory, Connected Labelling Component.

2. Semantic Abstraction, Image segmentation

4.2.2.1.1 Connected Labelling Components - Space Analysis

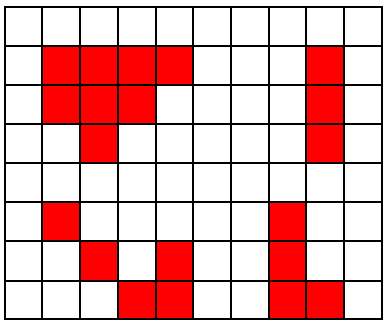
According to the previous chapter, section 2.3.2, a methodology based on generating clusters or groupings of cells based on image analysis algorithms is presented. For this case, the technique adopted was two-dimensional Connected Components Labelling, which consists of generating subgroups of patterns in an image based on its pixel value and its connectivity with neighbour cells or pixels.(Sheffield et al., 2009)

Figure 20 Connected Labelling Components

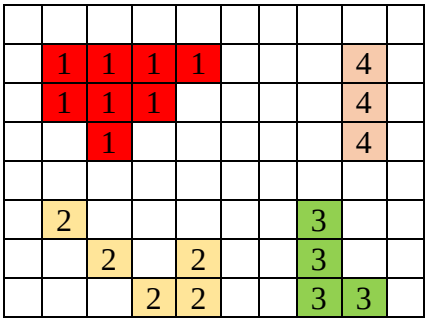


The algorithm is based on an algorithmic application of graph theory, where subsets of components are uniquely labelled based on a given heuristic. This heuristic could be scanning, pixel by pixel (top to bottom and left to right, identify connected pixel regions, i.e., regions of adjacent pixels that share the same set of intensity values. After labelling, a binary image will be transferred to a labelled image. Fig17

Initial image



Clustering



Picture Number	Label Id.
1	1
1	2
1	3
1	4

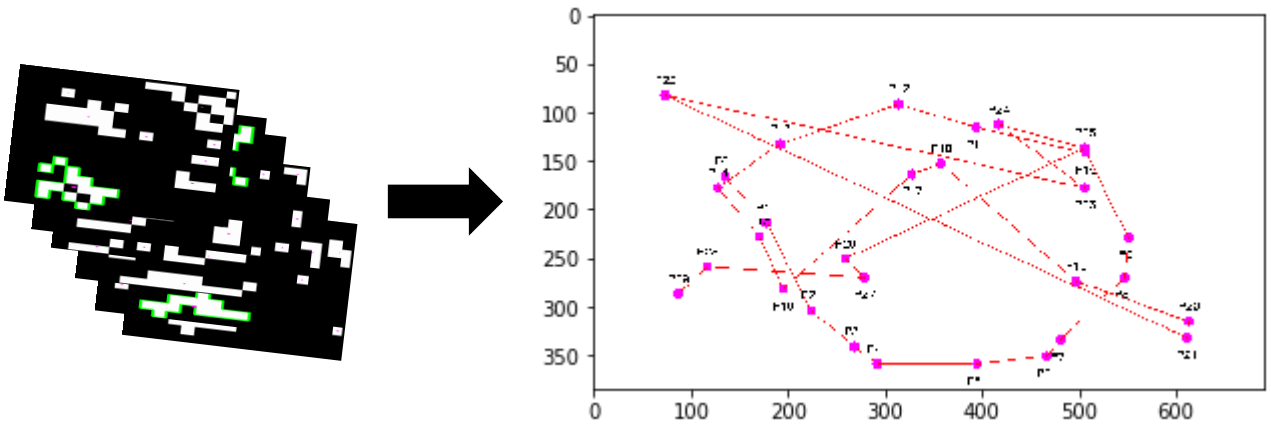
4.2.2.2 Step 2. Tracking: Trajectory Linkage

Figure 21 Features Extraction

	Picture Number	Label Id.	Centroid X	Centroid Y	Area	%Area on the image	Perimeter	Vertical longitude	Horizontal longitude
0	1	0	322	202	903	0.34	128	22	44
1	1	1	135	166	11944	4.49	535	142	129
2	2	0	366	250	966	0.36	130	24	43
3	2	1	178	214	12114	4.55	539	144	129
4	3	0	409	345	966	0.36	130	24	43
5	3	1	224	304	10977	4.13	490	119	129
6	4	0	269	341	8554	3.22	436	91	130
7	5	0	292	359	2642	0.99	257	44	87

Once the clusters (areas in drought) are identified, the largest cluster is identified at each time step t. Its geospatial properties such as centroid location, horizontal and vertical lengths, area percentage are then calculated (Figure 18). (Diaz et al., 2020a) Each centroid obtained by the cluster with the most significant area is analysed under the overlap percentage and Euclidean distance metric, which is defined as a union of events considering the minimum Euclidean distance and maximum overlap percentage. (Diaz et al., 2020b).

Figure 22 Conceptualisation Tracking Analysis



4.2.2.3 Step 3. Drought Characteristics

Finally, the information calculated throughout the previous steps allows describing the occurrence of the drought. In particular, information on the duration and severity. This is achieved according to section 2.3.2, in which the four steps to characterise drought events are presented.

4.3 Drought Forecasting

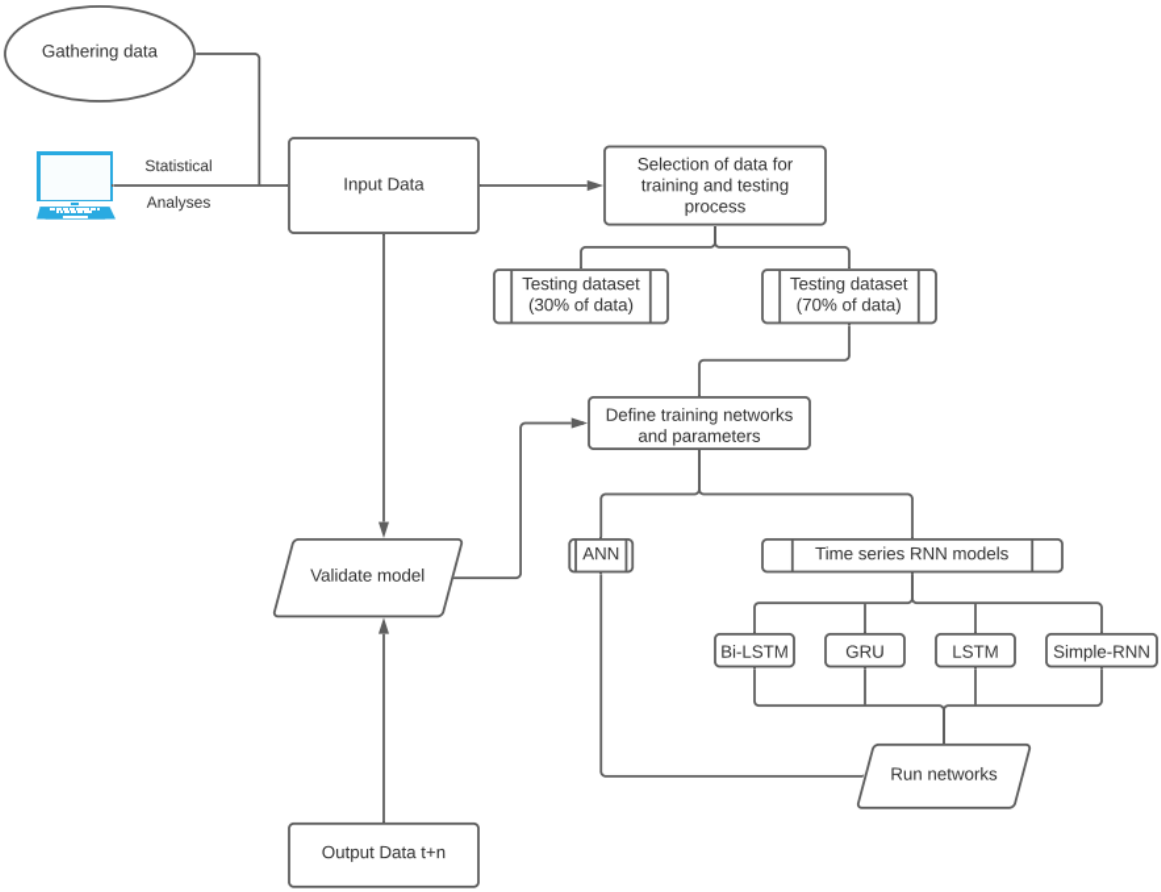
This research seeks to evaluate the potential of machine learning techniques applied to drought prediction in Central America. This framework will consist of the implementation of 5 deep and automatic learning algorithms such as SVR, ANN's, LSTM, as described in the previous chapter.

For this purpose, the following specific phases are generated:

1. Problem Definition

2. Feature Selection

3. Model Calibration and Validation



4.3.1 Problem Definition

In this section, the problem's definition is of great importance in developing any data-driven modelling context, as mentioned in section 2.4.2. in order to validate the hypotheses stated above in section 1.2.2.

1. Drought prediction (Drought Index) based on climatological information.

- $DI_{(t+1)}= f(DI_{(t)}, DI_{(t-1)},\ldots\ldots DI_{(t-n)}, P_{(t)}, P_{(t-1)},\ldots\ldots P_{(t-n)}, T_{(t)}, T_{(t-1)},\ldots\ldots T_{(t-n)},\ldots..)$

2. Prediction of spatiotemporal characteristics of drought, such as location and percentage of area.

- $Dcharacteristics_{(t+1)} = f(Dch_{(t)}, Dch_{(t-1)}, \ldots\ldots Dch_{(t-n)}, Centroid_{(t)}, Centroid_{(t-1)}, \ldots\ldots Area_{(t-n)}, \ldots..)$

4.3.2 Feature Selection

This stage is presented as an extension to section 4.1.1, in which we go deeper into the analysis of the characteristics required in the formulation of the forecasting problem. This section consists of generating a dataset whose properties allow the machine learning project to transfer information.

After data cleaning and data imputation, the relationship between inputs and expected outputs is required using Correlation techniques. The climate characteristics such as Temperature, Precipitation, Year, Month, Position X, Y, Potential Evapotranspiration were evaluated with the computed drought indexes of which only precipitation and temperature have a higher relationship concerning the others (R coefficient between 0.28-0.43).

The dataset is then divided into 80% for the training model and 20% for testing.

4.3.3 Model Building, Calibration and Validation

After the selection of potential and probable input variables according to the correlation, model development is based on model calibration, which allows finding a set of parameters or hyperparameters most appropriate to represent the data set.

These hyperparameters vary according to the structure of the model, i.e. autoregressive models require calibration of probabilistic parameters such as those mentioned above (p,q,P,Q,D), support vector machines (kernel type) and finally in neural networks (hidden neurons, connection weights and biases). In SVR, the hyperparameters are C, Kernel, , epsilon and gamma, where each one represents the Model calibration is not a simple task and is generally performed by an optimisation algorithm. Specifically, neural networks employ the backpropagation (BP) algorithm, which minimises the mean square error (MSE), for example.

Figure 23FlowChart SVR methodology

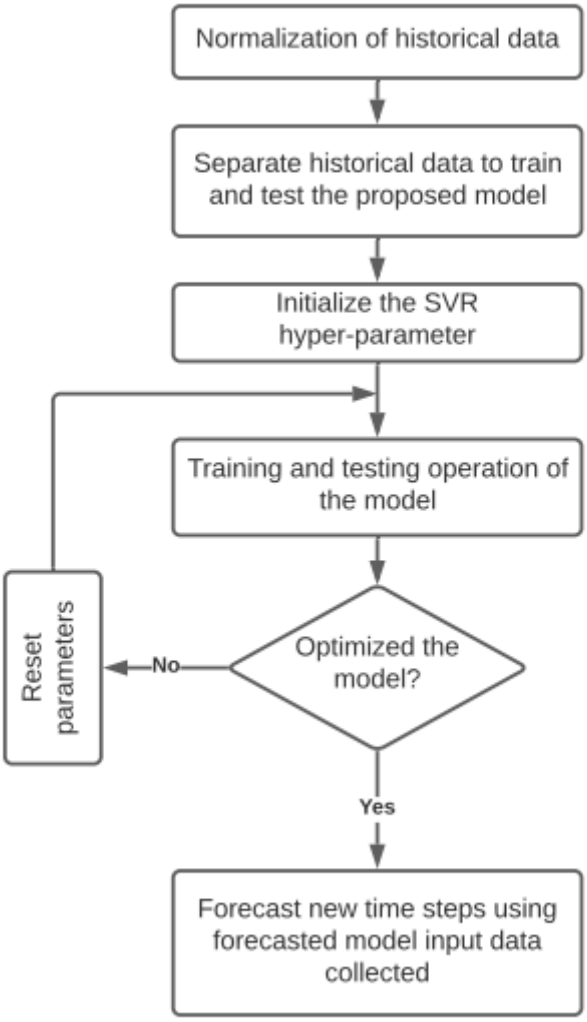
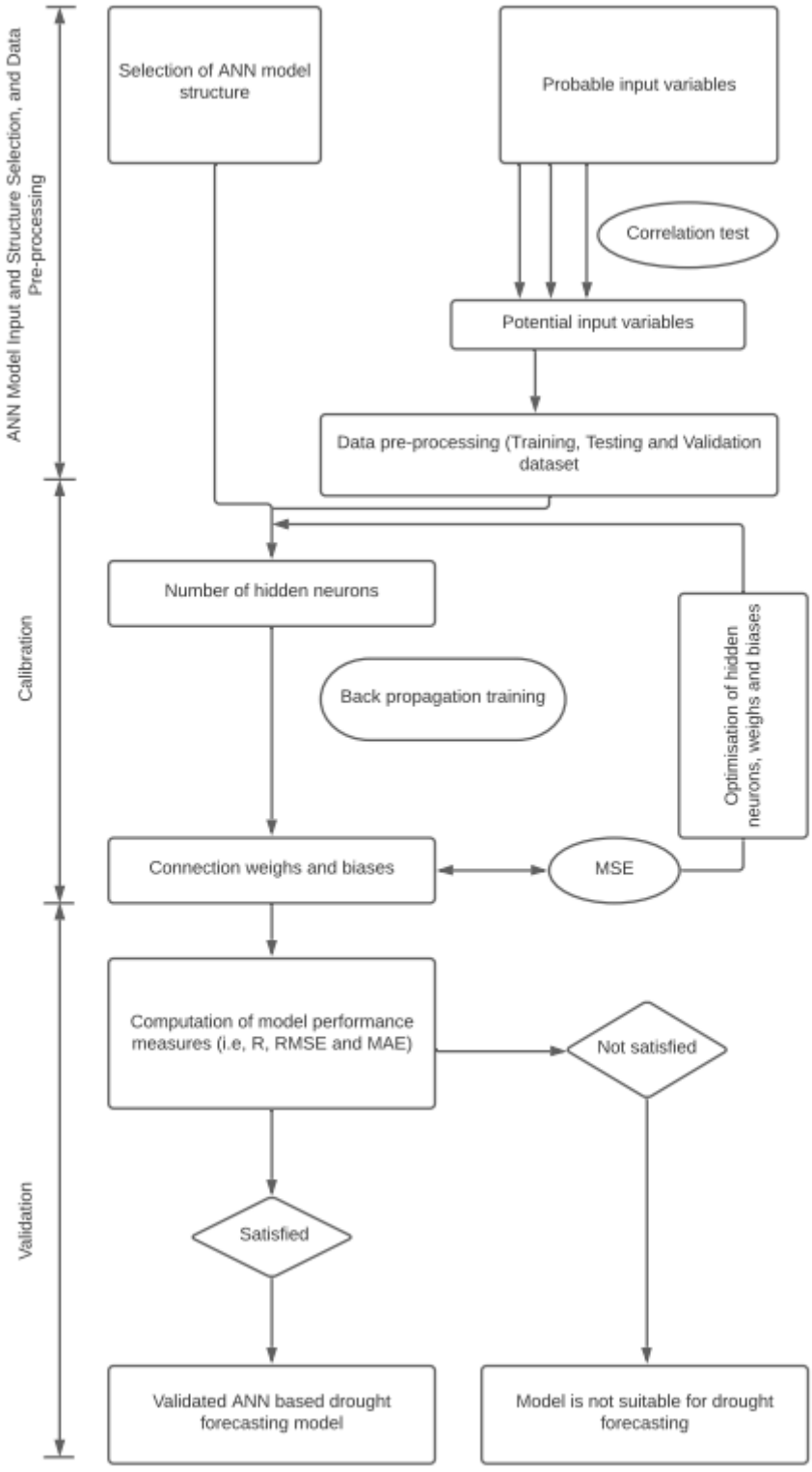


Figure 24 flowchart ANN model development based on Barua 2010

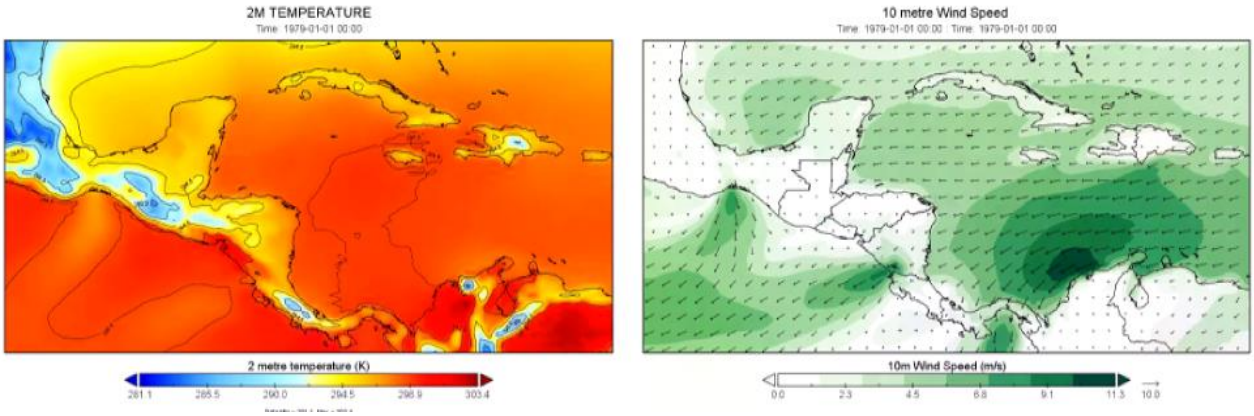


Chapter 5 Spatiotemporal Drought Assessment

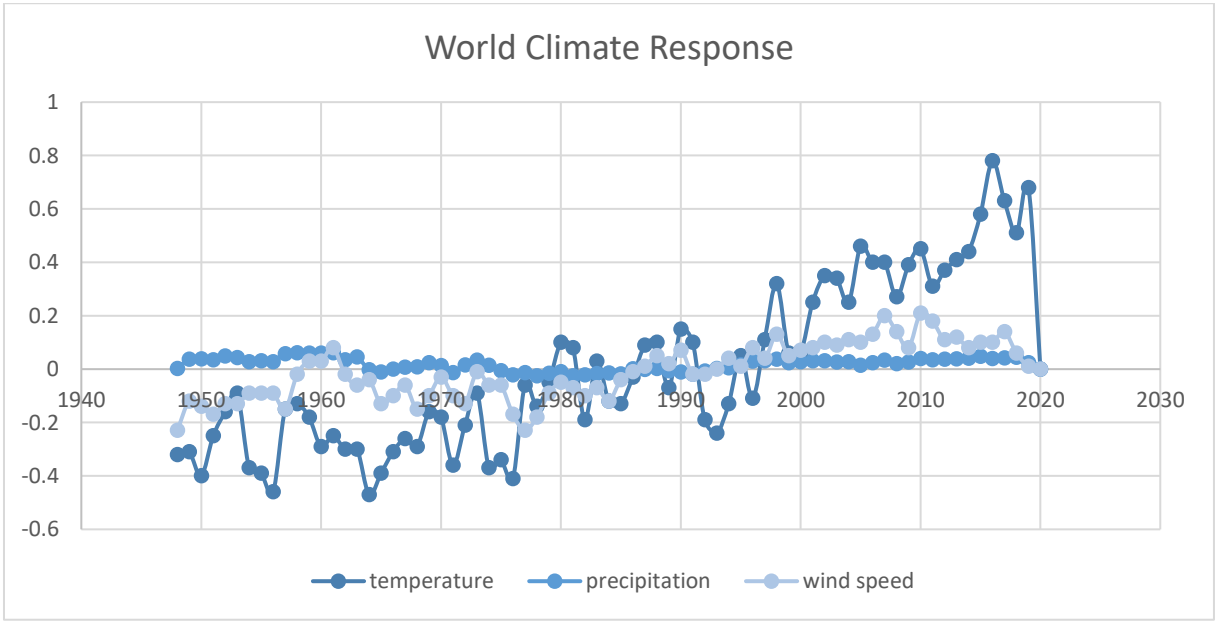
5.1 Data Acquisition

The climatological information required for calculating and evaluating the drought corresponds to the ERA 5 land monthly averaged reanalysis data with a resolution of 0.1° x 0.1°, resolution 9 km. Precipitation, Temperature, Winds, Atmospheric Pressure were analysed to extract maximum and minimum values and establish the area's hydrological regime.

Figure 25 Temperature and Wind ERA 5 map



We could observe the effect of climate variability causing the precipitation rate to decrease at a rate of 0.05 mm/year compared to the temperature, which has a trend close to 0.8 degrees/year.



Subsequently, the masks were extracted for the zone of interest (Central America) before the indices' computation.

About this point, it is important that, as noted previous, years like 1983, 1987, 1993, 1997, 1999-2001, 2000-2004, 2009-2011 and 2015-2019 have previously been associated with droughts, land anomalies and rainfall anomalies validate the high relationship between global dynamics with the land situation. That has been confirmed by checking historical SPI and SPEI reports in Central America and now warmer than normal temperature conditions. That is not to suggest that all the years that had drought also had higher than average temperatures, but instead, it is to bring out the argument that most droughts are often associated with higher than normal temperature conditions

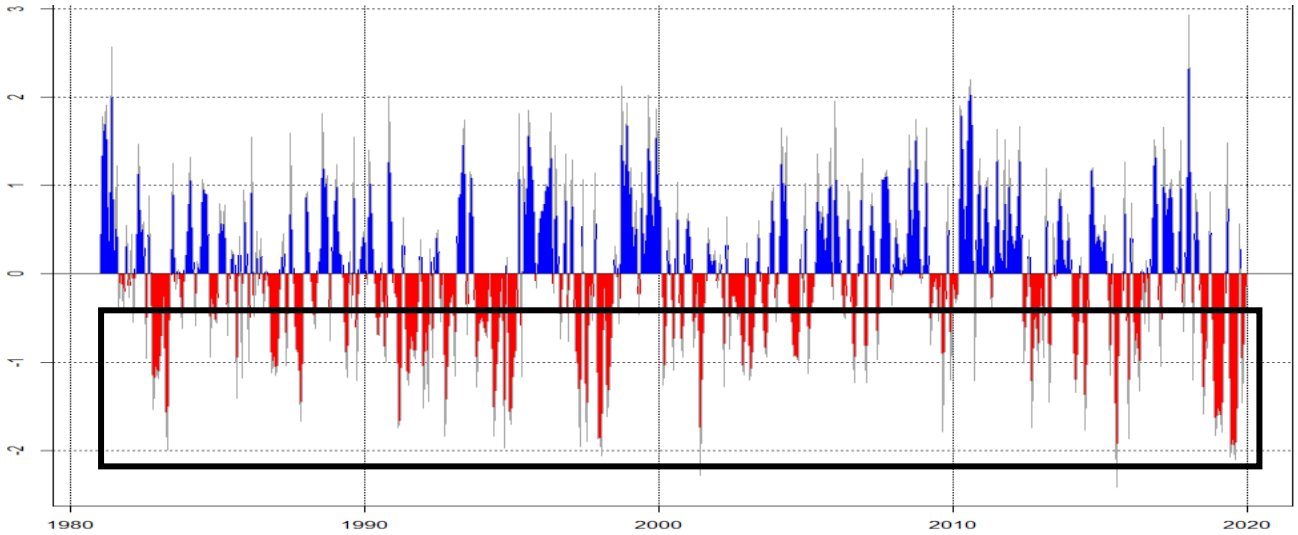
This allowed us to determine the characteristics of the variables mentioned above in the study area. Regarding precipitation and temperature, an annual average between 380 and 3500 mm per year was reported, and average temperatures between 20-26 degrees, establishing two rainfall regimes related to the dry season (on average from December to March) and the rainy season (April to November). Correlational analyses related the strong influence of temperature and winds (r=0.67) to an inverse precipitation condition (r= 0.32,0.18).

Also, since no evapotranspiration information was available, the Thornthwaite methodology (water balance) was adopted to generate it from the precipitation and temperature previously extracted.

5.2 Drought Indices SPI & SPEI

De acuerdo a lo mencionado en la metodología, índices estandarizados de sequia SPI y SPEI fueron calculados. Para ello se empleo lenguaje de programación Python para desarrollar las funciones requeridas para su calculo, empleando la distribución log Pearson III como función de probabilidad para la normalización de la precipitación y la evapotranspiración para periodos de acumulación de 1,3,6,9 meses (Appendix Cb).

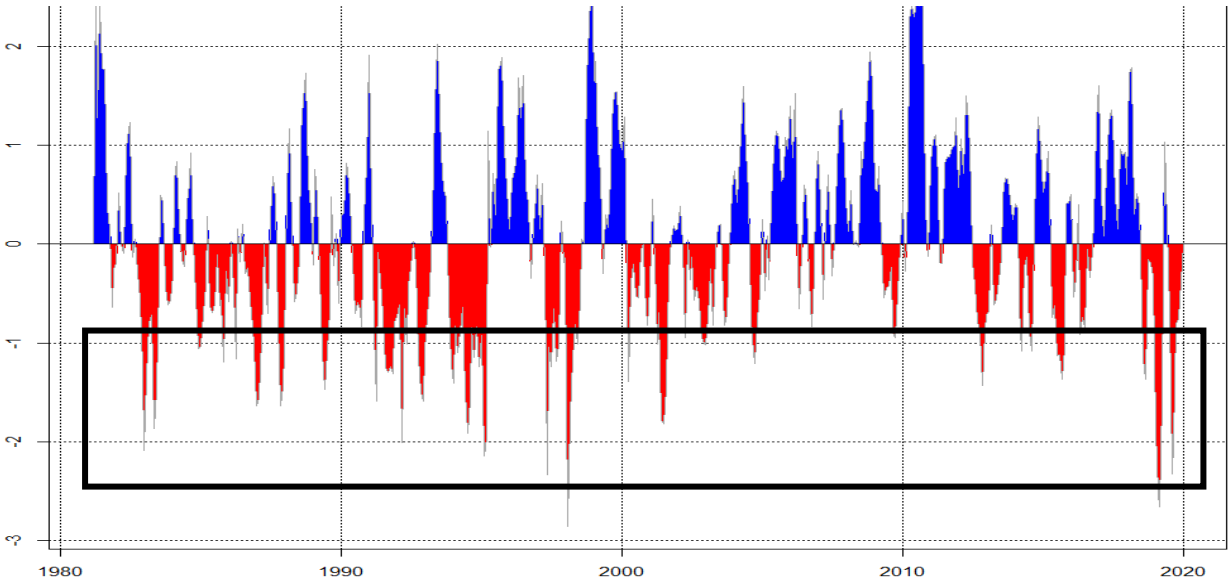
Figure 26 SPI 1 month



Next, an analysis of the time series obtained was developed, in which the concentration periods of 3 and 6 were established since they reflect the dynamic behaviour of the system and do not show abrupt alterations such as those reported in concentration periods of 1 month. Due to the study area's hydrological behaviour, if very short accumulation periods like 1 month or very long periods like 12 months are evaluated, we could be generating an incorrect evaluation of the drought, since between 3 and 6 months, the transition of the described phenomena is

presented. Its response to the propagation of meteorological drought (SPI) to agricultural drought (SPEI) is more appreciable (Figure 25-26).

Figure 27 SPI 3 month



Subsequently, heat maps and isolated time series are computed for the periods described, in which the behaviour of the index from 1981 to 2019 can be observed. The significant ENSO climatic phenomena show changes in the study area's water condition, mainly observable for 1983, 1987, 1993-1995, 2005-2015 and middle 2018-2019. These years, the drought index SPI and SPEI reported lower data (Extreme and Severe Dry) (Fig 27-34)..

Table 8 Drought Indices Correlation Matrix

	spei1	spei3	spei6	spei9	spei12	spi1	spi3	spi6	spi9	spi12
spei1	1.00	0.63	0.48	0.40	0.33	0.94	0.61	0.47	0.40	0.34
spei3	0.63	1.00	0.74	0.62	0.53	0.57	0.96	0.72	0.62	0.54
spei6	0.48	0.74	1.00	0.84	0.73	0.42	0.68	0.97	0.83	0.73
spei9	0.40	0.62	0.84	1.00	0.89	0.33	0.55	0.78	0.97	0.87
spei12	0.33	0.53	0.73	0.89	1.00	0.27	0.45	0.66	0.84	0.97
spi1	0.94	0.57	0.42	0.33	0.27	1.00	0.61	0.44	0.35	0.29
spi3	0.61	0.96	0.68	0.55	0.45	0.61	1.00	0.71	0.58	0.49
spi6	0.47	0.72	0.97	0.78	0.66	0.44	0.71	1.00	0.82	0.70
spi9	0.40	0.62	0.83	0.97	0.84	0.35	0.58	0.82	1.00	0.88
spi12	0.34	0.54	0.73	0.87	0.97	0.29	0.49	0.70	0.88	1.00

These drought values were also extracted by country (Appendix Cb), which indicate the countries with the most significant damage from extreme minimum hydrological disturbances for the study period. The most remarkable effects are reported in these countries for El Salvador,

Guatemala and Nicaragua Pacific, since, due to oceanic oscillations ENSO, POD and anomalies related to climate circulation patterns. They have been more vulnerable than Belize, whose effects are close to normal, under the most significant effects described above (Appendix Cb).

Figure 28 SPEI 3 Heatmap

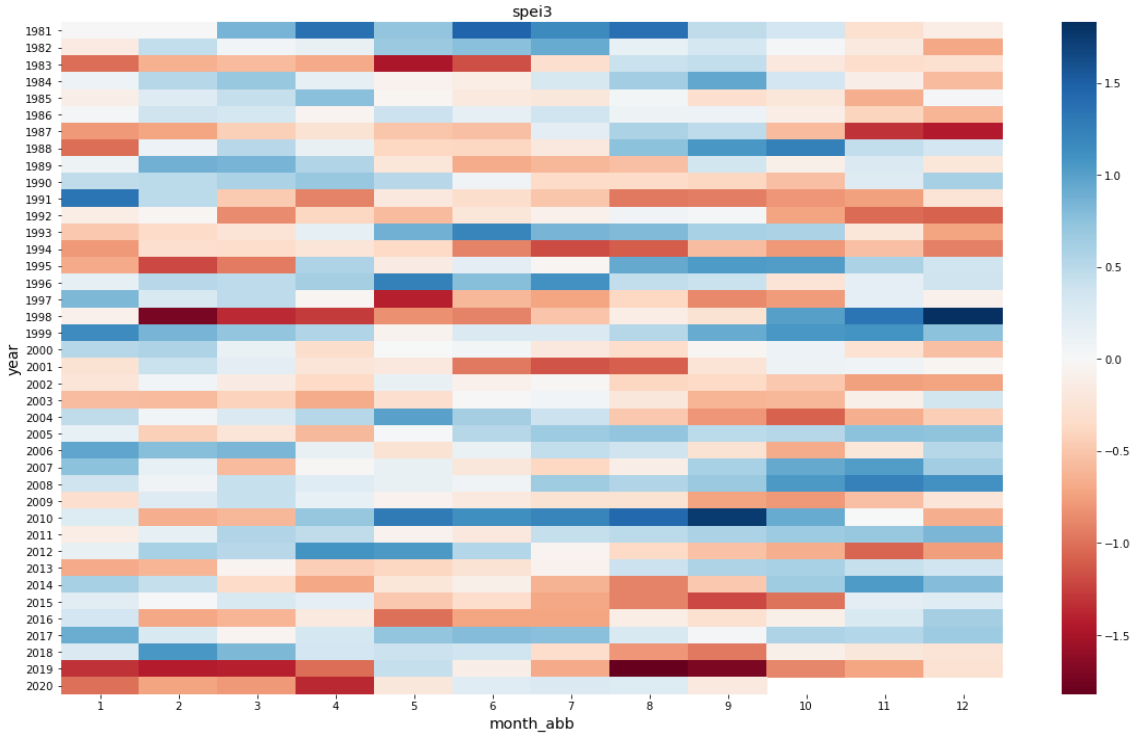
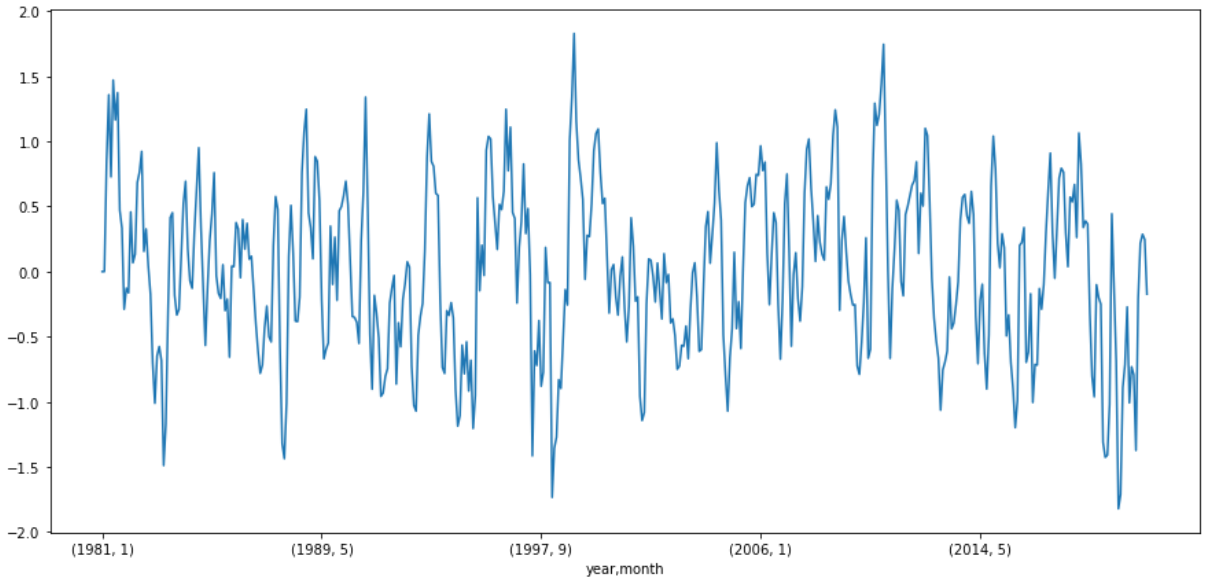


Figure 29 SPEI 3 Time Series



As it can be seen, under a 3-month accumulation period for SPEI, about 10 drought events can be observed, according to the established threshold (<-1.0).

Figure 30 SPEI 6 Heatmap

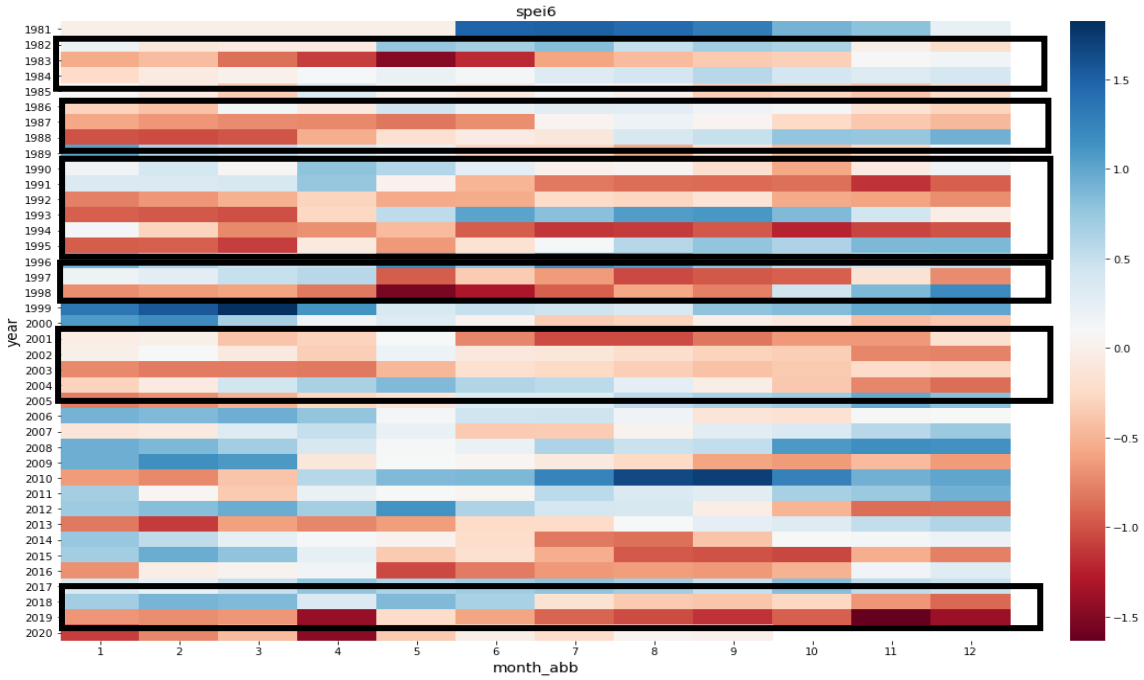


Figure 31 SPEI 6 Time Series

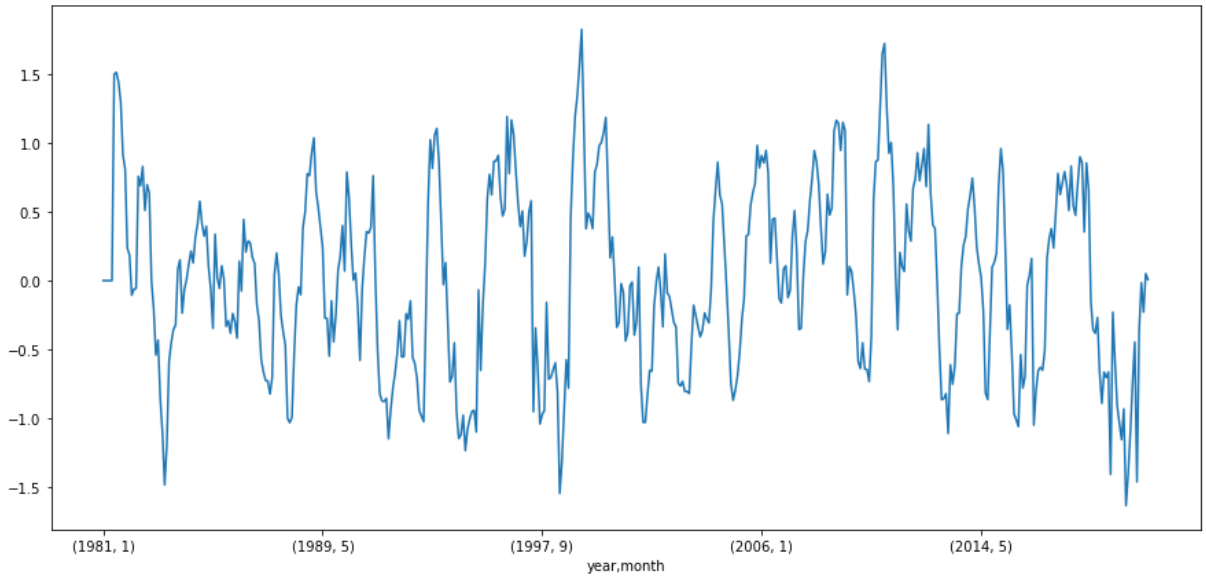


Figure 32 SPI 3 Heatmap

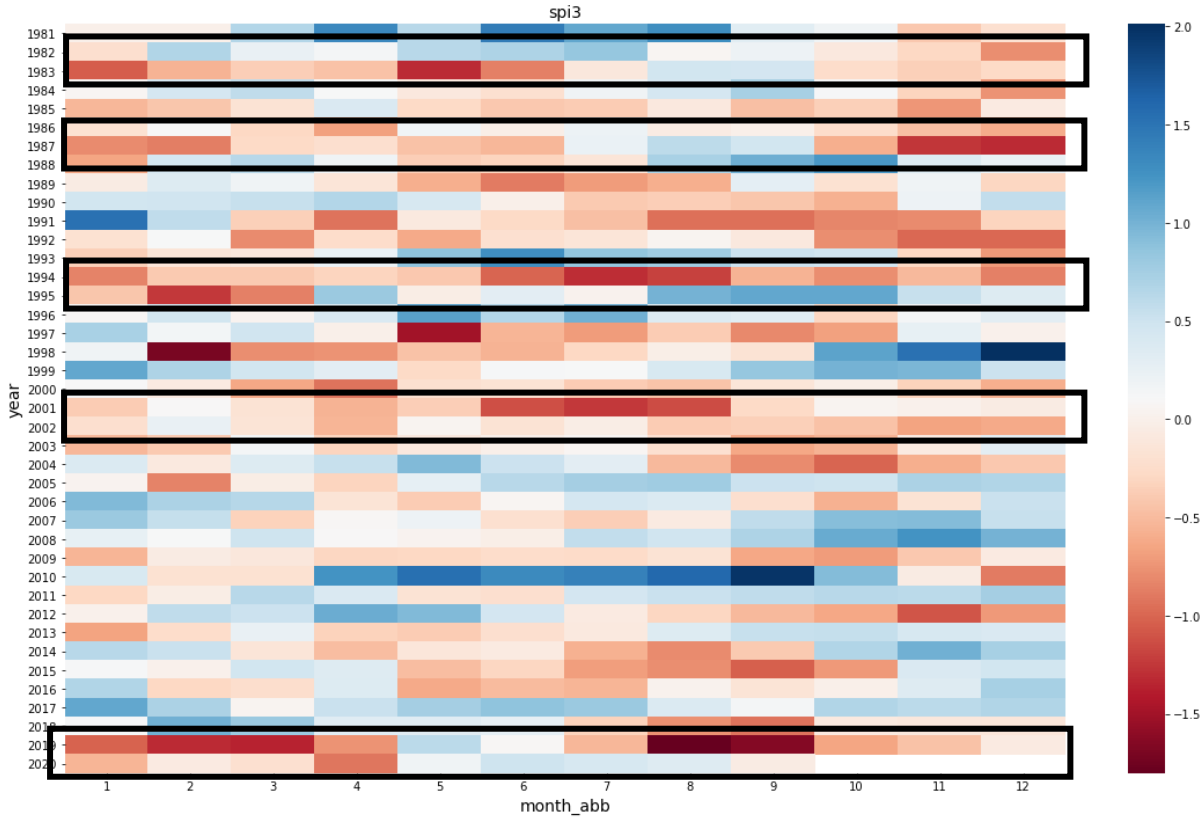


Figure 33 SPI 3 Time Series

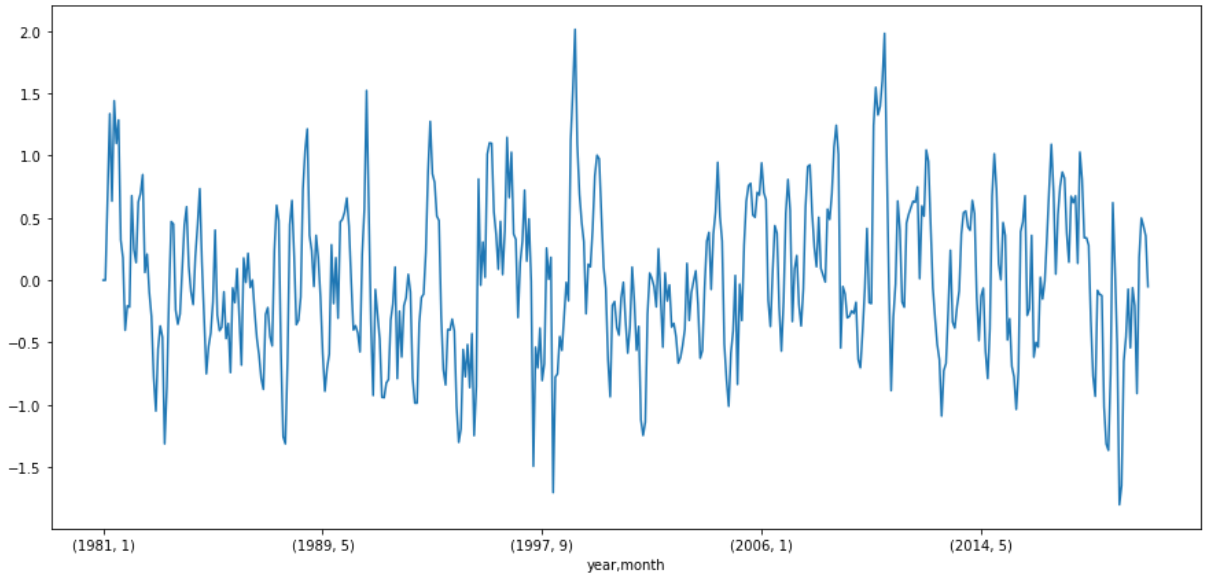


Figure 34 SPI 6 Heatmap

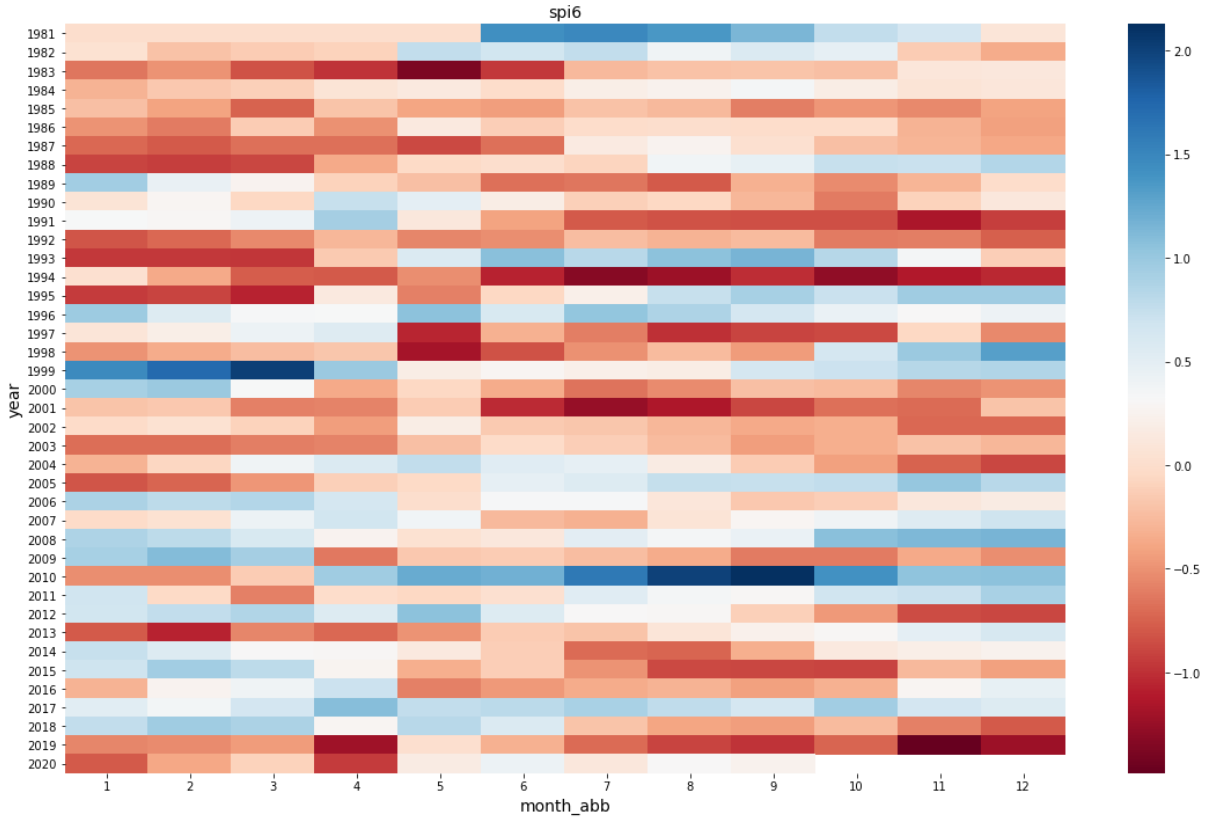
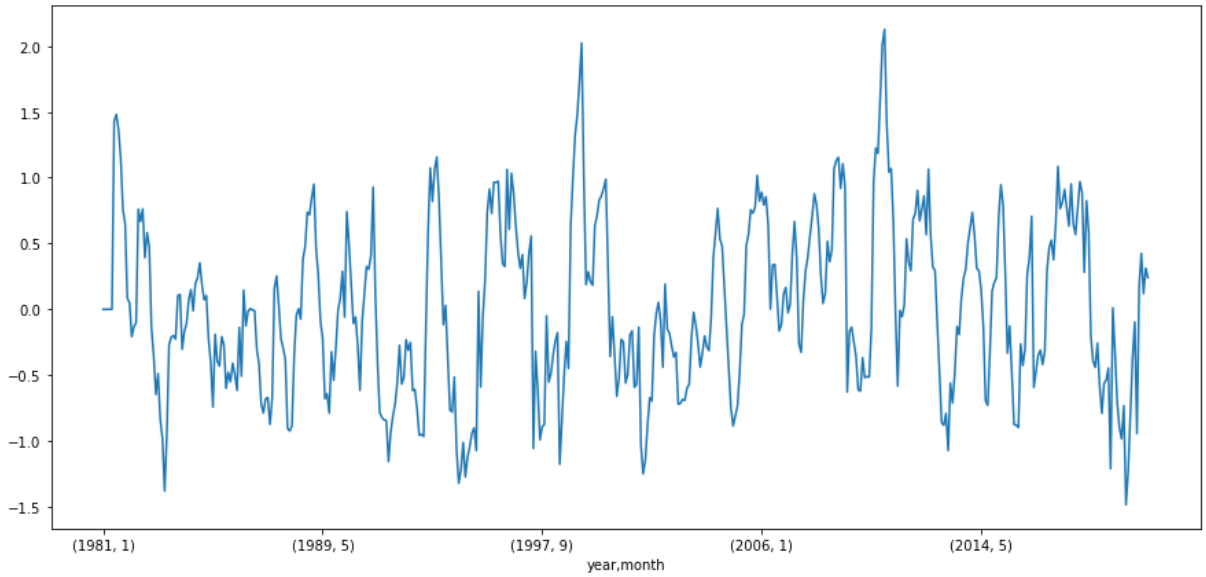


Figure 35 SPI 6 Time Series



Once concluded this analysis, the threshold was established to obtain the information required by the spatiotemporal analysis methodology described above. The established threshold was taken using as a base the drought reference of the index classification, i.e. values below -1.0.

These values allow extracting and punctuating more information regarding drought as a dynamic event (Figure 35-36). (Figure 35-36) which represents the SPEI index value with 3-month accumulation for the period 2015-10 (left) and 2015-11 (right), in which the thresholding is performed, and the drought scenarios are isolated (Figure 36).

Figure 36 SPEI 3 2015/10-11

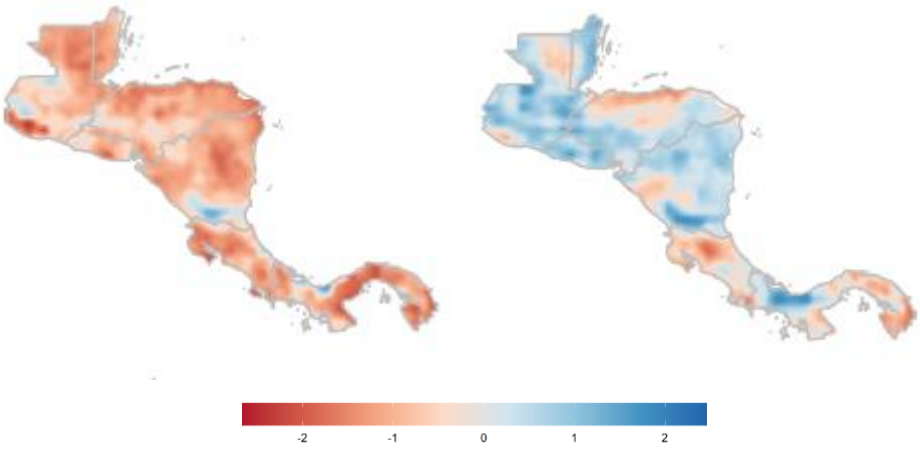
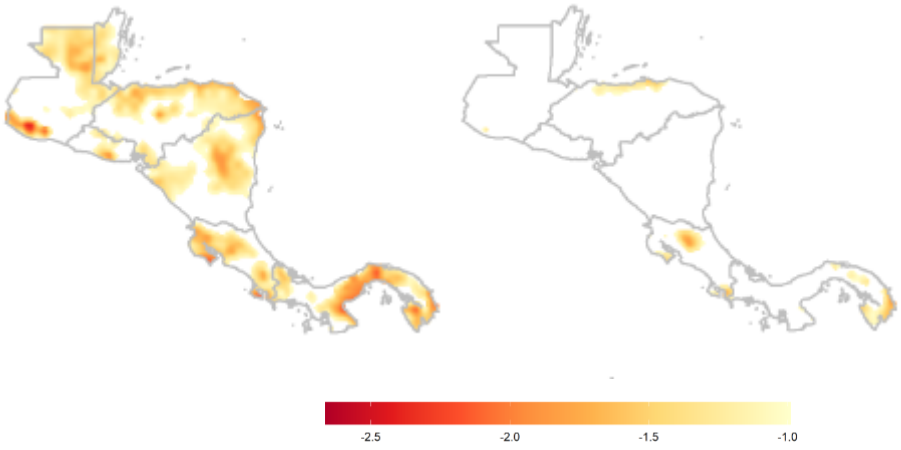


Figure 37 SPEI 3 2015/10-11 Thresholded



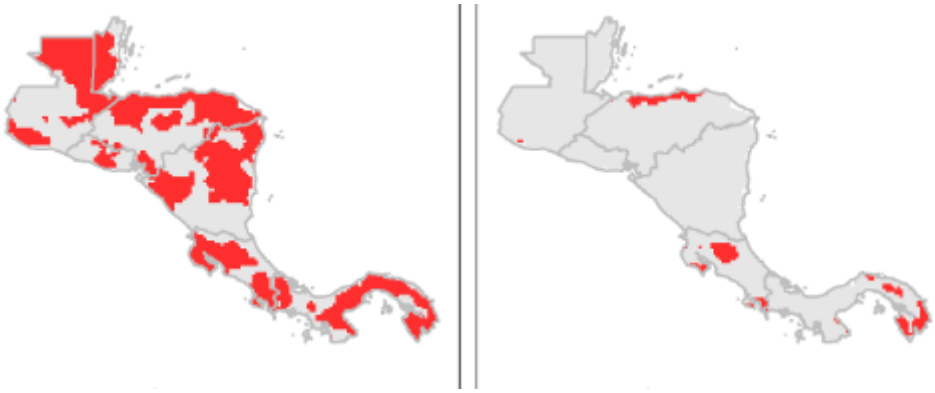
5.2.1 Spatiotemporal Analysis

Una vez efectuado el threshold, la información requerida para el análisis CDA esta completa. 468 imágenes generadas por índice (SPI,SPEI) en cada periodo de acumulación (3 y 6 meses) es generada

5.2.1.1 Image Pre-Processing Binarisation

This is the main process in terms of pattern analysis or detection of discrete objects in images. Using the previously isolated images, these are binarised, i.e., areas or pixels with a dryness value (any) differentiated from the empty cells, establishing the criteria of 1 and 0. Then implement them in the clustering or segmentation model used, as shown in Figure 37.

Figure 38 SPEI 3, 2015/10-11 Binarised

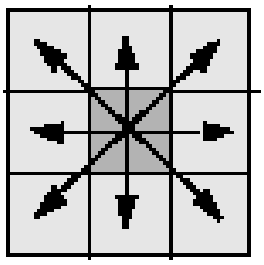


These images are exported in TIFF format, which will generate colour shifts (black and white).

5.2.1.2 Image Processing Clustering

The method of connected labelling components is implemented as part of image processing for the aggregation of contiguous areas in each time slot. Using libraries such as OpenCV, Skiimage, the algorithm was implemented for two-dimensional cluster analysis of 8 parameters, i.e. it takes into consideration their location and the location of the neighbors in the main and secondary cardinal orientations (Figure 38).

Figure 39 Connected Labelling Component 8 Parameters



This algorithm was used to process the SPI & SPEI 3 and 6 indexes, obtaining the description of the clusters found in each image, to finally obtain those with the largest area and perform the follow-up.

The following table outlines the outputs of the initial part of this model and exemplifies the location of the geospatial characteristics of the clusters found (Table 9).

Figure 40 Object Recognition

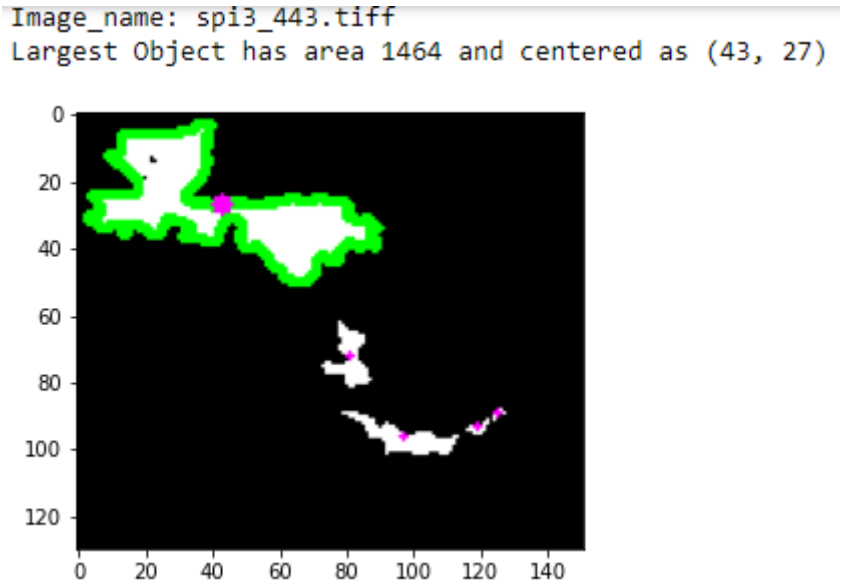


Table 9 Object Recognition Dataset

	Picture Name	Label Id.	Centroid X	Centroid Y	Area	%Area on the image	Perimeter	Vertical longitude	Horizontal longitude
0	spi3_2.tiff	0	0	0	0	0.00	2	2	1
1	spi3_2.tiff	1	67	72	4	0.02	8	3	4
2	spi3_2.tiff	2	30	36	3	0.02	6	3	3
3	spi3_2.tiff	3	32	28	20	0.10	21	8	6
4	spi3_2.tiff	4	14	28	70	0.36	43	11	13
...	...	...	...	...	...	...	...	...	...
4081	spi3_468.tiff	8	113	101	77	0.39	63	14	17
4082	spi3_468.tiff	9	129	92	3	0.02	8	3	4
4083	spi3_468.tiff	10	85	89	25	0.13	22	7	9
4084	spi3_468.tiff	11	49	28	80	0.41	80	7	30
4085	spi3_468.tiff	12	73	25	1	0.01	6	2	4

4086 rows × 9 columns

Once this information is extracted, the identified clusters acquire a Label ID indicator, which allows us to determine how many clusters are generated per time, as shown in Figure 40-43. Additionally, here we can estimate the characteristics of each of the clusters generated.

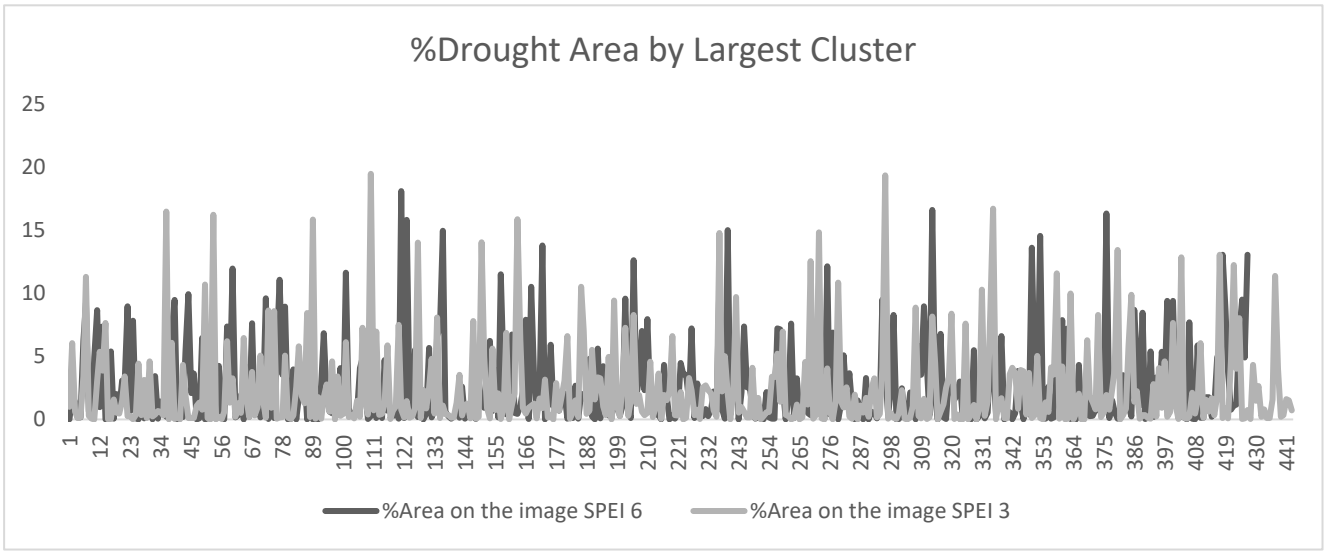
Figure 41 SPEI 3-Drought Area Percentage

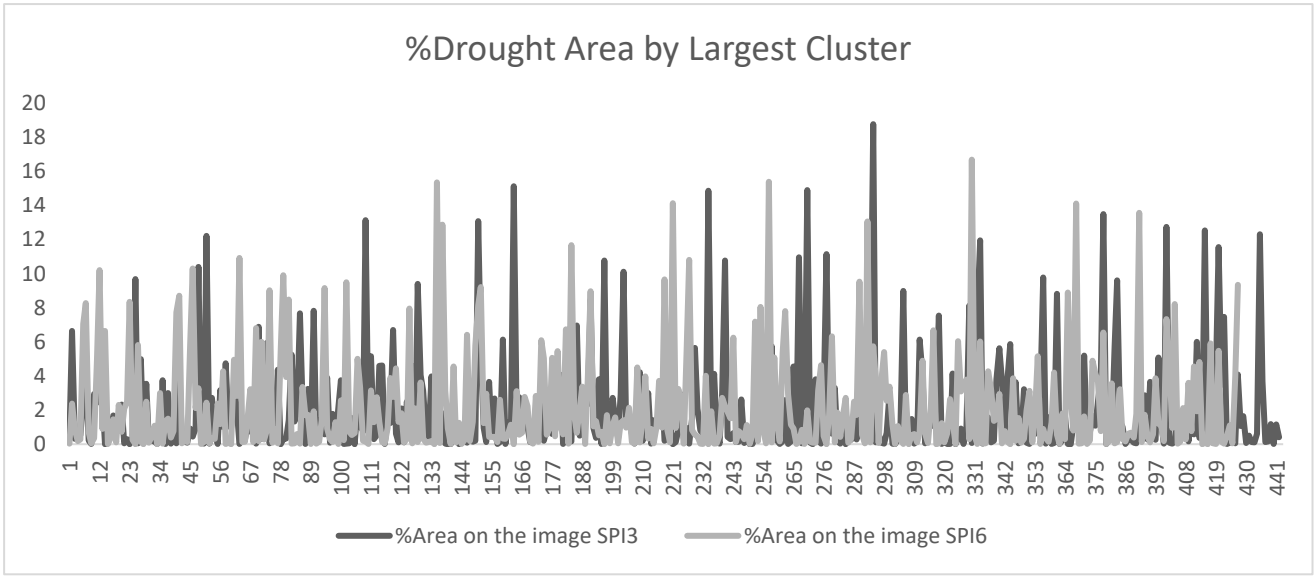
Figure 42 SPEI 6-Drought AreaPercentage

Figure 43 SPI 3-Drought Area Percentage

Figure 44 SPI 6-Drought Area Percentage

These indexes have a direct relationship, i.e. we can describe drought class propagation pattern (Meteorological to Agricultural). For SPI and SPEI their correlations are between 0.48-0.70. Besides, this propagation is temporally synchronous between the months of greater rainfall recession (December-March), where their correlations are higher. Regarding the observed frequency, the drought presented in the period 1981 to 2019, percentages of drought not greater than 20%, reflected in the periods of greater climatic magnitude ENSO described above and corresponding to cycles of recovery at the end of the hydrological cycle (September-October).



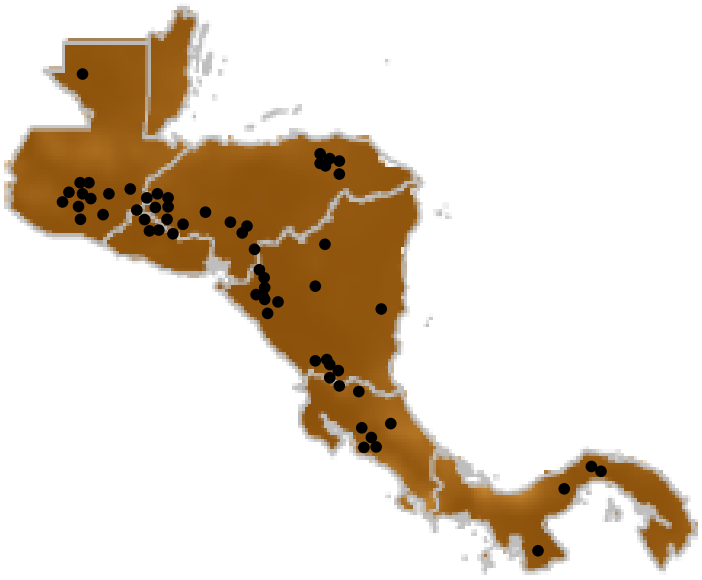


5.2.1.3 Tracking

Therefore, a hierarchy analysis was used to isolate the events with the largest area in each time step (Figure 45) , estimating the proximity criteria defined in chapter 3, Euclidean distance and overlap percentage. Tracking paths of each of the events consider the minimum distance and maximum overlap value as a criterion. With about 30% overlap, the events are related.

According to the above, an average of 90 dynamic events was found in the period 1981-2019, of which 85 drought events were reported for SPI3, 84 drought events for SPI6, 93 drought events for SPEI3 and 97 drought events for SPEI6, with a maximum duration of 2 to 7 months of consecutive drought. R squared 0.84 between total drought clusters and largest clusters

Figure 45 Localisation of most representative drought core



represent close relations, and it indicates the possibility to analyse drought using this methodology. The following graph mapped the main trajectories analysed, which have their beginnings in countries bordering the Pacific Ocean, such as El Salvador, Guatemala, Nicaragua, and their trajectory developed in southeast orientation in the maximum development of the drought and its recession in a southwest direction generally.

The reported trajectories emphasise that the development of the trajectories on average is located between 8 and 13 km. The droughts reported with greater extensions generally have their origin in Guatemala, were validated together with the trajectory of winds. This follows marked patterns in seasons of low atmospheric humidity and high temperatures. Additionally, it has been found that in the mountainous region located between Nicaragua and Costa Rica, thermal gradient phenomena and orography allow rewetting air masses, gradually decreasing the severity of the drought.

Figure 46 Principal Tracking Paths

5.2.2 General Discussion

Based on the findings, in the 1981-2019 period, an average of 84 large droughts out of 250 droughts reported in the study was identified. Of this total, 21 had a duration equal to two months, and 63 lasted more than two months.

Among the two-month droughts, 80% are developed in the period of water recession in the area. In comparison, the remaining 20% are generated according to extreme conditions that precede them, and gradually, their recovery is considered. According to the vegetation index (appendix bc), this is highly related to the increase in the cover that has presented stress in health and quality.

The reported changes in precipitation and temperature, as mentioned above, may partially explain the change in patterns of drought onset and magnitude under climate change scenarios. This suggests that in some way, the decrease in hurricane season rainfall and its concentration over time, with the consequent increase in the frequency of extreme rainfall events, has caused a modification in the spatial distribution of drought areas, thus causing a strong dynamic in later projections. On the other hand, the droughts' intensity that occurred in years before 1995 was 14.7%. Due to the effects of climate change, this variation in intensity has fluctuated and has managed to exceed about 10% more than reported.

An important aspect of droughts using computer vision techniques is that events can be generated sporadically according to the selected threshold, as it is for January, May and September. The drought can equal or exceed the value of -1, and its connectivity is interrupted.

According to the proposed trajectory, within the most extreme years of drought, we have the periods related to 1983, 1987, 1991, 1994, 1997, 1998, 2001, 2015, 2019, representing the years in which low drought index SPI and SPEI exceed the percentage of area in more than 48%.

In all cases, the drought areas overlap considerably in the later time steps, suggesting that the spatial extent perhaps remains more or less in the same region after reaching a considerable size. The permanence of large drought areas in the same region over time may explain the severity of droughts in these years.

Chapter 6 Drought Forecasting

In this chapter, the achieved results concerning the drought forecasting models are presented, considering the aspects of index forecasting and the spatiotemporal characteristics of the drought. For this purpose, each implemented tool's potential is presented and the most relevant aspects in its construction, calibration, and validation. For this reason, the objective of this chapter will be to succinctly present the development of the drought forecasting methodology using machine learning algorithms in order to evaluate possible potentials regarding the learning and forecasting capacity of erratic events such as drought.

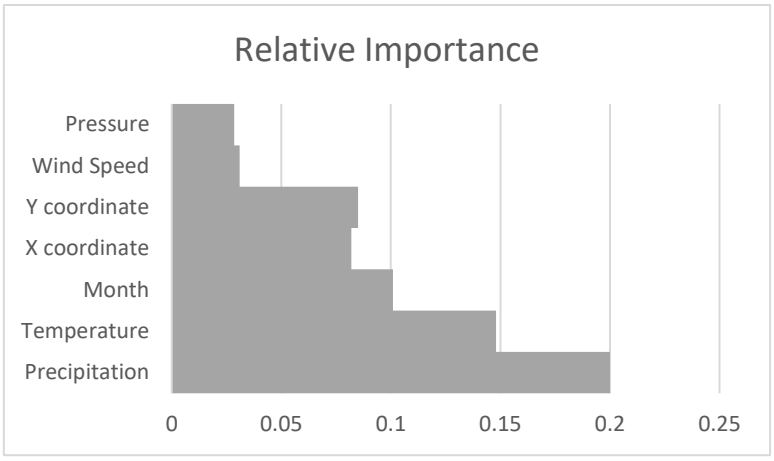
6.1 Problem Description

As mentioned above, the problem carried out should answer hypotheses 3 and 4 raised in chapter 1, alluding to the ability to forecast drought based on indexes and based on their spatiotemporal characteristics, which is why we have opted for the implementation of machine learning algorithms such as SVR, ANN, Deep learning RNN and under the conventional probabilistic forecasting model ARIMA mainly. The capacity of each of these algorithms to learn and generate forecasts in drought monitoring will be evaluated using error metrics such as RMSE and correlation coefficients.

- For drought index forecasting, $DI_{(t+1)}= f(DI_{(t)}, DI_{(t-1)},\ldots\ldots DI_{(t-n)}, P_{(t)}, P_{(t-1)},\ldots\ldots P_{(t-n)}, T_{(t)}, T_{(t-1)},\ldots\ldots T_{(t-n)})$, DI ={SPIx,SPEIx}
- For spatiotemporal drought characteristics forecast, $DArea_{(t+1)}= f(Area_{(t)}, Area_{(t-1)} ,\ldots.., Area_{(t-n)})$

6.2 Feature Engineering

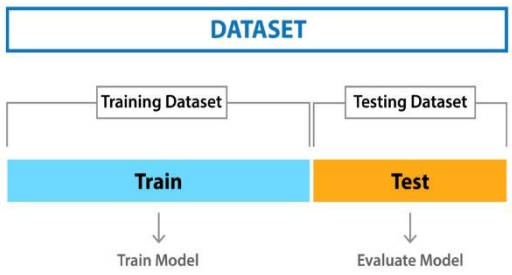
As mentioned above, two forecasting frameworks will be carried out. The selection of variables required in the machine learning project is made previously, and this depends on each model's specific requirements. Machine learning handles two types of variables, categorical or numerical, and according to its structure, the model will be feature classification or regressive, respectively.



In this case, regressive models will be implemented, i.e. both input and output variables will be continuous variables to which we wish to fit.

	gid	x	y	year	month	ppt	tmp	pet_month	pet_day	bal
count	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06
mean	2.113500e+03	-8.608024e+01	1.323483e+01	2.000377e+03	6.471698e+00	5.645655e+00	2.342787e+01	1.040628e+02	3.411273e+00	2.234382e+00
std	1.219941e+03	3.369686e+00	2.701234e+00	1.147526e+01	3.443729e+00	4.835318e+00	3.008655e+00	3.001289e+01	9.787763e-01	4.967173e+00
min	1.000000e+00	-9.210000e+01	7.300000e+00	1.981000e+03	1.000000e+00	4.444519e-04	6.789026e+00	3.316213e+01	1.133956e+00	-1.062001e+01
25%	1.057000e+03	-8.890000e+01	1.130000e+01	1.990000e+03	3.000000e+00	1.827579e+00	2.194909e+01	8.231855e+01	2.705163e+00	-1.291044e+00
50%	2.113500e+03	-8.590000e+01	1.400000e+01	2.000000e+03	6.000000e+00	4.573687e+00	2.402099e+01	1.022713e+02	3.350826e+00	1.247539e+00
75%	3.170000e+03	-8.420000e+01	1.520000e+01	2.010000e+03	9.000000e+00	8.273904e+00	2.543588e+01	1.220914e+02	3.998596e+00	4.791181e+00
max	4.226000e+03	-7.720000e+01	1.840000e+01	2.020000e+03	1.200000e+01	7.348044e+01	3.337982e+01	3.253021e+02	1.068262e+01	7.094318e+01

bal	spei1	spei3	spei6	spei9	spei12	spi1	spi3	spi6	spi9
2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06	2.015802e+06
2.234382e+00	7.514821e-03	5.441541e-03	5.398878e-03	4.475339e-03	3.791084e-03	1.000156e-02	5.090930e-03	3.686030e-03	2.943398e-03
4.967173e+00	9.856576e-01	9.814629e-01	9.779333e-01	9.743117e-01	9.710490e-01	9.709822e-01	9.726006e-01	9.679842e-01	9.656475e-01
-1.062001e+01	-7.184834e+00	-5.169174e+00	-9.042821e+00	-4.650128e+00	-3.933547e+00	-5.164133e+00	-5.214851e+00	-5.003726e+00	-4.483108e+00
-1.291044e+00	-7.076089e-01	-7.186585e-01	-7.274129e-01	-7.238612e-01	-7.153324e-01	-6.405586e-01	-6.699909e-01	-6.810055e-01	-6.766482e-01
1.247539e+00	-9.199277e-03	-3.279743e-03	0.000000e+00	0.000000e+00	0.000000e+00	-3.460428e-02	-2.948710e-02	-3.272556e-02	-1.993812e-02
4.791181e+00	7.105912e-01	7.220615e-01	7.258259e-01	7.338465e-01	7.247451e-01	6.325508e-01	6.498979e-01	6.444996e-01	6.575796e-01
7.094318e+01	3.105993e+00	3.804458e+00	3.520814e+00	3.516960e+00	3.638895e+00	4.843459e+00	4.405962e+00	4.240308e+00	4.004328e+00



Within the selected data framework, the characteristics evaluated to forecast drought index will be those that in chapter 5 were used in its analysis. We have Precipitation, Temperature, Location X, Y and time, for which SVR and ANN will be used for this case. Acquired spatiotemporal variables such as Location, Percentage of drought will be carried out using recurrent neural network models RNN.

Each dataset was split into 2 subsets, Training and Testing, for which the partition criterion is 70% and 30%, respectively.

6.2.1 Relating of Meteorological Data and Meteorological & Agricultural Droughts

As discussed in the previous chapter, the correlation between the potential predictors in the evaluated artificial intelligence models was estimated in pre-processing. The table below presents as a function of correlation the potential variables that can be adopted in the datasets of the models:

6.3 Drought Index Model

6.3.1 SVR

The regression problem is a generalization of the classification problem, in which the model returns a continuous-valued output, rather than a finite set output. In other words, a regression model estimates a continuous-valued multivariate function. SVR, uses the same process as SVM for classification, its marked difference lies in the type of variable to be used, categorical to numerical, so in the hyperplane it can have infinite solutions and a tolerance margin (epsilon) is established. The optimization problem consists of finding the maximum margin of separation of the hyperplane, correctly classifying as many training points as possible. The SVRs represent this optimal hyperplane with support vectors. These hyperparameters are:

C: It is the regularization parameter of the error term.

Kernel: Specifies the type of kernel to be used in the algorithm. It can be 'linear', 'poly', 'rbf', 'sigmoid', 'precomputed', or callable.

Degree: This is the degree of the polynomial kernel function ('poly') and is ignored by all other kernels.

Gamma: Is the kernel coefficient for 'rbf', 'poly', and 'sigmoid'.

To perform this task, we use GridSearchCV; it is a library function that is a member of sklearn model_selection package. It helps to loop through the predefined hyperparameters and fit their estimator (such as SVC) in our training set

```
param_grid = {'C': [70,80, 50,100,10], 'gamma': [0.01,1,0.1],'kernel': ['rbf'], 'epsilon': [0.06, 0.08, 0.07]}
grid = GridSearchCV(SVR(),param_grid,refit=True,verbose=2, scoring='neg_mean_squared_error')
```

Then we extracted the optimal values to make forecasting

```
SVR(C=100, cache_size=200, coef0=0.0, degree=3, epsilon=0.07,
gamma=0.01,
    kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False)
```

6.3.2 ANN

The neural network model's calibration determines the model's optimal parameters while maintaining the model's best learning capability. The ideal number of neurons in the hidden layer and the connection and connection weights and biases of the network are the parameters to be determined during the calibration process.

hyperparameter	values to optimize

scaling of predefined features	standard deviation, tanh, sqrt
Number of Hidden Units	1,2,5,10,15….
Number of Layers	1,2,3,4
Backpropagation Learning rate	0.001,0.05,0.01

```
{'activation': 'relu', 'alpha': 0.0001, 'batch_size': 'auto',
'hidden_layer_sizes': 500, 'learning_rate': 'invscaling',
'learning_rate_init': 0.001, 'max_iter': 1000, 'solver': 'adam'}
```

6.3.3 LSTM Model

The LSTM model was developed to analyze the predictive ability of the spatiotemporal properties of drought in different time variations (1,3,6,9), where the hyperparameters where stablished according with the calibration metrics, this configuration is shown below.

Model Set-Up

```
model = Sequential()
model.add(LSTM(n_neurons, batch_input_shape=(n_batch, X.shape[1], X.shape[2]),
stateful=True))
model.add(Dense(y.shape[1]))
model.compile(loss='mean_squared_error', optimizer='adam',metric=["accuracy"])
```

6.3.4 Calibration Metrics

After optimizing the model hyperparameters through calibration, it is necessary to validate the forecasting models using a validation data set. In this study, performance validation has been carried out by estimating: the correlation coefficient, the mean absolute error MAE and the root mean square error (RMSE), which are usually used for this case.

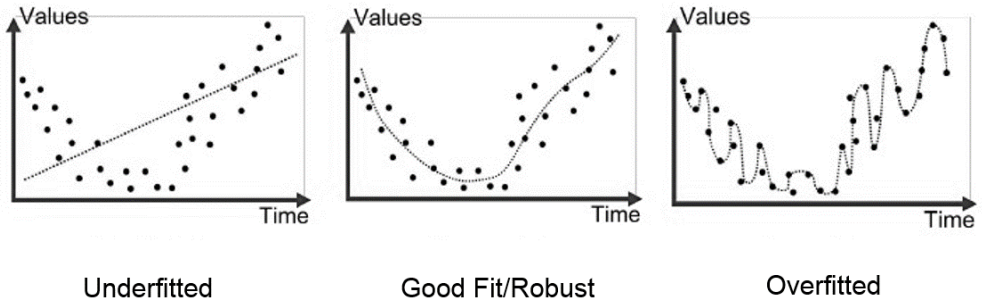
Figure 47 Results SVR, ANN 5 nodes

	SVR				ANN			
	SPI3	SPI6	SPEI3	SPEI 6	SPI3	SPI6	SPEI3	SPEI 6
R2	0,81	0,76	0,84	0,80	0,84	0,80	0,87	0,84
RMSE	1.487	2.061	1.541	1.982	1.157	1.941	1.237	1.846
BIAS Percentage%	35,2	7.3	24,9	28,1	68,1	−0,2	52,9	20,5

This is because changes in the MSE for the training and test data sets during calibration (epochs) are shown in the figure below. The training and test data sets during calibration (epochs) generate the analysis against each model's training type, as shown below.

Initially, the MSE for the subsets of data decreases; however, after a certain number of epochs, the MSE for the training set decreases. However, the MSE associated with the test data set

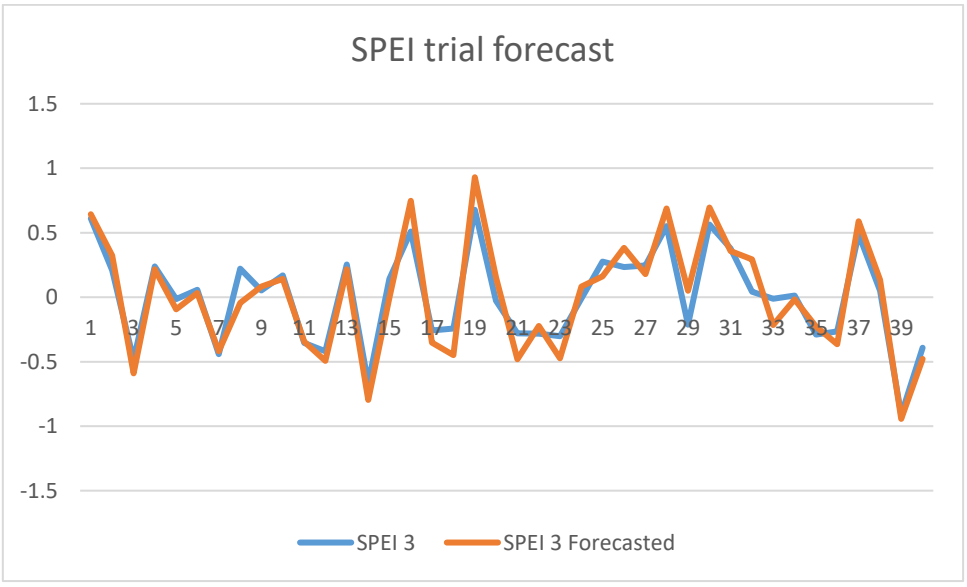
begins to increase. Test dataset begins to increase. This is an indication that further calibration is likely to overfit the network to the training data.



The MLP model was chosen with 230 nodes in the hidden layer (rounded to multiples of the sampling frequency) and one output node. The model was trained for 500 epochs with the "ReLU" for node activation, an ADAM optimizer for gradient descent and mean square error for the error metric.

Observed Drought	Modeled Drought
0.857607767	0.428167144
0.364343169	-0.51994802
-0.74550218	-1.01408602
1.474188514	0.231309251
-0.25223758	-1.46134169
-0.00560528	-1.17851647

The one-step-ahead forecasting results using MLP are shown on top, with forecasting RMSE average 2.126. The model forecasts are closely related to the out-of-sample test data with an RMSE of 8.15,11.18,1.27,3.58 for location, seasonality, precipitation and temperature, respectively.



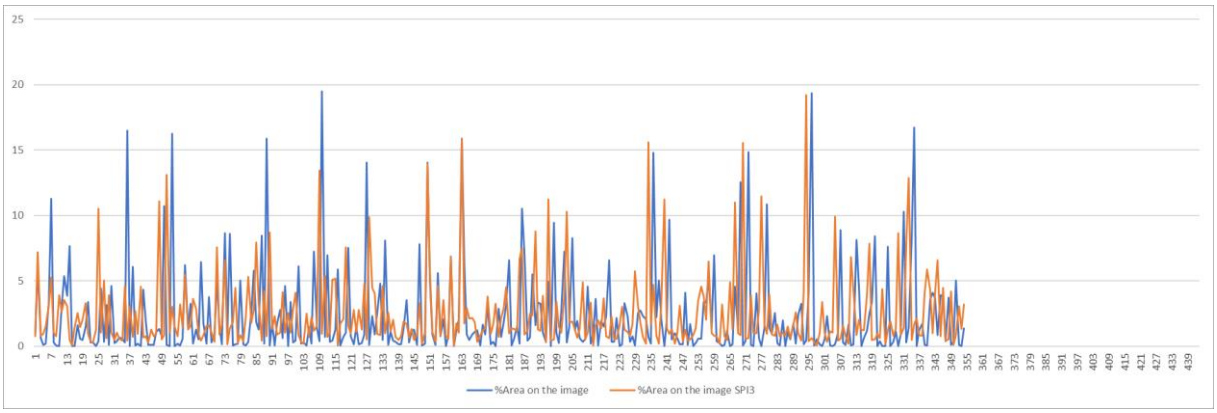
The artificial neural network models used in this study show the ability to model the drought process. This study shows the ability to model the drought process. Although they are simple

and easy to use, they have limitations in that they cannot forecast accurately when the data is noisy and have a nonlinear relationship. Therefore, the applicability of ANN for drought modelling has great potential for long-term drought forecasting.

6.4 Spatiotemporal characteristics forecasting Model

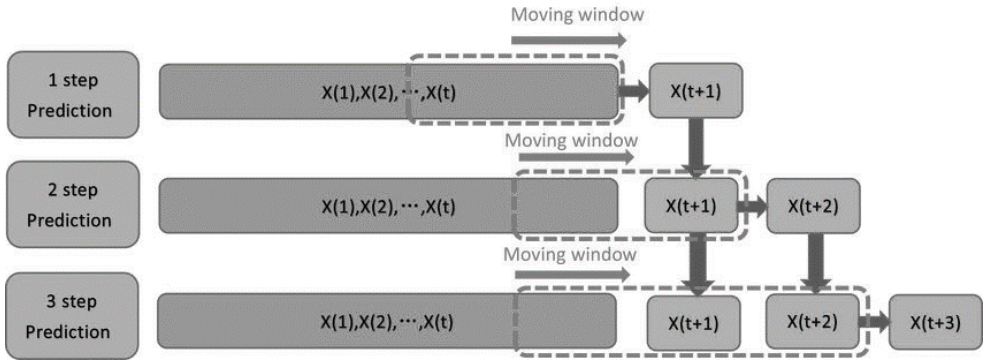
Recurrent neural networks are specifically designed to learn and predict sequential data. A recurrent neural network is a neural network in which the network's output from a one-time step is provided as input to the next time step. In this case the dynamic variable which is employed for this task is the area for largest drought event computed in the las chapter.

Figure 48Drought Area Percentage Training Dataset



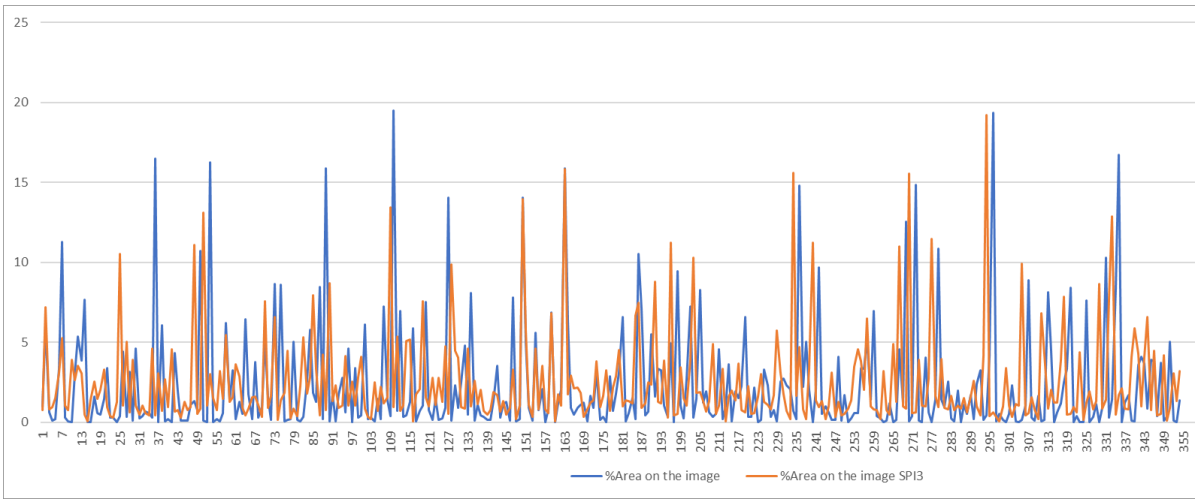
Subsequently, This allows the model to decide what to predict based on both the input from the current time step.

Figure 49 LSTM forecasting strategy based on Huang,



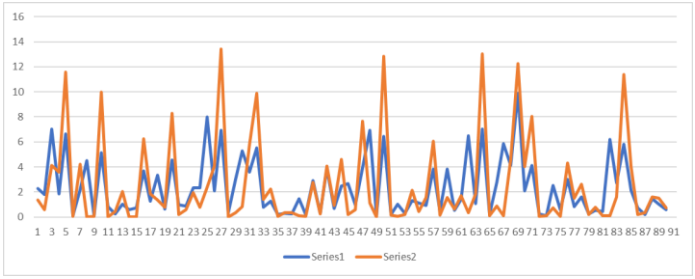
The model was developed to evaluate a 12-space window over time,

Input,	Predict
[Month 1]	Month 2
[Month 1 + Month2]	Month 3
[Month n + Month n+2 + Month n+3]	Month n+i



The model configuration presented below presents the results of this methodology, suggesting that there is an average time lag of 1 month at maximum drought area percentage values. This can be adjusted.

	t+1	t+2	t+3	t+4
R	0.75*	0.70*	0.64*	0.56*
	0.75*	0.70*	0.64*	0.58*
RMSE	0.61	0.67	0.76	0.93
	0.61	0.70	0.81	0.92
MAE	0.49	0.53	0.60	0.73
	0.49	0.56	0.64	0.72



Accordingly, as each timestep passes, the acquired error increases, which implies more significant uncertainty in the long term of use. Results show that depending on the SPI or SPEI index's accumulation period, the error will be a function of those reported events and limits its forecasting capacity. This can be improved by increasing the number of learning units. However, for this research, the results are partial. We are developing a greater capacity and design model in its architecture that combines the disaggregated analysis and the sequences articulated to images.

For the other features such as centroid location, these fluctuate, and correlation in terms of prediction does not exceed 0.35, which implies that this feature depends on image processing conditions. However, it is worth mentioning that this study's approaches are heuristic and purely experimental, which may differ according to the modeller's object. The work's current scope can be extended by evaluating other types of models or by complexifying the proposed structures based on different configurations and tuning the models using feature engineering.

This study proposes a baseline to understand how machine learning techniques can improve local management and monitoring conditions.

6.5 Limitations

Part 1. Drought Index Prediction:

Since the present study used only one reanalysis dataset for drought prediction models, it is necessary to perform analysis against the inputs' uncertainties originating. Depending on the dataset used, drought values may vary by resolution or time of measurement. Although models such as SVR and ANN allowed generating favourable learning results, it is necessary to evaluate the performance under extreme conditions. More significant biases were observed by hybridizing specific characteristics to conventional models. In front of the reported bias referring to the deviation, the longer the time is predicted, it is necessary to articulate another set of input variables such as relief and land cover that can generate better compressions in the model's learning.

Part 2. Prediction of Drought Characteristics:

The dataset was based on the previously reported drought conditions from a time series point of view. However, the results are appreciable. The bias varies according to the forecast time required (short, medium and long term). Therefore, articulating concatenated models could lead to the loss of learning from the individual time series. The use of articulated network models, on the other hand, can generate a more complete dynamics of the characteristics forecasting, i.e. the use of images or grids and their subsequent articulation to pattern detection techniques indeed allows an adequate spatiotemporal evaluation in terms of forecasting. This type of models is currently being developed by the author, in which he seeks to improve the limitations presented.

Chapter 7 Summary, Conclusions and Recommendations.

This research's main objective was to develop a methodology for the spatiotemporal analysis of drought events using artificial intelligence techniques to generate a starting point for monitoring and forecasting systems for future drought conditions in Central America. As observed, this study developed a detailed evaluation of climatic conditions' influence on the hydrometeorological response of the area and the application of spatiotemporal analysis, which allows seeing in detail both the drought relationships between classes and their dynamics in each space of time.

The execution of the methodology allowed the development of the activities that give fulfilment to the general objective, which consisted of :

- Selection of the study area and data collection and processing
- Review and evaluation of the drought indexes for meteorological and agricultural conditions and drought prediction techniques.
- Development of the spatiotemporal analysis methodology for drought events.
- Development some of the models of drought forecasting employing automatic learning.

1. Selection of the study area and data collection and processing

Dry Corridor of Central America was selected as a case study in this research because, in terms of climate variability, the conventional metrics of drought monitoring do not generate a dynamic conceptualization of the phenomenon. In addition to the fact that if it is considered in context at the water resources level, significant mortality and food insecurity are associated with the strong waves and drought phenomena presented.

The data used for the evaluation of the ERA-5 drought allows not only to easily generate a further analysis of the interaction between events but also to generate great flexibility of use in any region.

For the development of the drought evaluation, the main climatic variables that influence drought were analyzed, such as precipitation, temperature, winds, where discrete analyses allowed to analyze the seasonal dynamics of rainfall rate, temperature anomalies and their interrelations concerning humidity and the development of dry periods in the region. Therefore, it was identified to adopt an analysis of the meteorological and agricultural dynamics to see if these factors are highly related and how monitoring can be implemented with management uses and eco-social vulnerability management.

2. Review and evaluation of the drought indexes for meteorological and agricultural conditions.

Within the drought monitoring and forecasting context, several different methodologies were found and how each one contributes or differs in an integral aspect to characterise regional drought. On the one hand, the challenges of the definition of drought and its methodological aspects to quantify it, as well as the recognition of the great variety of tools that have allowed locally to analyze drought conditions in different regions and that under this context, the indices present particular utility depending on the context, required information, complexity and requirements of the investigation. The metrics refer to the quantitative characteristics that establish the phenomenon's seasonality (frequency, severity, duration).

3. Development of the spatiotemporal analysis methodology for drought events.

According to the information acquired, the drought index calculation allowed us to generate analyses regarding the synchronicity of global-scale events to drought events, which are influenced mainly by ENSO. Additionally, from the context of propagation, it is evaluated that in periods of accumulation of 3 and 6 months, these events make more sense from this context, there is a strong relationship between the conditions recorded at the meteorological and agricultural level, thus allowing to simplify further analysis wherefrom the meteorological conditions it is feasible to evaluate their implication according to changes in scale (hydrological, agricultural, socio-environmental).

Subsequently, the Spatiotemporal methodology proposed by Vitali 2018, regarding the analysis of drought events in aggregation is implemented. This method presents promising results for the study area since it allows corroborating the hypothesis of conceptualizing drought as dynamic events both in space and time and their interrelation. This methodology allowed us to analyze the characteristics of drought in Central America and validate these dynamics in historically recorded affectations. Within this analysis, it was evaluated which climatic variability factors are generating unfavourable alterations of drought, where it is reported growth in terms of extension of about 20% to the previous events and the evaluation of an average of macro-events in the period 1981 to 2019.

4. Development some of the models of drought forecasting employing automatic learning.

The drought forecast framework was developed using standardized indices. Variable selection, data cleaning and other feature engineering processes allowed the development of the datasets used for index forecasting and the spatiotemporal characteristics of the index, which were divided into 70-30 for the training and testing subsets. For the problem set on the insurability and accuracy of the drought index forecasting capability, SVR and ANN were implemented. Each of these models was optimized in its hyperparameters in order to determine the best model configuration. Optimal values of epsilon, gamma, C for SVR and number of nodes, learning rate, activation functions for ANN were established. Results showed good affinity for drought index forecasting in both models. However, ANN generated better results, mainly evaluated from RMSE and R2 metrics.

On the other hand, the percentage of drought area reported by the previous section was selected for the prediction of drought characteristics, which establishes the analysis based on time series using LSTM model. A multistep configuration is defined to evaluate the forecast accuracy according to the required time step, which yielded better results for 6-time steps in the future. Thus establishing the possibility of trend evaluation of a medium-term forecast at the evaluated scale. Finally, new frameworks currently under development are postulated, which aim to perform an aggregate spatiotemporal analysis.

7.1 Conclusions and Recommendations

- From the spatiotemporal analysis, findings were obtained regarding the spatial distribution of drought, showing that for Guatemala, Nicaragua and El Salvador, due to atmospheric circulation phenomena and geo-biographical conditions, there are areas of greater concentration of droughts on average. However, due to seasonality, countries such as Honduras and Panama have presented extreme but short-term severity events.
- It was established that the average duration of the maximum drought events is 5 months and that, on average, their extension does not exceed 30% of the region. However, according to the general analysis, this percentage of maximum recorded drought can exceed 50% coverage.
- For the years between 1983, 1987, 1991, 1993, 1994, 1997, 1998, 2001, 2015, 2019, there were higher percentages of drought areas validated from the reclamation reports registered by the sector. However, for 2019, the impact of drought had a maximum extent reported.
- Regarding the type of information acquired for the spatiotemporal analysis through the methodology presented, the scale factor plays a significant role in evaluating the functional drought areas, so it is necessary to implement downscaling techniques and integrate local records to strengthen the study, depending on the study area.
- The influence of wind and pressure on the distribution of atmospheric humidity and precipitation over the area was evaluated. However, they are not the predominant factors that lead to droughts. This is due to the observations of the pattern of precipitation and temperature anomalies that have a high correlation with the conditions and indices of reclamation reported in each context.
- The ANN model proved to be significantly better than the SVR model at all time scales (3 and 6 months). Also, in both cases, forecasting at longer time scales provided better r2 correlation values, suggesting that these models be used for long-term forecasting, however for extreme conditions at the typical mean, they exhibit much higher biases.
- The study used several categories of input variables. However, as noted, a proper study should be generated regarding the relative importance of these variables, which may interfere with the learning system's development—presenting the factors of location and temporality less affinity for precipitation or temperature. Additionally, it is suggested to incorporate parameters such as elevation, changes in coverage or remote sensing indexes since they are physically more related to drought and are easily accessible.
- According to the implementation of forecasting models, it was obtained that, on average, the models for drought index prediction and the LSTM model for drought characteristics are viable for short and medium-term forecasting. However, each model's suitability depends on the interests of the modeller and the availability of information.
- Within the deterministic and data-driven modelling frameworks, the implementation of statistical metrics (goodness of fit) is suggested to justify whether the model has the required functionality. Additionally, in spatiotemporal drought forecasting, despite generating a good approximation for the forecast of particular characteristics. More robust structures will allow the full integration of spatial and temporal characteristics, which more useful in evaluating dynamics and practicality.
- The forecasting framework has great potential for application in other regions, where information sources are limited. Although partial results of the full development of the approach were presented, it can be integrated into local monitoring and forecasting

systems and warning systems related to water use for agricultural and sustainable food production.

References

Abbot, J., & Marohasy, J. (2012). Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences*, *29*(4), 717–730. https://doi.org/10.1007/s00376-012-1259-9

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., & Hain, C. R. (2015). Remote sensing of drought: Progress, challenges and opportunities. In *Reviews of Geophysics* (Vol. 53, Issue 2, pp. 452–480). Blackwell Publishing Ltd. https://doi.org/10.1002/2014RG000456

Albergel, C., Dutra, E., Bonan, B., Zheng, Y., Munier, S., Balsamo, G., de Rosnay, P., Muñoz-Sabater, J., & Calvet, J. C. (2019). Monitoring and forecasting the impact of the 2018 summer heatwave on vegetation. *Remote Sensing*, *11*(5), 520. https://doi.org/10.3390/rs11050520

Anandhi, A., Srinivas, V. V., Nanjundiah, R. S., & Nagesh Kumar, D. (2008). Downscaling precipitation to river basin in India for IPCC SRES scenarios using support vector machine. *International Journal of Climatology*, *28*(3), 401–420. https://doi.org/10.1002/joc.1529

Anshuka, A., van Ogtrop, F. F., & Willem Vervoort, R. (2019). Drought forecasting through statistical models using standardised precipitation index: a systematic review and meta-regression analysis. In *Natural Hazards* (Vol. 97, Issue 2, pp. 955–977). Springer Netherlands. https://doi.org/10.1007/s11069-019-03665-6

Bacanli, U. G., Firat, M., & Dikbas, F. (2009). Adaptive Neuro-Fuzzy inference system for drought forecasting. *Stochastic Environmental Research and Risk Assessment*. https://doi.org/10.1007/s00477-008-0288-5

Belayneh, A., Adamowski, J., Khalil, B., & Ozga-Zielinski, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural networks and wavelet support vector regression models. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2013.10.052

Blain, G. C. (2012). Revisão da definição probabilística de seca: Qualidades, limitações e adaptação agrometeorológica. *Bragantia*, *71*(1), 132–141. https://doi.org/10.1590/S0006-87052012000100019

Corzo, G. (2019). *Framework for Spatio-temporal multi-objective optimization of Preventive Drought Management Measures PhD Research Proposal*. *May*.

Crocetti, L., Forkel, M., Fischer, M., Jurečka, F., Grlj, A., Salentinig, A., Trnka, M., Anderson, M., Ng, W. T., Kokalj, Ž., Bucur, A., & Dorigo, W. (2020). Earth Observation for agricultural drought monitoring in the Pannonian Basin (southeastern Europe): current state and future directions. In *Regional Environmental Change* (Vol. 20, Issue 4, pp. 1–17). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s10113-020-01710-w

Dai, A. (2011). Drought under global warming: A review. In *Wiley Interdisciplinary Reviews: Climate Change*. https://doi.org/10.1002/wcc.81

Deo, R. C., & Şahin, M. (2015). Application of the Artificial Neural Network model for prediction of monthly Standardized Precipitation and Evapotranspiration Index using hydrometeorological parameters and climate indices in eastern Australia. *Atmospheric Research*, *161–162*, 65–81. https://doi.org/10.1016/j.atmosres.2015.03.018

Diaz, V., Corzo Perez, G. A., Van Lanen, H. A. J., Solomatine, D., & Varouchakis, E. A. (2020a). An approach to characterise spatio-temporal drought dynamics. *Advances in Water Resources*. https://doi.org/10.1016/j.advwatres.2020.103512

Diaz, V., Corzo Perez, G. A., Van Lanen, H. A. J., Solomatine, D., & Varouchakis, E. A. (2020b). Characterisation of the dynamics of past droughts. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2019.134588

Dikshit, A., Sarkar, R., Pradhan, B., Segoni, S., & Alamri, A. M. (2020). Rainfall Induced Landslide Studies in Indian Himalayan Region: A Critical Review. *Applied Sciences*, *10*(7), 2466. https://doi.org/10.3390/app10072466

Djerbouai, S., & Souag-Gamane, D. (2016). Drought Forecasting Using Neural Networks, Wavelet Neural Networks, and Stochastic Models: Case of the Algerois Basin in North Algeria. *Water Resources Management*. https://doi.org/10.1007/s11269-016-1298-6

Dracup, J. A., Lee, K. S., & Paulson, E. G. (1980). On the definition of droughts. *Water Resources Research*. https://doi.org/10.1029/WR016i002p00297

Eckstein, D., Hutfils, M.-L., & Winges, M. (2017). *Germanwatch*. www.germanwatch.org

Fallah, A., Rakhshandehroo, G. R., Berg, P., O, S., & Orth, R. (2020). Evaluation of precipitation datasets against local observations in southwestern Iran. *International Journal of Climatology*, *40*(9), 4102–4116. https://doi.org/10.1002/joc.6445

FAO. (2019). Global report on food crises. *Food Security Information Network*.

Frieler, K., Lange, S., Piontek, F., Reyer, C. P. O., Schewe, J., Warszawski, L., Zhao, F., Chini, L., Denvil, S., Emanuel, K., Geiger, T., Halladay, K., Hurtt, G., Mengel, M., Murakami, D., Ostberg, S., Popp, A., Riva, R., Stevanovic, M., … Yamagata, Y. (2017). Assessing the impacts of 1.5 °C global warming – simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geoscientific Model Development*, *10*(12), 4321–4345. https://doi.org/10.5194/gmd-10-4321-2017

Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal Drought Prediction: Advances, Challenges, and Future Prospects. *Reviews of Geophysics*, *56*(1), 108–141. https://doi.org/10.1002/2016RG000549

Hassan, A., & Mahmood, A. (2017). Deep Learning approach for sentiment analysis of short texts. *2017 3rd International Conference on Control, Automation and Robotics, ICCAR 2017*. https://doi.org/10.1109/ICCAR.2017.7942788

Heye, A., Venkatesan, K., & Cain, J. (2017). Precipitation Nowcasting: Leveraging Deep Recurrent Convolutional Neural Networks. *Proceedings of the Cray User Group (CUG) 2017*.

Hudson, D., Alves, O., Hendon, H. H., & Marshall, A. G. (2011). Bridging the gap between weather and seasonal forecasting: Intraseasonal forecasting for Australia. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 673–689. https://doi.org/10.1002/qj.769

Keyantash, J., & Dracup, J. A. (2002). The quantification of drought: An evaluation of drought indices. In *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/1520-0477(2002)083<1191:TQODAE>2.3.CO;2

Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. In *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2010.07.012

Morid, S., Smakhtin, V., & Bagherzadeh, K. (2007). Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology*. https://doi.org/10.1002/joc.1498

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, *18*(9), 3511–3538. https://doi.org/10.5194/hess-18-3511-2014

Murakami, H., Vecchi, G. A., Villarini, G., Delworth, T. L., Gudgel, R., Underwood, S., Yang, X., Zhang, W., & Lin, S. J. (2016). Seasonal forecasts of major hurricanes and landfalling tropical cyclones using a high-resolution GFDL coupled climate model. *Journal of Climate*. https://doi.org/10.1175/JCLI-D-16-0233.1

Okal, H. A., Ngetich, F. K., & Okeyo, J. M. (2020). Spatio-temporal characterisation of droughts using selected indices in Upper Tana River watershed, Kenya. *Scientific African*, *7*, e00275. https://doi.org/10.1016/j.sciaf.2020.e00275

Podestá, G., Skansi, M., Herrera, N., & Veiga, H. (2016). Descripción de índices para el monitoreo de sequía meteorológica implementados por el Centro Regional del Clima para el Sur de América del Sur. *Reporte Técnico CRC-SAS*.

Poornima, S., & Pushpalatha, M. (2019). Drought prediction based on SPI and SPEI with varying timescales using LSTM recurrent neural network. *Soft Computing*. https://doi.org/10.1007/s00500-019-04120-1

Sachindra, D. A., & Kanae, S. (2019). Machine learning for downscaling: the use of parallel multiple populations in genetic programming. *Stochastic Environmental Research and Risk Assessment*, *33*(8–9), 1497–1533. https://doi.org/10.1007/s00477-019-01721-y

Sheffield, J., Andreadis, K. M., Wood, E. F., & Lettenmaier, D. P. (2009). Global and continental drought in the second half of the twentieth century: Severity-area-duration analysis and temporal variability of large-scale events. *Journal of Climate*. https://doi.org/10.1175/2008JCLI2722.1

Strazzo, S., Collins, D. C., Schepen, A., Wang, Q. J., Becker, E., & Jia, L. (2019). Application of a hybrid statistical-dynamical system to seasonal prediction of north american temperature and precipitation. *Monthly Weather Review*. https://doi.org/10.1175/MWR-D-18-0156.1

Ukkola, A. M., De Kauwe, M. G., Roderick, M. L., Abramowitz, G., & Pitman, A. J. (2020). Robust Future Changes in Meteorological Drought in CMIP6 Projections Despite Uncertainty in Precipitation. *Geophysical Research Letters*, *47*(11). https://doi.org/10.1029/2020GL087820

Unidas, N. (2008). *Climate Change in Central America: Potential Impacts and Public Policy Options*. www.cepal.org/en/suscripciones

Van Loon, A. F. (2015). Hydrological drought explained. *Wiley Interdisciplinary Reviews:*

Water. https://doi.org/10.1002/wat2.1085

Verdon-Kidd, D. C., & Kiem, A. S. (2014). Synchronicity of historical dry spells in the Southern Hemisphere. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-18-2257-2014

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., & Best, M. (2011). Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century. *Journal of Hydrometeorology*, *12*(5), 823–848. https://doi.org/10.1175/2011JHM1369.1

World Meteorological Organization. (2006). Drought monitoring and early warning : concepts , progress and future challenges. *World Meteorogical Organization*.

Zee Arias, A. van der, Zee, J. van der, Meyrat, A., Poveda, C., & Picado, L. (2012). *Estudio de caracterización del Corredor Seco Centroamericano*. 70. https://reliefweb.int/sites/reliefweb.int/files/resources/tomo_i_corredor_seco.pdf

Zhang, W., Villarini, G., Vecchi, G. A., Murakami, H., & Gudgel, R. (2016). Statistical-dynamical seasonal forecast of western North Pacific and East Asia landfalling tropical cyclones using the high-resolution GFDL FLOR coupled model. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1002/2015MS000607

Appendices

Appendix A. - Research Ethics Declaration Form

As of 2019, the IHE Delft MSc Students have to sign, submit and discuss with their mentors/supervisors a 'Research Ethics Declaration Form' (this form will be provided to all MSc students in October 2019). The Academic Board will specify who will endorse or approve this declaration.

The 'Research Ethics Declaration Form' aims to encourage all IHE Delft MSc Students to reflect on the potential ethical issues in their research proposal and later in their thesis. **All MSc students need to read the** Netherlands Code of Conduct for Research Integrity 2018 before signing this declaration.

United Nations Educational, Scientific and Cultural Organization
IHE DELFT
Institute for Water Education under the auspices of UNESCO

Research Ethics Committee
IHE Delft Institute for
Water Education
E ResearchEthicsCommitee@un-ihe.org

Date: 29 March 2021
To: Karel Aldrin Sanchez Hernandez
MSc Programme: WSE-IMHI
Approval Number: IHE-RECO 2020-108

Subject: Research Ethics approval

Dear Karel Aldrin Sanchez Hernandez,

Based on your application for Ethical Approval, the Research Ethics Committee (RECO), IHE Delft RECO has been approved ethical clearance for your research proposal Machine Learning methods for characterising and tracking spatiotemporal drought events in the Central America Dry Corridor.

This approval valid until April 16, 2021. You need to notify the RECO of any modifications to your research protocol. If you do not complete your research by the specified date, you should to contact RECO to request an extension.

Please keep this letter for your records and include a copy in the final version of MSc. Thesis, together with your personal reflection. Additional information is available at https://ecampusxl.un-ihe.org/course/view.php?id=1555&section=2.

On behalf of the Research Ethics Committee, I wish you success in the completion of your research.

Yours sincerely,

Dr. Angeles Mendoza Sammet
Acting Ethics Coordinator

Copy to: Archive.

Appendix B. - Data Assimilation

a. Data Adquisition Code

Using CDS API developed by Copernicus, the data retrieval was carried out according to this structure. These codes are made to download ERA5 land, single levels, monthly averaged data in NetCDF format.

```
import cdsapi
import xarray as xr
product = 'reanalysis-era5-single-levels'

c = cdsapi.Client()

c.retrieve(
    'reanalysis-era5-land-monthly-means',
    {
        'product_type': 'monthly_averaged_reanalysis',
        'variable': [
            '2m_temperature', 'total_precipitation',
        ],
        'year': [
            '1981', '1982', '1983',
            '1984', '1985', '1986',
            '1987', '1988', '1989',
            '1990', '1991', '1992',
            '1993', '1994', '1995',
            '1996', '1997', '1998',
            '1999', '2000', '2001',
            '2002', '2003', '2004',
            '2005', '2006', '2007',
            '2008', '2009', '2010',
            '2011', '2012', '2013',
            '2014', '2015', '2016',
            '2017', '2018', '2019',
            '2020',
        ],
        'month': [
            '01', '02', '03',
            '04', '05', '06',
            '07', '08', '09',
            '10', '11', '12',
        ],
        'time': '00:00',
        'format': 'netcdf',
    },
    'download.nc')
```

```
<xarray.Dataset>
Dimensions:    (latitude: 121, longitude: 171, time: 475)
Coordinates:
  * longitude  (longitude) float32 -93.0 -92.9 -92.8 -92.7 ... -76.2 -76.1 -76.0
  * latitude   (latitude) float32 19.0 18.9 18.8 18.7 18.6 ... 7.3 7.2 7.1 7.0
  * time       (time) datetime64[ns] 1981-01-01 1981-02-01 ... 2020-07-01
Data variables:
    tp         (time, latitude, longitude) float32 ...
Attributes:
    Conventions:  CF-1.6
    history:      2020-10-25 18:29:59 GMT by grib_to_netcdf-2.16.0: /opt/ecmw...
--------------------------------------------------------------------------------
```

Appendix C. - Drought Analysis by Country

a. Climatological Patterns (Precipitation and Temperature)

Figure 50 Costa Rica

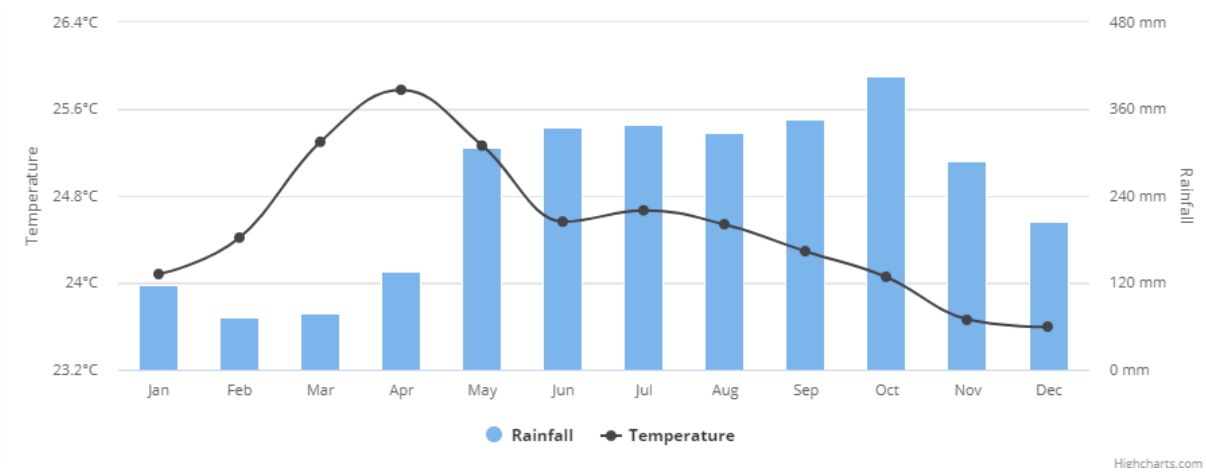


Figure 51 Panamá

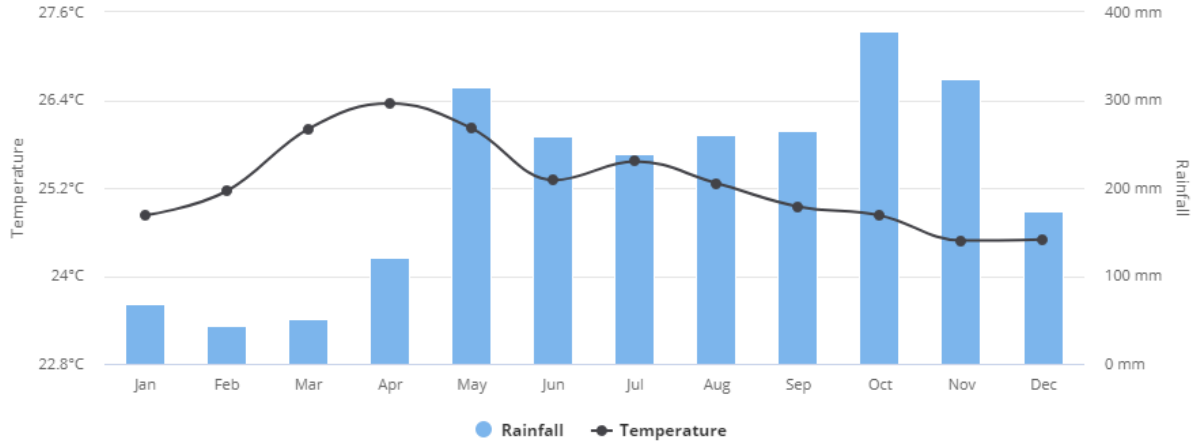


Figure 52 Nicaragua

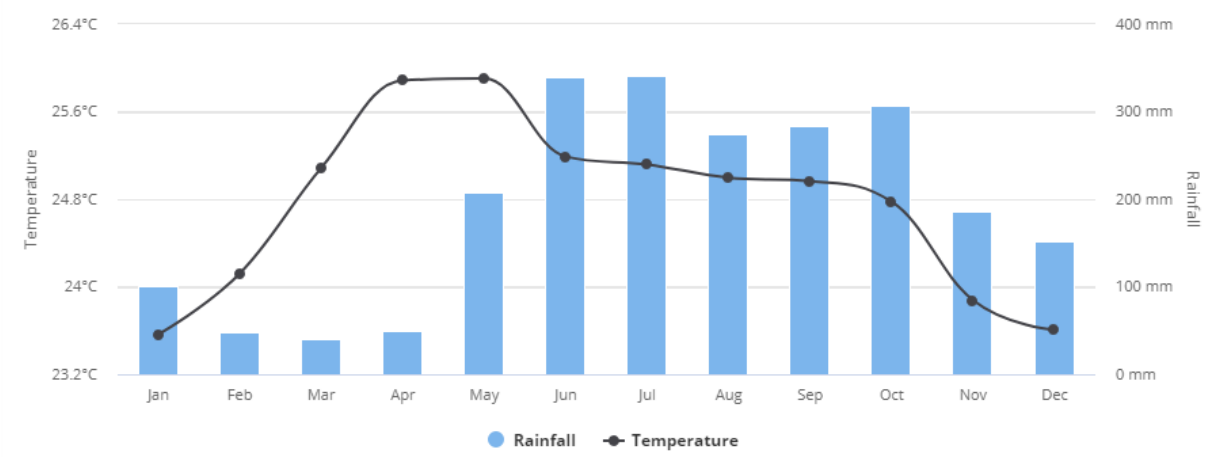


Figure 53 Honduras

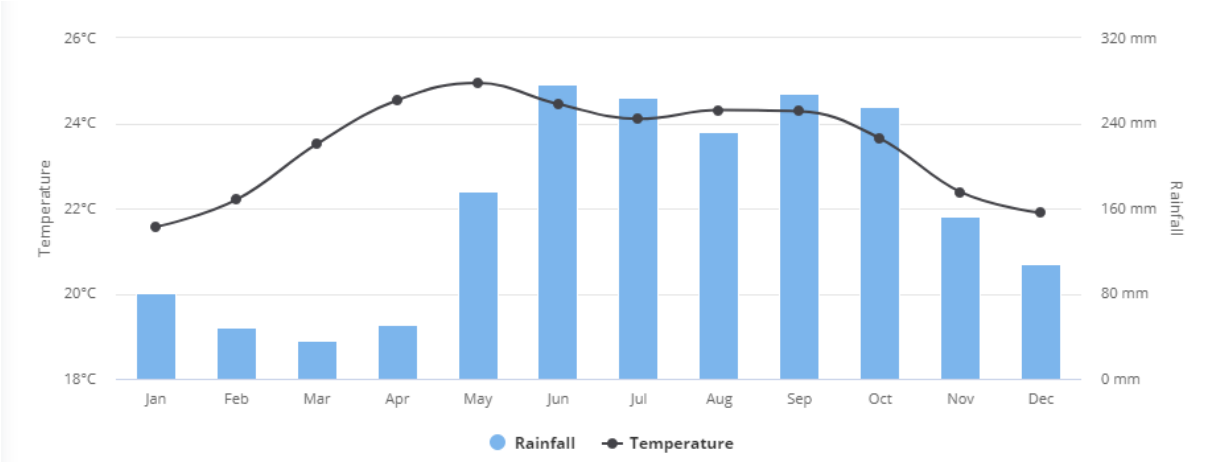


Figure 54 Guatemala

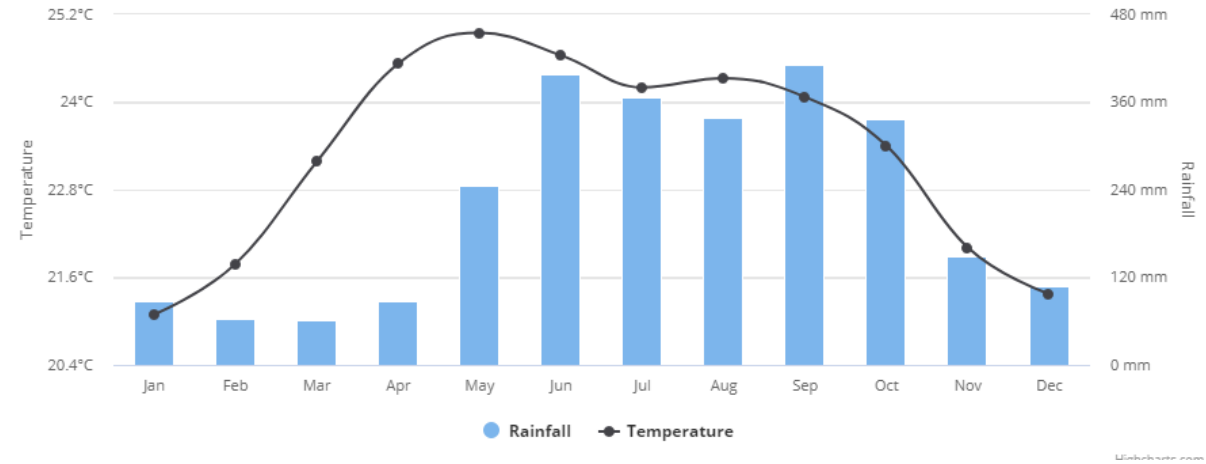


Figure 55 El Salvador

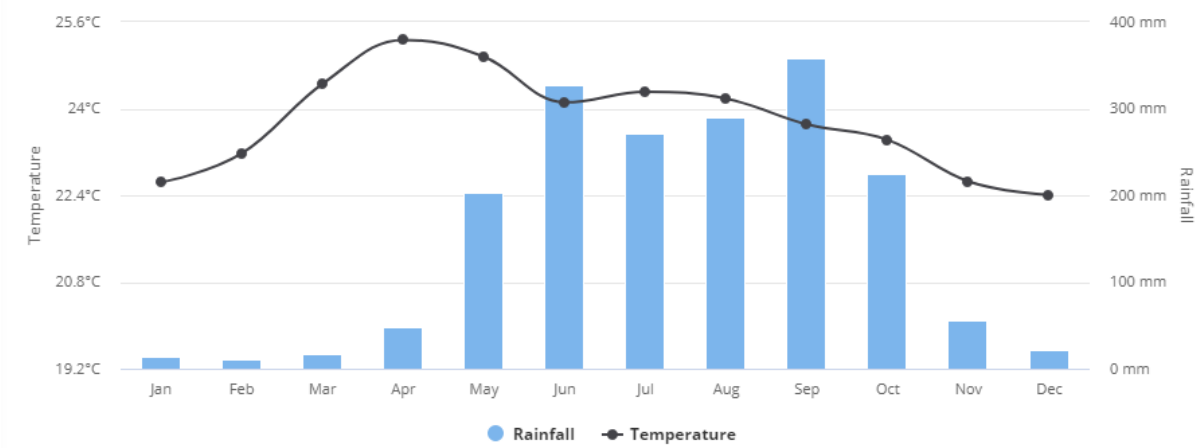
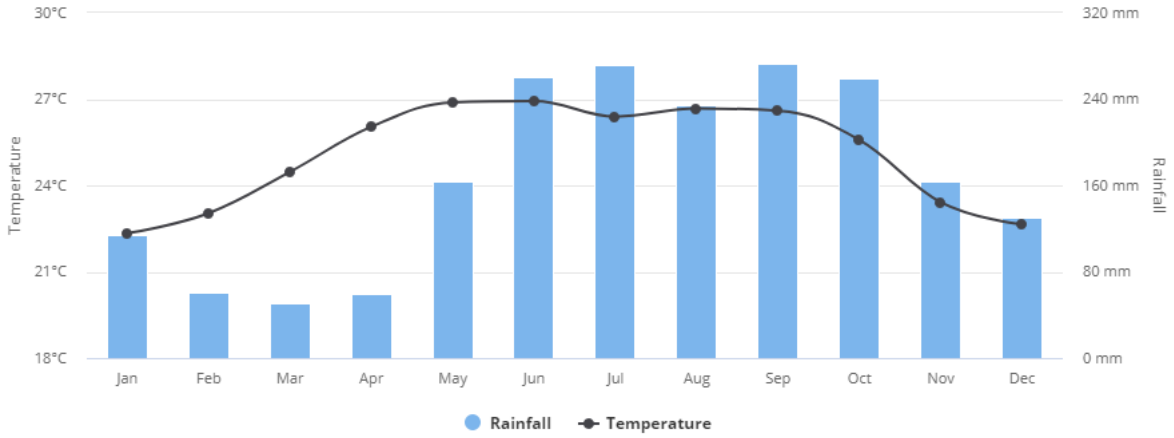


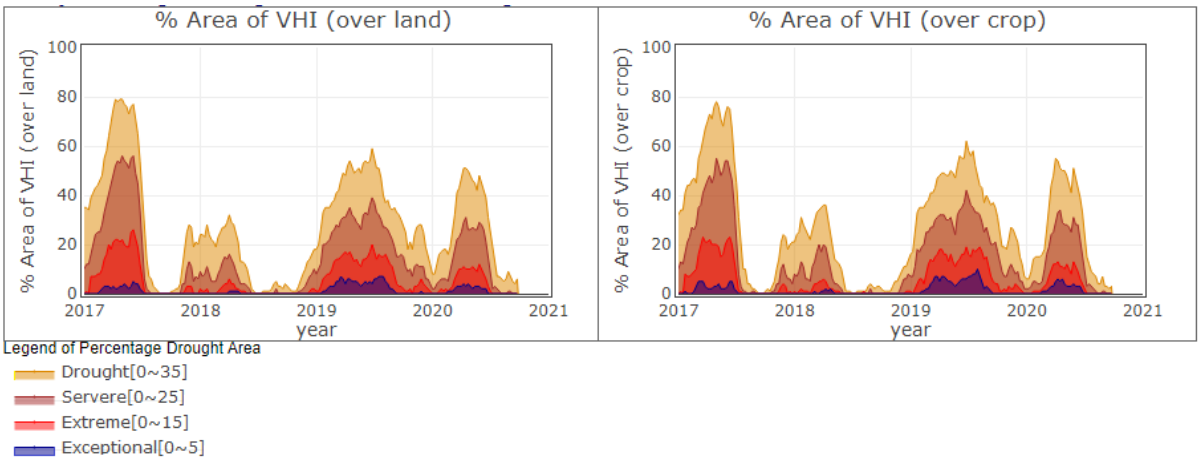
Figure 56 Belize



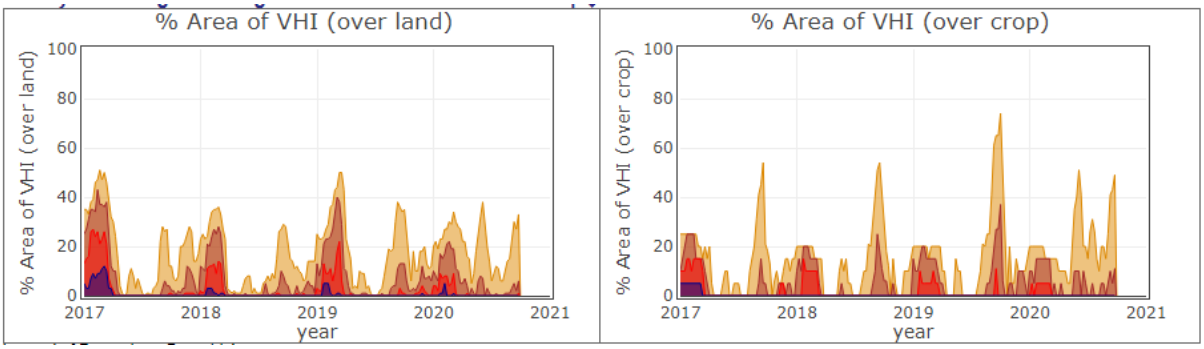
b. Drought in Vegetation Health Index

Values computed by NOAA and represent the VHI in the last 5 years, relating ENSO influence in Central America Dry Ecosystem (NOAA,2019)

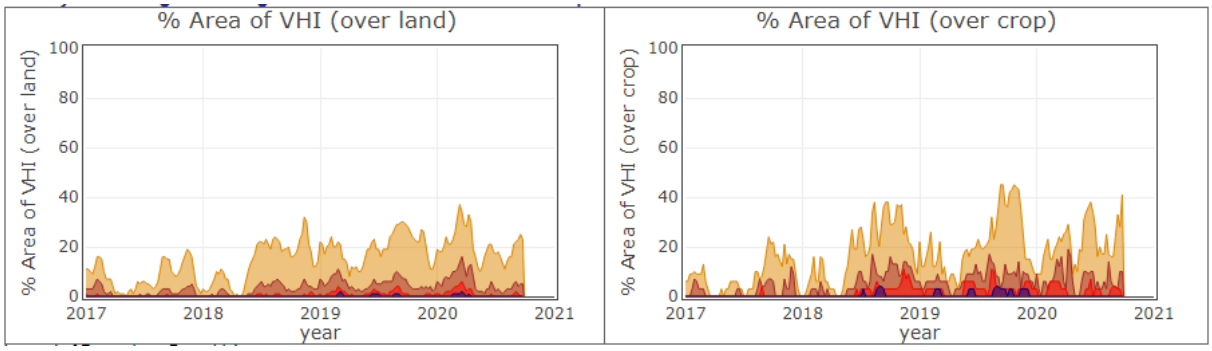
MEXICO



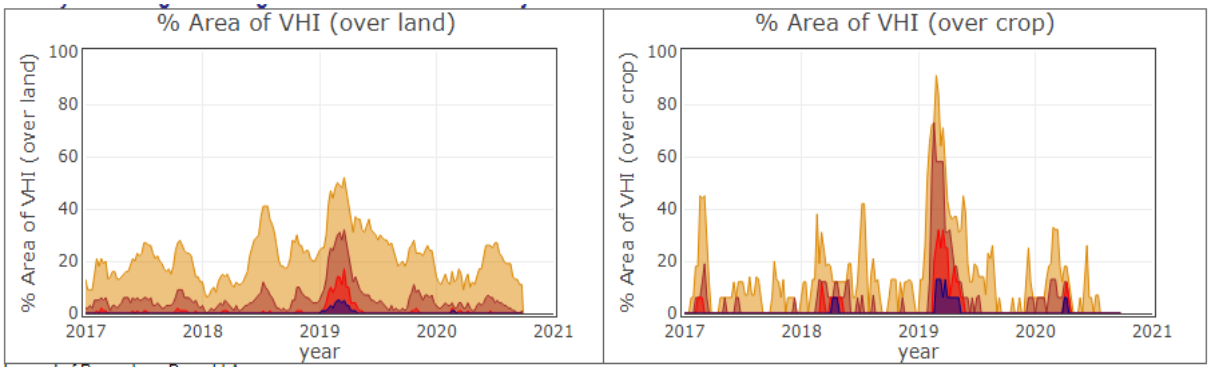
EL SALVADOR



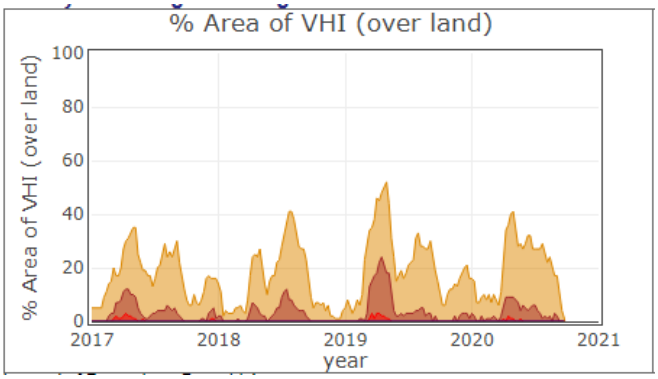
GUATEMALA



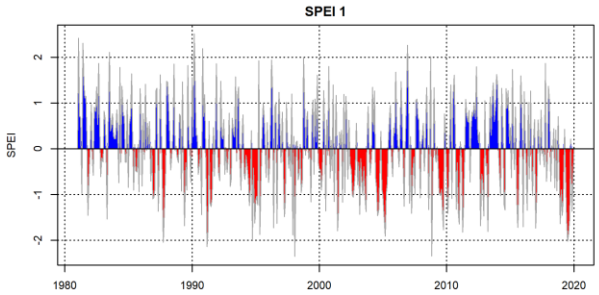
COSTA RICA

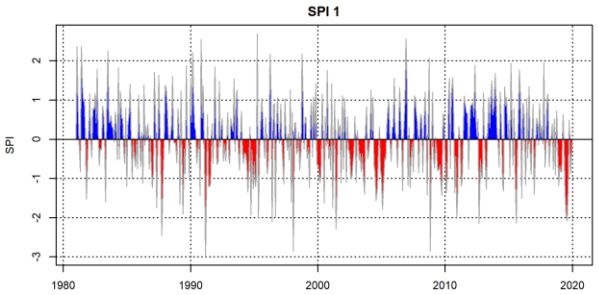


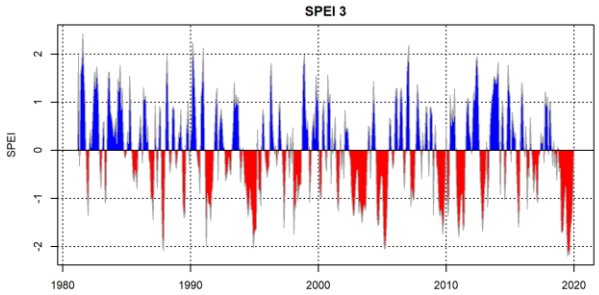
PANAMA

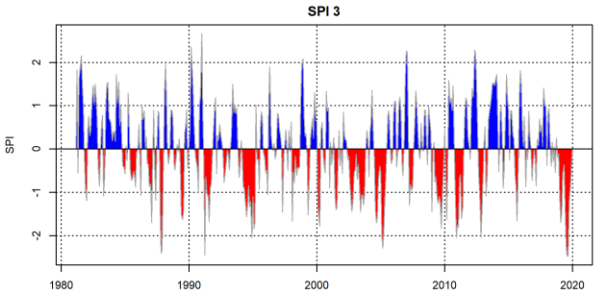


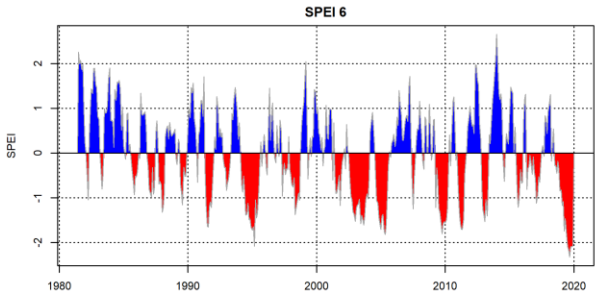
c. Drought Charts

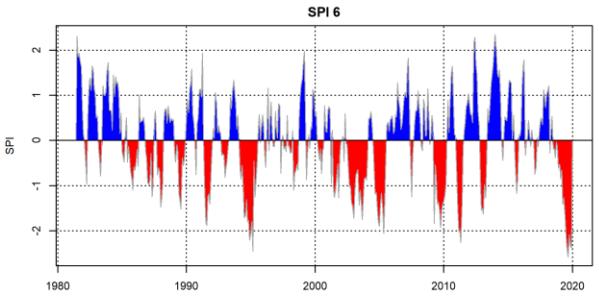


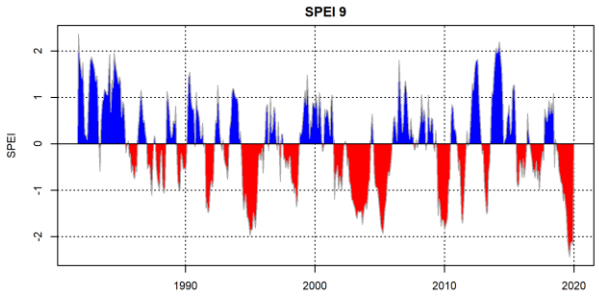


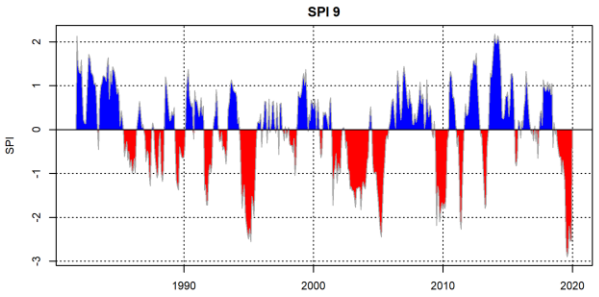


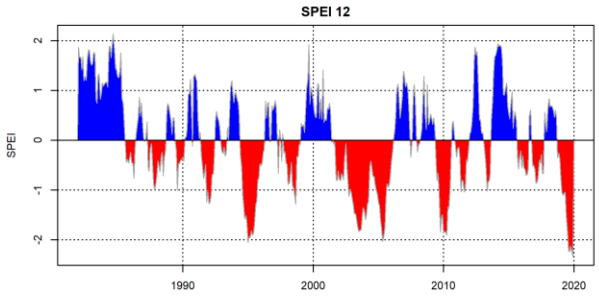


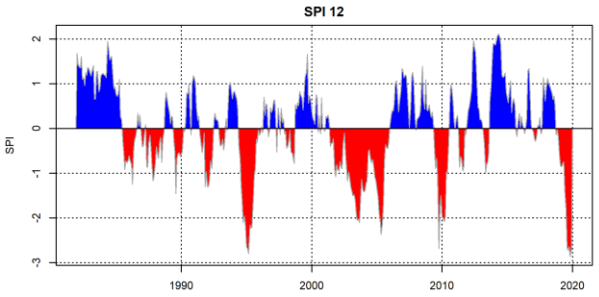








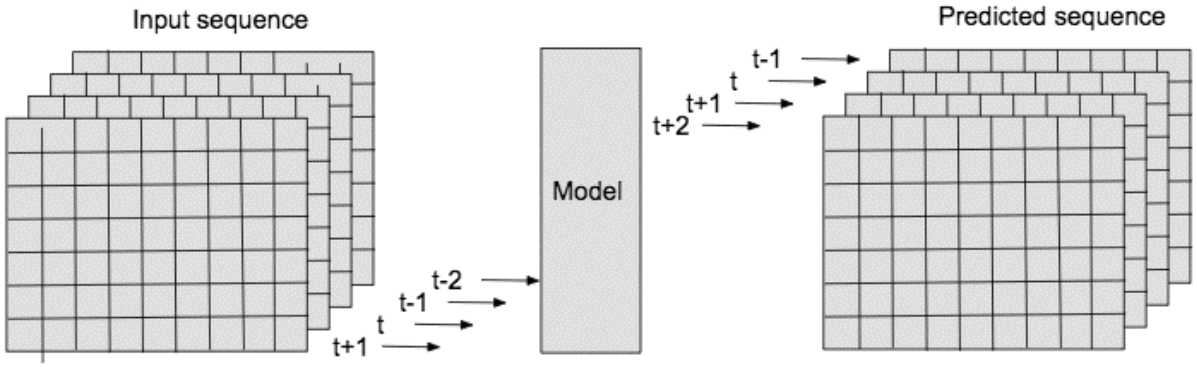


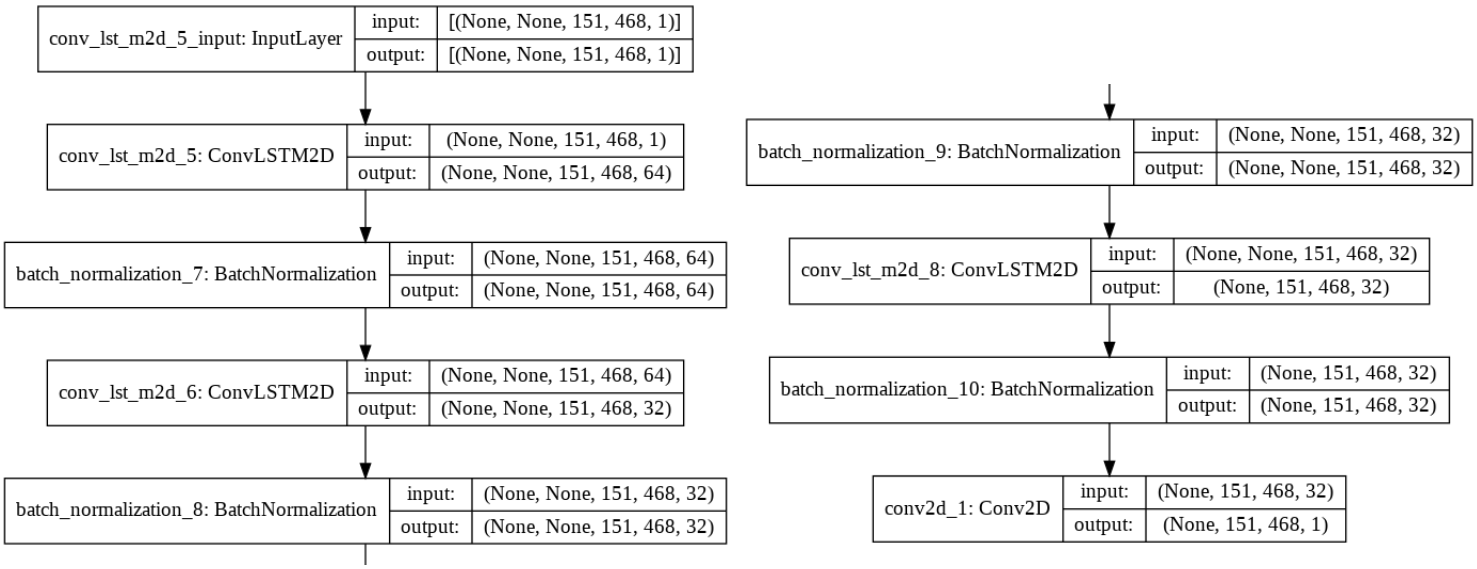


Appendix D. - Future Work

a. Advanced Drought Forecasting Models

Currently, the modelling framework for drought forecasting is being strengthened. From what has been seen in this paper, it has been necessary to generate a change of approach not to perform an individual analysis of the drought characteristics and allow an analysis against a dynamic sequence. For this reason, we are working on more robust architectures such as convolutional neural networks coupled to sequential models as shown below:





b. SpatioTemporal Drought Assessment Monitor

In order to generate and encourage this type of non-conventional analysis of drought monitoring and forecasting, the STDAM Spatiotemporal Drought Assessment Monitor platform has been proposed.This monitor's mission is to generate 1) Specific knowledge in terms of drought monitoring in any area, using reanalysis information 2) expose the spatiotemporal analysis

methodology as new techniques and tools that can generate a greater understanding of drought from the dynamic environment, thus generating more complex analysis and relationships in an easy and didactic way 3) generate a field of applied learning of the tools of artificial intelligence in the development of detailed hydrometeorological analysis.

Additionally, due to the complexity and technical requirements, this initiative will generate self-learning opportunities through tutorials and other tools that facilitate learning for any user.