POLITECNICO DI MILANO

Geoinformatics Engineering Master of Science
School of Civil Environmental and Land Management Engineering

Improving sub-seasonal drought forecasting via machine learning to leverage climate data at different spatial scales

Supervisor: Prof. Marcello Restelli

Co-Supervisors: Dr. Claudia Bertini,

Prof. Matteo Giuliani

Thesis by:
Francesco Bosso
Student ID: 965085

Academic Year   2021 / 2022

Table of Content

Abstract

Drought is one of the most dangerous natural extremes affecting society nowadays. These extremes are responsible for a yearly economic cost of about 9 billion euros in Europe and typically originate from precipitation shortfall, leading to water shortages, agricultural losses, and environmental deterioration. Even with all the work and recent improvements in weather and extreme weather event forecasting, it is still difficult to estimate rainfall reliably, especially at sub-seasonal lead times. The sub-seasonal time frame is indeed too short to be solely influenced by phenomena such as oceanic variability and, at the same time, too long to be mainly affected by the atmospheric initial conditions. Therefore, a soft contribution of both is present. Moreover, uncertainty persists regarding the relative contributions of local atmospheric conditions and climatic teleconnections to the development of total precipitation at the sub-seasonal scale. By advancing the Climate State Intelligence (CSI) framework and looking at how local atmospheric conditions and teleconnection patterns affect monthly total precipitation, we aim at addressing this lead time gap. To do so, this work is structured into three main Machine Learning (ML) model branches predicting total precipitation with a one-month lead time. The exploited ML algorithms include Extreme Learning Machines (ELM), Feed-Forward Neural Networks (FFNN), and Convolutional Neural Networks (CNN). We ultimately test the performance of our ML-based precipitation forecasts using the ECMWF Extended Range forecasts as a benchmark. The presented framework was created as part of the CLImate INTelligence (CLINT) project and used in the Netherlands' Rhine Delta region.

Abstract – Ita

La siccità è uno degli eventi estremi più pericolosi che interessa la società odierna. Questi eventi estremi sono responsabili di un impatto economico annuo di circa 9 miliardi di euro in Europa e, in genere, traggono origine dalla carenza di precipitazioni che porta a scarsità d'acqua, problemi nel settore agricolo e depauperamento ambientale. Nonostante i recenti miglioramenti nelle previsioni del tempo e degli eventi meteorologici estremi, è ancora difficile stimare le precipitazioni in modo affidabile, specialmente per *lead time* sub-stagionali. L'arco temporale sub-stagionale è infatti troppo breve per essere influenzato unicamente da fenomeni quali la variabilità oceanica e, allo stesso tempo, troppo lungo per essere influenzato principalmente dalle condizioni atmosferiche iniziali. Pertanto, è presente un tenue contributo di entrambi i fattori. Inoltre, vi è incertezza in relazione ai contributi relativi delle condizioni atmosferiche locali e delle teleconnessioni climatiche allo sviluppo delle precipitazioni totali a scala sub-stagionale. Attraverso l'estensione del Climate State Intelligence (CSI) framework e l'osservazione di come le condizioni atmosferiche locali e le teleconnessioni influenzano le precipitazioni totali mensili, miriamo a colmare questo divario presente alla scala sub-stagionale. Per fare ciò, questo lavoro è strutturato in tre diversi modelli di Machine Learning (ML) atti a prevedere la precipitazione totale con un lead time mensile. Gli algoritmi di ML sfruttati includono Extreme Learning Machines (ELM), Feed-Forward Neural Networks (FFNN) e Convolutional Neural Networks (CNN). La determinazione delle performances dei modelli addestrati rispetto allo stato dell'arte è stata eseguita mediante comparazione con "*ECMWF extended forecasts products*". Il framework presentato nasce come parte del progetto CLImate INTelligence (CLINT) ed è riferito alla regione del delta del Reno nei Paesi Bassi.

1 Introduction

In this section, an introductory overview of the problem is provided along with the main research questions addressed by the present work. Moreover, a brief introduction of the exploited methods is presented and the structure of the manuscript is introduced.

1.1 Context

With an annual economic cost of approximately 9 billion euros in Europe, droughts are one of the most severe natural disasters to society. A precipitation shortfall is typically the source of drought situations, which can then lead to water shortages, agricultural losses, and environmental deterioration. On that note, being able to forecast such events is crucial to timely prompt anticipatory drought management measures.

Drought forecasting is and has always been a challenging research area, especially for sub-seasonal lead times (i.e., 1 month) where the contribution of multiple factors is distributed across different spatiotemporal scales.

The present work is aimed at developing a Machine Learning (ML) framework made of a set of models able to perform precipitation prediction at sub-seasonal lead time. These precipitation predictions can then be translated into drought conditions thanks to the usage of drought indices such as the Standardized Precipitation Index (SPI).

1.2 Motivation and research questions

At sub-seasonal lead times, producing skillful forecasts represents a challenge for all weather-related variables. This is particularly true for precipitation that, compared with variables such as temperatures, is characterized by a less stable pattern. Moreover, forecasting precipitation is key to perform drought prediction by exploiting drought indexes such as SPI.

On top of that, in the sub-seasonal context, the considered period is short enough for the atmosphere to retain a memory of its initial conditions, but, at the same time, long enough for oceanic variability to affect atmospheric circulation. This is directly translated in the concurrent influence of multiple factors across different

spatiotemporal scales affecting the precipitation forecasting process (e.g., local variables, teleconnection patterns, global climatic variables, etc.). Therefore, variables coming from both global and local contexts should in principle be relevant to forecast precipitation at sub-seasonal lead time.

Another important aspect is the time domain on which the model is trained and is therefore able to produce predictions. For instance, ML models trained on data referred to a specific month, will be able to produce predictions only for that specific month. In the context of weather prediction this happens to be crucial because of the intrinsic meteorological variability encoded in each month due to seasonality. Moreover, models referred to shorter time domain have higher customization capabilities in terms of input variables than models referred to a wider time period. On top of that, as consequence of shrinking or extending the time domain of a model, the related dataset size decrease or increase accordingly (e.g., the dataset of a specific month will have fewer samples with respect to the complete dataset of the whole year)

Based on the described scenario, the questions addressed by this research are the following:

- Is the bridging of local and global contexts useful to increase the predictive skillfulness of the models?
- Is the provision of climatic information (e.g., climate indices of teleconnection patterns) beneficial for model performances?
- Which is the best tradeoff between model time-domain and number of training samples? (i.e., monthly-based models with few samples versus yearly-based model with many samples)

1.3 Proposed approach and original contributions

Due to the hybrid nature of the sub-seasonal lead time (i.e., variable from both local and global contexts happens to be relevant), the current study uses concurrent exploitation of global and local data to address precipitation forecasting. This concurrent exploitation is aimed at bridging global and local climate contexts thus enabling the provision of a more complete set of information to the models. Moreover, while investigating the meaningfulness of the global/local bridging process, the role of teleconnection patterns in enhancing that bridging has also been investigated.

This study is structured into three Machine Learning (ML) model branches aimed at forecasting the total precipitation in Rijnland (the study area in the Netherlands) with a one-month lead time. The exploited ML methods include Extreme Learning Machines (ELM), Feed-Forward Neural Networks (FFNN), and Convolutional Neural Networks (CNN). These three distinct branches are not conceived as a simple set of mirrored strategies. They will thus (i) utilize various sets of input features and (ii) be referred to various temporal domains. The different settings are designed to help determine which trade-off of type of input features, temporal domain, and dataset size is the most appropriate for precipitation prediction at the sub-seasonal scale. Moreover, for each branch, a different preprocessing framework derived or readapted from the existing literature is provided [1] [2] [3]. In the end, we use the ECMWF Extended Range forecasts as a benchmark to assess the quality of our ML-based precipitation forecasts.

1.4 Outline of contents

This study illustrates the design, implementation, and testing process of the three different ML model branches referred to the Rijnland case study. A literature review of the state of art is discussed in section 2. The case study and the data selection process are provided in section 3, followed by a detailed description of all the implemented approaches in section 4. The discussion of the obtained results is reviewed in section 5, followed by conclusions in section 6.

2 State of art

This part provides a review of the relevant literature as well as a more general introduction to the key theoretical underpinnings of droughts and weather forecasting. Starting with the fundamentals of weather and precipitation forecasting, we go on to the notion of drought before coming to the concept of teleconnection patterns.

2.1 Weather and precipitation forecasting at sub-seasonal lead-times

Weather forecasting consists in predicting atmospheric conditions over a certain location and for a certain time through the use of science and technology. The first step in forecasting weather is the gathering of observations from various sources, such as satellite and ground-based sensors, which provide information on the state of the atmosphere. This data can be projected into the future through an appropriate set of methodologies and/or models, providing a simulation of the evolution of weather conditions.

In the context of weather forecasting, the length of time between the issuance of a forecast and the occurrence of the predicted phenomena is called "lead time" [4]. In other terms, the lead time it is how far into the future the model is going to generate a forecast. Over the scientific literature, it is possible to identify different ways of classifying forecasting based on lead times. One of the possible classifications divides the forecasts into seven main categories [5]: nowcast, a term used to reference a forecast up to the next six hours; short range forecast, category of forecasts which are typically delivered for lead times in a range of one up to three days; medium range forecast, going from 4 up to 10 days; extended range forecast, also called sub-seasonal forecasts, characterized by a lead time ranging between two weeks and a month; seasonal range forecast, with lead time up to three months; long range forecast, characterized by wider lead times reaching the temporal scale of a year; decadal projections, with a lead time that can reach 15 years.

According to the desired lead time, different techniques could be exploited to produce weather predictions [6]:

- **Persistence forecast**: assumes that the current weather conditions are going to be constant in time, so that the future weather will be equal to the present. This assumption is very basic and works for lead times of a few minutes;

- **Extrapolative forecast or trend forecast**: assumes the presence of a changing pattern that is constant in time (e.g., a cold front motion will be assumed as constant in speed and direction). This assumption is more aligned with reality but is still working well only for lead times ranging from minutes to hours;

- **Numerical Weather Prediction (NWP)-based forecast**: it is by far the most used method for weather prediction nowadays. It is aimed at predicting future atmospheric conditions through a model that solves dynamics and physics equations designed to model the evolution of the atmosphere. These equations have a high degree of complexity, thus NWP models typically run on supercomputers able to handle such a high amount of computational needs. They can run either singularly, in the so-called Deterministic forecast, or as a multi-threaded setting, in the so-called Ensemble forecast. Ensemble settings typically improve long-term forecasting skills by initializing the NWP threads with a different set of initial conditions. This process allows for the mitigation of the endogenous uncertainty encoded in the initial conditions. The structure of the NWPs model is more adherent to reality compared to both persistence and extrapolative forecast and it works well for lead times from a few hours to a couple of weeks;

- **Climatological forecast**: assumes that the weather state will fall into average conditions statistically derived from several years of observations, i.e. the climatology. The forecast is going to be uniquely based on climatological statistics instead of a dynamic evolution of current atmospheric conditions. This type of forecasting method starts gaining significance with lead times from several weeks up to several months.

Analyzing the different forecasting methods, it is clear how lead time plays a fundamental role in weather forecasting. Persistence and extrapolative forecasts are well-performing if the goal is to predict precipitation at lead times of a few hours, Numerical Weather Prediction models are primarily used if the lead time is in the range of several days and climatological forecasts are very useful for longer lead

times up to several months but, at the same time, almost irrelevant for very short lead times.

Short-medium and Seasonal-long range forecasting are characterized by different sources of predictability. The high accuracy level of short-medium ranges forecasting models (around 80% [5]), primarily comes from: the rising quality of observations data, especially satellite-based, characterizing the last decades, the improved performance of NWP models, and, most importantly, the strong dependency between atmospheric circulation and initial state of the atmosphere at these time scales. On the other hand, seasonal-long range forecasts are primarily governed by large-scale, slowly varying features such as ocean variability and teleconnection patterns. In other terms, weather forecasts typically cover a time
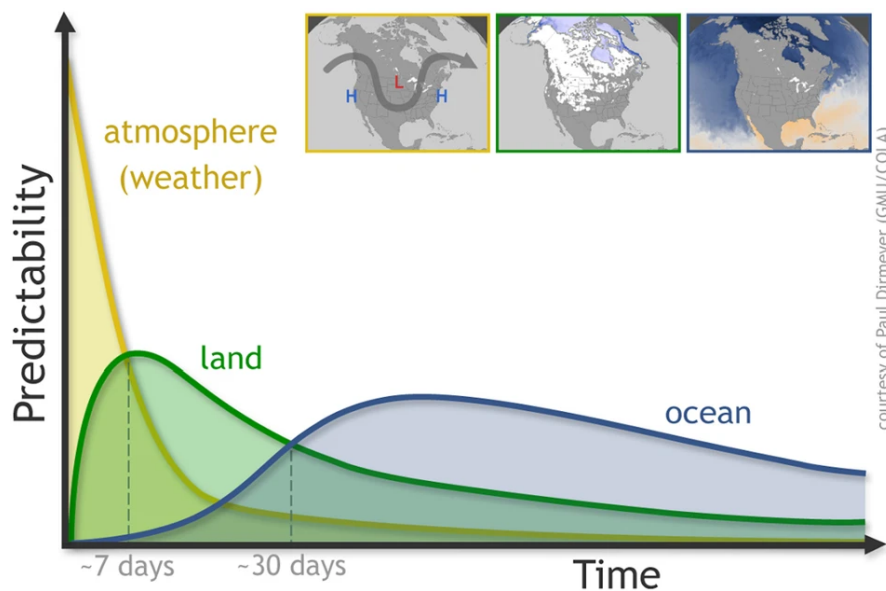
Figure 1: role of different earth system components at different lead time. [5]

interval extending up to 2 weeks, while climate forecasts start at 3-month time scales and go further in the future. The difference in the sources of predictability is reflected in the models used for weather and climate forecasting. Indeed, the traditional NWP systems used to perform short-medium term weather forecasting exploit high-resolution complex representations of the processes taking place in the atmosphere and are fed with the atmospheric initial conditions. In this case, the atmospheric dynamics can be skillfully predicted up to a limit of two weeks, as initial state-related atmospheric information is progressively lost [7].

On the other hand, the traditional models used for climate forecasting (Global Circulation Models) are usually designed and defined as "*coupled*", as they leverage the interaction between the atmosphere and the ocean. The exploitation of

this coupled interaction results in higher computational costs than NWP models, which is directly translated into the need for a reduction in data resolution [8]. Similarly, to reduce the computational needs, these "*coupled*" models rely on a less detailed representation of the atmospheric process with respect to classical NWP.

In between these two categories, at the so-called sub-seasonal scale, the lead time is short enough for the atmosphere to still have memory of its initial conditions, but it is also long enough for the ocean variability to start playing a role in the atmospheric circulation, as it can be seen from Figure 1. In this context, strictly atmospherically based weather models from one side, and climate low-resolution coupled models from the other, make weather and climate prediction not directly applicable to sub-seasonal lead time [8]. Hence, a weather-climate prediction gap originates [8].

The fundamental approach needed to produce skillful weather forecasts at sub-seasonal lead times is to generate a blending system able to exploit both atmospheric and climatic information. This system should be based on the analysis of the interaction between the atmosphere and slowly varying elements such as land and ocean [8]. In the last decades, different operational centers have tried to improve weather forecasts for the sub-seasonal lead times. The European Centre for Medium-Range Weather Forecasts (ECMWF), for instance, started in 2002 with an ensemble model of monthly forecasts. Within 10 years, the performance of this model in predicting 2-meter temperature anomalies was made comparable to the one with 15/25 days of lead time in 2002 [9]. Experiments highlight how the majority of the improvements in the model during these 10 years were strictly related to the fine-tuning of model physic parametrization and, therefore, how the improvements in medium-range forecasts have a direct impact in improving forecasting skills at sub-seasonal scale as well [9].

Starting from 2013, in order to concentrate the efforts and better study sub-seasonal lead times, the "sub-seasonal to seasonal prediction project" has been set up by a collaborative structure created by the World Weather Research Program (WWRP) and World Climate Research Program (WCRP). The main goal of the project was to bring together experts in the weather and climate fields to leverage their knowledge to improve forecasting skills and understanding of the sub-seasonal to seasonal (S2S) lead time scale [10]. This project was also characterized by a further declination towards extreme events such as tropical cyclones, droughts, and floods.

Another main aim of this S2S initiative was to compare sub-seasonal forecasting models of different operational centers to check for similarities or dissimilarities in skills improvement trends. Since the typical model structure is characterized by several ensemble members, and the original name of the S2S temporal range is extended-range, these models are formally called extended-range Ensemble Prediction Systems (EPSs).

S2S	Real time					Reforecast				
	forecast length	model resolution	ens. size	forecast freq.	data available period	system (model ver.)	ens. size	reforecast frequency	reforecast period	orig data grid
BoM (Australia)	D1 – 62 (00UTC)	T47L17	33	Sun Thu	2015.01.01 -	fixed (2014.01.01)	33	1st, 6th, 11th, 16th, 21st, & 26th of each month	1981 - 2013	T47 (144x72)
CMA (China)	D0 – 60 (00UTC)	A: T106L40 B: T266L56	4	A: daily B: Mon, Thu	A: 2015.01.01 - B: 2019.11.11 -	A: fixed (2014.05.01-) B: on the fly (2019.11.11 -)	4	A: daily B: Mon, Thu	A: 1994 - 2014 B: previous 15yrs	1.5x1.5
ECCC (Canada)	D1 – 32 (00UTC)	A: 0.45° x 0.45° L40 B&C: 0.35° x 0.35° L45 (A&B: uncoupled)	21	Thu	A: 2016.01.07 - B: 2018.09.27 - C: 2019.07.04 - D: 2021.12.09 -	on the fly	4	Thu	A: 1995 - 2014 B&C: 1998 - 2017 D: 2001 - 2020	1.5x1.5
ECMWF (Europe)	A: D0 – 32 B&C: D0 – 46 (00UTC)	A&B: TL639L91(≤D10) TL319L91(>D10) C: Tco639L91(≤D10) Tco319L91(>D10)	51	A: Thu B&C: Mon, Thu	A: 2015.01.01 - B: 2015.05.14 - C: 2016.03.08 -	on the fly A: 2015.01.01 - B: 2015.05.14 -	A: 5 B: 11	A: Thu B: Mon, Thu	previous 20yrs (e.g. 1999-2018)	1.5x1.5
HMCR (Russia)	D0 – 61 (00UTC)	1.125° x 1.40625° L28 (uncoupled)	20	A: Wed B: Thu	A: 2015.01.07 - B: 2017.06.08 -	on the fly A: 2015.01.07-, B: 2017.06.08- C: 2021.07.01-	10	A: Wed B&C: Thu	A&B: 1985 - 2010 C: 1990 - 2015	1.5x1.5
ISAC-CNR (Italy)	A&B: D0 – 31 C: D0 – 32 (00UTC)	0.75° x 0.56° L54 (a 'slab' ocean)	41	A: Mon B&C: Thu	A: 2015.11.09 - B: 2017.01.19 - C: 2017.06.08 -	fixed (2017.06.08)	5	every 5 days	1981 - 2010	1.5x1.5
JMA (Japan)	A: D0.5 – 33.5 B: D0.5 - 32.5 (12UTC)	A: TL319L60 B: TL479L100 (≤D18), TL319L100 (>D18) (uncoupled)	A: 25 B: 50	A: Tue, Wed B: Wed*	A: 2015.01.06 - 2017.03.15 B: 2017.03.22 -	fixed A: 2014.03.04-, B: 2017.01.31- C: 2020.03.31-, D: 2021.03.31-	A&B: 5 C&D: 13	A&B: 10th, 20th, & the last date of each month C&D: 15th & the last date of each month	A: 1981 - 2010 B: 1981 - 2012 C: 1981-2010 D: 1981 - 2020	1.5x1.5
KMA (Korea)	D0 – 60 (00UTC)	N216 (0.83° x 0.56°) L85	4	daily	2016.11.01 -	on the fly A: 2016.11.01- B: 2020.09.01-	3	1st, 9th, 17th, & 25th of each month	A: 1991 - 2010 B: 1991 - 2016	1.5x1.5
Met. France (France)	A. D0 – 61 B. D0 – 32 C. D0 – 47 (00UTC)	A&B: TL255L91 C: TL359L91	A&B: 51 C: 25	A. monthly (1st) B&C. Thu	A. 2015.05.01 - B. 2016.03.03 - C. 2020.10.22 -	fixed (A&B 2014.12.01 C: 2019.07.01)	A&B: 15 C: 10	A&B: 1st, 8th,15th, & 22nd of each month C: Thu	A&B: 1993 - 2014 C: 1992 - 2017	1.5x1.5
NCEP (US)	D0 – 44 (00UTC)	T126L64	16	daily	2015.01.01-	fixed (2011.03.01)	4	daily	1999 - 2010	1.5x1.5
UKMO (UK)	D0 – 60 (00UTC)	N216 (0.83° x 0.56°) L85	4	daily	2015.12.01 -	on the fly A: 2016.01.01 - , B: 2016.04.17 - C: 2017.03.25 - , D: 2018.09.01 -	A&B: 3 C&D: 7	1st, 9th, 17th, & 25th of each month	A: 1996 - 2009 B&C: 1993 - 2015 D: 1993 - 2016	1.5x1.5

Figure 2: details about available EPSs in the S2S project from different operational centers (updated in February 2022) [11]

Sub-seasonal lead time forecasting is today an active field of research and, although several efforts have already been done, because of the overlapping of different earth systems agents in this temporal range, the relative contribution of climate and atmospheric elements to the genesis of sub-seasonal forecasts remains unclear. Sub-seasonal forecasting is generally difficult for all weather-related variables, but it is more difficult for precipitation forecasting because it is generally less stable than variables such as temperatures. Moreover, the influence of ocean variability and teleconnection patterns on precipitation phenomena at this time scale is not clear. Thus, these aspects have to be further investigated.

2.2 Droughts

2.2.1 Background on droughts

Drought events are recognized as one of the most severe environmental issues and can be in the most generic way associated with a lack of water for an extended time period over a specific region. These phenomena could last for months or even years and are strictly related to variables such as temperatures, winds, low relative humidity, and rain amount during crop seasons [12]. The concept of drought is often confused with the concept of heat wave or aridity, although the three phenomena are quite different from each other. The main difference between drought and aridity is that a drought is a temporary event that could also last for a long time period but not forever, while aridity is a permanent region-specific feature. In other words, drought is a phenomenon that could happen, aridity is an endogenous condition that affects a specific region [13]. Also, in the case of the difference between drought and heat waves, the main feature to focus on is time. Indeed, if from one side droughts, as mentioned, last for months or even years, the time range of heat waves is on average in the order of a week [14].

The reason why drought as a phenomenon is becoming more and more relevant in the last decades is that its occurrence, in parallel with the occurrence of floods, is becoming extremely frequent. Indeed, relevant drought events have recently affected almost every continent [15]. On top of their increased frequency, droughts are among the most slowly developing and longly lasting extreme events and probably the least predictable [12]. The low predictability and the potential high-level impact that droughts can cause on the economy in general and on key sectors such as agriculture, lead to a boost in the research areas related to their assessment. Drought development mechanisms are very complex, especially because the atmosphere is not the only involved factor. What really acts as a proxy variable for droughts onsets is the interactions between the atmosphere and the hydrological processes providing it with moisture [12]. From a high-level perspective, it is possible to describe the mechanism of drought onset as follows: dry hydrologic conditions start, and as a consequence, the moisture content of upper soil layers decreases lowering the evapotranspiration rates and, in turn, lowering the relative humidity of the atmosphere. Low atmospheric relative humidity corresponds to a lower rainfall probability since it becomes more difficult to meet saturation

conditions [12]. The decrease in rainfall probability leads to an increase in the dry conditions, increasing the drought intensity and duration.

There are many reasons why droughts are different from other extreme events: firstly, drought phenomenon, in the pool of all the extreme events, is first for the number of affected people [16]; it is difficult to pinpoint when a drought starts or ends, due to the fact that its impact is typically slowly incremental and could last for years; there is no single definition of drought; unlike extreme events such as floods, earthquakes, tornados, and hurricanes, droughts impacts are defined as "*non-structural*" since they hardly cause damage to infrastructures and buildings; in contrast to the other extreme events, droughts can be potentially directly triggered by human actions, such for instance when over farming, when deforesting, when over-exploiting water [12].

Droughts are extremely complex phenomena and, in order to be able to deal with them, the scientific community needs some well-defined guidelines for their definition and for the estimation of their impacts. For this purpose, different definitions of droughts and different indices have been developed over the years. A list of the different definitions and indices is presented below.

2.2.2 Drought types

As previously mentioned, one of the issues related to drought monitoring and forecasting is the absence of a unique definition. Several authoritative sources define droughts in different ways: the World Meteorological Organization (WMO), for instance, defines droughts as a "*prolonged dry period in the natural climate cycle that can occur anywhere in the world. It is a slow on-set phenomenon caused by a lack of rainfall*" [17]; the National Integrated Drought Information System (NIDIS) states that drought is "*a deficiency of precipitation over an extended period of time (usually a season or more), resulting in a water shortage*" [18]. It is possible to identify four main drought categories [12]:

- **Meteorological drought**: this term defines a condition of scarcity in precipitation over a specific region and in a specific time period. Different precipitation data can be used to study this kind of drought. Some studies, for instance, analyze meteorological droughts by the difference in precipitation with respect to the regional average-monthly values [19]; other

studies, instead, exploit cumulative precipitation data aggregated over a specific period [20]. It is also important to remember that the identification of a meteorological drought must be thought of as a context-specific task. Indeed, the atmospheric-related conditions that bring a precipitation deficit over a specific place are widely different according to the geographical region;

- **Agricultural drought**: it is related to a period in which the soil moisture decreases, thus causing side effects such as crop failures. If a crop is used to receive a specific amount of water from the soil and this amount heavily decreases, the plant could be seriously affected;

- **Hydrological drought**: this term is usually referred to a condition in which the amount of surface water (surface runoff, streamflow, reservoir level) or groundwater available is not enough to satisfy water demand. As a consequence, this low availability does not allow the water resources management system to work properly and satisfy all the demands [12]. Hydrological droughts are typically lagged in time with respect to meteorological, because of the temporal offset between the occurrence of precipitation deficit and the actual repercussions over the hydrological system (streamflow, reservoir, etc.);

- **Socio-economic drought**: this concept is generated by conceiving water as a general-purpose good. Perfect market conditions are met whenever the supply of a good meets its demand. A good is said to be rare if the market asks for it more than its actual available quantity. The same concepts apply to socio-economic droughts: whenever a drought event takes place, water resources systems could experience difficulties in meeting the water demand, thus causing socio-economic impacts.

These different drought types are linked together such that the occurrence of one of them could, in turn, generate the onset of another one (for instance meteorological drought can generate hydrological drought). Moreover, the genesis of these

phenomena is strictly related to several climate and meteorological variables. A schematic of relationships and causalities among the different drought categories is provided in Figure 3 along with the relevant involved variables.
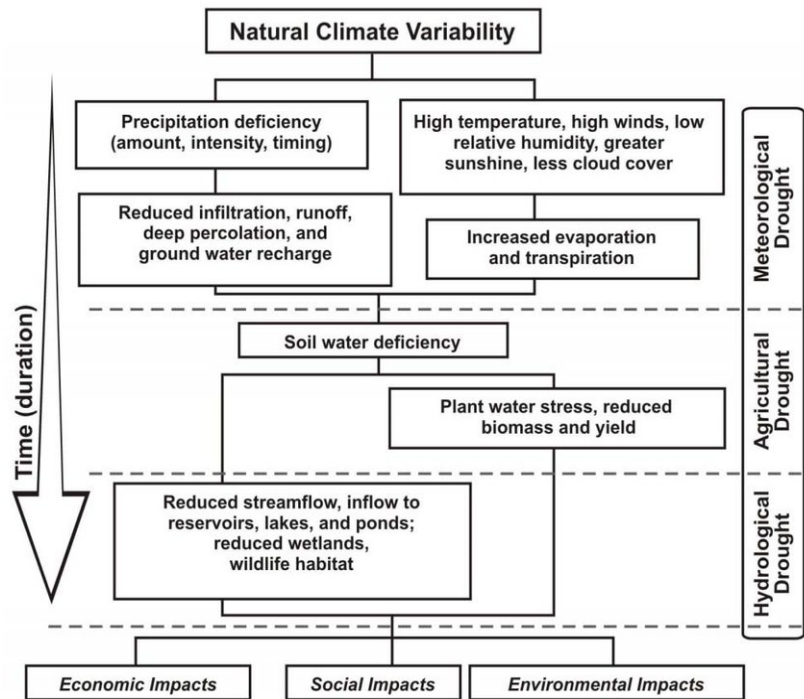

Figure 3: drought categories and sequence of its occurrence and impacts (NDMC) [73]

2.2.3 Drought indices

Over the last decades, several drought indices have been developed to define the intensity, duration, and spatial extent of drought events and to assess their effects. Because of the lack of a unique drought definition, several drought indices have been developed to help drought assessment, each one tailored for a specific aspect of the event. These indices involve individual or multiple hydroclimatic variables and are defined for different time scales, ranging from the monthly to the yearly [21].

For the case of meteorological drought, one of the most commonly employed indices is the Standardized Precipitation Index (SPI). It was developed by T.B McKee, N.J. Doesken, and J. Kleist in 1993 and it is computed based on precipitation data solely. The SPI is computed by (i) fitting a probability distribution to a long-term precipitation timeseries of a specific location, and (ii) transforming

it into a normal distribution such that the mean SPI is zero. The strength of SPI is its applicability at a variety of different time scales allowing, with the same index, the monitoring of short-term water-related phenomena (e.g., soil moisture monitoring, particularly important for agricultural purposes) and long-term water-related phenomena (e.g., reservoir levels). Across the literature, it is possible to find a vast variety of applications of this index to study drought phenomena in different ways, ranging from spatiotemporal analysis and climate impact assessment up to forecasting [12]. Based on the value of SPI, different drought intensity classes are defined to assess the severity of the event. It is worth noting that the SPI is used to assess in general the conditions of precipitation of a specific location, ranging from wet to normal to dry. The classes are presented in Table 1.

Table 1: SPI classes [22]

SPI	SPI class
$SPI \geq 2.00$	Extremely wet
$1.50 \leq SPI < 2.00$	Severely wet
$1.00 \leq SPI < 1.50$	Moderately wet
$0.00 \leq SPI < 1.00$	Mildly wet
$-1.00 \leq SPI < 0.00$	Mild drought
$-1.50 \leq SPI < -1.00$	Moderate drought
$-2.00 \leq SPI < -1.50$	Severe drought
$SPI < -2.00$	Extreme drought

Another very well-known drought index is the Palmer Drought Severity Index (PDSI), which uses temperature and precipitation data to estimate relative dryness [23]. The values of the index are standardized and typically represented in a range of $-4 \leq PDSI \leq 4$, although the real values could fall into a wider interval ranging from -10 up to +10. This index, as well as SPI, is employed for a variety of different applications such as the investigation of spatiotemporal drought characteristics, exploration of drought's periodic behavior, and drought forecasting [12]. Although the PDSI is computed using meteorological variables, it is particularly suitable for monitoring agricultural droughts, since it was aimed at estimating soil moisture

[12]. Table 2 shows the different drought classes that can be identified according to the obtained PDSI value.

Table 2: Classifications of Palmer Drought Severity Index [24].

PDSI	PDSI class
PDSI $\geq$ 4.00	Extremely wet
3.00 $\leq$ PDSI < 4.00	Very wet
2.00 $\leq$ PDSI < 3.00	Moderately wet
1.00 $\leq$ PDSI < 2.00	Slightly wet
0.50 $\leq$ PDSI < 1.00	Incipient wet spell
- 0.50 $\leq$ PDSI < 0.50	Near normal
- 1.00 $\leq$ PDSI < - 0.50	Incipient dry spell
-2.00 $\leq$ PDSI < -1.00	Mild drought
-3.00 $\leq$ PDSI < -2.00	Moderate drought
-4.00 $\leq$ PDSI < -3.00	Severe drought
PDSI < -4.00	Extreme drought

Still related to agricultural droughts, another common indicator developed it is the so-called Crop Moisture Index (CMI) [25]. Enforcing the concept that a high number of indices exist because of the necessity of verticalization of drought study, this index is specifically suited for evaluating short-term moisture conditions for crop monitoring on a weekly basis exploiting precipitation and temperature data to compute a moisture budget. The reason why this index is applicable to weekly time scales is that, due to its rapid response in adapting to short-term changes, applying this index to long terms scenarios could bring misleading information. Table 3 shows the different drought classes that can be identified according to the obtained CMI value.

Table 3: Classifications of Crop Moisture Index [26].

CMI	CMI class
CMI ≥ 3.00	Excessively wet
2.00 ≤ CMI < 3.00	Wet
1.00 ≤ CMI < 2.00	Abnormally moist
-1.00 ≤ CMI < 1.00	Slightly dry / Favorable moist
-2.00 ≤ CMI < -1.00	Abnormally dry
-3.00 ≤ CMI < -2.00	Excessively dry
CMI <-3.00	Severely dry

Although the indices presented so far only take into account precipitation and temperature data, drought indices could, in general, take as input a greater number of parameters. An example is provided by the Surface Water Supply Index (SWSI), developed starting from the PDSI drought index and particularly suited for monitoring hydrological droughts [27]. This index acts as a proxy variable for anomalies in surface water supply sources and is based on four different input variables: snowpack, streamflow, precipitation, and reservoir storage [12].

In the literature, it is possible to find a vast number of other indices, each one with specific purposes, pros, and cons (e.g., Standardized Runoff Index (SRI) [28], Soil Moisture Deficit Index (SMDI) [29]). In general, the fundamental aim of these indices is to provide a univocal and interpretable picture of drought occurrence in a specific place and over a specific time. By doing this, a drought index act as a key tool to transform raw data into meaningful information thus enabling an improved decision-making process.

2.2.4 Droughts over Europe

The impacts of droughts events around the world are evident and well-documented (e.g., for Europe [30] [31], for East Africa [32], for western Mediterranean [33]). Drought events have been spotted on almost half of the terrestrial surface, which corresponds to the same part of the earth on which the majority of the agricultural effort is spent [34]. Their occurrence has an impact on numerous facets of society

in addition to the environment. Among the numerous extreme events increasingly taking place over the last two centuries all around the world, drought has been recognized as the one having the most negative impacts [35] [36].

Across Europe, during the last 40 years, several relevant drought events have been reported and estimates indicate that, since 1991, the economic impact of droughts in Europe is, on average, about €5.3 billion every year [37].

In the period between 1975 and 1976, for instance, Europe experienced a relevant drought event. In that period, between the spring and summer seasons, a below-average precipitation regime occurred, affecting the central-western European zone. During the summer season, the drought event started spreading toward the northeast creating a uniform heavily affected zone over central Europe. Dry conditions echoed until August, when cold air coming from Scandinavia started a mitigation process [38]. During this event, a meteorological drought peak by means of SPI-6 (i.e., Standardized Precipitation Index which cumulates precipitations over 6 months) has been registered on July 27 (Figure 4) in almost every part of Europe except for both the Mediterranean zone and the coastal part of Norway.
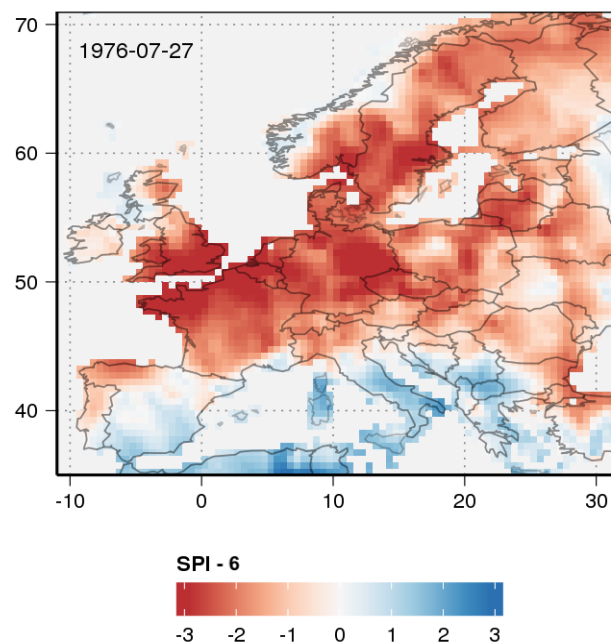

Figure 4: SPI-6 over Europe, meteorological drought peak, 1976 [35]

For what concerns the hydrological drought peak, instead, it has been registered on July 1st with the runoff across Europe below the 10th and 20th percentiles (red and orange respectively in Figure 5). The day of the hydrological drought peak is

explainable by the approximate drought onset, estimated to be around November 1975.
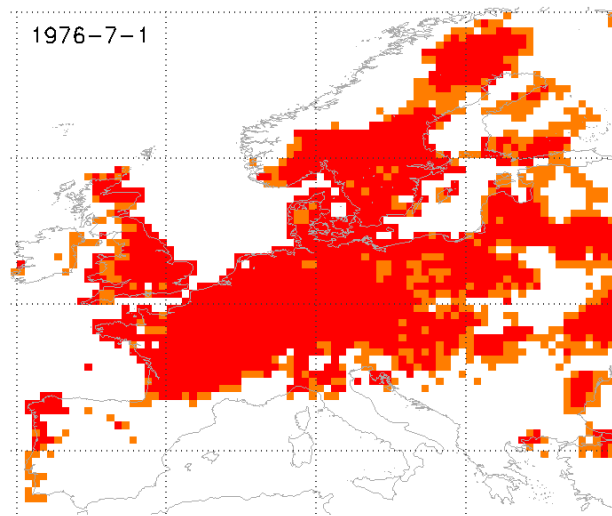
Figure 5: runoff - in red runoff values below 10th percentile, in orange below 20th percentile, 1976 [35]

Another severe drought event that affected Europe dates back to the period 1991-1995. In the time period 1991-1992, frequent drought events affected the entire Mediterranean zone, and the period 1992-1995 has been defined as one of the driest that the Iberian Peninsula has ever experienced in the century [38]. From a meteorological point of view, the whole period has been characterized by below-average precipitation conditions over southern Europe, with a focus on Spain and Portugal in 1992 and 1994. The meteorological drought peak by SPI-6 has been registered the 11/6/1993 affecting the southern European area (Figure 6)
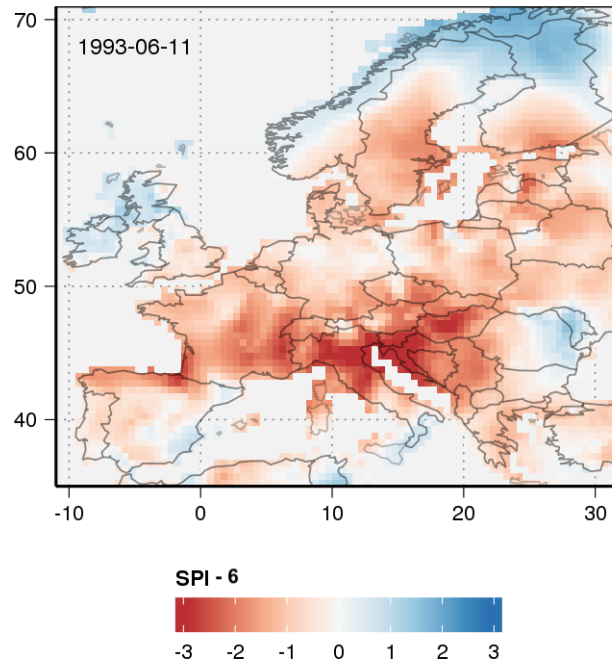
Figure 6: SPI-6 over Europe, meteorological drought peak, 1993 [27]

From the hydrological point of view instead, the peak has been registered around 5/5/1993 with a scattered configuration of runoff below the 10th and 20th percentiles (red and orange respectively in Figure 7) across central Europe.

Figure 7: runoff - in red runoff values below 10th percentile, in orange below 20th percentile, 1993 [35]

Among the several most recent events, it is possible to mention the droughts of 2018 and 2022.

Particularly, in 2018, a relatively short but intense drought event affected central and northern Europe starting from May. In that period, the persistence of anticyclonic conditions enabled the simultaneous occurrence of precipitation scarcity and above-average temperatures. These cond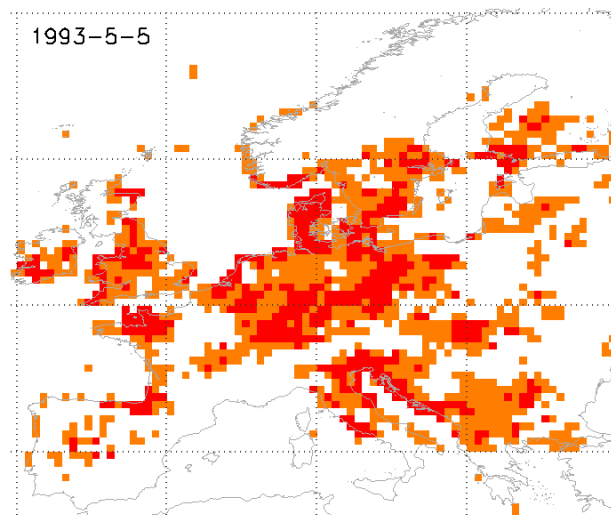itions have heavily affected the agricultural sector and have caused water supply disruptions or restrictions. On top of that, fire hazards increased in most of the European area and temperature-related phenomena, such as high water temperatures, led to a reduction in water quality and fish population [39] [40].

More recently, a severe drought event has affected Europe starting from the beginning of 2022. The blend between the prolonged lack of precipitation and the sequence of heatwaves from May 2022 was responsible for the establishment of dry conditions. After the establishment of this condition, several sectors have been affected. From the hydrological point of view, a sensible decrease in river discharges all over Europe was registered. Among the consequences linked with the hydrological side of this drought event, the reduction of the stored water volume has to be mentioned due to its severe impacts on the energy sector (hydropower generation, cooling systems of other power plants, etc.). Also, the agricultural sector has been heavily affected by a reduction in summer crop yields [41].

Drought events are becoming more frequent with time and projection analysis [42] endorses this trend also for the next years.

2.2.5 Drought forecasting

Being able to forecast drought phenomena in advance enables the possibility to mitigate the effect of their impacts. Unfortunately, due to the multiple factors that contribute to their development and evolution, forecasting droughts is a complex task. In addition, the different drought categories are very much connected to each other. Furthermore, as already discussed, droughts can be triggered by different factors (e.g. precipitation deficit, increased evapotranspiration, anthropogenic factors), have different starting modalities, different duration and spatial extent [43] [44]. Due to all these layers of complexity, drought forecasting is and has always been a challenging research area.

In order to build a framework able to forecast drought phenomena, three key components have to be defined: target, predictors, and method.

The target is the means through which it is possible to define a measure for drought conditions. It is typically a drought index (e.g., SPI), or one or multiple variables used to then compute one of the drought indices (e.g. cumulative precipitation over 6 months).

The predictor, or the predictors, is referred to the variables that are meaningful for forecasting a certain phenomenon. For forecasting in general, the identification of the appropriate set of predictors is essential for the good performance of the prediction framework developed. Since drought is a complex phenomenon chained with different atmospheric and climate components interacting with each other, the set of meaningful predictors is spread across different spatiotemporal scales. At the global scale, for instance, the Sea Surface Temperature (SST), has been recognized as one of the primary responsible for drought conditions at a large spatiotemporal scale and, for that reason, it has been widely used to improve drought forecasting capabilities [43]. Another important global-scale component that enabled a sensible improvement in drought prediction is the presence of teleconnection patterns linking hydroclimatic anomalies and global variables (e.g., SST, geopotential height, mean sea level pressure). These relationships are typically expressed in terms of climate indices and one major example of them is provided by El Niño Southern Oscillation (ENSO) [43]. Meaningful drought predictors are also present at the local scale, mainly for short-term forecasting. Several regional land surface features such as soil moisture, vegetation, and snow cover, have indeed shown an active role in enhancing or mitigating drought effects and persistency. Among the important predictors, it is also important to add external forcing factors such as variation in solar radiation and land use changes [43].

The most commonly used methods to perform drought prediction can be summarized in three main categories: statistical, dynamic, and hybrid methods.

The main idea for statistical drought prediction is to (i) identify a set of relevant predictors and (ii) pass them as input to a context-tailored statistical model. Predictors for statistical models usually come from two sources: historical observations and dynamic forecasts. Among the vast number of available predictors

coming from these sources, it is fundamental to choose the subset that explains most of the variance of the predictand. This selection process can be performed with a physical-based approach, i.e. by analyzing and understanding the physical mechanisms underlying the development of drought phenomena, or with a statistical-based approach, i.e. by applying techniques such as correlation analysis. In particular, while applying correlation analysis, is extremely important to take into account plausible mechanisms between analyzed variables. The reason why is that brute-forcing correlation analysis through a big number of predictors with no plausibility-based pre-selection could lead to false correlation [45]. Also, a set of high-correlated plausibility-based predictors is not guaranteed to bring high predicting performances. An example of a statistical model could be a regression model, which aims at building a mathematical relationship between predictors and predictand thus enabling the inferential process. Typical regression models work in the continuous case, meaning that they are capable of predicting numeric quantities having an infinite number of values between any two values.

Dynamical meteorological drought prediction is based on the so-called General Circulation Models (GCMs). The purpose of these models is to run a simulation of the physical processes taking place in the atmosphere, ocean, and land in order to allow inferential processes over climate events (e.g., droughts). The two fundamental pillars for the good performance of such models are climate system understanding and computational resources. Indeed, to obtain reliable results from these models a deep understanding and a good reproduction of the physical mechanisms underlying climate behavior are needed. On top of that, these types of models are particularly demanding from the computational point of view.

The hybrid drought prediction approach is aimed at merging statistical and dynamical methods to take advantage of both. Numerous studies have shown how these two methods are complementary [43]. On one side, statistical methods are simple to implement and less computationally demanding but, at the same time, they do not consider physical processes; on the other side, dynamical methods account for physical processes (thus allowing for the exploitation of nonlinear interaction of atmosphere, land, and ocean [46]) but, at the same time, they are computationally demanding and affected by bias [47]. Therefore, the combination of these two methods, if well structured, can result in sensible performance gains. The way in which this merging process is performed is composed of a two-step

procedure: (i) climate forecast calibration, to correct the bias coming from GCM, and, subsequently, (ii) multiple forecast merging, to merge the predictions coming from both statistical and dynamical methods. As depicted in Figure 8, the merging step could be obtained with different techniques such as regression and Bayesian distribution and, once the merging has been performed, the result can be exploited to obtain a probabilistic prediction.
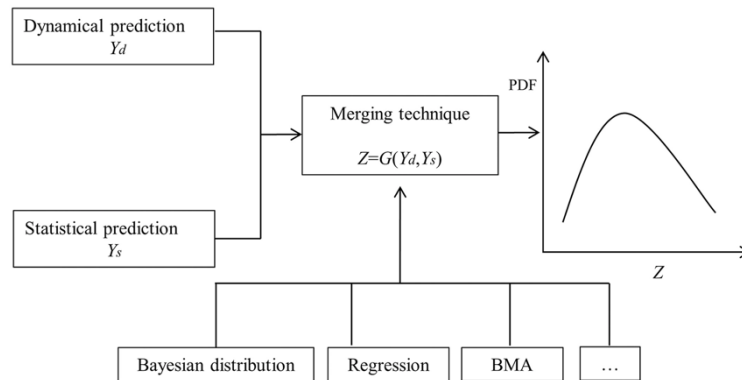

Figure 8: Hybrid drought prediction based on drought indicator Z. Merging step performed with function G applied on dynamical forecast (Yd) and statistical forecast (Ys) [40].

An additional category for forecasting, often also classified as a subcategory of statistical methods, is Artificial Intelligence (AI). AI methods have been proven to be extremely helpful in establishing complex relationships among variables, and a great variety of them is used in the drought forecasting field (e.g., [48], [49]). Some well-known examples are: **Neural Networks (NNs),** able to model highly complex nonlinear relationships between inputs and outputs. Its structure is composed of an input layer, a hidden layer(s), and an output layer made of processing units called neurons. These neurons receive a set of values from the preceding layer, add up all these values, add a bias term, and shrink the result by means of an activation function. Finally, the output is multiplied by a weight and passed to the next layer. NN is one of the most used AI methods in the hydrologic forecasting research field [50] and some experiments showed, for instance, how for arid and semi-arid locations they are even outperforming statistical methods [51]. Moreover, their design complexity is much lower with respect to dynamical methods; **Support Vector Machine (SVM),** usually exploited to solve classification tasks, can also be used to solve regression problems in the form of a Support Vector Regressor (SVR).

The general concept of SVM is to maximize the distance between the supporting planes of different classes. An example of SVR application in drought forecasting, for instance, has been developed to predict drought indices in different locations in Australia and it turned out to be very efficient [52]; **Extreme Learning Machine (ELM)**, structurally similar to NNs, is characterized by a reduced training time. In ELMs weights and biases are randomly initialized and the learning procedure consists of a simple mathematical operation instead of an iterative and computationally demanding process.

2.3 Teleconnection patterns and climate indices

The physical mechanisms underlying the evolution of climatic conditions take place as a consequence of the interactions of all the different Earth's system components. These interactions often result in the creation of climatic patterns, such as pressure or temperature configurations, that are repeated and cyclical over time. These recognizable patterns are associated with specific meteorological implications such as above/below-normal temperatures or precipitations. Teleconnection patterns can be therefore defined as terms used to reference specific patterns of global climatic variables, such as pressure or temperatures, with the aim of formalizing causality between climatic configurations and meteorological events. Several teleconnection patterns have been formalized, each of them having different meteorological implications affecting specific regions of the planet. Moreover, all these patterns are characterized by "*phases*", a term used to reference different instances of the same teleconnection pattern which translate into different meteorological consequences (e.g., two phases of a teleconnection pattern involving locations x and y could be low-pressure over x / high-pressure over y and high-pressure over x / low-pressure over y). Among the most known teleconnection patterns, there are North Atlantic Oscillation (NAO), SCAndinavian oscillation (SCA), East Atlantic oscillation (EA), East Atlantic/Western Russian oscillation (EA/WR), and El Niño Southern Oscillation (ENSO). The term climate index is instead referred to the quantitative description of the teleconnection pattern (i.e., timeseries of values).

The NAO pattern is formally defined as the difference in surface air pressure over two specific regions: Iceland and the Azores. It is essentially characterized by two phases, positive and negative. The negative phase of NAO is characterized by moderately high pressure over the Azores and moderately low pressure over Iceland. The positive phase, instead, presents a very high pressure over the Azores and a very low pressure over Iceland (so that the pressure difference over the two locations is wider). The observed implications of positive NAO phases are, from one side, above-normal temperatures in the eastern United States and northern Europe and, from the other, below-normal temperatures across Greenland and southern Europe. They are also linked with above-normal precipitation over the northern European zone and below-normal precipitation over southern and central Europe. For what concerns negative NAO phases, they usually exhibit an opposite set of implications. NAO climate index can be computed and, starting from the
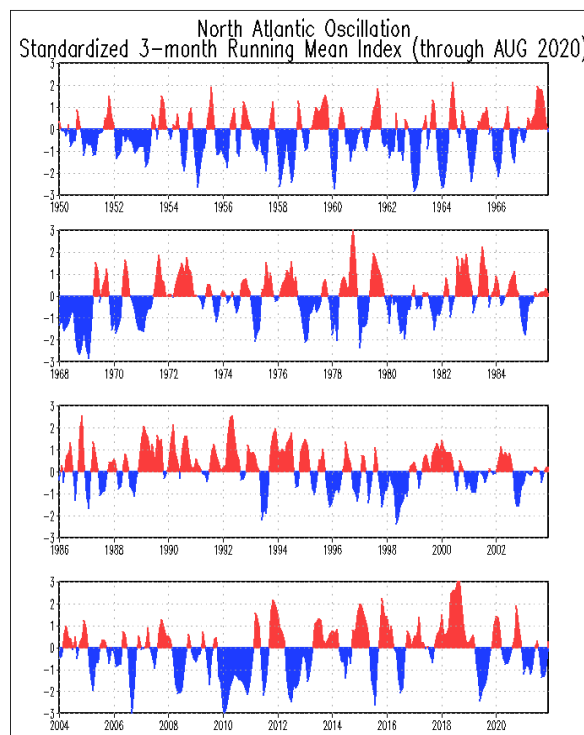

Figure 9: Oscillating trend (positive [red]/negative [blue]) of North Atlantic Oscillation. [75]

timeseries, it is possible to associate positive phases with positive values and negative phases with negative values. As it is possible to observe from Figure 9 (depicting the NAO climate index), the time span between the positive and negative phases is, on average, intra-annual.

The SCA pattern, also called SCAND, is composed of three main circulation centers. The main one is located over Scandinavia and the other two, weaker and of opposite sign, over eastern Russia / western Mongolia and western Europe. As mentioned for NAO, also SCA is mainly characterized by two phases. The positive phase of SCA is associated with high pressure over northern Europe and low pressure over southern Europe [53]. During SCA negative phases, instead, the pressure over northern Europe is low and the pressure over southern-central Europe becomes high (Figure 10). The observed implications of positive SCA phases are related to below-average temperatures over central Russia and western Europe. For what concerns precipitations, it is associated with above-average precipitations
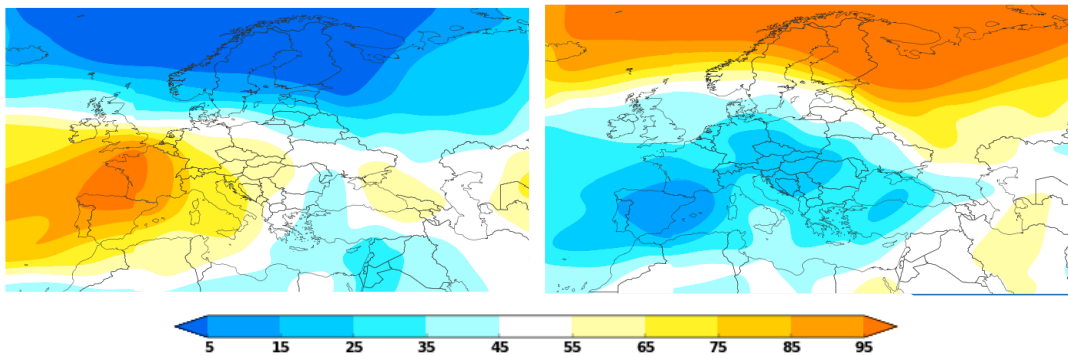
Figure 10: 500mb height during negative (left) and positive (right) SCA phases [50]

across central and southern Europe and below-average precipitation over the Scandinavian area. SCA climate index can be computed and, starting from the timeseries, it is possible to associate positive phases with positive values and negative phases with negative values. As it is possible to observe from Figure 11 (depicting the SCA climate index), the time span between a positive and a negative phase is on average biannual / intra-annual.

Figure 11: Oscillating trend (positive [red]/negative [blue])
of SCAndinavian pattern [76]

The EA oscillation, known as the climate pattern with the second lower frequency in its alternation over north Atlantic [54], is often associated with the NAO pattern due to its similar but d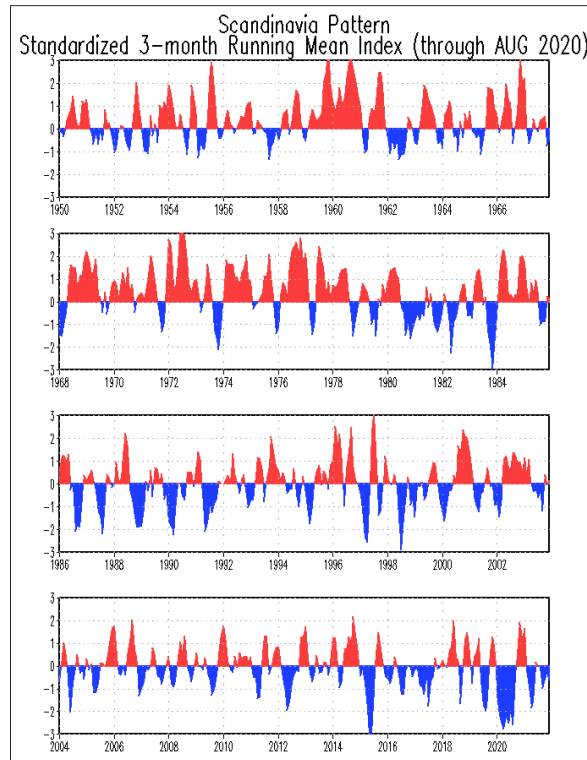own-shifted structure. If NAO, on one hand, is defined as the pressure difference between the Azores and Iceland, EA, on the other, is the difference in pressure between the northern and the tropical part of the Atlantic Ocean. It is characterized by positive and negative phases. The positive phase is associated with hot and dry air flow all over the Mediterranean area and with above-average precipitations over the northern European zone. The negative phase instead, especially during winter seasons, is associated with low temperatures over the entire Mediterranean zone. EA climate index can be computed and, starting from the timeseries, it is possible to associate positive phases with positive values and negative phases with negative values. As it is possible to observe in Figure 12 (depicting the EA climate index), the time span between a positive and a negative phase involves, on average, several years. However, in some short periods the switching frequency increase until becoming infra-annual.

The EA/WR pattern, also called EATL/WRUS, is one of the most affecting patterns for the Eurasian region. It is composed of four main anomaly locations: Europe,
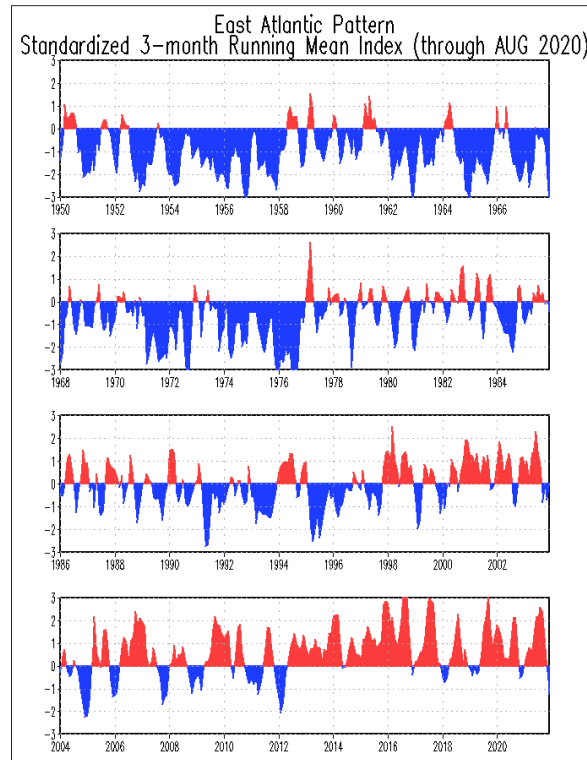
32

Figure 12: Oscillating trend (positive [red]/negative [blue])
of East Atlantic pattern [77]

Northern China, the Northern-Central Atlantic region, and the northern Caspian Sea. It is characterized by a positive phase occurring as a consequence of positive height anomalies over two of the four anomaly locations: the northern-Central Atlantic region, and the northern Caspian Sea. The negative phase, instead, is characterized by an opposite configuration. The meteorological implication of the positive phase consists of above-average temperatures for the eastern Asian region and below-average temperatures for western Russia and northeast Africa. For what concerns precipitation, it is usual to experience above-normal precipitation in eastern China and below-normal precipitation in central Europe. EA/WR climate index can be computed and, starting from the timeseries, it is possible to associate positive phases with positive values and negative phases with negative values. As it is possible to observe from Figure 13 (depicting the EA/WR climate index), the time span between the positive and negative phases for EA/WR is, on average, annual/intra-annual.

The ENSO oscillation pattern, probably the most well-known oscillation pattern, is a periodic fluctuation (i.e., every 2/7 years) in the Sea Surface Temperature (SST) and in the above pressure. The counterpart of this phenomenon is called La Niña and the interchanging mechanism between El Niño e La Niña is the following: the equatorial zone in the Pacific Ocean becomes warm and the SST in that region starts
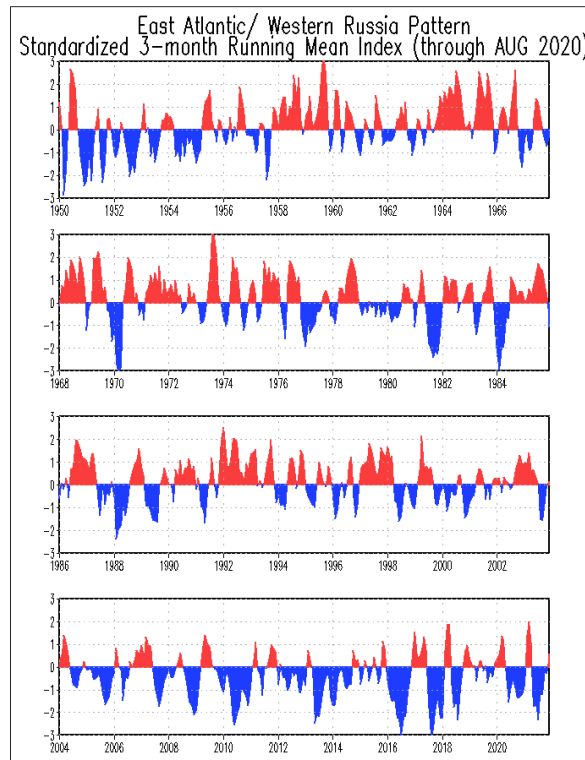
Figure 13: Oscillating trend (positive [red]/negative [blue])
of East Atlantic / West Russian pattern [78]

increasing. Due to the equatorial currents caused by the Coriolis's effect, this warm water is pushed westward (in the vicinity of Australia), and this SST pattern composed of warm water at the west and cooler water at the east represents La Niña. When the equatorial currents start decreasing in intensity, instead, the warm zone shifts eastward forming the SST configuration corresponding to El Niño. During La Niña events, warm water is coupled with low pressure and high precipitation in the Australian zone, and cold water is associated with high pressure and dry weather in South America.
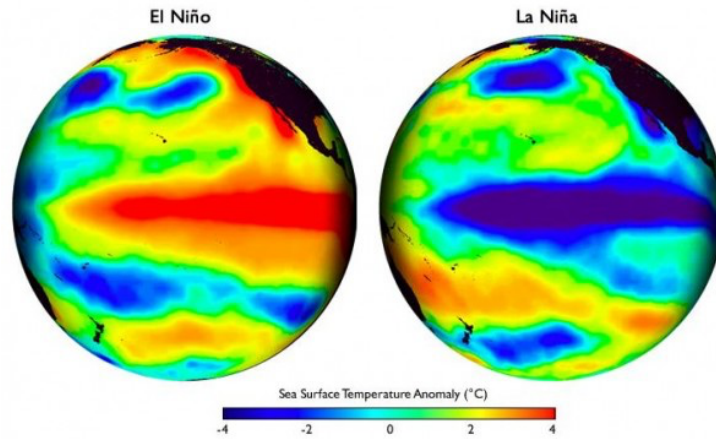
Figure 14: Sea Surface Temperature pattern describing El Niño and La Niña [79]
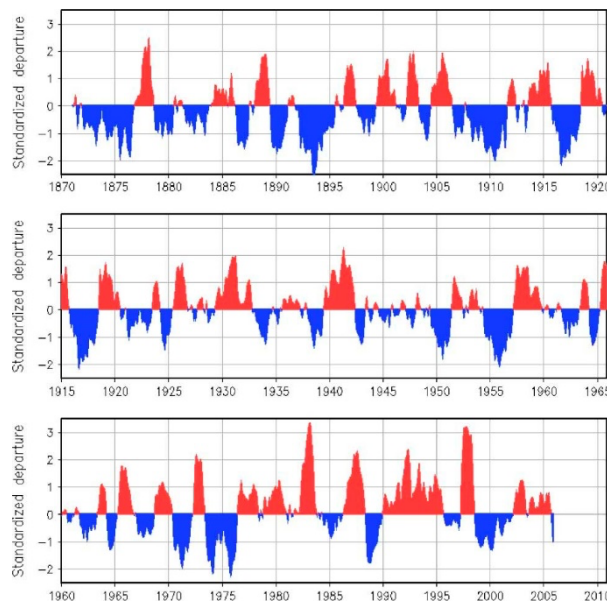
Figure 15: Oscillating trend (positive [red]/negative [blue])
of El Niño Southern Oscillation pattern [80]

ENSO climate index can be computed and, starting from the timeseries, it is possible to associate positive phases with positive values and negative phases with negative values. As it is possible to observe in Figure 15 (depicting the ENSO climate index computed through the Multivariate Enso Index method), the time span between a positive and a negative phase is, on average, more than a year.

Understanding climate patterns and the way in which they influence the occurrence of meteorologic events is fundamental for a better understanding of the mechanisms behind extremes such as droughts. The usage of some of these indices, coupled with

climatic global variables, could represent a valid approach for the enhancement of drought prediction at subseasonal lead times by bridging climatic variables with the occurrence of meteorologic events.

3 Data and case study

In this section, a context-related description of the problem is provided regarding the study area of Rijnalnd. After that, a careful and detailed description of the process of variable selection is provided along with the main pre-processing steps performed to prepare the data for the framework that will be introduced in the subsequent section.

3.1 Case study

The case study is the region of Rijnland, located in the Netherlands. It is a small sub-catchment of 1000 Km^2 at the end of the Rhine delta (Figure 16). More than 70% of the area is below sea level, and, for that reason, it now has a highly monitored drainage system governed by the Rijnland water board. The drainage system includes a series of pumps and locks that are used to discharge the excess water into the Rhine River, shipping canals, or directly into the North Sea, but also to flush freshwater from the Rhine into the channels to counter salinity intrusion. The majority of the Rijnland land use involves agricultural crops among which some high-cash crops are also present. The area is not only exploited for agricultural purposes, as several cities, for a total of 1.3 million citizens, and strategic infrastructures, such as the international airport Amsterdam Schiphol, are located in Rijnland.
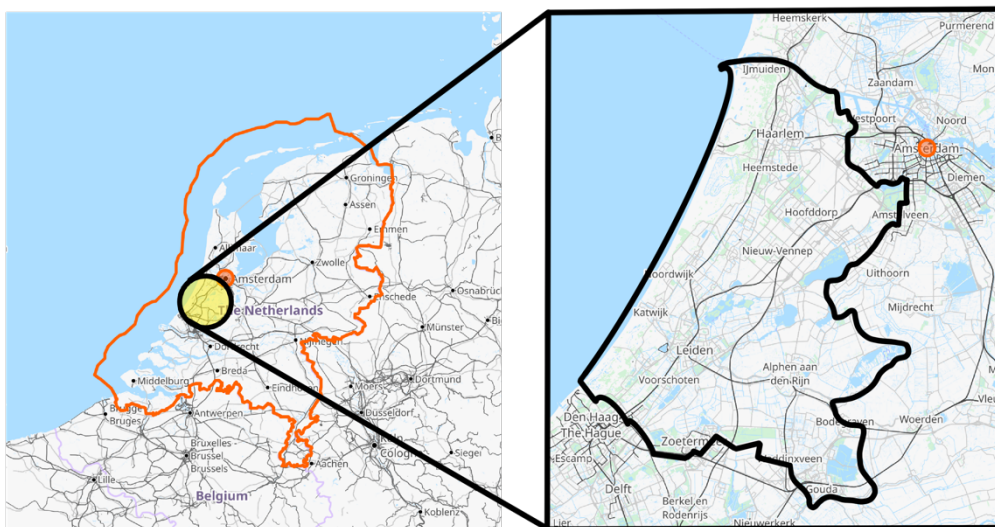

Figure 16: Water system boundaries of the Rijnland water board

Rijnland is characterized by an average yearly amount of rain of 842 mm (Schiphol meteorological station), and by a temperature pattern consisting of much lower temperatures during winter than in summer. This relevant temperature difference makes the evapotranspiration rate increase during the summer season, such that it often exceeds precipitation. The yearly average Makkink evapotranspiration is estimated to be around 602 mm.

The region is affected by summer drought events, which have been occurring more frequently in the past years. Droughts in the area are usually characterized by low flows in the Rhine, low precipitation, and high evapotranspiration rate. Over time, the Rijnland water board came up with its own context-specific definition of drought, which occurs when the two following conditions are simultaneously met:

- Cumulative precipitation deficit greater than *150 mm*. The cumulative precipitation deficit is computed as the difference between precipitation and potential Makkink evapotranspiration, and it is cumulated from the 1st of April to the 30th of September. The calculation stops during the remaining months of the year, when usually the precipitation rate is higher than evapotranspiration.

- Discharge at Lobith station, where the Rhine enters the Netherlands, lower than *1100 m^3/s*.

The concurrent existence of these two conditions makes the amount of fresh surface water available decrease, causing salinity intrusion in the irrigation channels and dike instability issues. As a consequence, dike inspections have to be performed more frequently and an alternative supply of fresh water has to be planned in collaboration with the surrounding municipalities and water boards, which can provide more freshwater.

To monitor and potentially issue early-warnings for the occurrence of droughts, the Rijnland waterboard produces weekly or monthly-based reports from April to October, with information related to the observed and forecasted streamflow in Lobith and cumulative precipitation deficit over the Netherlands. The forecasts are now available with a lead time of two weeks and are based on:

- 2-week ensemble predictions of precipitation and temperature from ECMWF;

- 2-week discharge prediction of the Rhine at Lobith, provided by Rijkswaterstaat, the Directorate General of the Ministry of Infrastructure and Water Management of the Netherlands.

In 2021 the Rijnland water board started to use SPI as a monitoring tool, although not yet for issuing warning, and included the SPI maps provided by the Royal Netherlands Meteorological Institute (KNMI) SPI in the summer reports.

The current forecasting capability of the Rijnland water board does not allow for a timely early warning of drought occurrence. Indeed, in order to design more efficient maintenance plans for dikes and a more coordinated system of freshwater sharing among the surrounding municipalities, the water board would need to extend the current lead time of 2-week forecasts to a month.

3.2 Dataset description

The sub-seasonal lead time represents a critical horizon in weather forecasting in general, since it is a time window in which the relevance of all the potential predictors (global and local) is mixed across different spatiotemporal scales. When approaching sub-seasonal lead times, indeed, valid information for short-medium term forecasting such as atmospheric initial conditions start becoming less relevant, but their contribution is still present. At the same time, large-scale phenomena, for lead times between 10 and 46 days, start becoming more and more relevant [55], with the slow ocean variability affecting atmospheric circulation and becoming a relevant source of information for forecasting potential extreme events, such as droughts. Moreover, teleconnection patterns could also play a role in the genesis of extreme events and in atmospheric circulation at the sub-seasonal lead times [56]. For these reasons, the set of input variables selected for this work involves local initial atmospheric conditions, global climate variables, and climate indexes to search for the best compromise of predictors that is able to increase the model accuracy. The details on the set of climate indices, global variables and local variables selected as input for this study are presented in the following sections.

3.2.1 Climate indices

To select the teleconnection patterns and the related climate indices to use, first a summary of the teleconnection patterns with meteorological implications across the

northern European zone has been produced. Table 4 highlights the teleconnection patterns selected for the present work with information related to the switching phase frequency. NAO, SCA, and EA are characterized by oscillations that can be kept in sub-seasonal time frames and with meteorological consequences over or in the surroundings of northern Europe.

*Table 4:summary table of the selected climate indices. **NE surroundings** and **NE** respectively represent the meteorological implications in the surrounding of Northern Europe and in Northern Europe. **Annual/sub-Annual** refers to the alternation frequency between the positive and negative phases. **Multi-annual *** refers to alternation time greater than 2 years but that could be anyway interesting for seasonal/sub-seasonal forecasting*

	Implications	Switch frequency		
	NE surroundings	NE	Annual/sub-Annual	Multi-annual *
NAO	X	X	X	
SCA	X		X	
EA	X	X		X
ENSO				

All the climate signals falling in at least one of the two green columns are worth to be considered as a potential input variable. The ENSO as well, however, is considered among the selected indices, even if does not respect this condition. The reason for this is that ENSO is recognized as the dominant multiannual signal in the global climate context and little to no clue is provided in terms of effects over the European area [2]. Moreover, even though its switching frequency is between 2 and 7 years, there is no reason to a priori exclude that such slow variability could affect sub-seasonal forecasting, especially while considering several decades of data as in the present work. The timeseries related to the climate indices associated with these four patterns have been obtained from the National Oceanic and Atmospheric Administration (NOAA) web platform [57] [58] [59] [60] and are in the form of a unidimensional timeseries of float values with a monthly time aggregation.

3.2.2 Global variables

After defining the most influential teleconnection patterns and associated climate indices for the case study, a set of global variables related to these phenomena was selected as well. For instance, as the NAO is a teleconnection pattern due to the difference of pressure in two different locations, the Mean Sea Level Pressure (MSLP) has been selected as one of the potential input variables. For a similar

reason, the Sea Surface Temperature (SST) was selected because of ENSO. Table 5 presents the list of the climate patterns/indices considered with their corresponding global variable.

Table 5: couples climate indices / global variables based on the definition of the climate indices

Climate signal	Global variable	
	SST	Pressure
NAO		X
SCA		X
EA		X
ENSO	X	

In addition, the geopotential height at 500 hPa (Z500) was selected as the third global variable, since it is proven to be representative of the average state of the atmosphere.

The global variables come from the fifth generation of ECMWF reanalysis data, the ERA5 dataset, and have been downloaded from the Copernicus Climate Data Store (CDS). They are gridded timeseries of data covering the entire Earth's surface and, for the purpose of the present work, have been selected with a daily temporal resolution and with a spatial resolution of 1.5 degrees. Even though the original spatial resolution of ERA5 is 0.25 degrees, the time series have been re-gridded on the coarser resolution of 1.5 degrees, to meet the spatial resolution of ECMWF forecasts, which are considered as the benchmark forecasts currently available.

3.2.3 Local variables

The initial atmospheric variables are, in this thesis, often referred to simply as local variables because of the intent of considering them only for the area of Rijnland. However, later in the manuscript, another instance of this exact same set of variables will take the name of "medium-scale" information, since their spatial extension will be increased.

The local variables were initially selected following the work of Felsche et.al. [49]. This study dealt with the exploitation of Artificial Neural Networks (ANN) for the prediction of drought occurrence in Lisbon and Munich with a one-month lead time and it was chosen as an example to follow because of the similarity of the problem addressed.

Among the local variables considered by Felsche et al. [49], only a subset of variables that could be meaningful for the northern European context has been selected. Moreover, because the original study used data coming from the Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE), the subset of variables has been further filtered based on ERA5-translatable criterion. The ERA5 dataset was chosen instead of the CRCM5-LE to keep consistency with the global variables data source. Finally, a total number of ten variables was extracted, as presented in Table 6 and Table 7. Particularly, Table 6 is referred to the variables belonging to the ERA5 pressure level realization (with pressure level fixed at 500 hPa), and Table 7 refers to variables belonging to the ERA5 single levels realization. In this case, the choice also of the pressure level was needed and, for all the variables, has been decided to fix it ad 500 hPa.

All these variables are taken at the daily resolution, in the period 1979-2021. In terms of spatial resolution and extension, in a first case (when called local variables), they correspond to simple unidimensional timeseries of float values located in Rijnland with a spatial resolution of 1.5 degrees, while when referred to as the medium-scale variables they are downloaded and extracted for the whole European region.

*Table 6: CRCM5-LE variables coupled with respective ERA5 **pressure** levels variables. For each of the two data sources name and measurement unit is reported. In the first column, the abbreviation of the variables used in the present work. The column "pressure level" represents the reference pressure level for the selected variable*

	CRCM5-LE		ERA5 pressure Levels		Pressure level
UW	Eastward Near-Surface Wind	m/s	U-component of wind	m/s	500 hPa
VW	Northward Near-Surface Wind	m/s	V-component of wind	m/s	500 hPa
RH	Near-surface relative humidity	%	Relative humidity	%	500 hPa
SH	Near-surface specific humidity	1	Specific humidity	1	500 hPa

Table 7: CRCM5-LE variables coupled with respective ERA5 single levels variables. For each of the two data sources name and measurement unit is reported. In the first column, the abbreviation of the variables used in the present work.

	CRCM5-LE		ERA5 Single Levels	
TCC	Total cloud cover	%	Total Cloud Cover	%
MER	Evaporation	$Kg\,m^2/s$	Mean Evaporation Rate	$Kg\,m^2/s$
MSSHF	Surface Upward Sensible Heat Flux	W/m^2	Mean Surface Sensible Heat Flux	W/m^2
TCWV	Water vapor path	Kg/m^2	Total Column Water Vapour	Kg/m^2
tp	Precipitation	Kg/m^2s	Total precipitation	m
t2m	Surface temperature	K	2m temperature	K

A final summary of the pool of considered variables is reported in Table 8.

Table 8: Summary of the considered variables. For local variables, a double sample's shape is specified in order to account also for their medium-scale instance

Acronyms	full name	spatial scale	sample's shape	source
NAO	North Atlantic Oscillation	Global	1×1	NOAA
SCA	SCAndinavian oscillation	Global	1×1	NOAA
EA	East Atlantic oscillation	Global	1×1	NOAA
ENSO	El Niño Southern Oscillation	Global	1×1	NOAA
SST	Sea Surface Temperature	Global	121×240	ERA5
MSLP	Mean Sea Level Pressure	Global	121×240	ERA5
Z500	Geopotential height at 500 hPa	Global	121×240	ERA5
TCC	Total Cloud Cover	Local	$1 \times 1 / 31 \times 39$	ERA5
MER	Mean Evaporation Rate	Local	$1 \times 1 / 31 \times 39$	ERA5
MSSHF	Mean Surface Sensible Heat Flux	Local	$1 \times 1 / 31 \times 39$	ERA5
TCWV	Total Column Water Vapour	Local	$1 \times 1 / 31 \times 39$	ERA5
UW	U-component of Wind	Local	$1 \times 1 / 31 \times 39$	ERA5
VW	V-component of Wind	Local	$1 \times 1 / 31 \times 39$	ERA5
RH	Relative Humidity	Local	$1 \times 1 / 31 \times 39$	ERA5
SH	Specific Humidity	Local	$1 \times 1 / 31 \times 39$	ERA5
tp	Total Precipitation	Local	$1 \times 1 / 31 \times 39$	ERA5
t2m	2m Temperature	Local	$1 \times 1 / 31 \times 39$	ERA5

3.3 Data pre-processing and cleaning

Before training the different machine learning models, a pre-processing and cleaning step is needed. Specifically, global and local variables have been submitted to the different and tailored pre-processing procedures described below.

Global variables coming from ERA5, as also mentioned in section 3.2.2, have by default a spatial resolution of 0.25 degrees. For computational reasons related to specific steps in the processing pipeline and for compatibility with baseline forecast products (ECMWF Extended-range forecasts), it has been decided to use a coarser resolution of 1.5 degrees. On top of that, the ERA5 dataset makes available these global variables with a default temporal resolution of an hour, however, the temporal resolution needed for the development of the present work is of 1 day.

Thus, two main pre-processing steps have to be applied: (i) a grid upscaling from 0.25 to 1.50 degrees, and (ii) a temporal aggregation from an hourly resolution to a daily-based one. To avoid the need to locally perform these two pre-processing steps for the whole timeseries of data going from 1979 up to 2021, it has been decided to exploit the Copernicus Data Store Toolbox (CDS Toolbox). This Toolbox allows for the specification of a set of instructions to directly obtain the set of pre-processed data without the need to add additional time-consuming processing layers to the pre-processing pipeline.

For what concerns local variables, the pre-processing steps consist of considering only land-referred ERA5 grid cells, and re-gridding data with respect to the ECMWF Extended-range forecasts grid. As it was for global variables, a re-gridding (upscaling) procedure from ERA5 to ECMWF has been performed. The process started by taking ERA5 grid cells over the Netherlands and by applying a Thiessen weighted average to determine the average rainfall. Specifically, since the study area is a small sub-catchment of 1000 Km^2, and since most of the Rijnland-located 0.25 degrees cells belonging to the 1.5 degrees ECMWF cell are on the sea - where no ground sensor is present – Thiessen was applied assuming zero weight for sea-located grid cells. The result after applying Thiessen was the upscaled Rijnland-referred cell. In this case, since the procedure is extremely tailored to the study area, all the data have been first obtained and then locally processed. For the medium-scale instance, instead, the only needed pre-processing steps were an upscaling of ERA5 data from 0.25° to 1.5° and a temporal aggregation from hourly resolution to a daily one. These operations were directly performed by means of the CDS Toolbox.

Table 9 provides a summary of all the mentioned processing steps for each group of variables.

Table 9: pre-processing steps summary

Data component	Initial data	Pre-processing steps	Final data
Global variable	0.25 degrees, hourly resolution	upscaling and time aggregation	1.50 degrees, daily resolution
Local variable	0.25 degrees, Equal weights, all Netherlands	ECMWF regridding, sea cells weighting, Rijnland extraction	ECMWF grid 1.5 degrees, sea cell zero weighted, Rijnalnd cell only
Medium-scale variables	0.25 degrees, hourly resolution	upscaling and time aggregation	1.50 degrees, daily resolution

4 Methodology

The purpose of the present work is to forecast cumulative monthly precipitation using Machine Learning and investigate, at the same time, the importance of climate information for the forecasts. For this purpose, three ML branches have been designed, each considering a different level of detail in terms of climate information provided, ranging from providing information on teleconnection patterns and phases to not giving any teleconnection pattern information at all. Where no teleconnection information is provided, a different dataset structure is exploited in order to provide the specific model with more samples. This is aimed at investigating if an increased dataset size could counterbalance the absence of climatic information.

The present work is based on the creation of a diverse set of models with quite different characteristics under several points of view: methodology for dataset creation, dataset structure, algorithm type, and architecture. Therefore, the usage of different types of ML algorithms does not correspond to a set of mirrored procedures with a simple algorithmic change. Instead, the different algorithms were provided with the needed specific care in terms of dataset and hyperparameters to allow them to work properly for the purpose of the work. Specifications about model-tailored pipelines will be provided in section 4.2, where each one of the used algorithms is explained in detail.

The high-level structure of the work can be decomposed into three main ML algorithms and corresponding approaches used to forecast cumulative monthly precipitation: Extreme Learning Machine (ELM), Feed-Forward Neural Network (FFNN), and Convolutional Neural Network (CNN).

The approach based on ELM builds up from the Niño Index Phase Analysis (NIPA) developed by Zimmerman et al [1] and the Climate State Intelligence (CSI) framework developed by Giuliani et.al. (2019) [2], which aims at exploiting information related to climate and teleconnection patterns to forecast precipitation. In this framework, 12 different ELM forecasting models are built (one for each calendar month) and trained with a set of climate variables, climate indices, and local atmospheric variables. Each ELM is trained with a different set of input

variables, i.e. the one that provides the highest model performance. This approach has the advantage of including a high level of climate information and of creating a tailored model for each calendar month, but at the same time, it has the drawback of having very few samples available for training the ELM models, thus reducing the possibility of including several input variables at the same time. To investigate the benefits of having more climate information over higher number of input variables and samples, the framework with FFNN has been developed.

The framework based on FFNN is aimed at creating a single model to predict precipitation all across the year, instead of having twelve month tailored models. As one only model is built for the whole time period, it is not possible to include information related to climate patterns (see details in sections 4.1 and 4.2). Different combinations of input variables (global climate data and local variables) are used to train the FFNN model and the combination which provided the best model performance was selected as the final input dataset to train and test the FFNN. This approach on the one hand has the advantage of sensibly increasing the number of available samples in the dataset but, on the other hand, it has the disadvantage of not including teleconnection patterns information and of having to incorporate the intrinsic diversity of the climatic processes occurring each month in one single model.

Finally, to investigate if the loss of teleconnection patterns information could be balanced out by the addition of multiple local atmospheric variables as input features, the CNN approach has been developed. The approach based on CNN is very similar to the FFNN case, as it is aimed at developing one single model that exploits the same set of variables of the FFNN. The difference with the previous approach lies in the spatial extension of the input features considered, which is now extended to the whole Europe rather than to just one cell.

Once the precipitation forecasts are obtained with the three above mentioned approaches, the cumulative precipitation is in turn transformed into drought classes according to the SPI index introduced in section 2.2.3.
The details on the NIPA and CSI frameworks, on the models development and application to the case study are presented in the following sections.

4.1 Niño Index Phase Analysis and Climate State Intelligence

The aim of the Climate State Intelligence (CSI) framework developed by Giuliani et.al (2019) [2] is to exploit the behavior of global climate variables during specific phases of teleconnection patterns to forecast seasonal precipitation in Lake Como, Italy. In this thesis, the original framework has been readapted to forecast the cumulative precipitation of the upcoming 30 days for the region of Rijnland, in the Netherlands. To do so, a chain of processing steps is involved. Firstly, the framework searches for the presence of relevant climate teleconnections through the Niño Index Phase Analysis (NIPA) framework developed by Zimmerman et al [1]. Then, the information extracted by NIPA is used to train 12 different Extreme Learning Machines (ELMs) – one for each month - targeted on seasonal local average precipitation on Lake Como.

As already mentioned, the goal of NIPA is to detect which climate teleconnections could be relevant to predict the total precipitation of the upcoming month. In this sense, this framework operates as an input feature selection and dimensionality reduction method. The way in which it performs this operation is by (i) analyzing multiple climate indices, (ii) labeling and dividing the global and local variables according to the phase of the provided climate index, and (iii) identifying the regions of the Earth where the considered global variable is statistically significantly correlated with the local precipitation. The NIPA framework can be divided into six main steps: input data preparation, data extraction, phase segmentation, correlation analysis, Principal Component Analysis (PCA), and output reshaping.

In order to run, the framework requires two kinds of data: data streams and setting parameters. Regarding the data streams, the framework asks for three main inputs: local target, global variable, and climate index. In this specific case, the local precipitation represents the target of the forecasting framework, i.e. the timeseries of monthly cumulative precipitation in the region of Rijnland. The climate index and global variable are those associated with the teleconnection pattern or climate state which may have an influence on the genesis of the local target. As discussed in sections 3.2.1 and 3.2.2, the NAO, SCA, EA, and ENSO were selected among

the teleconnection patterns, while SST, MSLP, and Z500 were selected among the global variables. It should be noted that NIPA processes one couple consisting of climate index and global variable at the time. Because the final goal of the framework is to define the (climate) input features to use in each monthly ELM model, also the analysis in NIPA is performed for each calendar month. Because of this and as the climate indexes of the selected teleconnections are available at a monthly resolution, also the global and local precipitation need to be at the same temporal aggregation. Therefore, the first step of the NIPA/CSI framework consists in preparing the input data, by aggregating to the monthly scale both local precipitation and global variables that originally had a daily temporal resolution. The aggregation has been carried out by means of a monthly cumulation for local precipitation and of a monthly average for SST, MSLP, and Z500. Every single run of NIPA is referred to a specific month. Fixed the month, NIPA extracts meaningful information contained in the global variable for all the instances of that specific month across the provided timeseries. In each run NIPA searches for the correlation between the local precipitation of a certain month of the year and the n preceding month of the selected global variable, where n is the "*aggregation level*". Both the month and the aggregation level are the parameters needed to run the NIPA framework. More specifically, the "*month*" parameter is referred to the month of the target precipitation. It can be a value between 1 and 12 (e.g., 1 = January, 2 = February, etc.), which defines for which month NIPA has to produce the output. The "*aggregation level*" parameter can be a number from 1 to 3 and it represents the number of previous months of the input variables to consider with respect to the month of the target. For instance, if for a NIPA run the setting parameters are *month = 1* and *aggregation level = 1*, the method will look for correlation between the local precipitation of January and the selected global variable of December; instead, for *month = 1* and *aggregation level = 2*, the method will take into account the local precipitation of January and the selected global variable of November and December. It should be noted that, when $n > 1$, the considered global variable is averaged over the n months, and the same happens for the climate index.

The second step of the NIPA/CSI framework is the data extraction, which is performed on the whole timeseries (e.g., 1979-2021) of climate index and global variable. In this phase, the climate index and the global variables of the n months

prior to the month stated by the *month* parameter are extracted and merged to form a tuple. The final result consists of 42 tuples (one for each year), each one containing two timely matched samples: one float value for the climate index and one gridded sample for the global data. Note that this is a year-based operation, therefore, from one side there will be a tuple for each year, but from the other, that tuple is representative of a specific month throughout the whole timeseries (e.g. all the Decembers). Figure 17 depicts the entire procedure.
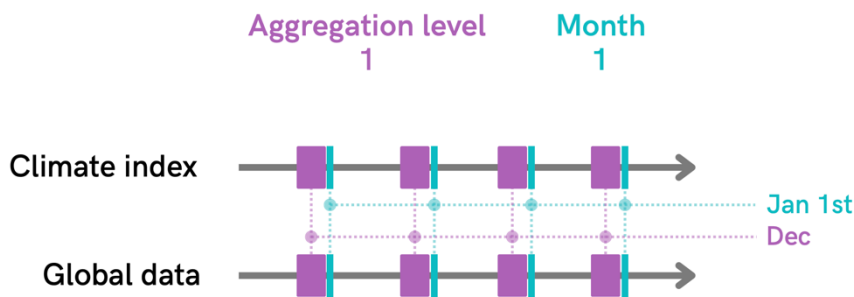

Figure 17: Data extraction step with month = 1 (light blue) and aggregation level = 1 (purple). In this picture the month parameter is only used as a baseline to extract the preceding data.

In the segmentation step, the just extracted data tuples are divided into two datasets, one for the positive and one for the negative phase of the climate teleconnection pattern considered. The division in phases is done by looking at the value of climate index and labeling the global variable with the same phase of the climate index, as depicted in Figure 18.

The correlation analysis step aims at computing the correlation between the selected global variable and the cumulative local precipitation in Rijnland, and at filtering
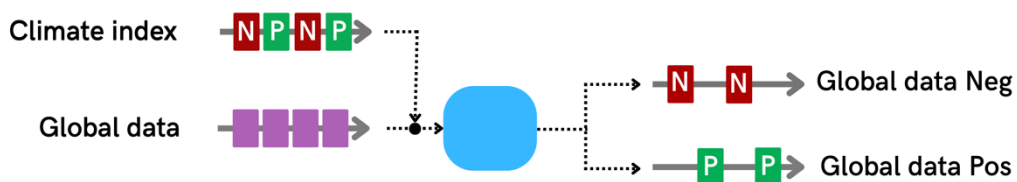

Figure 18: Phase segmentation step. On the left: climate index data (phase-labeled) and global data. On the right: negative and positive sub-chunks of global data according to climate index labeling

51

the resulting correlation map with a set of three filtering conditions. Firstly, thanks to the *month* parameter specified in the inputs step, NIPA is able to perform the same data extraction procedure performed for global variables and climate index also for the local precipitation. Therefore, the local precipitation of all the instances of the same month across the timeseries is selected (e.g., *month* = 1, extraction of the local precipitation of all the Januarys) and temporally matched with the global data now divided into two different sub-chunks. As a consequence, two different datasets are now present: one containing positively classified global and local data, and the other containing negatively classified ones. For each of these two datasets, a correlation analysis between each grid cell of the global variable with the grid cell of the local precipitation is performed. The result consists of two correlation maps, one for each phase, as depicted in Figure 19.
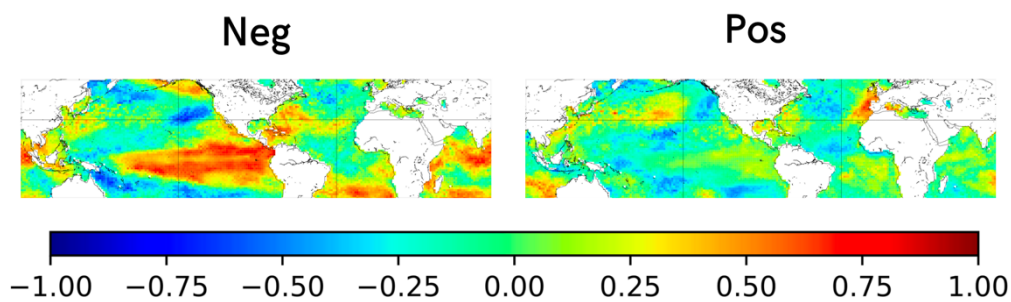
Figure 19: correlation maps for positive and negative phases before filtering. In the image the correlation has been computed between local precipitation in Rijnland and SST data according to ENSO.

The two obtained correlation maps are then filtered by means of a triplet of filtering conditions:

- **95% significance level**: this condition was also applied by Giuliani et.al [2] in the CSI framework and consists in masking out all the correlation pixels with a significance level lower than 95%.
- **Minimum correlation threshold**: after the 95% masking, all the remaining correlated pixels are checked against a fixed minimum correlation threshold and eventually masked out. During the development of the present work, two different minimum thresholds were tested out: 0.6 and 0.5.
- **3x3 contiguous area check**: the output of the previous two filters is then only checked for the presence of at least one 3x3 contiguous window of

correlated pixels. If the condition is verified, the process continues, otherwise, the process stops, and the output is not produced. This checking condition has been added to avoid considering results originating from spurious correlation.

An example of the result of this process is shown in Figure 20. For each of the remaining correlated pixels (or cells) it will be extracted the corresponding time series of global climate variable selected.
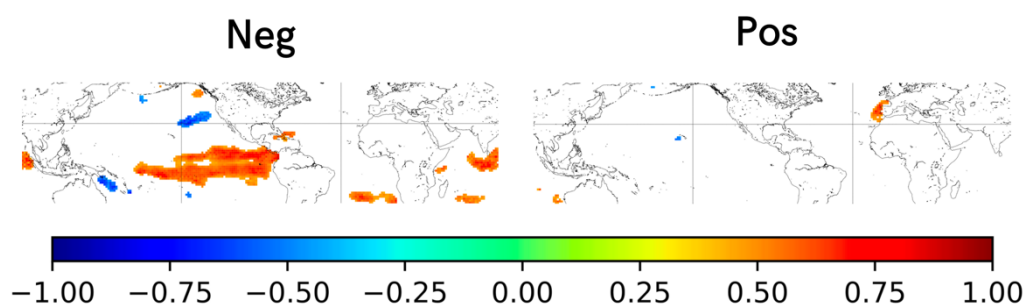

Figure 20: correlation maps for positive and negative phases after filtering. In the image the correlation has been computed between local precipitation in Rijnland and SST data according to ENSO.

 The timeseries of the global variables are allocated into two matrices having a number of rows equal to the number of years falling in each of the two categories (e.g., positive or negative), and a number of columns equal to the number of correlated pixels. Once the two matrices are obtained, they are used to perform a Principal Component Analysis, which aims at reducing the number of considered features from the number of pixels to one. Finally, only the first principal component of each of the two datasets is considered.

Finally, the output reshaping step is performed by concatenating in a single timeseries the two obtained principal components and by specifying for each of these values the corresponding climate index's phase. Therefore, the resulting dataset is going to be composed of two features: one corresponding to the first principal component of the global gridded data provided as input, and the other corresponding to the phase of the climate index in which it was classified during the phase segmentation step. The NIPA framework is run for each month, for each

combination of local precipitation and global variable (e.g., SST, MSLP, and Z500), for each aggregation level of the global variable (e.g., 1,2, and 3), and for each climate signal (e.g., ENSO, NAO, SCA, and EA). Therefore, the total number of potential combinations for which NIPA performs the whole procedure given the data considered in this work is 432.

However, of these 432 potential combinations, only those meeting the filtering conditions mentioned above were selected.

4.2 Machine Learning algorithms

In this section, each of the Machine Learning (ML) algorithms exploited in the present work is introduced by first providing a theoretical background and subsequently discussing their case study application.

4.2.1 Extreme Learning Machine

Extreme Learning Machine (ELM) is a machine learning algorithm that has been introduced with the aim of improving some of the limitations of classical Feed-Forward Neural Networks (FFNN). The fundamental building block of an ELM is called "neuron" and can be interpreted as a processing unit progressively transforming inputs values into outputs values. This transforming process is based on a mapping between features and target that the algorithm learns and materializes by assigning weights to the connections between neurons. The training process of ELMs consists of a simple inverse operation used to compute the output weights connecting hidden and output layers. The one-time nature of this training procedure directly translates, for the majority of the cases, into a significantly higher learning speed. On top of that, further analysis showed how ELM have higher chances to reach the global optimal solution with respect to traditional Single-hidden Layer Feedforward Neural networks (SLFNs) [61].

Analytically, ELM can be expressed as follows:

Given N training samples (x_i, t_i) where x_i is the input and t_i the target and given an ELM with $\tilde{N}$ hidden nodes, the output vector can be determined with

Equation 1

$$o_i = \sum_{j=1}^{\tilde{N}} \beta_j \, f(x_i \cdot w_j + b_j) \quad i=1, \dots, N$$

Where w_j and b_j are respectively the weights connecting the input layer to the hidden layer and the biases of the hidden layer, β_j is the vector containing all the links between the hidden layer and output layer and $f(\cdot)$ is the activation function. Equation 1 can be compactly written as

Equation 2

$$H\beta = O$$

Where:

$$H = \begin{bmatrix} f(w_1 \cdot x_1 + b_1) & \cdots & f(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ f(w_1 \cdot x_N + b_1) & \cdots & f(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{\tilde{N}} \end{bmatrix}_{\tilde{N} \times 1}, \quad O = \begin{bmatrix} o_1 \\ \vdots \\ o_{\tilde{N}} \end{bmatrix}_{\tilde{N} \times 1}$$

H is the output matrix of the hidden nodes and contains all the input x_i multiplied by the weights w_j (connecting the input layer and the hidden layer) and summed to the biases b_j of each hidden layer's neuron. Each of these terms is then wrapped into the neuron's activation function $f(\cdot)$. In order to minimize the loss function during the ELM training process, weight and biases are randomly assigned without taking into account the input data. Therefore, Equation 2 becomes a linear system allowing the output weights vector $\hat{\beta}$ to be analytically determined with a least-square solution as in Equation 3

Equation 3

$$\hat{\beta} = H^+ \, T$$

with H^+ being the Moore-Penrose generalized inverse of the matrix H and T being the vector containing all the targets of the training set.

In the past decades, FFNN has been widely used to model natural and artificial phenomena. One of the reasons why relates to their capabilities to adequately approximate highly non-linear relationships for contexts in which a parametric approach becomes hard to implement [62]. The majority of the learning algorithms used to train FFNN rely on iterative processes that are not guaranteed to reach the globally optimal solution due to the presence of local minima in which gradient descent approaches can get stuck [62]. Extreme Learning Machine (ELM) is aimed at improving the above aspects for SLFNs.

4.2.1.1 Case study application

In the present work, 12 different ELM models have been created, one for each month. The datasets used for the creation of the models are structured as a combination of the local variables described in section 3.2.3, and the global information extracted by NIPA from global variables and climate indices. As previously said, NIPA is a statistical framework that performs feature selection and dimensionality reduction on timeseries of data referred to all the instances of the same month across the whole period. Blending that information with the intention to build 12 different models as in the work of Giuliani et.al. [2] the drawback of using NIPA relates to the reduction of the total available samples for training. Each of the 12 models, indeed, results in having one sample for each year of the timeseries for a total number of 42 samples per model. Such a scarcity in the dataset size generates two main issues related to phenomena generalization and model testing. The first issue is addressed by considering ELM as the algorithm. One of the main advantages of using ELMs is indeed the reduced number of parameters that they need to learn in order to reach good generalization performances. For what concerns the testing part, in normal circumstances the procedure to be applied is quite straightforward: (i) divide the entire dataset into two chunks, one for training and one for testing purposes, (ii) further detach a part of the training set for validation purposes, (iii) perform the training, (iv) select the best-performing model based on validation error, and (v) perform the final test of the best-performing model. In the present scenario, it is clear that the reduced number of samples makes

56

unfeasible a proper training-testing splitting of the dataset. Thus, as in the work of Giuliani et.al. [2], the performance metrics of the 12 models have been determined with a Leave One Out Cross-Validation (LOOCV) procedure. Moreover, the datasets have been constrained to a maximum number of 4 variables (plus an extra feature indicating the phase label of the climate index if one or more global variables are considered). In these datasets, the maximum number of considerable global variables for a single dataset has been fixed to 2. The reason for that relates to the method in which more global variables are integrated into the same dataset. As mentioned in section 4.1, indeed, the output produced by NIPA is composed of two features: one corresponding to the first principal component of the global gridded data provided as input, and the other corresponding to the phase of the climate index in which it was classified during the phase segmentation step. In case more than one global NIPA output is considered in the same dataset, the two distinct phase label information of the two NIPA outputs are merged into a single one. The resulting labels provide, instead of the pure phase information, an integer representing the unique combination of the two different phase labels (i.e., climate state) as shown in Table 10.

Table 10: merging methodology for more than 1 NIPA output belonging to the same dataset. N.B: the real dataset's column is a time series of integers values; here phase names are used for ease purposes.

Phase label 1	Phase label 2	Climate state
Positive	Negative	PN
Negative	Positive	NP
Negative	Negative	NN
Positive	Positive	PP

This directly implies an internal segmentation of the set of data, which, if extended to more than two variables, can have a further mitigating effect on the semantic size of the dataset.

4.2.2 Feed-Forward Neural Network

The advancement in computational performances that took place in the previous decades has enabled the diffusion of several computationally demanding Machine Learning algorithms that were theorized in the past. One of the most evident examples of this computational enablement is represented by the highly cross-

fielded application of Artificial Neural Networks (ANN) that took place in the recent past. ANNs are today widely exploited in many diverse contexts such as pattern recognition, classification, signal processing, and stock market prediction [63]. The basic building block composing an ANN is called "*neuron*" and can be thought of as a processing unit progressively transforming inputs values into outputs values. This transforming process is based on a mapping between predictors and predictand that the algorithm learns and materializes by assigning weights to the connections between neurons. Connected neurons are typically organized in different layers and, considering a Feed-forward Fully Connected Neural Network, each neuron is connected to all the other neurons of the preceding and subsequent layer. Each neuron belonging to the network receives, as input, a set of weighted information coming from the neurons of the previous layer and produces, as output, a value determined by passing the sum of all these inputs into an "*activation function*" just after the addition of a bias term. Once the information went through the whole network, it reaches the output layer which produces the final prediction. In order for that prediction to be meaningful, the set of weights (i.e., the connection between neurons) has to be properly tuned to allow the network to transform the input numerical values into output predictions. This tuning process can be achieved thanks to the usage of different learning algorithms of which the most popular and used is called "*back-propagation*".

The essence of back-propagation consists of the update of network weights based on a backward network propagation of a gradient vector where each item is defined as the derivative of a loss function with respect to a weight [63]. In order to compute the error based on a loss function, a set of "true" targets is needed, thus making the described setting a supervised learning approach. Back-propagation is an iterative process that progressively adjusts the network's parameters by means of a gradient vector, analytically this is translated in what follows:

Defined the error $e_j(n)$ as the difference between the desired (true) output and the actual (predicted) output, and fixed the Residual Sum of Squares (RSS) as the loss function, it is possible to write

Equation 4

$$RSS(n) = \frac{1}{2} \sum_{j \in Q} e_j^2(n) = \frac{1}{2} \sum_{j \in Q} \left(y_j - \hat{y}_j(n) \right)^2$$

where y_j represents the desired output for the neuron j, $\hat{y}_j(n)$ the real output of neuron j with the network's weight configuration at iteration n, and Q the set of all the outputs neurons. Supposing to exploit the backpropagation algorithm in a "*sequential mode*" (i.e., weight update after each training example provided during the training phase), the mathematical formulation of back-propagation can be described as follows:

Considering the equation of the output of neuron j,

Equation 5

$$\hat{y}_j(n) = f\left(b_j + \sum_{i=1}^{m} w_{ji}(n)\, y_i(n)\right)$$

where $f(\cdot)$ is the activation function, b_j is the bias term of neuron j and w_{ji} are the weights of the connections between neuron j and each of the m neurons i. The updates of all weights w_{ji} connected to the neuron j are proportional to the derivative of the *RSS(n)* with respect to the same weights:

Equation 6

$$\frac{\partial RSS(n)}{\partial w_{ji}(n)}$$

Considering Equation 4 and Equation 5, and applying the chain rule to Equation 6, it is possible to obtain what follows:

$$\frac{\partial RSS(n)}{\partial w_{ji}(n)} = \frac{\partial RSS(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial \hat{y}_j(n)} \frac{\partial \hat{y}_j(n)}{\partial w_{ji}(n)}$$

where

Equation 7

$$\frac{\partial RSS(n)}{\partial e_j(n)} = e_j(n)$$

59

Equation 8

$$\frac{\partial e_j(n)}{\partial \hat{y}_j(n)} = -1$$

Equation 9

$$\frac{\partial \hat{y}_j(n)}{\partial w_{ji}(n)} = f'\left(b_j + \sum_{i=1}^{m} w_{ji}\, y_i(n)\right) y_i(n)$$

Replacing Equation 7, Equation 8, and Equation 9 in Equation 6, the result is

$$\frac{\partial RSS(n)}{\partial w_{ji}(n)} = -e_j(n) f'\left(b_j + \sum_{i=1}^{m} w_{ji}\, y_i(n)\right) y_i(n)$$

Finally, the weight correction $\Delta w_{ji}(n)$ to be applied to $w_{ji}(n)$ can be defined by exploiting the delta rule as follows:

$$\Delta w_{ji}(n) = -\alpha\, \frac{\partial RSS(n)}{\partial w_{ji}(n)}$$

Where α is said to be the learning rate which determines the step size of the movement toward the loss minimization.

4.2.2.1 Case study application

This case study application is aimed at building one single Feed-Forward Neural Network (FFNN) with 30 days of lead-time. In contrast with the approach introduced in section 4.2.1.1, in this case, the aim is to produce one single model able to predict cumulative local precipitation in Rijnland all across the year exploiting the same set of input features. Therefore, the provided set of features should be carefully chosen to be informative enough to allow the model to generalize the mapping between features and target across all months of the year.

If considering the same dataset size as ELMs, the need for a fixed set of features covering the entire year reasonably pushes toward a decrease in generalization performances. This is due to the intrinsic seasonal variability affecting each month and to the higher difficulty in finding a set of features that is representative of all the months at one time. For that reason, an increase in the dataset size is needed. However, as mentioned in the data section, the timeseries of the considered climate indices have a monthly temporal resolution, meaning that it is not possible to temporally match them with data having a temporal resolution different from one month. The implications of that are twofold:

- It is no more possible to explicitly embed climatic information in the pool of features;
- It is no longer possible to exploit NIPA as a pre-processing framework due to the absence of climatic information.

In order to compensate for the increased difficulty of the mapping task between predictors and predictand, and the complete absence of explicit climatic information, the dataset for the present work's branch has been built based on a 30 days moving aggregation thus increasing the number of samples. This aggregation window slides over the whole daily data timeseries producing a total number of samples greater than 15000. The base idea of this compensation process is to investigate if a considerably higher number of samples could counterbalance the complete absence of climatic information by allowing the FFNN to learn the underlying relationships solely based on local and global variables.

As it is possible to infer, this work's branch is not a simple mirroring of the monthly-based ELMs introduced in section 4.2.1. Here the complete pipeline is modified by: (i) considering just one model for prediction across the whole year instead of twelve different models, (ii) not considering the climate index in the pool of input variables, (iii) not exploiting NIPA as pre-processing framework anymore, (iv) having a considerably higher number of samples. Even if this approach could apparently seem disconnected with respect to the previous branch, it is actually meant to investigate if, at sub-seasonal lead times, the exploitation of a climatic pre-processing framework such as NIPA brings some benefits to the final model by better bridging local and global climate contexts.

Since the aim is still to bridge local and global information and since FFNNs need 1D timeseries as input features, careful planning of how to build the final datasets is needed.

Starting from local Rijnland variables, they already come in the form of a unidimensional timeseries because referred to a single grid cell.

Global data, instead, is still in the form of a timeseries of gridded data covering the entire world. To be able to embed these global data in the final datasets, a dimensionality reduction has to be performed to shrink these 2D timeseries into 1D timeseries. For doing so the global data are subjected to a Principal Component Analysis (PCA) after which just the first principal component is considered. In order to be as much consistent as possible with the pre-processing applied by NIPA in the previous branch, the mentioned PCA is performed just after a pseudo-replication of the NIPA correlation step. Specifically, the applied pre-processing pipeline is the following:

1. 2D timeseries of global data are considered.
2. Pixel-wise means subtraction to center the data with respect to the mean.
3. Application of the 30 days moving average on the 2D timeseries.
4. Computation of spatial correlation between local Rijnalnd precipitation and global gridded variables (as performed in NIPA).
5. Filtering of the resulting correlation map based on a 95% significance level (as performed in NIPA).
6. Reshaping of the 2D timeseries - now masked based on the filtered correlation map - into a matrix with a number of columns equal to the number of correlated pixels and a number of rows equal to the number of considered timestamps.
7. Application of PCA procedure.
8. Selection of the first principal component (which will represent the global information in the final dataset).

In essence, the performed process can be considered a simplified version of the NIPA framework in which no climatic information is provided. From now on it will be referenced with the name "*pseudo-NIPA*".

Once clarified how different spatial scales are reconciled in the same dataset, it is finally possible to articulate the dataset structure with three main points. The dataset can indeed be composed of:

- Timeseries of 30 days moving average for all the local variables (but total precipitation).
- Timeseries of 30 days moving cumulation of total precipitation.
- Timeseries of global data PCA values coming from pseudo-NIPA pre-processing.

Given the complex procedure underlying the reconciliation of different spatial scales inside the same dataset, a careful process has to be planned in order to allow for a coherent splitting procedure of training and testing data. Indeed, to be coherent, the mask derived after the correlation step has to be exclusively based on training data but equally applied to testing. On top of that, the PCA procedure applied in pseudo-NIPA must be exclusively applied to training data, and testing data have then to be reprojected in the same PCA space. This is because computing correlation or performing the PCA on the whole dataset and splitting afterward, intrinsically adds some latent information of the testing set in the training set and vice versa. This causes the final testing procedure to be not completely based on data that the model has never seen before. In addition to that, further issues can arise from the operational point of view since the pipeline is not going to be reproducible for new samples, and the training procedure has thus to be repeated.

To accommodate a dataset creation process that considers all these steps, the blueprint in Figure 21 has been planned.
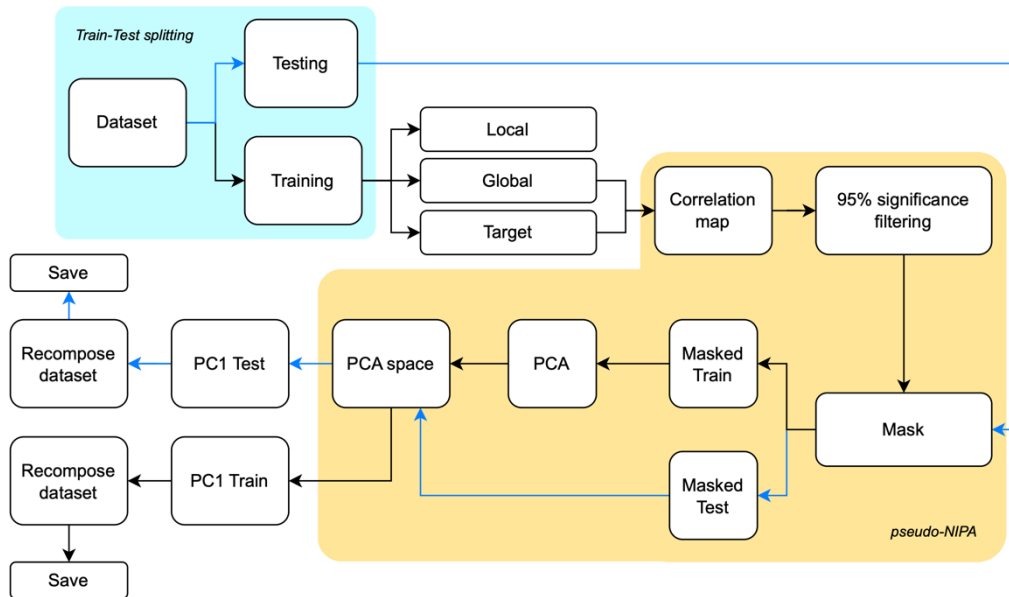
Figure 21: dataset creation blueprint

Firstly, once all the variables to be included in the dataset have been identified, the dataset is split into training and testing. At this point in the process, timeseries are still in their original form, meaning that global data are still 2D timeseries of gridded samples. The training data stream is then considered and further segmented into local data, global data, and target (i.e., cumulated precipitation in Rijnland). At that point, pseudo-NIPA computes the spatial correlation between global variables and the target in order to produce a correlation map subsequently filtered by applying a 95% significance level mask. The global variables belonging to the test data stream created at the beginning are now considered for applying a filtering procedure based on the just obtained mask. At the end of the masking step, there are still two different data streams: masked training data and masked testing data. The (global) masked training data are then passed into the PCA step of pseudo-NIPA and the first principal component is extracted. Thus, the testing data stream is now considered and projected in the previously determined PCA space for the extraction of the first principal component for testing. Finally, for both streams, global and local variables, now dimensionally matching, are reconciled to create the final dataset.

The final goal is to compare different ML algorithms coming from the different branches of the present work to gain some insight into which is a fair approach to

follow while creating a model for sub-seasonal precipitation prediction. Given this, and remembering that the branch described in section 4.2.1.1 is composed of 12 different models, appears evident that there is not an immediate fair comparison between the two settings. In order to enable the comparison of the single model against the twelve different monthly-based models, a proper structuring procedure of the test set has been embedded into the blueprint of Figure 21. As depicted in Figure 22, the training-test splitting divides the data into two main hyper-categories: "clean" and "dirty" samples. The term *clean* is used to reference samples averaged (or cumulated in case of tp) with the 30 days moving windows positioned such that it comprises the first day of a specific month and the following 29 days (i.e., it averages or cumulates the first 30 days of the month). Whatever other sample is instead said to be *dirty*.
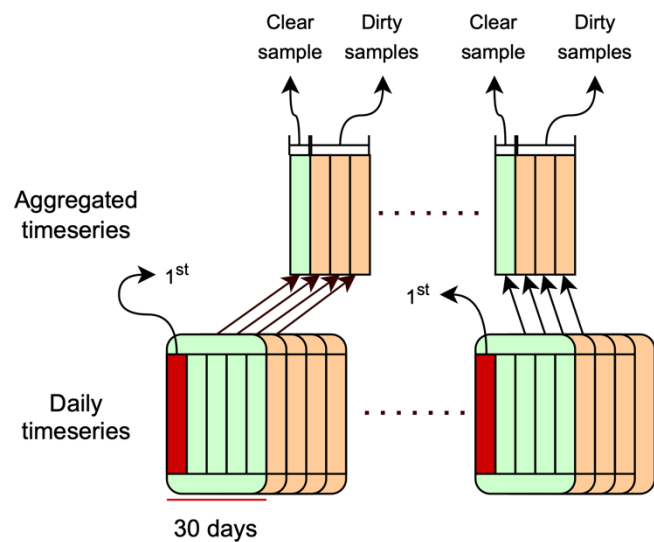

Figure 22: definition of clear and dirty samples after the 30 days moving windows aggregation process

It is clear that if the aim is to perform a month-by-month comparison of the two branches, the test set of the present branch has to be randomly sampled with a high-level pattern ensuring an even distribution of samples across all the months. At the same time, given that each sample of this branch is obtained through a moving average or cumulation, considering dirty samples could be misleading for the identification of the actual monthly-based performance of the model. Therefore, the monthly test procedure will be based on an equal number of clear samples spread

across the twelve months. The test set extracting procedure depicted in Figure 23 is straightforward: after the average/cumulation, each resulting sample is labeled with the day and month of the first element in the previously applied moving window. At that point, fixed at 80-20 the training-test split, for each month i :

- all the samples labeled with month i and day 1 are selected and 20 samples out of 42 are randomly extracted.
- all the samples labeled with month i and day j with $j \neq 1$ are selected, and 241 samples are randomly extracted.

That results in having the exact same number of clean and dirty samples for each month.

At the end of the training of the best-performing model, the testing procedure is therefore going to be twofold: on one side the "global" model performances over the whole year are tested with the entire test set, on the other, a month-by-month testing procedure based only on clear samples is performed to allow for the comparison with the branch introduced in section 4.2.1.1.
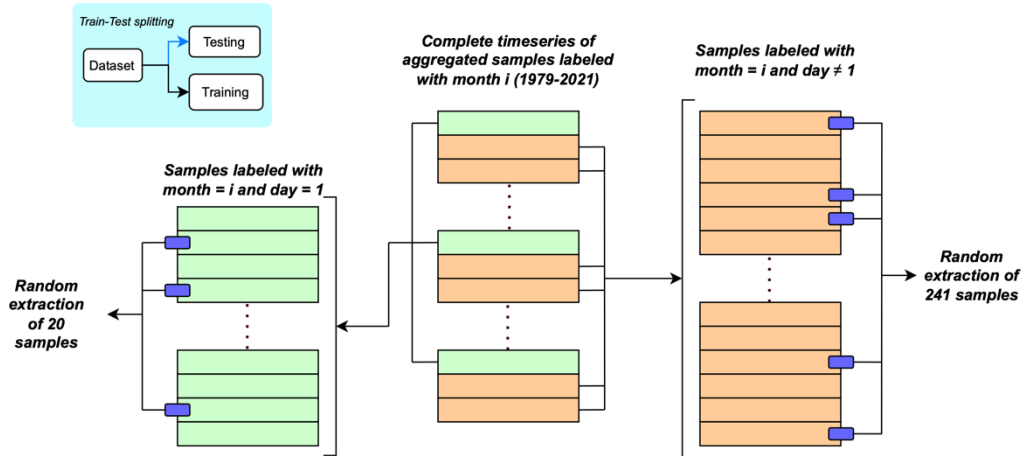

Figure 23: Training-Testing splitting procedure

The above-described dataset-building process is applied for all the combinations of variables with a number of features ranging from 5 to 10. Each of these datasets is then fit to a 2-layered FFNN whose hyperparameters are tuned according to a bandit-based optimization approach called Hyperband [64] based on the validation loss. The validation loss is computed on a further 20% detachment of samples from

the training data. High-performing hyperparameter settings are founded for the 2-layered FFNNs of each dataset, and the training of models based on these settings is performed. All these training processes are based on an early stopping condition monitoring validation loss to prevent it to rise after reaching its minimum. The best-performing model is finally selected by looking at the last epoch of the validation loss.

4.2.3 Convolutional Neural Network

When thinking of classical FFNN, their structure is typically composed of the main elements introduced in section 4.2.2. Thus, the main building blocks are essentially one input and one output layer, plus one or more hidden layers in between. The input layer takes an input vector X, the network performs some function F on it, and the output layer returns the output vector Y as depicted in Equation 10 [65].

Equation 10

$$F(X,W)=Y$$

In the Convolutional Neural Network (CNN) setting, although the general idea and the information flow do not change, the building blocks are different. A CNN model mainly consists of a concatenated sequence of convolutional layers and pooling layers. In this layered structure, the convolutional layer takes a single or multi-channeled image (e.g., grayscale or RGB) as input and applies the convolution operation by sliding on it a weight vector horizontally as well as vertically. This weight vector, most commonly named filter or kernel, is usually squared shaped (e.g., 3x3, 5x5, etc.) and, by sliding on the input image, produces the so-called feature maps [65]. Essentially, considering a squared kernel with dimension k, we can interpret the convolution operation as a single neuron sliding over the entire image and computing a weighted sum of all the pixels. The weights exploited in this operation are the value stored in the kernel itself. The just computed $k \times k$ weighted average is now a single value that will be stored in the feature map mentioned before. More than one kernel can be used in the same convolution layer, thus producing a number of different feature maps equal to the number of considered kernels. The output a_{ij} produced after a convolution operation and stored

in the feature position *(i,j)* of the produced feature map, is computed as in Equation 11:

$$a_{ij} = \sigma((W \star X)_{ij} + b)$$

where $\star$ represents the convolution operation, *X* the input, *W* the kernel weights, and σ the activation function.

At that point, the feature maps produced by the convolution layer are passed to the pooling layer. The general advantages of using this layer relate to the great reduction in the number of parameters and in the introduction of translational invariance [65]. The pooling layer simply takes the obtained feature maps, selects a window in each of them (i.e., pooling window), and applies a pooling function to the pixels of the feature maps lying in that window. Different pooling functions exist such as *MaxPooling* or *AveragePooling*. An example of the application of *MaxPooling* on a 4x4 feature map with a 2x2 pooling window is shown in Figure 24.
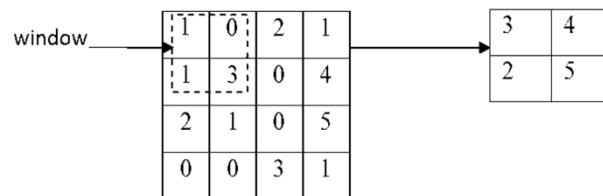

Figure 24: example of application of MaxPooling function.

Filter weights, as happens for connection between neurons in the context of FFNN, are pre-initialized. The initial CNN prediction performances, based on these pre-initialized sets of kernel weights, are expected to be low. The training process is aimed at adjusting these parameters thus allowing the model to learn an appropriate set of weights. Having a properly tuned set of weights enables the extraction of relevant features from the input images thus increasing prediction skillfulness. The learning process is performed in the same way as FFNN: the output of each layer is determined, and the error introduced by each of these outputs in the last layer is

computed. The network optimization is then executed by backpropagating the obtained gradients [65] as reported in section 4.2.2.

4.2.3.1 Case study application

The present work branch can be thought of as a mirroring of the FFNN branch introduced in section 4.2.2.1. Also in this case, indeed, the aim is to build a single model predicting cumulative precipitation with a 30 days lead time across all twelve months of the year. On top of that, since the CNN model training is highly computationally demanding, has been decided not to explore different combinations of variables as done in the case of FFNN. Therefore, CNN training will be exclusively performed on the set of features considered by the best-performing model resulting from the FFNN branch.

Although input variables are the same as the best FFNN, their structure is adapted to the new nature of the algorithm. A CNN, indeed, is meant to handle image data. In the present case, this translates in passing from considering local and global variables as unidimensional timeseries, to considering them in gridded form. Specifically: global variables are considered as gridded data covering the entire earth's surface with a resolution of $1.5° \times 1.5°$; local variables, that for this branch are going to be called "*medium-scale variables*", are considered as gridded data covering the European area with the same resolution of $1.5° \times 1.5°$. The aim of this branch is to map these global and medium-scale information to the same target as before: the 30 days cumulative precipitation in Rijnland. In the previously introduced branches, both NIPA and pseudo-NIPA were basically aimed at extracting the most relevant features from the global data image, and dimensionally reducing them with a PCA. The PCA output was then a unidimensional timeseries of global information to be used for FFNN training together with the local unidimensional timeseries. Since CNNs are meant to handle image data, the dimensionality reduction step is now no more needed, and no pre-processing framework such as NIPA or pseudo-NIPA is used. Since the set of variables used for the creation of the CNN model is derived from the best-performing FFNN, even in this case no explicit climatic information (i.e, climate index timeseries) is provided to the network. Reasonably, this can potentially increase the difficulty of the mapping process by leaving the model searching for a meaningful connecting

pattern between global and medium-scale variables. In order to compensate for the increased difficulty of the mapping task, the dataset for the present work's branch is built with the same process as the FFNN case. The only difference relates to the exploitation of image data. The dataset is indeed based on a 30 days moving aggregation (i.e., cumulation for precipitation and average for the other variables) sliding over the whole daily timeseries and producing a total number of samples greater than 15000.

The main idea behind the blend of removing explicit climatic information and increasing the number of available samples is the same as in the FFNN branch: to investigate if a considerably higher number of samples could counterbalance the complete absence of climatic information by allowing CNN to learn the underlying relationships solely based on medium-scale and global variables.

As mentioned before, differently from the FFNN branch, the extension of the medium-scale variable comprises now a wider area. Local variables for FFNN were indeed exclusively referred to the Rijnland grid cell, while medium-scale variables are now referred to the whole European area. The reason for that originates from the idea of gradually shifting from the global context to the local one passing through a medium-scaled bridging point. In addition, as mentioned in section 3.1, in the pool of information that the Rijnland waterboard inserts in the drought reports, also meteorological observations related to other specific locations are present. Therefore, considering medium-scale information allows for the implicit consideration of variables not only related to Rijnland but also to its wider surroundings.

The final goal is always to monthly compare the ML models coming from the different branches of the present work. To enable such a comparison the same structuring procedure of the test set used for FFNN has been implemented. Therefore, the data are divided into clean and dirty samples as shown in Figure 22, and randomly sampled with a high-level pattern ensuring an even distribution of samples across all the months as shown in Figure 23. The proportions are the same of the FFNN case.

Moreover, to enable hyperparameter tuning of the model built on the only provided dataset, a further 20% of data is detached from the training set for validation purposes. The hyperparameters tuning process is performed by exploiting Hyperband [64], the same optimization approach used for FFNN.

Given the different extents of global and medium-scale gridded data, the input variables of the CNN cannot simply be stacked and passed to the convolutional layers. On one hand, global data are presented with a shape of 121×240, on the other, medium-scale information covering the European area has a shape of 31×39. In order to reconcile these two different data extents, a double-steamed CNN setup was thought. Therefore, as shown in Figure 25, global and medium-scale variables are passed as input into two distinct CNN streams for each of which a customized hyperparameters setting is performed. Moreover, for the global data CNN stream, 4 distinct convolutional + max pooling layers are applied, for the medium-scale data CNN stream, instead, just 2 of these layers were used. Once both data streams have reached the end of their respective CNN, the final output feature maps are flattened and connected to a dense layer with n neurons to allow for a final merging of the features extracted from the global and the medium-scale context. Finally, the previous layer is connected to the last dense layer provided with a single neuron to produce the final prediction. In addition, in between the two dense layers, a dropout layer is inserted in order to add a regularization effect preventing possible overfitting phenomena. The operation performed by this dropout layer is to randomly set units connecting to the last layer to 0 at a frequency previously specified by a rate representing the fraction of the input units to be switched off (e.g., 0.1 translates in switching off 10% of the units). This process is performed at each training step and inputs that have not been switched off are appropriately scaled by *1/(1 - rate)* so that the sum over all inputs remains unchanged [66].

To adequately learn local weather patterns, it is important to provide CNNs with data standardized by accounting for the local spatial difference in variabilities [67]. In order to enable this accounting process, a proper standardizing procedure has to be applied. From this point of view, it is possible to identify two opposite standardization approaches:

- **Global standardization**: consists of applying a single value for both mean and standard deviation over the entire area. This approach could be problematic thinking about the huge differences that different regions of the world (e.g., equator and poles) can have with respect to the same variable.

That could bring poor accuracy, especially for CNN which is translation-invariant [3].

- **Pixel-wise standardization**: consists of applying different means and standard deviations coefficients pixel-wise thus having several sets of parameters equal to the number of pixels covering the area. This approach can translate into losing relevant information which would otherwise be represented through the local spatial differences in variability [3]
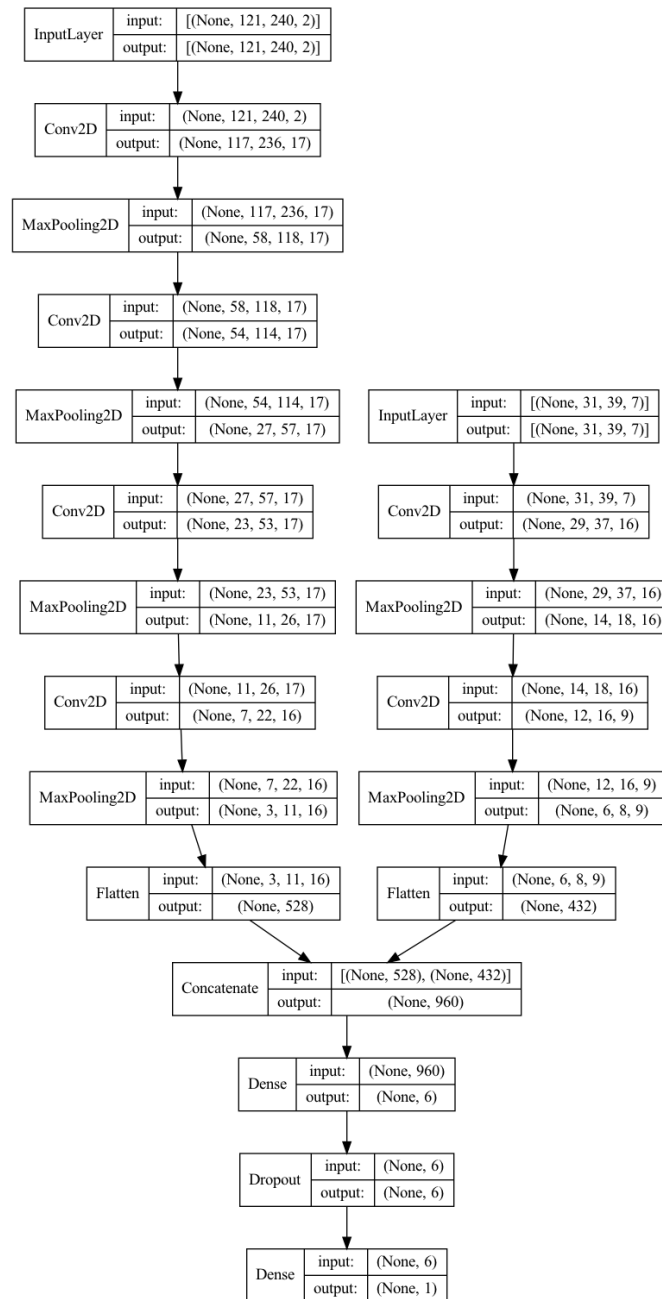

Figure 25: structure of the CNN model. For each layer input and output shapes are specified. "None" is referred to the batch size.

In order to mitigate the drawbacks of these two opposite approaches, the Local Area-wise Standardization (LAS) procedure proposed by Grönquist et.al [3] has been applied (Figure 26). The first step of LAS is to apply moving average and moving standard deviation filters to the training set. The filter's shape is matched with the shape of the kernel that will be used in the convolutional layers. The resulting mean and standard deviation maps have now reduced dimensions with respect to the original data. In order to make the two maps' dimensions match the data shape, LAS pads the latitude-related missing grid cells using edge values and wraps around the longitude [3]. Finally, a Gaussian filter is applied to the just obtained padded maps.

Essentially, sequencing the just mentioned padding and blurring steps, the LAS approach allows for conserving the spatial differences in variability between pixels.
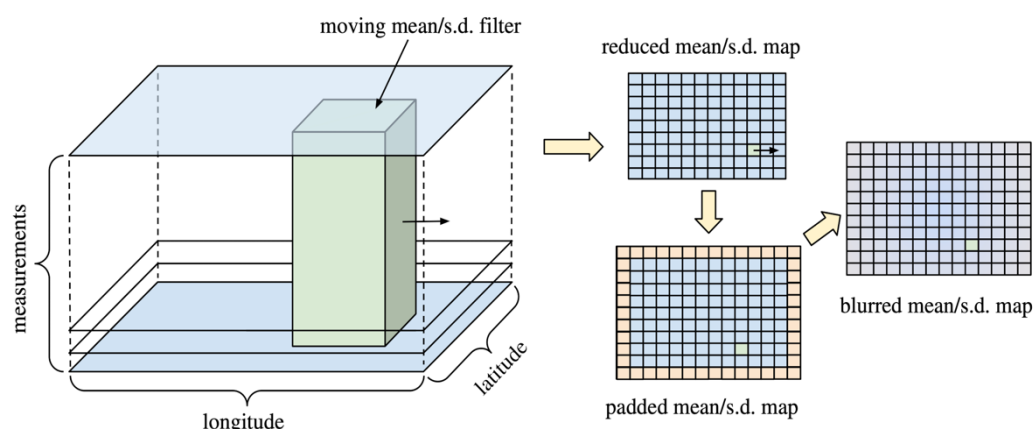

Figure 26: Local Area-wise Standardization (LAS) schematics. The depicted procedure is repeated two times, one for mean and the other for standard deviation filter thus obtaining two different "standardization maps" [54]

4.3 Drought forecasting with SPI

As anticipated in section 2.2.3 drought conditions can be estimated by exploiting drought indices. In general, the aim of these indices is to provide an interpretable picture of drought occurrence in a specific place, and to act as a key tool to transform raw data into meaningful information for an improved decision-making process. In order to create a direct mapping between the cumulated monthly-based precipitation predicted by all the previously described models, and the Rijnland drought conditions, the present work employs the Standardized Precipitation Index

(SPI). The SPI, often identified with "*SPI-n*" where *n* corresponds to the rainfall accumulation period, is a statistical indicator that compares the total precipitation registered in a given place during a period of *n* months, against the long-termed precipitation distribution for the same location and time frame [68].

The main reason behind the choice of the SPI index relates to its capability to be determined thanks to the sole usage of cumulative precipitation data. Therefore, no further information with respect to the model output is needed to predict the drought class in the context of the present work. On top of that, in 2010 the World Meteorological Organization (WMO) recognized SPI as a key meteorological drought indicator that should be operationally generated by every meteorological center [68], and it has been recently included in the drought monitoring reports produced by the water board of Rijnland. The SPI computation process is articulated in the following steps: first, a Probability Density Function (PDF) is fitted to the frequency distribution of cumulative precipitation referred to a specific place and cumulation time range (e.g., 1 month). Then, the just obtained probability distribution is passed into the standardized normal distribution with zero mean and variance one to obtain the SPI value (Figure 27). Based on the SPI values, seven different classes can be identified, each describing a different precipitation
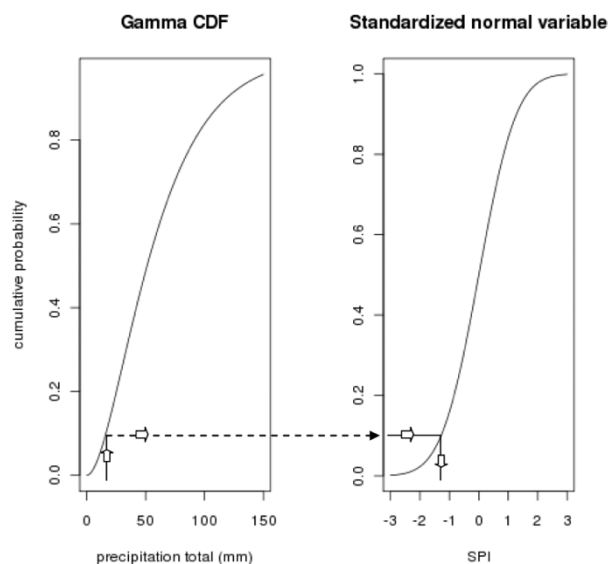

Figure 27: transformation of predicted precipitation into SPI class via Gamma Cumulative Distribution Function (CDF) and standardized normal variable [57]

condition. The SPI class is nothing but a value representing the units of the standard

deviation of the considered model output with respect to the long-term mean of the standardized distribution [68] (Figure 28).
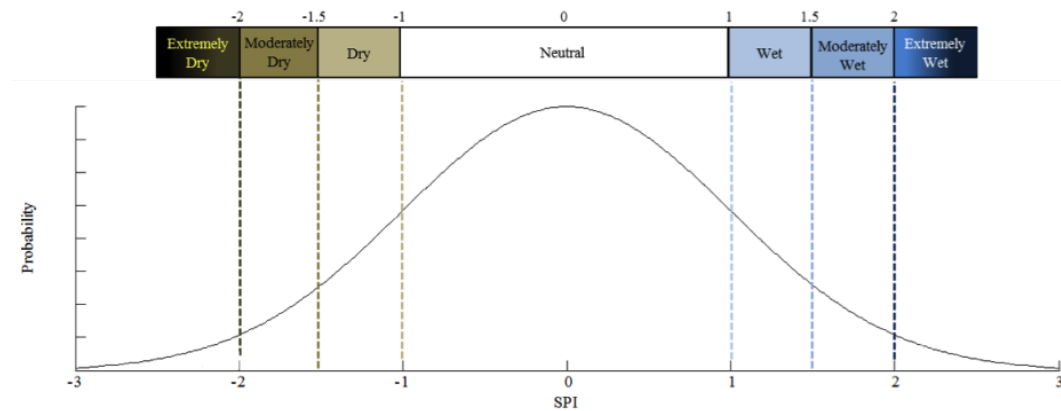

Figure 28: Relationship between SPI classes and normal distribution [71]

In the present case, to forecast drought conditions it is used the SPI-1, since the ML-based forecasts produced are referred to the total cumulative precipitation of the upcoming one month. Once all the best models for each of the three ML frameworks have been trained, the corresponding precipitation forecasts of the test set are used to assess the performances of drought forecasting by means of SPI-1. In this scenario, the ML-based forecasts, the target observations, and the ECMWF forecasts are converted to SPI drought classes and compared against each other to obtain the classification success rate for each month (i.e., number of correctly classified samples / number of total samples).

To do so, the operative procedure followed consists of (i) fitting a gamma distribution to the timeseries of ERA-5 observed 1-month cumulative local precipitation; (ii) saving the parameters of the gamma distribution ; (iii) computing the probability of the forecasted precipitation values by using the parameters of the gamma distribution previously saved; (iv) transform the gamma probability into the normal distribution with 0 mean and variance 1 to obtain the SPI value. The procedure is repeated on the target and on the predicted vectors of each model test set and for each month. The classes obtained by the target are then used as true labels and are compared with the classes obtained from the vector of prediction to compute the success rate (# correct classification / # total test samples).

5 Results and discussion

This section presents the results of the precipitation and drought forecasts obtained with the three different frameworks based on machine learning. In the following sections, the results of the ML-based approaches and a comparison of models performances are presented first, followed by a discussion on the influence of climate information on the forecasting capabilities of the models. Then, a comparison of the produced forecasts against the baseline ECMWF products is shown, followed by the assessment of the forecasts in terms of drought prediction and by a final discussion.

5.1 Machine Learning models' performances

In this study, three different machine learning models were created: a set of twelve ELMs models predicting cumulative precipitation of the upcoming month, for each month; a single FFNN model predicting cumulative precipitation across the entire year with a lead time of 30 days; a single CNN model predicting cumulative precipitation across the entire year with a lead time of 30 days. In the following sections, the results of the NIPA framework and in terms of machine learning models architecture are presented first, followed by a comparison of performances among all the models and a linear regression model.

5.1.1 Niño Index Phase Analysis and Climate State Intelligence

The Niño Index Phase Analysis (NIPA) has been applied as an input feature selection and dimensionality reduction method to detect climate teleconnection patterns that might be relevant to forecast one-month precipitation. The outputs of the framework are the principal components (one per each teleconnection phase) of the global variables correlated with the precipitation of the target location, during a specific teleconnection pattern phase.

As already discussed in section 4.1, in the application of NIPA, only the areas highly correlated with the precipitation of the target locations were considered to run the PCA on the global variables. More specifically, two different minimum correlation threshold levels were tested, first of ± 0.6, and second of ± 0.5. The application of these two different thresholds produced quite different results, both

in terms of correlated areas and of produced outputs, and hence of input feature combinations.

The 0.6 minimum threshold trial resulted in a total number of produced outputs of 34, meaning that more than 92% of the combinations were discarded due to a 3x3 contiguous area check failure. Looking at Figure 29 and keeping in mind that the aim is to produce one model for each month, it is possible to observe how specific months have little to no presence of NIPA output. This means that in the months in which there is no output at all (e.g., January, May) it is unfeasible to exploit any kind of global information for the model training, causing the bridging between the global and local context to be inexistent.

The coefficients in Figure 29, are Pearson coefficients coming from a linear correlation analysis between local precipitation and global variable. The reason why it has been decided to perform a second trial with a 0.5 threshold is precisely related to the linear nature of the Pearson coefficients. Indeed, the Machine Learning algorithm fed with the information coming from NIPA is going to search for relationships among predictors and predictand that could be highly non-linear. Therefore, imposing a high threshold could be unnecessarily severe, especially if resulting in a small to null pool of results. On the other side, completely removing the correlation threshold would increase the number of outputs issued by NIPA resulting in too many predictors among which choose to build the ELM models. Therefore, a middle ground approach has been applied.

Group 1 — Jan–Jun

Aggregation level 1

	Jan SST	Jan MSLP	Jan Z500	Feb SST	Feb MSLP	Feb Z500	Mar SST	Mar MSLP	Mar Z500	Apr SST	Apr MSLP	Apr Z500	May SST	May MSLP	May Z500	Jun SST	Jun MSLP	Jun Z500
SCA_N	0	0	-0,67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	-0,66	0	0	0	0	0	0	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAO_N	0	0	0	0	0	0	0	0	-0,67	0	0	0	0	0	0	0	-0,71	0,75
SCA_P	0	0	0	0	0	-0,69	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	-0,64	0	0	0	0	0	0	0
ENSO-mei_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAO_P	0	0	0	0	0	-0,64	0	0	0	0	0	0	0	0	0	0	-0,64	-0,65

Aggregation level 2

	Jan SST	Jan MSLP	Jan Z500	Feb SST	Feb MSLP	Feb Z500	Mar SST	Mar MSLP	Mar Z500	Apr SST	Apr MSLP	Apr Z500	May SST	May MSLP	May Z500	Jun SST	Jun MSLP	Jun Z500
SCA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,7	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAO_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SCA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,62	0
ENSO-mei_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NAO_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Aggregation level 3

	Jan SST	Jan MSLP	Jan Z500	Feb SST	Feb MSLP	Feb Z500	Mar SST	Mar MSLP	Mar Z500	Apr SST	Apr MSLP	Apr Z500	May SST	May MSLP	May Z500	Jun SST	Jun MSLP	Jun Z500
SCA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,7	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,68	0,72
NAO_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SCA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,65	0
ENSO-mei_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,69	-0,63
NAO_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Group 2 — Jul–Dec

Aggregation level 1

	Jul SST	Jul MSLP	Jul Z500	Aug SST	Aug MSLP	Aug Z500	Sept SST	Sept MSLP	Sept Z500	Oct SST	Oct MSLP	Oct Z500	Nov SST	Nov MSLP	Nov Z500	Dec SST	Dec MSLP	Dec Z500
SCA_N	0	0	0	0	0,75	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0,84	0	0	0	0	0	-0,7	0	0	0	0	0	0	0
ENSO-mei_N	0	0,78	0	0	0,71	0,7	0	0	0	-0,82	0	0	0	0	0	0	0	-0,79
NAO_N	0	0	0	0	0,81	0	0	0	0	0	0	0	0	0	0	0	0,74	0
SCA_P	0	0	0	0	-0,66	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0,69	0	0	0	0	0	0,68	0	0	0	0	0	0	0
ENSO-mei_P	0	0,66	0	0	-0,87	0,66	0	0	0	0,8	0	0	0	0	0	0	0	-0,66
NAO_P	0	0	0	0	0,73	0	0	0	0	0	0	0	0	0	0	0	0,69	0

Aggregation level 2

	Jul SST	Jul MSLP	Jul Z500	Aug SST	Aug MSLP	Aug Z500	Sept SST	Sept MSLP	Sept Z500	Oct SST	Oct MSLP	Oct Z500	Nov SST	Nov MSLP	Nov Z500	Dec SST	Dec MSLP	Dec Z500
SCA_N	0	0	0	0	0,72	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,73	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	0	0,65	0	0	0	0	0	0
NAO_N	0	0	0	0	0,84	0	0	0	0	0	0	0	0	-0,65	0	0	0,73	-0,67
SCA_P	0	0	0	0	0,65	0	0	0	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,66	0
ENSO-mei_P	0	0	0	0	0	0,76	0	0	0	0	0	0,7	0	0	0	0	0	0
NAO_P	0	0	0	0	-0,71	0	0	0	0	0	0	0	0	-0,66	0	0	0,74	0,72

Aggregation level 3

	Jul SST	Jul MSLP	Jul Z500	Aug SST	Aug MSLP	Aug Z500	Sept SST	Sept MSLP	Sept Z500	Oct SST	Oct MSLP	Oct Z500	Nov SST	Nov MSLP	Nov Z500	Dec SST	Dec MSLP	Dec Z500
SCA_N	0	0	0	0	0	0	0	0,66	0	0	0	0	0	0	0	0	0	0
EA_N	0	0	0	0	0	0	0	0	0	0	0	0	0	-0,77	0	0	-0,78	0
ENSO-mei_N	0	0	0	0	0	0	0	0	0	0	0	-0,7	0	0	0	0	0	0
NAO_N	0	0	0	0	-0,67	0	0	0	0	0	0	0	0	0	0	0	0	0
SCA_P	0	0	0	0	0	0	0	-0,76	0	0	0	0	0	0	0	0	0	0
EA_P	0	0	0	0	0	0	0	0	0	0	0	0	0	0,63	0	0	0,74	0
ENSO-mei_P	0	0	0	0	0	0	0	0	0	0	0	-0,71	0	0	0	0	0	0
NAO_P	0	0	0	0	-0,73	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 29: summary of the outputs issued by the NIPA run with 0.6 Pearson correlation threshold. The two groups of tables are organized in six main columns (one for each month) and three main rows (one for each possible value of the aggregation level parameter from 1 (top) to 3 (bottom)). Each table shows all the combinations between each phase of the considered climate indices and each global variable. Every cell shows the Pearson correlation value of a specific combination, if zero (red) NIPA has not produced the output for that combination, if different from zero (green shaded) the output for that combination has been produced.

Jan–Jun group

Aggregation level 1

	Jan SST	MSLP	Z500	Feb SST	MSLP	Z500	Mar SST	MSLP	Z500	Apr SST	MSLP	Z500	May SST	MSLP	Z500	Jun SST	MSLP	Z500
SCA_N	-0,66	0,8	0,74	0,78	0	-0,63	0,71	0	0	0	-0,63	0	0	0	-0,62	0	0	0
EA_N	0	0,63	0,7	0	0	0,68	0	0	0	0,7	-0,67	0	0	-0,54	-0,56	0	-0,73	0
ENSO-mei_N	0	0,7	-0,69	0,79	0	0,72	0	0	0	-0,7	0,55	0	-0,64	-0,58	-0,67	0	-0,58	0
NAO_N	0	0,63	-0,66	0	0	0	0	0,52	0,7	0,69	-0,58	0	0	0	0	0	0,72	0,73
SCA_P	0,68	-0,53	0,54	-0,76	0	0,69	-0,79	0	0	0	0,56	0	0	0	0,64	0	0	0
EA_P	0	-0,65	0,69	0	0	0,54	0	0	0	-0,75	-0,66	0	0	-0,69	-0,71	0	-0,57	0
ENSO-mei_P	0	0,7	0,71	0,74	0	-0,62	0	0	0	0,74	-0,76	0	0,83	-0,59	-0,7	0	0,73	0
NAO_P	0	-0,72	0,67	0	0	0	0	0,67	-0,67	0,79	-0,71	0	0	0	0	0	-0,68	0,65

Aggregation level 2

	Jan SST	MSLP	Z500	Feb SST	MSLP	Z500	Mar SST	MSLP	Z500	Apr SST	MSLP	Z500	May SST	MSLP	Z500	Jun SST	MSLP	Z500
SCA_N	0	0	0	-0,68	-0,69	-0,61	0	0	-0,55	0	0,71	-0,68	0	0,53	0	0	0,65	-0,53
EA_N	0	0,53	-0,71	-0,68	0	0,59	0	-0,53	0	0	-0,59	0	0	-0,55	0,61	0,76	0,83	-0,82
ENSO-mei_N	0	0,57	0,54	-0,77	-0,59	-0,56	0	-0,62	0,56	0	0,55	0	0	-0,55	0,66	0	0,63	0,61
NAO_N	0	0	0	-0,67	0	-0,73	0	0	0,67	0	-0,64	0,72	0	0	0	0	0,66	-0,62
SCA_P	0	0	0	-0,75	-0,61	-0,76	0	0	-0,63	0	0,59	0,61	0	0,65	0	0	-0,68	-0,61
EA_P	0	-0,69	-0,67	0,75	0	0,54	0	-0,64	0	0	0,74	0	0	-0,63	0,72	-0,8	0,66	0,72
ENSO-mei_P	0	0,66	-0,56	0,73	0,58	-0,75	0	0,61	-0,53	0	0,72	0	0	-0,75	0,71	0	-0,68	0,59
NAO_P	0	0	0	-0,77	0	-0,7	0	0	-0,55	0	-0,58	-0,7	0	0	0	0	0,58	0,72

Aggregation level 3

	Jan SST	MSLP	Z500	Feb SST	MSLP	Z500	Mar SST	MSLP	Z500	Apr SST	MSLP	Z500	May SST	MSLP	Z500	Jun SST	MSLP	Z500
SCA_N	0	0	-0,66	0,72	0	0	0	0	-0,63	0,59	0,69	0,71	0	0,6	0,65	0	0,62	0
EA_N	0	0,68	0,61	-0,63	0	0	0	0,51	0,57	0,8	0	0,66	0	0,67	0	0,77	-0,7	0,63
ENSO-mei_N	0	0,56	0	0	0,77	-0,71	0	0	0	0,67	-0,71	0,57	0	0,54	0	0	0,69	-0,71
NAO_N	0	0,56	0	0	0	0,66	0	0,61	0,6	0	0	0,69	0	0	0	0	0,67	0,72
SCA_P	0	0	-0,64	0,81	0	0	0	0	-0,62	0,75	-0,63	-0,67	0	0,59	0,57	0	0,58	0
EA_P	0	0,62	0,6	0,76	0	0	0	-0,6	0,54	-0,66	0	0,61	0	0,59	0	-0,71	-0,64	0,57
ENSO-mei_P	0	0,56	0	0	0,61	0,7	0	0	0	0,71	0,72	-0,72	0	-0,65	0	0	0,67	0,7
NAO_P	0	0,57	0	0	0	-0,66	0	-0,63	0,6	0	0	0,65	0	0	0	0	0,64	0,65

Jul–Dec group

Aggregation level 1

	Jul SST	MSLP	Z500	Aug SST	MSLP	Z500	Sept SST	MSLP	Z500	Oct SST	MSLP	Z500	Nov SST	MSLP	Z500	Dec SST	MSLP	Z500
SCA_N	0	-0,7	0	0	0,83	0	0	0	0,66	0	-0,61	0,77	0	-0,65	0	0	0,65	0,67
EA_N	0	0	0	0,68	0,83	0,59	0	0,53	-0,58	0	0,72	0	-0,67	0	0	0	0	-0,8
ENSO-mei_N	0	-0,77	-0,71	0	0,73	0,74	0	0	0,61	0,76	-0,7	0,67	0	0,74	0	0	0,69	0,78
NAO_N	-0,74	0,74	-0,6	0	-0,86	-0,78	0	-0,6	0,68	0	0,64	0	0	0,6	0,59	0,54	0,58	-0,63
SCA_P	0	-0,73	0	0	0,78	0	0	0	-0,64	0	0,59	0,6	0	0,55	0	0	-0,76	0,69
EA_P	0	0	0	-0,74	-0,84	-0,59	0	0,59	0,82	0	0,71	0	-0,59	0	0	0	0	0,68
ENSO-mei_P	0	0,64	0,64	0	0,84	0,68	0	0	-0,69	0,79	-0,6	0,67	0	-0,64	0	0	0,55	0,69
NAO_P	0,72	-0,74	-0,71	0	-0,79	-0,59	0	0,65	0,69	0	-0,62	0	0	-0,64	-0,66	-0,55	-0,71	0,65

Aggregation level 2

	Jul SST	MSLP	Z500	Aug SST	MSLP	Z500	Sept SST	MSLP	Z500	Oct SST	MSLP	Z500	Nov SST	MSLP	Z500	Dec SST	MSLP	Z500
SCA_N	0	0	0	0	-0,75	-0,67	0	0,56	0	0	0	0,65	0	-0,7	0,55	0	-0,68	0,67
EA_N	0,57	0,65	-0,72	0	0,69	0,7	0	0	0,73	0	0	-0,81	0,72	0,56	0,59	0	-0,7	0
ENSO-mei_N	0	0	0	0	-0,62	-0,78	0	-0,68	0	0	-0,59	-0,66	0	-0,63	0	0	0	0
NAO_N	0	0	0	0	-0,79	-0,77	0	0	0	0	0,62	0	0,59	0,63	-0,75	0	-0,66	-0,7
SCA_P	0	0	0	0	-0,7	0,69	0	-0,76	0	0	0	0,77	0	-0,58	-0,68	0	-0,67	0,57
EA_P	-0,68	-0,6	0,59	0	-0,61	-0,63	0	0	0,72	0	0	-0,67	0,67	0,7	0,67	0	0,65	0
ENSO-mei_P	0	0	0	0	-0,82	0,75	0	-0,57	0	0	0,63	-0,69	0	0,68	0	0	0	0
NAO_P	0	0	0	0	-0,75	-0,59	0	0	0	0	-0,56	0	-0,58	0,65	-0,65	0	-0,67	-0,77

Aggregation level 3

	Jul SST	MSLP	Z500	Aug SST	MSLP	Z500	Sept SST	MSLP	Z500	Oct SST	MSLP	Z500	Nov SST	MSLP	Z500	Dec SST	MSLP	Z500
SCA_N	0	0	0	0	-0,56	0	0,72	-0,67	0	-0,66	0	0	0	-0,62	0	0	0,74	-0,73
EA_N	0	0,67	0	0	0	0	0	-0,78	0	0	-0,75	0,81	0	0,85	0	-0,79	0,84	0
ENSO-mei_N	0	-0,7	0,58	0	0,59	-0,61	0	0,58	0	0	0	0,7	0	0,73	0	0	0,64	-0,56
NAO_N	0	0	0	0	0,58	0	0	0	0	0,76	0	-0,58	-0,59	-0,61	0	0	0,69	0,63
SCA_P	0	0	0	0	0,74	0	-0,62	-0,75	0	0,73	0	0	0	0,58	0	0	-0,66	-0,57
EA_P	0	-0,57	0	0	0	0	0	0,66	0	0	-0,59	-0,88	0	-0,6	0	-0,64	0,77	0
ENSO-mei_P	0	0,63	-0,56	0	0,87	-0,69	0	-0,85	0	0	0	0,67	0	0,68	0	0	-0,63	0,64
NAO_P	0	0	0	0	0,81	0	0	0	0	0,59	0	0,64	0,65	-0,63	0	0	0,67	0,71

Figure 30: summary of the outputs issued by the NIPA run with 0.5 Pearson correlation threshold. The two groups of tables are organized in six main columns (one for each month) and three main rows (one for each possible value of the aggregation level parameter from 1 (top) to 3 (bottom)). Each table shows all the combinations between each phase of the considered climate indices and each global variable. Every cell shows the Pearson correlation value of a specific combination, if zero (red) NIPA has not produced the output for that combination, if different from zero (green shaded) the output for that combination has been produced.

As can be seen in Figure 30, the 0.5 minimum threshold trial resulted in a total number of outputs equal to 212, meaning that almost 50% of the total number of producible outputs have been issued by NIPA. The reason behind such a rising in the number of produced outputs is relatable to the widening effect that the lower correlation threshold has on the correlated areas remaining on the map, as it can be seen from Figure 31. Wider correlated areas directly translate in greater chances to positively pass the 3x3 contiguous area check condition and produce the output.
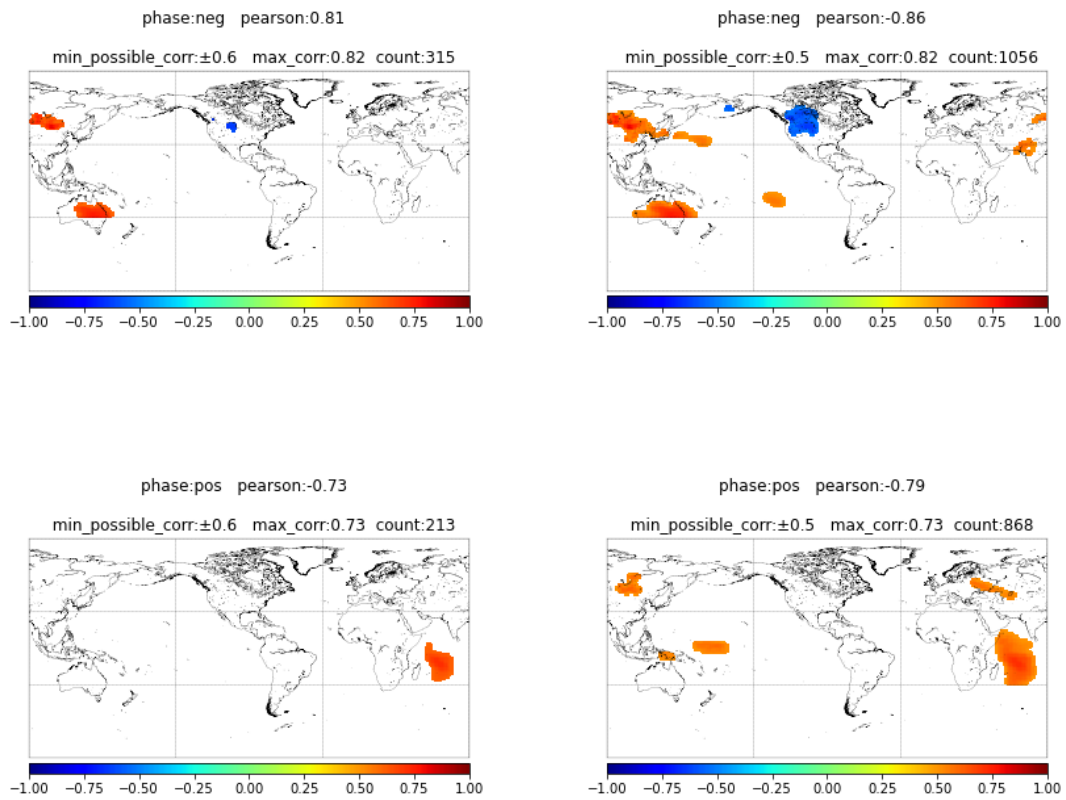
Figure 31: widening effect of the correlated areas by considering a lower correlation threshold. The NIPA inputs considered in the plot are NAO index and MSLP global variable with aggregation level 1 in August. Results based on 0.6 (left) and 0.5 (right) thresholds for both positive and negative phases.

The results obtained in the 0.5 trial are going to be the ones considered in the design of the ELMs models.

5.1.2 Extreme Learning Machine

In the approach based on ELM, twelve models were produced to forecast precipitation for each calendar month. Each model is characterized by a specific configuration in terms of the number of neurons, activation function, set of input features, which has been determined in the training and validation phase. The summary of the hyperparameters and best sets selected for each month is provided in Table 11.

Table 11: Hyperaparameters configuration of the 12 trained Extreme Learning Machines. In the "Dataset" column the global information is highlighted in orange and has the following syntax: climate index _ global variable – aggregation level. In the column named Activation, the different options for the activation functions are the sigmoid function (sigm) and the rectified linear unit (relu).

Month	Dataset	Neurons	Activation
Jan	t2m tp *EA_Z500-2 SCA_Z500-1*	8	*relu*
Feb	t2m TCWV *ENSO_MSLP-3 SCA_SST-3*	9	*sigm*
Mar	MSSHF *SCA_SST-1 NAO_MSLP-1*	10	*sigm*
Apr	tp *NAO_Z500-3 NAO_SST-1*	4	*sigm*
May	TCC TCWV *EA_Z500-2 ENSO_Z500-1*	8	*sigm*
Jun	*EA_Z500-2 NAO_Z500-1*	10	*sigm*
Jul	UW VW *SCA_MSLP-1 NAO_Z500-1*	8	*sigm*
Aug	*SCA_MSLP-1 NAO_MSLP-2*	9	*sigm*
Sep	MER *EA_Z500-2 NAO_Z500-1*	8	*sigm*
Oct	RH SH *EA_MSLP-1 ENSO-mei_SST-1*	8	*sigm*
Nov	SH *SCA_MSLP-2 EA_SST-1*	12	*sigm*
Dec	SD TCWV *NAO_MSLP-3 EA_MSLP-3*	8	*sigm*

5.1.3 Feed Forward Neural Network

For the approach based on FFNN, a single FFNN model predicting cumulative precipitation across the entire year has been produced. The best-performing model has been selected by building one 2-layered model for each possible combination of variables with a total number of features ranging from 5 to 10. The maximum number of input features was limited to 10 for two main reasons: (i) to reduce the computational effort. Indeed, a rising in the upper limit of considerable input features translates into more combinations to be considered and, therefore, more models to be trained; (ii) to maintain the number of input features in a reasonable range with respect to the dataset size to prevent overfitting. The hyperparameters of all these models are then tuned with the Hyperband implementation of keras-tuner [69] in order to determine the customized best setting for each set of input variables. The degrees of freedom allowed during the tuning process are reported in Table 12 together with their best set according to Hyperband.

Table 12: degrees of freedom for the hyperparameters tuning process of Feed-Forward Neural Network and best-resulting setting according to Hyperband

Hyperparameter	Degree of freedom type	Possible values	Best values
Activation function	*Choice*	*Relu, Sigmoid*	*Sigmoid*
Neurons layer 1	*Integer*	*5 - 30*	*30*
Neurons layer 2	*Integer*	*5 - 30*	*28*
Initial learning rate	*Choice*	*0.001, 0.01, 0.1*	*0.01*

After the execution of all the tuning processes, the best-performing model is selected based on the validation loss of the last training epoch. In order to enable this model selection approach, an "*Early stopping*" condition during training was
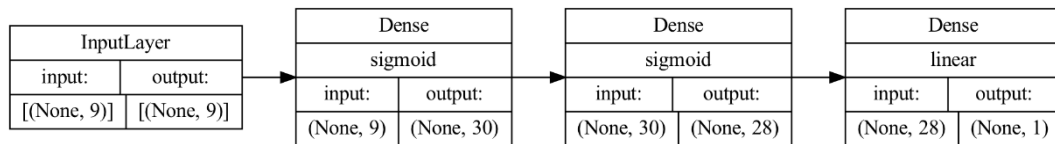
Figure 32: Feed-Forward Neural Network structure. "None" corresponds to the batch size, 9 to the number of input variables, 30 and 28 to the number of neurons and 1 to the final output

used to prevent the validation loss to rise after reaching its minimum. The structure of the best-performing model is depicted in Figure 32 and is based on the dataset composed by the variables specified in Table 13.

Table 13: Variables exploited for the training of the best Feed-Forward Neural Network model and their aggregation method

Variable acronym	Variable name	Aggregation method (30 days)
MSLP	Mean Sea Level Pressure	Average
Z500	Geopotential height at 500 hPa	Average
MSSHF	Mean Surface Sensible Heat Flux	Average
SH	Specific Humidity	Average
t2m	2m temperature	Average
TCC	Total Cloud Cover	Average
tp	Total Precipitation	Cumulation
UW	U component of wind	Average
VW	V component of wind	Average

82

5.1.4 Convolutional Neural Network

In the approach based on CNN, a single model predicts cumulative precipitation across the entire year. As already mentioned in section 4.2.3.1, the best hyperparameter setting was searched with respect to the same set of variables provided in Table 13, hence no further investigation has been carried out to determine the best set of input features. The degrees of freedom for the search are in this case mainly referred to kernel number and size, learning rate, dropout rate, activation function, and number of neurons of the last but one dense layer (CNN structure is provided in Figure 25). Their summary, as well as the best-resulting setting, are reported in Table 14.

Table 14: degrees of freedom for the hyperparameters tuning process of Convolutional Neural Network and best-resulting setting according to Hyperband

Hyperparameter	Degree of freedom type	Possible values	Best values
Activation function	*Choice*	*Relu, Sigmoid*	*Relu*
Filter size - global	*Choice*	*3, 5, 7*	*5 (5x5)*
Filters ConvLayer 1 - global	*Integer*	*1 - 20*	*17*
Filters ConvLayer 2 - global	*Integer*	*1 - 20*	*17*
Filters ConvLayer 3 - global	*Integer*	*1 -20*	*17*
Filters ConvLayer 4 - global	*Integer*	*1 - 20*	*16*
Filter size - global	*Choice*	*3, 5, 7*	*3 (3x3)*
Filters ConvLayer 1 - medium	*Integer*	*1 - 20*	*16*
Filters ConvLayer 1 - medium	*Integer*	*1 -20*	*16*
Filters ConvLayer 1 - medium	*Integer*	*1 -20*	*9*
Dense layer neurons	*Integer*	*5-10*	*6*
Dropout rate	*Choice*	*0.1, 0.2, 0.3, 0.4, 0.5*	*0.1*
Initial learning rate	*Choice*	*0.0001, 0.001, 0.01, 0.1*	*0.0001*

5.1.5 Comparison against a naïve model

Once the three approaches are developed, their performance is compared against each other but also against the performance of a naïve linear regression model, built

exploiting the exact same datasets, in order to understand if the higher complexity introduced by ML models is justified by a performance gain. Since the ELM setting is composed of monthly-based models, the comparison will be performed on a monthly basis. As shown in Figure 33, both FFNN and CNN models are meant to produce predictions across the entire year. In this case, as mentioned in section 4.2.2.1, an even set of monthly-based samples have been randomly extracted to perform, in addition to the proper test, also a monthly-based testing procedure for model comparison.
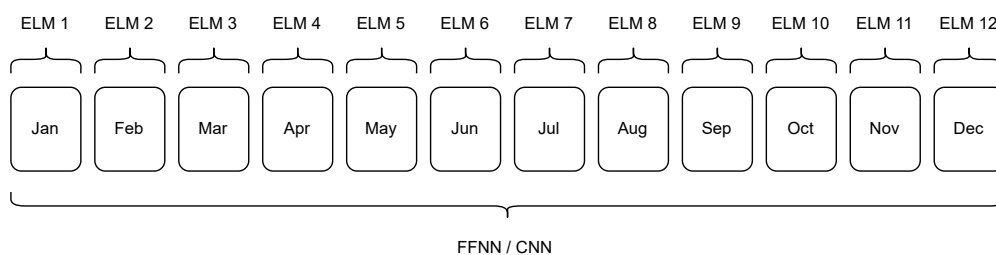
Figure 33: temporal coverage of the three different branches of trained Machine Learning models: ELMs, FFNN and CNN.

Table 15 reports all the model performances in terms of Mean Squared Error (MSE) with respect to the prediction of cumulative local precipitation [mm]. Thirteen different linear models to be compared against their ML counterparts have been built.

Specifically, twelve of them are monthly-based models built on the same dataset as the twelve ELMs. As mentioned in section 4.2.1.1, given the restricted pool of available samples due to the pre-processing step performed by NIPA, the ELMs performances reported in Table 15 have been obtained with a LOOCV procedure. In order to allow for a fair comparison with linear models, the exact same procedure was performed to determine also their performances. The comparison of the two groups of models highlights how the whole set of ELMs outperforms all the linear models ensuring that the relationship bonding predictor and predictands can be skillfully mapped by exploiting models as ELM.

The remaining linear model, one for the entire year, is built on the same dataset of the FFNN with 30 days moving aggregation. In this case, as mentioned in section 4.2.2.1, the dataset was produced based on a high-level random sampling pattern ensuring an even distribution of samples across all the months. This splitting

approach enables a monthly-based testing procedure both for ML and linear models. The performances of the linear model are compared against FFNN and CNN since they both share the same set of input variables. Starting from the FFNN branch, it outperforms the linear model both in the yearly and monthly testing domain, ensuring that the exploitation of an ML model in this setting has been effective. For the CNN branch, instead, although most of the tests resulted in outperforming the linear model, the set of months comprising November and June obtained better results in the linear case. It must be noted also how in that case, given the complex nature of the data used for the CNN training (gridded data covering different spatial extensions), the benchmarking has been performed against a linear model built on the FFNN dataset. Therefore, although it is built on the exact same set of variables as the CNN, the linear model exploits 1D timeseries instead of gridded data.

Table 15: model's MSEs across the twelve months. For yearly-based models (FFNN, CNN and related Linear model) also the global MSE is reported. MSE for ELM and related linear models are computed with a LOOCV procedure, for the remaining models a proper testing procedure with a separated set has been performed

Testing period	ELM	Linear models	FFNN	CNN	Linear model
Year	/	/	538.19	796.72	921.39
Jan	345.35	583.49	555.04	641,81	1379.66
Feb	189.20	220.80	308.45	856,94	951.66
Mar	344.41	546.88	512.22	723,75	763.67
Apr	261.36	362.45	327.65	603,9	1139.72
May	293.82	450.56	409.15	709,47	840.51
Jun	242.66	412.92	495.77	947,97	743.96
Jul	539.56	730.55	978.36	1176,39	1716.77
Aug	333.38	453.24	353.21	1085,39	1166.03
Sep	340.40	393.58	719.84	539,66	1332.39
Oct	271.95	305.27	547.09	909,47	1415.54
Nov	214.30	266.91	189.62	1027,37	386.66
Dec	234.29	359.46	450.75	532,54	1145.63

The comparison between ML models shown in Figure 34 at the bottom-left enables the following considerations. On average it is evident that FFNN results are worse than ELMs ones. The reason for that probably originates from the fact that the

FFNN branch has been thought of as one single model predicting precipitation across the entire year. This reasonably increases the difficulty encountered by the model in mapping inputs with target mostly because the generalization of the intrinsic variability of each month seems to be difficult to reach with a single model. On top of that, ELM models find much of their skillfulness inside the tailored monthly-based dataset configuration which sensibly reduces the error by considering a set of few and only relevant features for each specific month. This setting is of course inapplicable for the single FFNN model that, instead, has to rely on a fixed set of features to produce predictions all across the year.
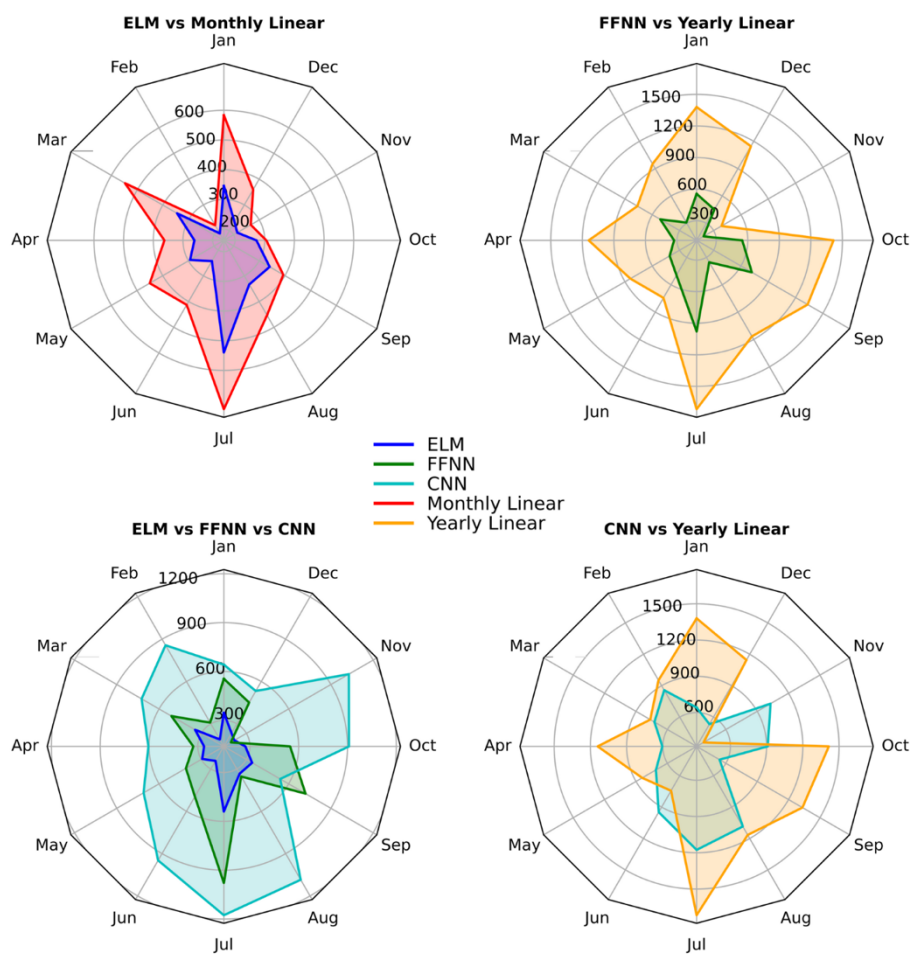
Figure 34: comparison of the MSEs of each of the three ML models with each other (bottom left plot) and with their respective linear counterpart (rest of the plots). NOTE: the scale of the top-left and bottom-left plots are different from the others

The results obtained with CNN are on average worse than the ones obtained with the other ML models. A potential reason for that relates to the blend between the number of considered variables and the data format used for CNN training. As

acknowledged all across the scientific literature, CNNs are known for the sensible reduction in the number of needed parameters with respect to FFNN. This relates to the fact that, essentially, considering a squared kernel with dimension k, allows for interpreting the convolution operation as a single neuron sliding over the entire image and computing a weighted sum of all the pixels within the $k \times k$ kernel. This implies that the weights stored in the kernel are kept constant after the training process, meaning that there is no constraint of having a weight for each pixel of the image as would be to FFNN. Although this is factual, in the present case, while shifting from FFNN to CNN, the algorithmic change was accompanied by a complete dataset restructuring. Indeed, although the considered variables are exactly the same, the data structure exploited for the FFNN training was composed of unidimensional timeseries. This means that each variable, either global or local, was represented by a list of float values each one referring to a different time range. In the case of CNN instead, each sample was composed of a grid storing a float value for each location falling in it, and each of these values was therefore simultaneously referred to a specific location and time range. In other words, the two models are referred to the same set of variables but not to the same data, resulting in an overall increase in the number of features considered in the CNN with respect to the FFNN model. In addition, this features increase is not backed by any kind of increment in the number of samples. Therefore, the only way through work's consistency was (i) to test if the number of available samples was appropriate for the training of a CNN model with a considerably higher number of features and, if not, (ii) to search for the best-performing hyperparameters configuration that could be obtained with that specific setting. The best configuration is reported in Table 14 and the testing MSEs are shown in Table 15.

Looking at the results it can be noted how ELM outperforms all the other settings for all the months but November, which, instead, shows a lower MSE in the FFNN setting. Despite that, ELM can be identified, on average, as the best model setting for the present work. Moreover, looking at Figure 34, it is possible to recognize a common error pattern across the months for both ELM and FFNN. This can be further confirmed by looking at the MSE reported in Table 15. For the ELM case, indeed, the set of four worse results comprises July, January, March, and September. On the other side, FFNN presents his three worse results in July,

September, and January, highlighting a similar but scaled error distribution across the year.

The CNN setting presents worse results than ELM for every month, and worse results than FFNN for all but one month. The only exception is referred to September, which, as mentioned, is also one of the worse results of the entire FFNN setting.

Figure 35, Figure 36, and Figure 37 respectively depict monthly plots of predicted against observed precipitation on monthly-basis for ELM, FFNN, and CNN.
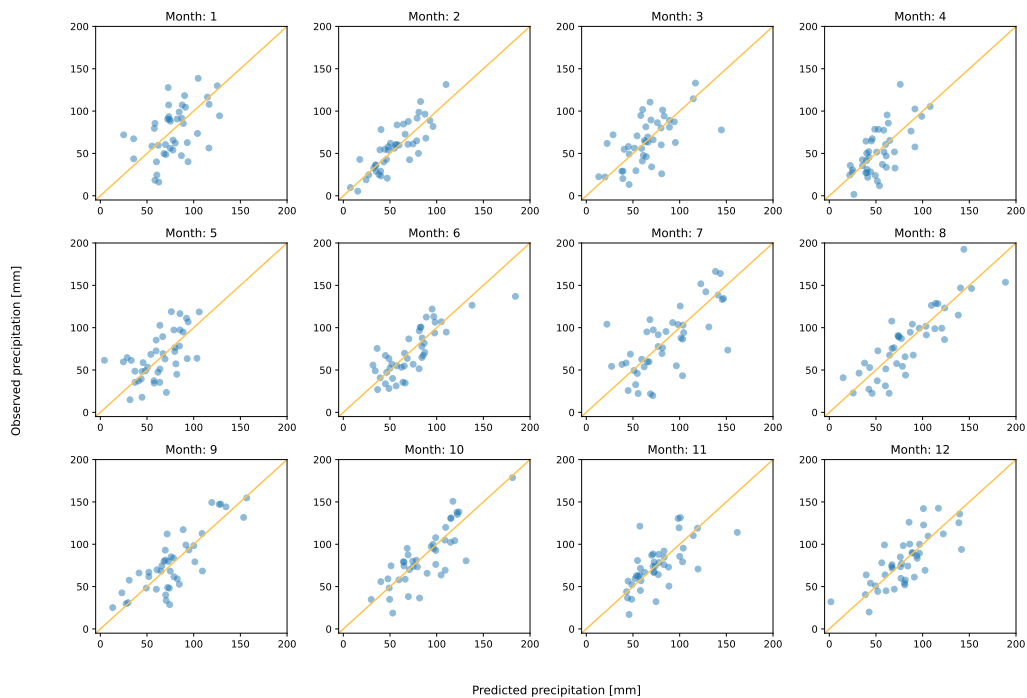
Figure 35: ELMs Leave One Out precipitation predictions against true observed precipitation values. Each plot is referred to a specific month.
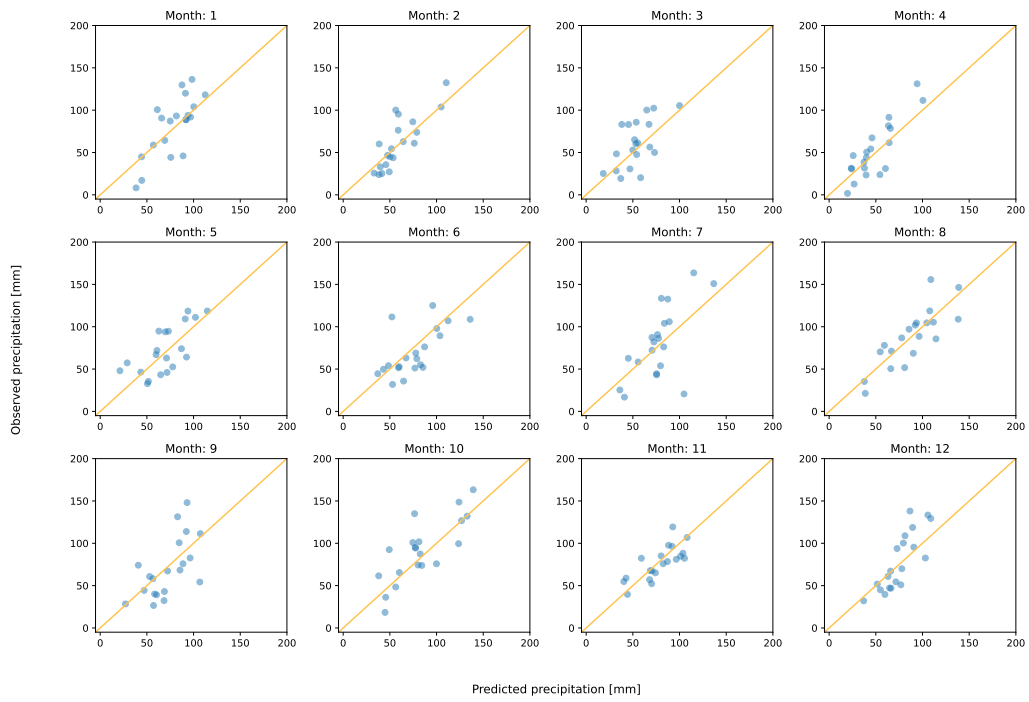
Figure 36: FFNN testing precipitation predictions against true observed precipitation values. Each plot is referred to a specific month.
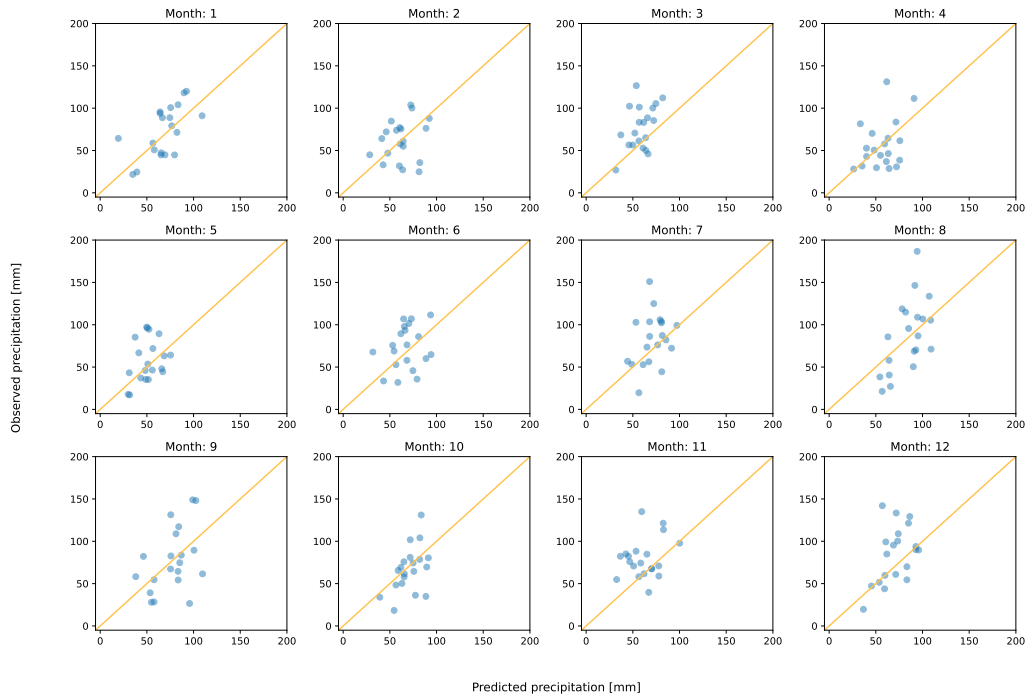
Figure 37: CNN testing precipitation predictions against true observed precipitation values. Each plot is referred to a specific month.

5.2 Influence of climate data and local atmospheric data

The model selection processes performed on the different model branches resulted, in most cases, in simultaneously accounting for both global climatic and local atmospheric data. For instance, the model selection process performed on the FFNN branch has been planned to leave a degree of freedom also related to the type of considered variables. Indeed, no constraint was present based on the variables that should be used for training, and, therefore, the choice of training a model exclusively based on global or local variables was a possibility. Despite that, the best-performing model of the FFNN branch resulted in being based on a blend of information coming from both the global and local contexts, highlighting a certain convergence of the process toward datasets accounting for both information. This tendency was also confirmed by the ELM branch, where, for most of the months, the exploited training sets were merging local information and global data coming from NIPA.

Although this tendency is on average quite evident, exceptions still arise in the context of monthly-based ELMs. In this case, indeed, two months resulted in being addressed by ELMs models exclusively based on global climatic information coming from NIPA: June and August. As can be observed from Table 11, the June ELM is trained on a set of two predictors (out of four possible) coming from NIPA. These two considered features, on one side, are based on two different climate indices (EA and NAO) and on two different aggregation levels (2 and 1 month(s) before the target), on the other, are based on the same global variable (Z500). Therefore, this setting shows how, apart from the variation in the considered climate index, just a single variable has been considered within the pool of the 13 possible global and local predictors. This behavior suggests the other considered variables as not relevant for the prediction of precipitation in June. Enforcing this line of reasoning, looking at Figure 34, the best-performing FFNN model, which considers Z500 among its two global features, presents performances aligned with the ELM branch (but scaled due to the error introduced by the yearly-based nature of the model).

For the August ELM, the exact same pattern is repeated, this time by considering MSLP as the global feature (with NAO and SCA as climate indices, and 2 and 1 as

aggregation levels). Even in this case, FFNN performances are pretty aligned with the August-based ELM, and the second global feature considered in the best-performing FFNN model is MSLP.

At the same time, in the bottom-left part of Figure 34, it can be noted as for both the June and August ELM models, the predictive skillfulness is higher compared to the FFNN counterpart. This can be relatable to two main reasons:

- The NIPA capability of embedding into global data also information related to different climatic patterns. Indeed, the consideration of climatic patterns can act as a bridging point for the mapping between global variables and local precipitation (target).
- The yearly-based nature of the FFNN branch that increases the generalization difficulty by fixing a set of features to generate predictions across all twelve months.

Always during the summer season, also July shows another interesting behavior. In this case, the MSE of both models increases if compared to the surrounding months. If looking exclusively at the July-based ELM, it is possible to note how the number of features matches the previously fixed maximum limit of allowed predictors for a single model (four). This, blended with the relatively high MSE (July produced the highest MSE of the ELM branch), pushes toward the idea that, in that specific month, an improvement in performances could be reached by considering a higher number of more informative features. Clearly, this cannot be done in the ELM context due to reduced dataset dimension and, therefore, to the risk of overfitting that adding further predictors brings. However, the consideration of a higher number of features can be performed in the FFNN case. By considering the July-based testing results of the FFNN case, it is possible to realize how performances show the same pattern as the ELM counterpart (even if, again, scaled due to the error introduced by the yearly-based nature of the model). This enforces the idea that, among the pool of considered variables, some relevant July-related predictors are missing.

5.3 Comparison against baseline forecast products

The benchmarking of the obtained results was done with respect to the ECMWF Extended-range forecasts providing cumulative precipitation predictions with a lead time of 30 days. As can be observed from Table 16 and Figure 42, all the trained ML models show sensible improvements with respect to the ECMWF benchmark for all the months. Specifically, ELM models stood out for all but one month as the best-performing models of the work. Just in November, the FFNN seems to be the most skillful.

Table 16: comparison of the MSEs of the ML and linear models with respect to the ECMWF extended-range forecasts. The green column contains the ECMWF MSEs. Cells highlighted in orange are referred to the best-obtained results for each specific month.

Testing period	ELM	Linear models	FFNN	CNN	Linear model	ECMWF subseasonal model
Jan	345.35	583.49	555.04	641.81	1379.66	1103.93
Feb	189.20	220.80	308.45	856.94	951.66	930.93
Mar	344.41	546.88	512.22	723.75	763.67	550.90
Apr	261.36	362.45	327.65	603.9	1139.72	958.27
May	293.82	450.56	409.15	709.47	840.51	752.60
Jun	242.66	412.92	495.77	947.97	743.96	812.00
Jul	539.56	730.55	978.36	1176.39	1716.77	2201.04
Aug	333.38	453.24	353.21	1085.39	1166.03	3702.95
Sep	340.40	393.58	719.84	539.66	1332.39	1626.13
Oct	271.95	305.27	547.09	909.47	1415.54	2310.86
Nov	214.30	266.91	189.62	1027.37	386.66	1313.64
Dec	234.29	359.46	450.75	532.54	1145.63	1922.19

In Figure 38 and Figure 40 the comparison between ELM and ECMWF predictions is provided. Specifically, Figure 38 shows the prediction trend comparison of ELM and ECMWF for all the June instances across the period 2001-2020. The figure is divided into two different plots because of the different targets of ELM and ECMWF. ELM target is indeed the cumulative monthly precipitation, while ECMWF target is the 30 days cumulative precipitation. Although the target difference, the filling between prediction and target curves in the two plots highlights how ELM seems to be slightly more adherent to its target than ECMWF.

The same behavior can be observed in Figure 40, which depicts the same comparison of Figure 38 but focused on July.

Figure 39 and Figure 41, instead, show the comparison of the ECMWF model respectively against FFNN and CNN. In this case, since the dataset of FFNN and CNN originates from a 30 days moving aggregation window, their target coincides with the ECMWF one. Therefore, only one plot per month (June and July) is created and the three trends are directly compared.
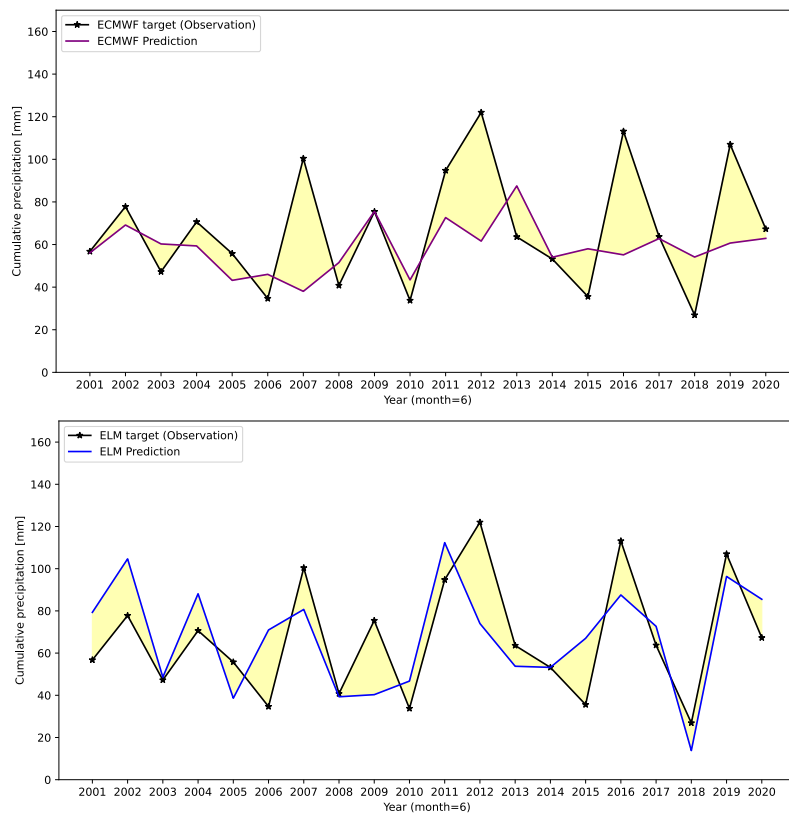
Figure 38: June prediction trend comparison: ECMWF / ECMWF target (top) and ELM / ELM Target (bottom)
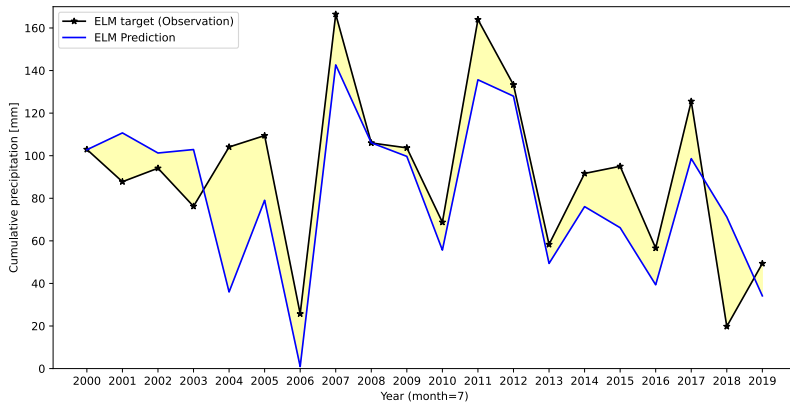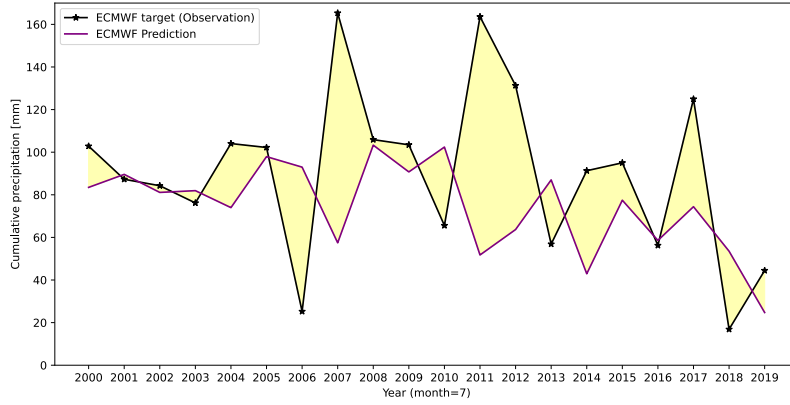
Figure 40: July prediction trend comparison: ECMWF / ECMWF target (top) and ELM / ELM Target (bottom)
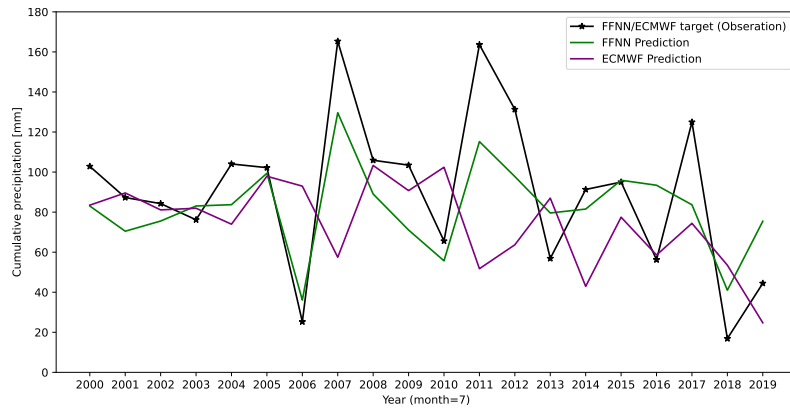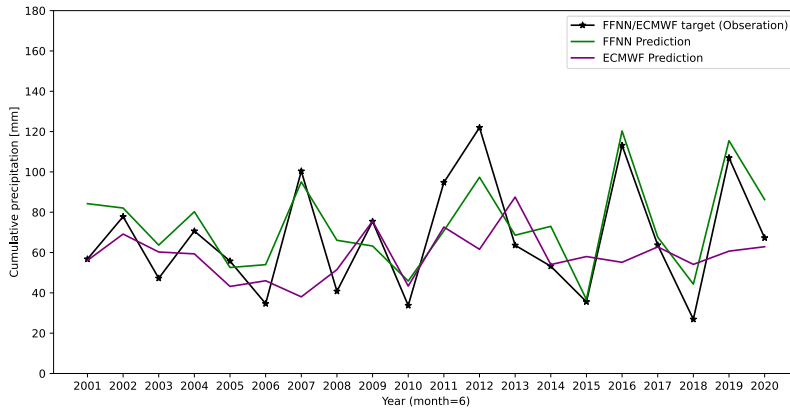
Figure 39: Prediction trend comparison: FFNN / ECMWF/ target. From top to bottom the plots are referred to June and July
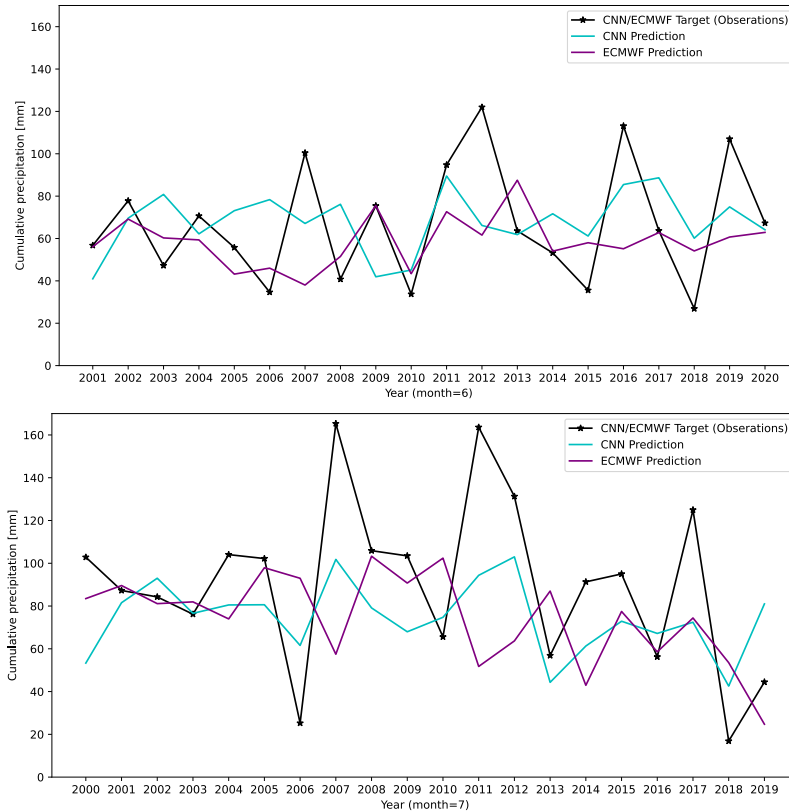
*Figure 41: Prediction trend comparison: CNN / ECMWF/ target. From top to bottom
the plots are referred to June and July*

June and July have been considered for the previous plots since they respectively represent the best and worse ML settings obtained in the range April-September (the range of interest for the Rijnland waterboard). Furthermore, the 2001-2020 range of the plot originates from the ECMWF model instance considered in the present work. The considered ECMWF model, indeed, provides hindcast values only in that time range.

Another interesting aspect to analyze is the comparison between linear models and ECMWF forecasts. Starting from the monthly-based linear models of the ELM branch, it is possible to note how they have lower MSEs than ECMWF forecasts across all the months. Even the yearly-based linear model setting shows better or aligned performances. The reason behind these results originates from considering different sets of variables tailored to the case study for the creation of the linear models. The tailored set of variables provided the models with meaningful and customized features thus allowing them to outperform the ECMWF benchmark.
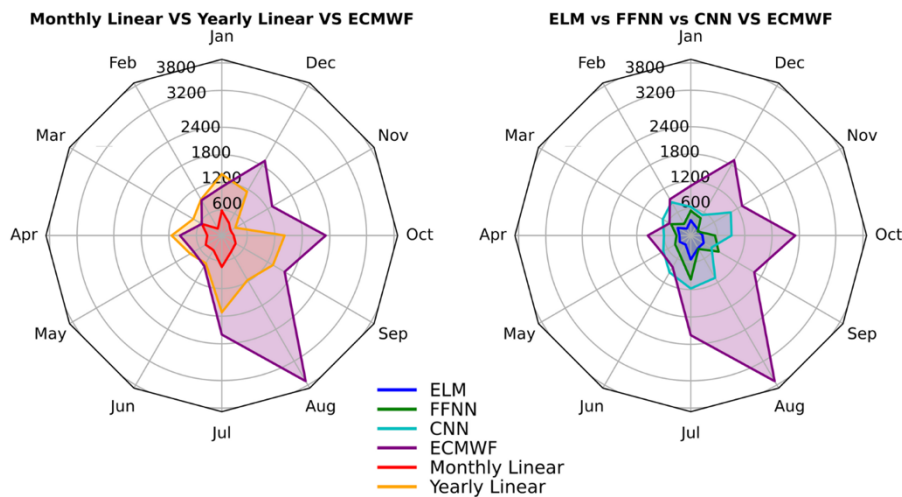
Figure 42: comparison of the MSEs of ML and linear models with ECMWF Extended-range forecasts (right and left charts respectively)

5.4 Performances of drought forecasting

Exploiting the validation set of ELM, the test sets of FFNN and CNN, and the ECMWF prediction values, the methodology specified in section 4.3 was applied to compute the classification success rate. The rates of each model (columns of Table 17) are computed by comparing the drought classes obtained from the prediction with the ones obtained from the related target.

In this case, the pattern of the best rates is more scattered across the branches with respect to the best MSEs of the cumulative precipitation presented in the previous section. Despite that, even in that case, the ELM setting seems to behave slightly better in terms of monthly predominance by outperforming the rest of the models for 6 months out of 12. Just after, the CNN setting presents the best results for March, April, July, and October and, finally, FFNN has produced the best rates for both January and August. Although the scattered results, the predominance of the ML models with respect to ECMWF is confirmed also in the drought prediction for all the months.

Table 17: Rates of correct classification with respect to SPI categories for each month and model. The rate's value ranges between 0 and 1 (where 1 corresponds to all correct classifications). Highlighted in orange is the best rate for each month; highlighted in green is the column containing the benchmark's rate

month	ELM	FFNN	CNN	ECMWF
Jan	0,41	0,55	0,45	0,35
Feb	0,57	0,35	0,45	0,4
Mar	0,46	0,4	0,5	0,45
Apr	0,4	0,3	0,45	0,25
May	0,46	0,45	0,25	0,35
Jun	0,51	0,35	0,2	0,45
Jul	0,43	0,45	0,65	0,5
Aug	0,49	0,6	0,35	0,25
Sep	0,46	0,3	0,35	0,3
Oct	0,51	0,55	0,6	0,3
Nov	0,58	0,55	0,3	0,35
Dec	0,57	0,4	0,5	0,4

5.5 Discussion

The idea behind the creation of the three different model branches was based on the research for the correct balance among three main elements: model's temporal domain, number of samples, and climatic information. On one side, ELMs were built on a short temporal domain, with few samples, and with explicit providing of climatic information by means of NIPA. On the other, FFNN and CNN were based on a much wider temporal domain and on a bigger dataset without climatic information. Each of the two configurations presents some pros and cons from the theoretical point of view. The ELM setting can benefit from the monthly-based configuration that enables the set of variables to be customized and more informative for each specific month. At the same time, the low number of samples composing the dataset can act as an obstacle to generalization. The FFNN and CNN settings were aimed at investigating the possible counterbalancing effect of having a much wider dataset for a much wider model's temporal domain.

The results discussed in this section highlight how both for FFNN and CNN, the sensible increase in the dataset size has not resulted in performance gaining if compared to the ELM branch (best-performing setting). The reason for this can be articulated in two main points:

- The yearly-based nature of FFNN and CNN models. Because it reasonably increases the difficulty of finding a proper set of input features able to generalize the problem. Therefore, the increased dataset size seems to not counterbalance the drawback introduced by keeping the set of predictors constant all over the year.

- The absence of climatic information and of a proper climatic pre-processing framework as NIPA. Climate index data, as specified in section 2.3, could play a fundamental role in sub-seasonal precipitation forecasting since they act as bridging points between the local and the global climate context. The not explicit provision of that information during the training process leads models such as FFNN and CNN to not fully capture these links even with sensibly increased dataset sizes.

Another interesting behavior can be derived from the bottom-left plot of Figure 34. Looking at the radar plot shape of the CNN model, it can be noted that it differs from the much more similar ELM and FFNN shapes. The reason why could be that, although the set of considered features is the same as FFNN, CNN exploits gridded data and an algorithmic structure that is conceptually more different. This brings the distribution of the model's MSEs across the months to not follow the same pattern presented by the two other model branches.

At the same time, ELM and FFNN seem to produce a similar but scaled distribution of MSEs. Even if the input features considered in the two branches are not the same, their structure is much more similar with respect to the CNN data. In both these two cases, indeed, variables are provided in the form of unidimensional timeseries. The wider FFNN error area is probably linked to the error added by the yearly-based nature of the model and to the absence of climatic information.

Globally, all the produced models result in better performances than ECMWF extended-range forecasts, with the only exception of March and June of the CNN model. Even linear baselines are, for most of the months, outperforming the ECMWF model. These performance differences probably originate from the feature selection process performed in the present work with:

- A first preselection of global and local variables respectively based on climatic influence and Felsche et.al. [49];
- A model selection step performed on models based on all the combinations of variables. This allows for the selection of the best-performing subset of predictors.

Therefore, the selection of context-related features resulted in being a useful step to ease and empowers the learning process carried out by the ML models.

For what concerns the obtained SPI drought classification rates, they do not allow to identify an ML branch clearly outperforming the others (as it was for the results obtained from the testing of the ML models). Indeed, although ELM seems still to be the best performing setting, Table 17 shows a scattered pattern of best performing rates distributed across all the three different branches. The reason for these results can be related to a lot of different reasons. For instance, the structure of the test set of both FFNN and CNN, as specified in section 4.2.2.1, was specifically thought to be functional for an even testing across all the different months of the year. Structuring the test set with this ratio, clearly locks the degree of freedom related to an equal distribution of drought classes all over the months. Moreover, the uneven distribution of the drought classes has to be added to the already small size of the monthly-referred test set due to the "*clear*"/"dirty" sample logic introduced in section 4.2.2.1.

Another reason could be also that the test sets of a specific month i for different models are, on one hand, composed of samples referred to the same month i but, on the other, these samples are not guaranteed to be referred to the same years. This is due to the randomness added to the train-test splitting procedure. Indeed, the monthly-referred test sets are randomly sampled to avoid the bias that could be introduced by a predefined selection. This attention with respect to the models testing procedure brings to the side of SPI classification some uncertainty. That uncertainty, for the structure of the present work, is unavoidable since the real target of the developed models is the cumulative monthly precipitation and not the SPI drought class.

6 Conclusions

The aim of this work was to investigate the influence of climate information referred to different spatial scales (global and local) for the training of ML models with target monthly cumulative precipitation. The reason was to see if appropriate bridging of global and local climatic contexts could bring some benefits in terms of models' predictive skillfulness. The search for the meaningfulness of this bridging has been performed with respect to three main model branches: ELM, FFNN, and CNN. The main difference between ELM and FFNN/CNN branches is summarized into three points: model's temporal domain (monthly for ELM and yearly for FFNN and CNN), dataset size (sensibly increased in FFNN and CNN), and provision of climatic information (only present in the ELM setting). Obtained results show how the best-achieved tradeoff among these three key points is obtained in the ELM setting, where the temporal domain of the model is fixed to a single month and the dataset dimension is restricted at one sample for each year of the considered timeseries. Moreover, another important peculiarity of ELMs with respect to the other branches is the provision of meaningful climatic information for training. The ELM setting is indeed backed by NIPA, a statistical framework able to embed climate indices information into global data and further transform them into unidimensional timeseries for ELM training. In this work, results highlight how the exploitation of such a pre-processing framework has led to the creation of a set of models able to outperform ECMWF extended-range forecast products all across the year. Another reason for the low ELMs MSEs can be founded in the tailoring capabilities of the monthly-based model setting. This setting allows for the choice of a customized set of monthly-related input features as well as for a customized hyperparameter configuration. Thus, it enables the creation of an ad hoc setting accounting for the endogenous monthly variability caused by seasonality.

Acknowledged the good results, the drawback of using such a narrow temporal domain setting as ELMs relates to the small resulting dataset size. The NIPA pre-processing step, indeed, creates a constraint in terms of the model's temporal domain by imposing it on a single month. Therefore, this significantly reduces the dataset size and, with it, the number of maximum considerable input features. This makes prohibitive the usage of more complex models with a higher number of

parameters, which is the reason why, following the work of Giuliani et.al. [2] ELM was used. Additionally, the small dataset dimension does not allow for the application of a proper testing procedure, resulting in using LOOCV to determine model performances.

For FFNN and CNN branches, the considerably higher number of samples seemed to not able to compensate for the lower tailoring capabilities caused by the yearly-based model setting. In these two settings, indeed, the same set of input features has been used and kept constant all over the year. Higher MSEs scores can be also relatable with the complete absence of explicit climatic information provision, subtracting the model with a potential bridging point between global and local climate contexts.

In summary, we investigated two pole-apart situations: monthly-based models with few samples and yearly-based models with increased dataset size.

One further branch that could be interesting for future works, is to deeper investigates the in-between seasonal configuration. Creating a set of four models, each one referring to a specific season, could act as a valid tradeoff between all the before mentioned elements. If compared to the ELM setup, seasonal-based models may have access to a significantly higher number of samples while still retaining most of its customization capabilities by allowing the selection of different season-related variables for each model. The higher number of samples can also allow for the application of a proper testing procedure and for the increase of the considered number of features.

Interesting could be also to investigate the effect of considering the Madden Julian Oscillation (MJO) as a model feature. Its sub-seasonal influence is indeed recognized across the literature, making it known as one of the dominant modes for the variability of sub-seasonal climate at the global scale [56].

Finally, a further experiment could consist in building classification models instead of regressors. This would allow (i) having a direct mapping between input and drought class, (ii) would enable the possibility to use more sophisticated metrics for classification performance assessment, and (iii) would be beneficial for the reliability of the classification results since the entire procedure will be exclusively based on a classification logic.

7 References

[1] B. G. a. V. D. J. a. B. P. J. Zimmerman, «Utilizing the state of ENSO as a means for season-ahead predictor selection,» *Water Resources Research,* vol. 52, n. 5, pp. 3761-3774, 2016.

[2] M. a. Z. M. a. C. A. a. D. G. a. B. P. Giuliani, «Detecting the State of the Climate System via Artificial Intelligence to Improve Seasonal Forecasts and Inform Reservoir Operations,» *Water Resources Research,* vol. 55, n. 11, pp. 9133-9147, 2019.

[3] P. G. a. C. Y. a. T. B.-N. a. N. D. a. P. D. a. S. L. a. T. Hoefler, «Deep learning for post-processing ensemble weather forecasts,» *The Royal Society,* vol. 379, n. 2194, 2021.

[4] American Meteorological Society (AMS), "ametsoc.org," 26 01 2012. [Online]. Available: https://glossary.ametsoc.org/wiki/Forecast_lead_time#:~:text=The%20leng th%20of%20time%20between,American%20Meteorological%20Society% 20(AMS)..

[5] U. C. Mohanty, «Extended Range Prediction System of Indian Summer Monsoon,» in *Monsoon Monograph - Vol II*, Ajit Tyagi, G.C. Asnani, U.S. De, H.R. Hatwar, A.B. Mazumdar, 2012, pp. pp413-467.

[6] B. Shaw, «GLOBAL AG TECH INITIATIVE,» 14 11 2017. [Online]. Available: https://www.globalagtechinitiative.com/digital-farming/data-management/weather-forecasting-how-does-it-work-and-how-reliable-is-it/.

[7] «Atmospheric Predictability as Revealed by Naturally Occurring Analogues,» *Journal of Atmospheric Sciences,* vol. 26, n. 4, pp. 636-646, 01 07 1969.

[8] A. Mariotti, «Progress in subseasonal to seasonal prediction through a joint weather and climate community effort,» *npj Climate and Atmospheric Science,* vol. 1, 2018.

[9] V. Frédéric, «Evolution of ECMWF sub-seasonal forecast skill scores over the past 10 years,» *ECMWF Technical Memoranda,* 02 2013.

[10] W. M. Organization, «public.wmo.int,» [Online]. Available: https://public.wmo.int/en/projects/subseasonal-seasonal-prediction-project#:~:text=The%20Subseasonal%2Dto%2DSeasonal%20prediction% 20project%20(S2S)%20started,seasonal%20to%20seasonal%20time%2Dsc ale..

[11] J. University of Tsukuba, «http://gpvjma.ccs.hpcc.jp/,» 11 01 2022. [Online]. Available: http://gpvjma.ccs.hpcc.jp/S2S/.

[12] A. K. Mishra, «A review of drought concepts,» *Journal of Hydrology,* vol. 391, pp. 202-216, 2010.

[13] D. A. Wilhite, Preparing for Drought: A Guidebook for Developing Countries, Nairobi, Kenya: Programme, United Nations Environment, 1992.

[14] F.-C. Chang, «Meteorological Conditions during Heat Waves and Droughts in the United States Great Plains,» *Monthly Weather Review,* vol. 115, pp. 1253-1269, 01 07 1987.

[15] D. L. Comte, «Highlights Around the World,» *Weatherwise,* vol. 47, n. 1, pp. 23-26, 1994.

[16] G. Obasi, «WMO's Role in the International Decade for Natural Disaster Reduction,» *Bulletin of the American Meteorological Society,* vol. 75, n. 9, pp. 1655 - 1662, 1994.

[17] W. M. Organization, «public.wmo.int,» [Online]. Available: https://public.wmo.int/en/our-mandate/water/drought#:~:text=Drought%20is%20a%20prolonged%20dry, use%2C%20increase%20vulnerability%20to%20drought..

[18] N. I. D. I. System, «drought.gov,» [Online]. Available: https://www.drought.gov/what-is-drought/drought-basics.

[19] W. J. Gibbs, « Drought, its definition, delineation and effects,» *Lectures Presented at the 26th Session of the WMO,* n. 5, pp. 3-30, 1975.

[20] M. a. P. D. a. M. M. Estrela, «Multi-annual drought episodes in the Mediterranean (Valencia region) from 1950–1996. A spatio-temporal analysis,» *International Journal of Climatology,* vol. 20, n. 13, pp. 1599-1618, 2000.

[21] Z. a. S. V. P. a. X. Y. Hao, «Seasonal Drought Prediction: Advances, Challenges, and Future Prospects,» *Reviews of Geophysics,* vol. 56, n. 1, pp. 108-141, 2018.

[22] M. Karabulut, «Drought analysis in Antakya-Kahramanmaraş Graben, Turkey,» *Journal of Arid Land,* vol. 7, 06 2015.

[23] N. C. f. A. Research, «ncar.ucar.edu,» [Online]. Available: https://climatedataguide.ucar.edu/climate-data/palmer-drought-severity-index-pdsi.

[24] M. Janssen, «Population Aggregation in Ancient Arid Environments,» *Ecology and Society,* 06 2010.

[25] W. C. Palmer, «Keeping Track of Crop Moisture Conditions, Nationwide: The New Crop Moisture Index,» *Weatherwise,* vol. 21, n. 4, pp. 156-161, 1968.

[26] NOAA, «cpc.ncep.noaa.gov,» 28 01 2023. [Online]. Available: https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/regional_monitoring/cmi.gif.

[27] I. D. M. Programme, «droughtmanagement.info,» [Online]. Available: https://www.droughtmanagement.info/surface-water-supply-index-swsi/.

[28] S. a. W. A. W. Shukla, «Use of a standardized runoff index for characterizing hydrologic drought,» *Geophysical Research Letters,* vol. 35, n. 2, 2008.

[29] B. N. a. R. Srinivasan, «Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring,» *Agricultural and Forest Meteorology,* vol. 133, n. 1, pp. 69-88, 2005.

[30] E. D. Centre, «geo.uio.no,» [Online]. Available: https://www.geo.uio.no/edc/droughtdb/edr/DroughtEvents.php.

[31] E. C. -. JRC, «Drough in Europe,» European Commission, 2023.

[32] E. C. -. JRC, «Drought in East Africa,» European Commission, 2022.

[33] E. C. -. JRC, «Drought in western Mediterranean,» European Commission, 2022.

[34] USDA, «Major world crop areas and climatic profiles,» in *Agricultural Handbook*, World Agricultural Outlook Board, US Department of Agriculture, 1994, pp. 157-170.

[35] B. J.P., «Natural disaster reduction and global change,» *Bulletin of the American Meteorological Society,* vol. 75, n. 10, pp. 1831-1835, 1994.

[36] U. O. f. D. R. Reduction, «Human cost of disaster. An overview of the last 20 years. 2000-2019,» UN Office for Disaster Risk Reduction, 2020.

[37] C. o. t. E. Communities, «Addressing the challenge of water scarcity and droughts in the European Union,» Brussels, 2007.

[38] E. D. Centre, «geo.uio.no,» [Online]. Available: https://www.geo.uio.no/edc/droughtdb/edr/DroughtEvents/_1976_Event.php#:~:text=Drought%20Event%20Summary,west%20(Norway)%20were%20unaffected..

[39] E. C. -. JRC, «Drought in Central-Northern Europe – July 2018,» European Commission, 2018.

[40] E. C. -. JRC, «Drought in Central-Northern Europe – August 2018,» European Commission, 2018.

[41] E. C. -. JRC, «Drought in Europe August 2022,» European Commission, 2022.

[42] J. a. V. J. V. a. N. G. a. B. P. a. D. A. Spinoni, «Will drought events become more frequent and severe in Europe?,» *International Journal of Climatology,* vol. 38, n. 4, pp. 1718-1736, 2018.

[43] V. P. S. ,. Y. X. Zengchao Hao, «Seasonal Drought Prediction: Advances, Challenges, and Future Prospects,» *Reviews of Geophysics,* vol. 56, n. 1, pp. 108-141, 2018.

[44] J. a. S. J. a. W. E. F. Kam, «Changes in drought risk over the contiguous United States (1901–2012): The influence of the Pacific and Atlantic Oceans,» *Geophysical Research Letters,* vol. 41, n. 16, pp. 5897-5903, 2014.

[45] C. a. W. M. N. Brown, Managing Climate Risk in Water Supply Systems, IWA Publishing, 2013.

[46] A. a. W. Q. J. a. R. D. E. Schepen, «Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall,» *Journal of Geophysical Research: Atmospheres,* vol. 117, n. D20, 2012.

[47] T. S. a. O. A. a. G. B. a. M. D. a. Y. D. a. A. K. a. K. K. a. W. L. a. S. M. a. P. N. a. A. S. a. O. T. a. W. Yun, «Understanding and Predicting Seasonal-to-Interannual Climate Variability - The Producer Perspective,» *Procedia Environmental Sciences,* vol. 1, pp. 55-80, 2010.

[48] A. Kikon, «Artificial intelligence application in drought assessment, monitoring and forecasting: a review,» *Stochastic Environmental Research and Risk Assessment,* vol. 36, p. 1197–1214, 2022.

[49] E. a. L. R. Felsche, «Applying machine learning for drought prediction using data from a large ensemble of climate simulations,» *Natural Hazards and Earth System Sciences Discussions,* vol. 2021, pp. 1-20, 2021.

[50] Y. S. a. C. K. a. Y. H. a. K. Fung, «Application of artificial intelligence models for the prediction of standardized precipitation evapotranspiration index (SPEI) at Langat River Basin, Malaysia,» *Computers and Electronics in Agriculture,* vol. 144, pp. 164-173, 2018.

[51] E. a. T. M. a. K. H. a. F. M. Azizi, «Evaluating the efficiency of the neural network to other methods in predicting drought in arid and semi-arid regions of western Iran,» *Arabian Journal of Geosciences,* vol. 12, 2019.

[52] S.-S. S. C.-C. L. S.-M. B. Deo RC, «Drought prediction with standardized precipitation and evapo- transpiration index and support vector regression models,» *Integrating disaster science and management,* pp. 151-174, 2018.

[53] W. C. Service, «worldclimateservice.com,» [Online]. Available: https://www.worldclimateservice.com/2021/09/06/scandinavian-pattern/.

[54] NOAA, «cpc.ncep.noaa.gov,» [Online]. Available: https://www.cpc.ncep.noaa.gov/data/teledoc/ea.shtml.

[55] ECMWF, «ecmwf.int,» [Online]. Available: https://www.ecmwf.int/en/forecasts/documentation-and-support/extended-range-forecasts/justification-ENS-extended.

[56] D. a. B. L. Specq, «Do subseasonal forecasts take advantage of Madden–Julian oscillation windows of opportunity?,» *Atmospheric Science Letters,* vol. 23, n. 4, p. e1078, 2022.

[57] NOAA, «psl.noaa.gov,» [Online]. Available: https://psl.noaa.gov/enso/mei/data/meiv2.data.

[58] NOAA, «ftp.cpc.ncep.noaa.gov,» [Online]. Available: https://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/nao_index.tim.

[59] NOAA, «ftp.cpc.ncep.noaa.gov,» [Online]. Available: https://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/scand_index.tim.

[60] NOAA, «ftp.cpc.ncep.noaa.gov,» [Online]. Available: https://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/ea_index.tim.

[61] J. Wang, «A review on extreme learning machine,» *Multimedia Tools and Applications,* vol. 81, pp. 41611-41660, 2022.

[62] S. Ding, «Extreme learning machine and its applications,» *Neural Computing and Applications,* vol. 25, pp. 549-556, 2014.

[63] M. Sazli, «A brief review of feed-forward neural networks,» *Communications Faculty Of Science University of Ankara,* vol. 50, pp. 11-17, 2006.

[64] L. L. a. K. J. a. G. D. a. A. R. a. A. Talwalkar, «Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization,» *Journal of Machine Learning Research,* vol. 18, n. 185, pp. 1-52, 2018.

[65] S. I. a. A. K. G. a. S. M. a. P. Asopa, «Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach,» *Procedia Computer Science,* vol. 132, pp. 679-688, 2018.

[66] F. Chollet, *Keras,* https://keras.io, 2015.

[67] J. A. a. D. D. R. a. C. R. Weyn, «Can Machines Learn to Predict Weather? Using Deep Learning to Predict Gridded 500-hPa Geopotential Height From Historical Weather Data,» *Journal of Advances in Modeling Earth Systems,* vol. 11, n. 8, pp. 2680-2693, 2019.

[68] E. Commission, «Product fact sheet: SPI - Africa,» European commission, 2013.

[69] T. a. B. E. a. L. J. a. C. F. a. J. H. a. I. L. a. o. O'Malley, *KerasTuner,* 2019.

[70] R. W. Reeves, Interviewee, *Edward lorenz revisiting the limits of predictability and their implications.* [Intervista]. 6 11 2007.

[71] N. E. A. (. e. S. A. f. D. a. C. (. Karim Bargaoui, «Methodology for drought hazard Assessment, Modelling, and Mapping for Georgia,» p. 8, 2021.

[72] N. D. M. C. (NDMC), «drought.unl.edu,» University of Nebraska, 2017. [Online]. Available: https://drought.unl.edu/Education/DroughtIn-depth/TypesofDrought.aspx.

[73] U. S. P. a. T. C. SHARMA, «Challenges in drought research: some perspectives and future directions,» *Hydrological Sciences Journal,* vol. 47, n. sup1, pp. S19-S30, 2002.

[74] NOAA, «cpc.ncep.noaa.gov,» [Online]. Available: https://www.cpc.ncep.noaa.gov/data/teledoc/nao_ts.shtml.

[75] NOAA, «cpc.ncep.noaa.gov,» [Online]. Available: https://www.cpc.ncep.noaa.gov/data/teledoc/scand_ts.shtml.

[76] NOAA, «cpc.ncep.noaa.gov,» [Online]. Available: https://www.cpc.ncep.noaa.gov/data/teledoc/ea_ts.shtml.

[77] NOAA, «cpc.ncep.noaa.gov,» [Online]. Available: https://www.cpc.ncep.noaa.gov/data/teledoc/eawruss_ts.shtml.

[78] C. C. school, «news.climate.columbia.edu,» [Online]. Available: https://news.climate.columbia.edu/2014/07/02/el-nino-the-basics/.

[79] K. a. T. M. S. Wolter, «El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext),» *International Journal of Climatology,* vol. 31, n. 7, pp. 1074-1087, 2011.

[80] C. C. D. Store, «cds.climate.copernicus.eu,» [Online]. Available: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview.

[81] C. C. D. Store, «cds.climate.copernicus.eu,» [Online]. Available: https://cds.climate.copernicus.eu/cdsapp#!/home.

[82] Z. a. S. V. P. a. X. Y. Hao, «Seasonal Drought Prediction: Advances, Challenges, and Future Prospects,» *Reviews of Geophysics,* vol. 56, n. 1, pp. 108-141, 2018.

[83] N. P. S. Laboratory, «psl.noaa.gov,» [Online]. Available: https://psl.noaa.gov/enso/mei/.

[84] M. a. P. F. a. C. A. Giuliani, «Making the most of data: An information selection and assessment framework to improve water systems operations,» *Water Resources Research,* vol. 51, n. 11, pp. 9073-9093, 2015.

[85] F. Chollet, *Keras,* https://keras.io, 2015.

[86] J. Keyantash., «climatedataguide.ucar.edu,» [Online]. Available: https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-index-spi.

[87] T. a. B. E. a. L. J. a. C. F. a. J. H. a. I. L. a. o. O'Malley, *KerasTuner,* 2019.

List of figures

List of tables