IURI GORENSTEIN

Investigation of the Multidecadal Variability of the South Atlantic Region Hydroclimate since the Mid-Holocene Using a Non-Supervised Neural Network

São Paulo
2023

IURI GORENSTEIN

Investigation of the Multidecadal Variability of the South Atlantic Region Hydroclimate since the Mid-Holocene Using a Non-Supervised Neural Network

Dissertação apresentada ao Instituto Oceanográfico da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Ciências, Programa de Oceanografia, área de Oceanografia Física

Orientador: Prof.ª Dr.ª Ilana Wainer
Coorientador: Prof.ª Dr.ª Luciana Figueiredo Prado

São Paulo

2023

GORENSTEIN, Iuri. **Investigation of the Multidecadal Variability of the South Atlantic Region Hydroclimate since the Mid-Holocene Using a Non-Supervised Neural Network**. Dissertação (Mestrado) apresentada ao Instituto Oceanográfico da Universidade de São Paulo para obtenção do título de Mestre em Ciências, Programa de Oceanografia, área de Oceanografia Física.

Aprovado em: ___/___/_____.

Banca Examinadora

Prof. Dr. _____ Instituição_____

Julgamento_____Assinatura_____

Prof. Dr. _____ Instituição_____

Julgamento_____Assinatura_____

Prof. Dr. _____ Instituição_____

Julgamento_____Assinatura_____

Dedico esse trabalho a quem passar por ele.

Agradecimentos

Quero agradecer à Ilana por me ceder uma mesa de seu laboratório e apontar um caminho; à Luciana pela ajuda a cada passo; à Natalia por seus códigos de Matlab; à Fernanda por me dar dicas de como escrever agradecimentos; ao Mineiro por sua paciência, ao Bruno pelo incentivo; à Marina, Laura, Paulo, Eduardo, Lucas, Pedro, Francesco, Myriam, Elder e muitos outros.

Aos professores da graduação de Física, à quem nunca tive a chance de agradecer.

Agradeço ao meu pai, minha mãe, minha irmã, minha avó, tios, tias, primas, primos e amigos. Em especial à Ligya e ao Toni, comigo quase todos os dias desses últimos dois anos.

Agradeço à USP e à todos os gigantes de ombros largos,

vai Corinthians.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

"*E, afora este mudar-se cada dia,*
Outra mudança faz de mor espanto:
Que não se muda já como soía."

Luiz Vaz de Camões

RESUMO

GORENSTEIN, Iuri. **Investigation of the Multidecadal Variability of the South Atlantic Region Hydroclimate since the Mid-Holocene Using a Non-Supervised Neural Network**.. 2023. 219 f. Dissertação (Mestrado) – Instituto Oceanográfico, Universidade de São Paulo, São Paulo, 2023.

Anomalias de precipitação (PPT) na escala decadal interferem nos níveis de reservatórios, na fauna e na vegetação, além disso, elas podem influenciar eventos de alta frequência, como secas e alagamentos. A chuva nos trópicos é associada à Zona de convergência Intertropical (ZCIT), que por sua vez tem suas anomalias decadais controladas pelos modos oceânicos de variabilidade. As populações do Nordeste do Brasil (NE) e da África Ocidental (AO), duas regiões adjacentes ao Oceano Atlântico, tem economias agrárias altamente dependentes de mudanças na posição da ZCIT e, portanto, do ciclo de variabilidade decadal do Oceano Atlântico. Usando as redes neurais não supervisionadas Self Organizing Maps (SOM), o clima da região do Atlântico tropical foi analisada; e a variabilidade decadal de suas anomalias de temperatura de superfície de oceano (TSO) e PPT foi estudada usando: dados observacionais de satélite de 1979-2015; dados da reanálise HadISST de 1870-2019; e, finalmente, os modelos climáticos do EC-Earth, CESM e GISS com rodadas Pré-industriais e diferentes tipos de cenários com parâmetros de insolação do Holoceno-médio (HM) e de vegetação representado o período do Green-Sahara (GS).

O SOM foi bem sucedido na redução dimensional dos dados climáticos. As anomalias de TSO, pressão e ventos são intrinsecamente conectadas em escalas decadais. Juntas, elas controlam a anticorrelação decadal de precipitação entre o NE e a AO, encontradas nas séries de Standard Precipitation Index decadais dessas regiões de 1979-2015. O HadISST dataset de 1870-2019 mostrou uma periodicidade de 40 a 50 anos na qual a TSO do oceano Atlântico oscilou durante o século vinte. Ao usar a mesma metodologia com as simulações dos modelos numéricos dos períodos do PI, HM e GS, emergiram padrões de TSO em escalas decadais que se assemelham aos ciclos encontrados nos dados observacionais. Usando a entropia de Shannon como medida de variabilidade das rodadas dos modelos, a variabilidade de PPT no modelo EC-Earth apresentou dependência apenas na forçante de emissão de partículas (rodada de GS com redução de poeira teve a menor entropia dentre os cenários), enquanto a variabilidade de TSO nesse modelo parece ter sido mais influenciada pela presença de vegetação no Sahara (rodadas do GS e GS com redução de poeira tiveram menor entropia). O

modelo GISS mostrou a menor mudança de variabilidade de TSO entre diferentes cenários e os modelos CESM indicaram grande variabilidade interna de PPT com cenários iguais apresentando valores divergentes de entropia (GS com full vegetation, rodadas 1 e 2). Portanto, a utilização de large ensembles se vê necessária para atribuir incertezas da variabilidade interna dos modelos à entropia e aumentar a robustez dos achados.

Palavras-chave: Nordeste Brazil, Africa Ocidental, anti-correlação de chuva, precipitação decadal, Self Organizing Maps, EC-Earth, GISS, CESM.

ABSTRACT

GORENSTEIN, Iuri. **Investigation of the Multidecadal Variability of the South Atlantic Region Hydroclimate since the Mid-Holocene Using a Non-Supervised Neural Network**.. 2023. 219 f. Dissertação (Mestrado) – Instituto Oceanográfico, Universidade de São Paulo, São Paulo, 2023.

Decadal precipitation (PPT) anomalies are related to water reservoirs, affect biota, and may interfere with higher frequency events such as floods and drought. Rainfall in the tropics is mostly associated with the Intertropical Convergence Zone (ITCZ), which in turn has its decadal displacements and anomalies controlled by Oceanic modes of variability. Populations from Northeast Brazil (NE) and West Africa (WAF), two regions adjacent to the Atlantic Ocean, have mainly agricultural economies dependent on ITCZ shifts and, consequently, on the Atlantic decadal variability cycle. Using Self Organizing Maps (SOM) non supervised neural networks, the tropical Atlantic region climate was analyzed and its sea surface temperature (SST) and PPT decadal variability was studied using: satellite observational data from 1979-2015; Reanalysis HadISST product from 1870-2019; and, finally, EC-Earth, CESM and GISS numerical climate models from Pre-industrial (PI) runs and different scenarios with mid-Holocene (MH) insolation and vegetation representing the Green-Sahara (GS) period. The SOM has successfully reduced the dimensionality of climate data. The Atlantic Ocean SST, pressure and wind anomalies are entangled at decadal scales. Together, they control the decadal PPT anomaly anti-correlation between NE and WAF regions, depicted by the standard precipitation index series from 1979 to 2015. The 1870-2019 HadISST reanalysis dataset has shown a 40 to 50 years periodicity, representing the full Atlantic decadal SST anomaly cycle in the 20th century. With the numerical model simulations from PI, MH and GS, SST anomaly structures that closely resemble the observational data cycles appeared at decadal scales. Using Shannon's Entropy as an analogue of the model runs' decadal variability, the EC-Earth PPT variability showed a dependency in dust emission (GS with dust reduction had the lowest entropy of all), while the SST variability in this model seems to be affected more by the presence of vegetation in the Sahara (GS and GS with dust reduction shown lower entropy than the PI and the MH runs). The GISS model presented the lowest SST variability change between different scenarios and the CESM model point to a large PPT internal variability, with identical scenarios showing

divergent entropies (GS full vegetation runs 1 and 2). Therefore, large ensembles are necessary if we want to attribute the uncertainties of internal variability from the models' entropies and achieve more robust results.

Keywords: Northeast Brazil, West Africa, precipitation anti-correlation, decadal precipitation, Self Organizing Maps, EC-Earth, GISS, CESM.

List dof Figures

List of Tables

List of Acronyms

Acronyms	Full Name
WAF	West Africa
NE	Northeast Brazil
SOM	Self Organizing Maps
SPI	Standardized Precipitation Index
DBI	Davies Bouldin Index
ITCZ	Intertropical Convergence Zone
SST	Sea Suface Temperature
PPT	precipitation
SSA	Subtropical South Atlantic
SA	South Atlantic
NA	North Atlantic
PI	pre-Industrial
MH	mid-Holocene
GS	Green-Sahara
GSdr	Green-Sahara with dust reduction
GSsl	Green-Sahara with soil and lake inputs
GSNA	Green-Sahara with North Africa vegetation
GSEX	Green-Sahara with extra tropical vegetation
GSALL	Green-Sahara with full vegetation

Summary

Chapter 1

Introduction

1.1 The tropical Atlantic variability at decadal scales.

"The fight against climate change is inseparable from the fight against poverty and inequality" (da Silva, 2022). Studying climate's natural variability is crucial to fully comprehend its equilibrium. Both West Africa and Northeast Brazil (WAF and NE - Figure 1.1) have an agricultural-based economy, making them highly dependent of rainfall (Nobre e Shukla, 1996; Polo et al., 2008; Guenang e Kamga, 2014; Cunha et al., 2019; IPCC, 2020; Arias et al., 2021b). These are sensitive regions to climatic changes and presented large decadal precipitation variability in recent history (Nash et al., 2016; Cunha et al., 2019; Arias et al., 2021b). The Failed State Index for 2018, which is a social, political and economic index shows that West African countries are either in the alert or high alert category. In its most recent report (AR6), the Intergovernmental Panel on Climate Change (IPCC) concluded that drought is expected to double over North Africa, the western Sahel and Southern Africa (Arias et al., 2021b). In WAF, Urban populations exposed to severe droughts are projected to increase 65.3±34.1 million by 2050. Furthermore, considering an expected 1.7°C global warming, 17–40 million people could migrate internally in sub-Saharan Africa (>60% in WAF). This region exhibits severe droughts caused by irregular decadal rain anomalies, where the understanding of what is causing climate variability and its effects are still subject of investigation (Rice e Patrick, 2008; Guenang e Kamga, 2014; Arias et al., 2021b). Also according to the IPCC fifth and sixth Assessment Report, NE appears as one of the most vulnerable region with respect to droughts, fires, agricultural losses, climatic-related migrations and displacements (Jimenez-Muñoz et al., 2016; Cunha et al., 2019; Ajjur e Al-Ghamdi, 2021; Torres et al., 2021; Arias et al., 2021a). 70,000 km^2

have reached a point at which agriculture is no longer possible. Intense droughts have triggered migration to urban centers in and outside NE. More than 10 million people have been impacted by the drought of 2012/14 in the region, which was responsible for water shortage and contamination (*Government of Brazil*, 2020). On the other hand, a 1.5°C increase in global projection would result in an increase of 100–200 % of the population affected by floods in Brazil (Arias et al., 2021a).

Rainfall in the tropical Atlantic and adjacent continents is mostly related to the band of intense convection and surface wind convergence known as the Intertropical Convergence Zone (ITCZ) (Dhrubajyoti et al., 2019). The meridional displacements of the ITCZ is the main driver of rainy and dry seasons for NE and WAF regions. Bringing the rainy season in WAF from June to September and to the South Hemisphere in late November (Hagos e Cook, 2005). Wind, pressure and sea surface temperature gradients are able to move this convergence zone across the equator into the warmer hemisphere (Deser et al., 2010; Hounsou-Gbo et al., 2019). All regions dependent on the ITCZ are affected by these anomalies and consequently, the Atlantic decadal variability cycle (Nobre e Shukla, 1996; Hounsou-Gbo et al., 2019; Cunha et al., 2019).

The entanglement of ocean and atmosphere interactions arise from non-linear dynamical processes with large scale climate phenomena and its influences (Deser et al., 2010, 2012). These processes are also understood as modes of variability that can have large impacts in local economies, associated with changes in regional and global rainfall (Santos, 2006; Nnamchi e Li, 2011; Ham et al., 2013; Rojas et al., 2014). The sea surface temperature (SST) structure of the Tropical Atlantic affects the meridional position of the ITCZ, influencing the decadal rainfall distribution from the adjacent regions (Wainer e Soares, 1997; Villamayor, 2020). Therefore, at decadal time scales we may have an anti-correlation between NE and WAF precipitation (PPT) anomalies, similar to the interannual anti-correlation from the wet and dry seasons of these regions, but with slight deviations due to the Pacific and Indian Oceans, which have also been partially associated with the PPT in these regions (Villamayor, 2020; Cai et al., 2020).

Using observational data, the Atlantic modes of climate variability and its relations to changes in rainfall at decadal time-scales in both sides of the tropical Atlantic are explored. Based on the 1870-2019 SST anomaly, the Atlantic Ocean decadal cycles for the last century are also depicted.

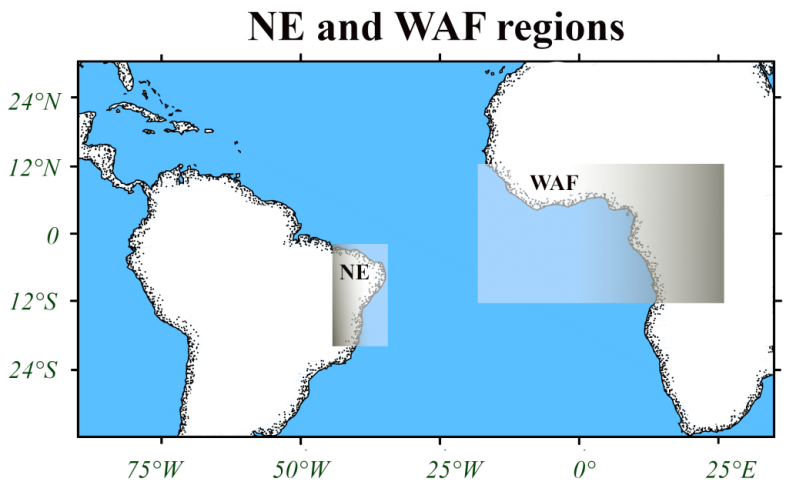

Figure 1.1: NE (50W-35W and 15S-0N) and WAF (20W-25E and 11.4S-15N) box regions highlighted according to the IPCC criteria in South America and Africa map (IPCC Assessment Report 5, 2020).

1.2 Neural Networks and the problem of dimensionality reduction.

The Big Data era can be defined as the recent revolution in measurement technologies which enhanced the quality and quantity of available data. These advances are based on the recent improvement of sensors, hardware (for high-performance computing and storage) and processing algorithms (Hey et al., 2009). This era is the stage where statistical learning and machine learning continue to develop (Breiman, 2001; Brunton et al., 2021). The former one focuses on the development of interpretable models of data, while the latter one finds its objective either in classifying or making attributions to the system, maximizing accuracy via supervised or unsupervised learning.

Although many refer to machine learning algorithms functioning as black boxes, artificial intelligence have already produced ground breaking performance metrics in areas such of computer vision, speech process, image recognition and clustering (Howard et al., 2017; Vaswani et al., 2017; Tozadore et al., 2018; Azencot et al., 2020).

In most problems, the minimum number of parameters needed to account for the observed properties of the data is unknown (Samarasinghe, 2016). Dimensionality reduction is the transformation of high-dimensional data into a lower dimensional meaningful representation of it (Burges, 2005; Van Der Maaten et al., 2009). There are many modern algorithms using neural networks that promote dimensionality reduction, some of them use encoder networks to transfer data from the input space into a lower dimensional feature

space. In this feature space the problem can be solved with lower computational cost and recovered via a decoder network back to the input space (Azencot et al., 2020; Brunton et al., 2021; Chadebec et al., 2022a).

The reduced representation of a system should have dimensions that correspond to the minimum number of parameters needed to account for the observed properties of the data. The ocean and atmosphere are an entangled nonlinear system, ideal for applying dimensionality reduction since its data usually has high dimensionality and complexity.

1.2.1 The Self Organizing Maps

Ocean modes can be identified using linear methods (e.g., EOF), in this case, they do not account for the nonlinearity of intrinsic internal variability (Liu et al., 2006; Hounsou-Gbo et al., 2009; Costa et al., 2021). Self Organizing Maps (SOM), on the other hand, are useful tools as they capture nonlinear aspects associated with climate modes. SOM is a machine learning technique where an unsupervised neural network is created from a given amount of data and a predefined architecture (Liu et al., 2015). The SOM algorithm is able to identify connections between the data points and recreate underlying patterns present within the data set. The patterns obtained from the SOM are subsequently used to understand the original data links, in our case, with respect to climate-related variables such as the spatial distribution of temperature, wind and pressure anomalies (Samarasinghe, 2016; Gibson et al., 2017).

In this study, we used the SOM to understand the physics controlling the decadal PPT and variability in the NE and WAF regions. After training the SOM, the monthly climate data was projected into its non-linear feature space, where the dimensionality reduction was applied via clusterization.

1.3 The mid-Holocene

The mid-Holocene (MH, period between 5000-7000 years Before Present) was characterized by lower summer insolation in the Southern Hemisphere and higher summer insolation in the Northern Hemisphere, resulting in the so-called "Holocene Thermal Maximum" (Berger, 1988; Liu et al., 2002; Bova et al., 2021). While it may be seen as a conceivable geological analog for the future (Burkea et al., 2018; Kaufman et al., 2020),

the warming associated with these changes was neither global nor year-round, with large differences across the hemispheres and seasons (Berger, 1978; Zhao e Harrison, 2012; Huo et al., 2021).

Climate changes during the MH vary across the globe. While the direct effects of the MH insolation pattern produced an enhancement in the Northern Hemisphere monsoon, the Southern Hemisphere PPT experienced a reduction when compared to present day (Liu et al., 2004; Wanner et al., 2008; Smith e Mayle, 2017). Other MH climate forcings, such as the presence of vegetation on contemporary deserted locations in North Africa and Asia, have been associated to changes in global-mean surface temperature, global precipitation and to an increase of the Atlantic Meridional Overturning Circulation (Harrison e Bartlein, 2012; Pausata et al., 2017; Griffiths et al., 2020; Ding et al., 2021; Huo et al., 2021; Zhang et al., 2021). MH climate conditions have been linked to vegetation changes in southwest Amazonia, southeast Brazil and NE, regions reliant on South American summer monsoon derived precipitation (Smith & Mayle, 2017; Gorenstein et al., 2022a).

Paleorecord studies show that the NE semi-arid region frontier with the Amazon was drier than present, while its coast had higher than present precipitation during the MH (Gorenstein et al., 2022a). The WAF region was highly influenced by precipitation in the Sahel region, where the presence of shrub vegetation during the MH motivated the definition of this period as the Green-Sahara (GS). The presence of past vegetation in present deserted regions is a great opportunity to study the variability of Earth's climate and test different forcing attributions with the use of numerical climate simulations.

1.3.1 Numerical Models

Climate models are a numerical parameterization able to simulate past, present and future climate from a set of initial conditions and constraints. They are used to study internal atmospheric and oceanic variability in different scenarios (Deser et al., 2006, 2012). However, different numerical models carry uncertainties in their internal variability representation due to structural differences (Lehner et al., 2020). To study multiple climate models with different scenarios and make plausible attributions, we need a large ensemble of simulation runs from each model to systematically characterize their uncertainties (Ferrero et al., 2021). Even when looking at a single model, a large ensemble is needed to characterize its internal variability and underlying uncertainties before making different

scenario attributions.

Even though a large ensemble of model runs is not present in this study to determine the uncertainties of the models, numerical simulations were used to test the SOM's efficiency in representing climate data variability and making attributions. Using different forcings from EC-Earth, CESM and GISS numerical models, the dimensionality reduction using the SOM was aplied to SST and PPT decadal anomalies in the Pre-industrial (PI), MH and different GS simulations.

Chapter 2

Data and Methods

2.1 Data

2.1.1 Observational Data

Precipitation data

The Global Precipitation Climatology Center (GPCC) uses the global data collections of the Climate Research Unit (CRU) (NCAR, 2020), Food and Agriculture Organization (FAO), and GHCN (NOAA/OAR/ESRL, 2020), as well as other precipitation data collections from international regional projects, while the CPC Merged Analysis of Precipitation (CMAP) (Arkin et al., 2020) and GPCP are the most widely recognized and used merged data sets. The GPCP is based on the sequential combination of microwave, infrared, and gauge data. From 1979 to 2020 and offer globally complete satellite only precipitation estimates (Sun et al., 2017).

Since we are interested in showing both continental and ocean precipitation and identifying its patterns during different climate regimes over the Atlantic we used the GPCP Version 2.3 Combined Precipitation dataset (Adler et al., 2003), in a monthly, 2,5°x2,5° grid for the 1979-2015 climate patterns recognition and NE/WAF decadal SPI (described below).

Pressure and Wind data

The Pressure and wind stress (tauu and tauv) observational data used in this study are both from NCEP-NCAR Reanalysis 1, a 1,875°x1,875° monthly dataset (Kalnay, 1996). The time period used to create the Atlantic patterns was the same as for the precipitation

limits, from 1979 to 2015. A decadal filter was applied.

Sea surface temperature data

The temperature observational data comes from HadISST1. A monthly 1°x1° dataset (Rayner et al., 2003). The time interval examined is from 1870 to 2019. The sea surface temperature anomaly was calculated with respect to the 1979-2015 period. To separate the Atlantic SST anomaly pattern of variability from any global warming signal the mean global temperature anomaly was also calculated and subtracted from each pixel's SST anomaly time series (Zhang et al., 1997; Mantua et al., 1997; Bonfils e Santer, 2011).

2.1.2 Model Data

The numerical model outputs used in this study are all monthly data. They have all been regridded to a 1x1 grid and had decadal filters applied.

EC-Earth

The European Consortium Earth System Model Version-3 (ECE3) scenarios analyzed in this study were: PI (100 years run B405 and 200 years run B400), MH (100 years run Z6KA and 200 years run B6KA), GS (100 years run G105 and 50 years run G100) and GS with dust reduction (100 year run 506 and 200 year run G501).

ECE3 standard configuration consists of the atmosphere model IFS including the land surface module HTESSEL and the ocean model NEMO3.6 with the sea ice module LIM3. Coupling variables are communicated between the different component models via the OASIS3-MCT coupler (Döscher et al., 2021). The ECE3 model is used to contribute to CMIP6 in several configurations, for example, the EC-Earth3-Veg configuration which couples the LPJ-Guess dynamic vegetation model (Smith et al., 2014) to the atmosphere and ocean model; however, the performance of EC-Earth3 and EC-Earth3-Veg is very similar (Wyser et al., 2020).

CESM

The Community Earth System Model (CESM) outputs from different scenarios used were from CESM-Toronto: PI, MH, GS and GS with soil and lake input (100 years run each); and iCESM: PI, MH, GS (100 years run each).

The CESM models used here are from the Cmip6 multi-model ensemble. The CESM-Toronto simulations is a PMIP experiment for the mid-Holocene with Green Sahara and mid-Holocene with soil and lake inputs made by the University of Toronto (UofT), Canada. The model configuration was made by UofT-CCSM4 (2014), atmosphere from CAM4 (finite-volume dynamical core; 288 x 192 longitude/latitude; 26 levels; top level 2 hPa); ocean: POP2; seaIce: CICE4; land: CLM4. The iCESM runs used in this study were intended for PMIP oxygen isotope tracer experiments, also from Cmip6 multi model ensemble.

GISS

The scenarios from the Goddard Institute for Space Studies Model E2 coupled with the Russel ocean model (GISS-E2-R) were: PI, MH and GS with North Africa vegetation only, GS with Extra-Tropical vegetation only and two runs of GS with Full vegetation (100 years run each).

All runs except GS Full Vegetation run 1 are updated aerosol/ozone input runs for non anthropogenic simulations using Nancy Kiang's regressions (GISS's veg modeler). GS Full Vegetation run 1 uses Koepen Gieger regression scripts. Several experiments have been set up for the last millennium with GISS due to uncertainties in past forcings and their effects, with different combinations of solar, volcanic, and land use/vegetation (Colose et al., 2016; LeGrande et al., 2015; Bühler et al., 2022).

Table 2.1 - Data used in this study

Source	Variable	Time interval	Reference
HadISST1	Sea Surface Temperature	1870-2019	Rayner et al. (2003)
GPCP Version 2.3	Precipitation	1979-2015	Adler et al. (2003)
NCEP-NCAR	Pressure and Wind	1979-2015	Kalnay (1996)
Model	**Scenarios**	**Rounds size (years)**	**Reference**
EC-Earth	PI,MH,GS,GS dust reduction	50, 100 and 200	Döscher et al. (2021)
CESM	PI,MH,GS and GS soil and lake	100	Kay et al. (2015)
GISS	PI,MH,GS, GS North Africa, GS Extra-Tropical, GS Full vegetation	100	Schmidt et al. (2014)

2.2 Methods

2.2.1 The Standardized Precipitation Index (SPI)

The SPI was designed to quantify the precipitation deficit for multiple timescales (McKee, 1995). It is an easy to calculate index that uses only one variable (precipitation). It measures both positive (flood) and negative (drought) precipitation anomalies periods and it has been used in several studies to quantify the precipitation rate in different regions of the world (Guttman, 2007; Xie et al., 2013; Ionita et al., 2016; Saada e Abu-Romman, 2017).

Different time scales can be inferred for the SPI calculation. To study the low frequency fluctuations on precipitation we used a decadal filter (Figure 3.1). SPIs of this timescale are usually tied to streamflows, reservoir levels and even groundwater levels (Svoboda et al., 2012).

The SPI is a normalized time series (Figure 3.1), it can be calculated using a predeveloped algorithm (Adams, 2017) and has a simple numerical interpretation: SPI values between -1 and 1 indicate a near normal precipitation regime, from 1 to 1.49 (-1 to -1.49) moderately wet (dry), 1.49 to 1.99 (-1.49 to -1.99) indicates very wet (dry) and above 2 (below -2) extremely wet (dry).

2.2.2 Self Organizing Maps

A Neural Network is a machine learning technique where a collection of interconnected neurons incrementally learn from their environment (data) to capture essential linear and nonlinear patterns and trends in complex data. Used in many areas of study, the neural networks help represent information and characteristics from a main problem (Haykin, 2009; Samarasinghe, 2016).

The SOM is a non-supervised neural network. Its goal is to map a high dimensional space of data into a lower dimensional space, without disrupting the original data topology, creating a function that projects the input data in the SOM map and back to the input space. The SOM algorithm was created during the 80s by Teuvo Kohonen, at that time professor in the university of Helsinki, Finland (Kohonen, 1982, 1995), inspired in Willshaw & Malsburg (1976) work. The conception of his algorithm is an abstract form of data manipulation and it has been used to resolve a range of different problems, from Satellite Imagery Patterns used to predict interannual and decadal climate variability (Richardson et al., 2003; Chang et al., 2020; Gu e Gervais, 2021) to analyzing the spatial and temporal spread of COVID-19 (Melin et al., 2020; Galvan et al., 2021).

The SOM analysis for this work was implemented in both single variable and multivariable forms, with two different time intervals for the observational datasets and another time with the numerical models simulation outputs. In the first analyses, precipitation, SST, wind and pressure monthly anomalies of the Tropical Atlantic region (65W-25E, 29N-59S) from 1979 to 2015 were given as the input data to the multivariable SOM algorithm to generate a eight hundred neuron grid. In Figure A.1 this input data appears as the x's. Random values starting weights were used to initialize the eight hundred neuron grid representing the Atlantic region climate variables from 1979 to 2015. The arcs from x's in Figure A.1 represent those weights, the sum between products of weights and input data creates the neuron grid points (y's from Figure A.1). The SOM algorithm will then use the input data in a loop to compare and improve its representations of the data in the neuron grid (presented after the training process in Figure A.2).

To classify the number of hidden classes from an input data we can separate the neuron grid into different clusters. In many practical situations the number of clusters from the data set is unknown. There are various k-means methods to define how many hidden

groups there are in a neuron grid (MacQueen et al., 1967). In this first step of the study, we used the Silhouette methods to find the optimal number (Wang et al., 2017). This method consists of making a comparison between the average distance between each point and the points inside his cluster and the average distance between each point and the points outside of his cluster, normalized by the largest distance between all points. This creates a Silhouette metric that goes from -1 to 1. Where the closer the Silhouette value is to 1, the closer you are to the optimal number of clusters.

With a fixed number of clusters, the hidden patterns from the neuron grid can be expressed as in Figure 3.2, with the Atlantic region cluster patterns evolving in time.

After defining the number of clusters in the SOM that best represents our data, the variance explained by each SOM pattern and its evolution time is calculated. When comparing two images, pixel by pixel, a value of R-squared between them indicates the percentage of variance explained from an observational monthly Atlantic map by an Atlantic map from the neuron grid. This way the variance from each one of the Atlantic patterns was calculated (Figure 3.3). The evolution maps from the Tropical Atlantic (Figures 3.2 and 3.3) are used to elucidate some of the ocean and atmospheric physics behind the rainfall regime in NE and WAF regions for the last four decades (Figure 3.1).

Following the exact same SOM methodology the 1870-2019 SST anomalies from the south tropical Atlantic region were used to create an eight hundred neuron SOM (Figure B.1) and the pattern evolution of the SST anomaly cycles (Figure 3.4).

With the combined data from numerical climate models able to simulate past scenarios with different vegetation inputs shown in section 2.1.2 (Table 2.1), two single variable neural-networks were trained using the SOM algorithm. These two networks have 20.000 neurons each (rectangular grid of 200x100) and they were trained separately with the decadal PPT anomaly and decadal SST anomaly model data from the Atlantic ocean. These two model data SOM were used to project the decadal SST and PPT anomalies from the different runs of all models into a common clusterized feature space. Used in the training process, the model data had a combined time dimension of 28.000 monthly SST anomaly and 28.000 monthly PPT anomaly. Its clustering is discussed next.

2.2.3 Entropy as an analogue for climate variability

Shannon's Entropy (equation 2.1) is used in this study to measure the variability from each model's scenario by attributing its time series with an entropy value.

$$H = -\sum_{k=1}^{N} P(x_k) ln(P(x_k)) \tag{2.1}$$

$P(x_k)$ is the probability that our system x is found in the k state. Therefore, the entropy is the sum from the probability times its logarithm over all possible states. The essence being, the probability distribution from microstates (the Atlantic SST or PPT pattern, a.k.a SOM cluster) can be used to determine macro properties of a system (the Atlantic ocean SST and PPT variability).

2.2.4 Proposed approach for measuring uncertainty

There are a few sources of uncertainty in calculating the entropy from a numerical climate model output.

First, numerical models carry three types of uncertainty by themselves: the models internal variability; the different numerical discretizations responsible for the physical representation of the model's dynamics; and, the different scenarios uncertainty (Lehner et al., 2020). Since we do not have a large ensemble, the conclusions led by this study are to be considered only for the model runs referenced in section 2.1.2 Model Data. With this in mind, the uncertainty that emerges from the models' entropies are coming from the clusterization of the feature space.

Clustering of the model data SOM

The Techniques for calculating the correct number of clusters in data sets are many: the K-means, Silhouette, Calinski Harabasz, Dunn, Davies Bouldin and more (Desgraupes, 2013; Roushangar et al., 2020). After defining a metric from the feature space, most methods consist of measuring which clustering minimizes the distance from members inside the same cluster (compact clustering), while also maximizing the distance between clusters (separation of different classes).

For the model data SOM clustering, two methods were applied: the Silhouette and the Davies Bouldin Index (DBI) (Davies e Bouldin, 1979; Wang et al., 2017). Their results can

be seen in Figure C.1, where the curves represent the index value for each SOM clustering. The Silhouette curve in Figure C.1 is actually the negative value for the Silhouette index, therefore, by the indexes definition, the minimum value for the indexes indicate the ideal number of clusters in the feature space. However, highly dimensional climate data is not simple to clusterize and the curves show no clear minimum. In these cases, we look for a point where improving the number of clusters is not in fact generating useful new classes. The elbow method is usually applied in K-means' Sum Squares Error, but its concept is useful for other indexes as well. In this approach we get to a point where elevating the number of clusters is no longer worth it (the clusterization improvement we get from raising the number of clusters gives us hardly any advantage from the previous number of clusters), this value sometimes can be visually interpreted as an elbow in the graph due to the sharp decay from the 1st derivative of the index per number of clusters. Therefore, the ideal number of clusters in the model data SOM was defined as the Silhouette and DBI curves' elbow region.

Entropy's Mean and Standard Deviation

From the 20.000 neurons model data SOM clustering arose the entropy's main uncertainty. Changing the number of clusters, even by one, can generate large entropy differences for the same time series (Figure 3.7). As discussed before, the entropy value is dependent on the number of possible states from the system (a.k.a its clusters).

In each scenario run for each model, an interval of ideal numbers of clusters near the Silhouette and DBI elbows were applied to the SOM. The number of clusters used were from 150 to 200 for the SST SOM, and from 300 to 350 for the PPT SOM (Figure C.1). The entropy value attributed to each time series was calculated by the mean of the entropies from the different clusterizations, while its standard deviation was used to define its uncertainty for the 95% confidence interval (Table 3.1).

Chapter 3

Results

3.1 Observational data Analysis

Using precipitation data from NE and WAF regions (Figure 1.1), the SPI evolution from 1970 to 2015 was calculated (Figure 3.1). Then, a multivariable SOM with 800 neurons using decadal SST, PPT, pressure and winds anomalies for the Tropical Atlantic Ocean was implemented (Figure A.2). Its clusterization yielded seven cluster patterns which summarizes the 432 months of the multivariable time series (Figure 3.2). An analysis of the variance evolution from each variable was used to discuss their entanglement in the Atlantic decadal variability (Figure 3.3). Finally, searching for the decadal variability cycle of the Atlantic ocean, the 20th century decadal SST anomaly was used to create a 800 neuron single variable SOM. The clustering yielded nine clusters that show the variability envelope the Atlantic Ocean decadal SST oscillated from 19870 to 2019 (Figure3.4).

3.1.1 1979-2015 SOM Patterns and SPI series

The decadal SPI series for NE and WAF (Figure 3.1) starts at 1989 (the first SPI value for January 1989 is computed with the precipitation mean from 1979 to 1989). In this decadal time scale the anti-correlation between NE and WAF precipitation emerges. From 1989 to 1995 the SPI index of NE is above 1 (indicating very wet conditions) while WAF shows lower than -1 SPI. From 1995 to 2005 NE and WAF precipitation anomalies swap places, while holding their anti-correlation. With NE SPI reaching its most negative value in 2000, overtaking -2 (severe drought), and WAF reaching its most wet period not far after (2004). From 2005 to 2015 the WAF SPI series stays inside the [-1, 1] interval and negative from 2010 on, while the NE reaches a new precipitation high around 2010.

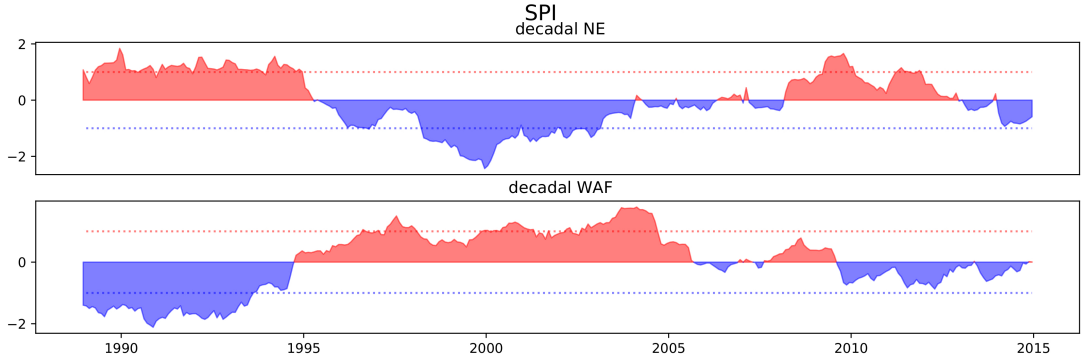

Figure 3.1: The Standardized Precipitation Index series with a decadal filter from 1989 to 2015, with the GPCP data set (Adler et al., 2003). NE region on the top and WAF region on the bottom. The dashed line indicates the +1 and -1 SPI limit for wet and dry conditions.

To identify the SST patterns and Atlantic modes associated with the out-of-phase decadal rainfall distribution for the Tropical Atlantic adjacent regions, a neuron-network using the SOM algorithm with hexagonal geometry, Gaussian neighboring, PCA initialization and eight hundred (20x40) neuron-grid points was created (Figure A.2). The SOM was then trained using the 1979-2015 observed anomalies of precipitation, sea surface temperature, winds and sea level pressure anomalies for the Atlantic region (65W-25E, 29N-59S). When retrieving their projection on the SOM's feature space, the K-means and Silhouette algorithm yielded seven clusters (Figure 3.2).

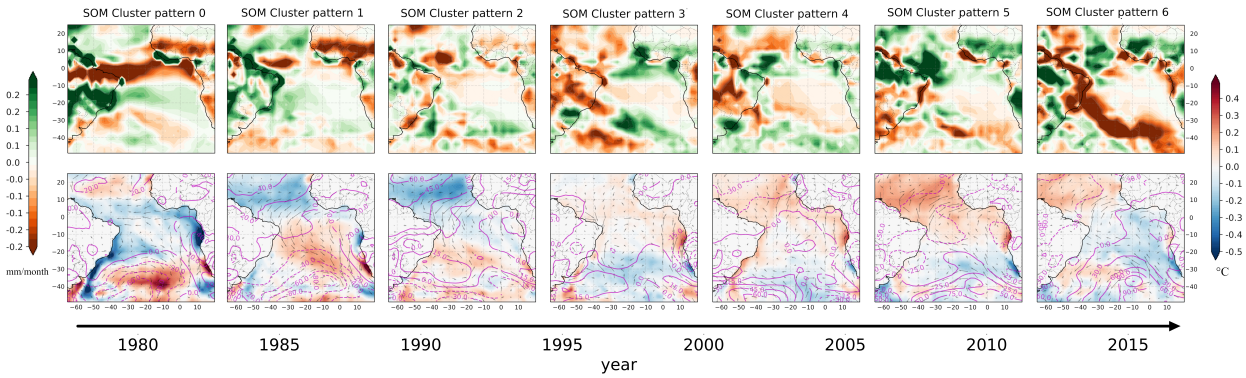

Figure 3.2: SOM cluster patterns evolution from 1979 to 2015. The top row: precipitation anomaly cluster patterns. The bottom row: SST (shading), wind (vectors) and pressure anomalies (contours) each cluster patterns. The patterns are organized by their evolution in time.

After the network has converged, the observational data used in the training was compared with the SOM neurons once again, creating a curve of the percentage of variance explained by each SOM pattern in the time series (Figure 3.3). For example, the SST anomalies from SOM Cluster pattern 0 (Figure 3.2 - bottom left) show positive correlation with the SST anomalies from the observational data from 1979 to 1985, reaching its peak in 1983, when it represented nearly 90% of its variance (top left curve from Figure 3.3).

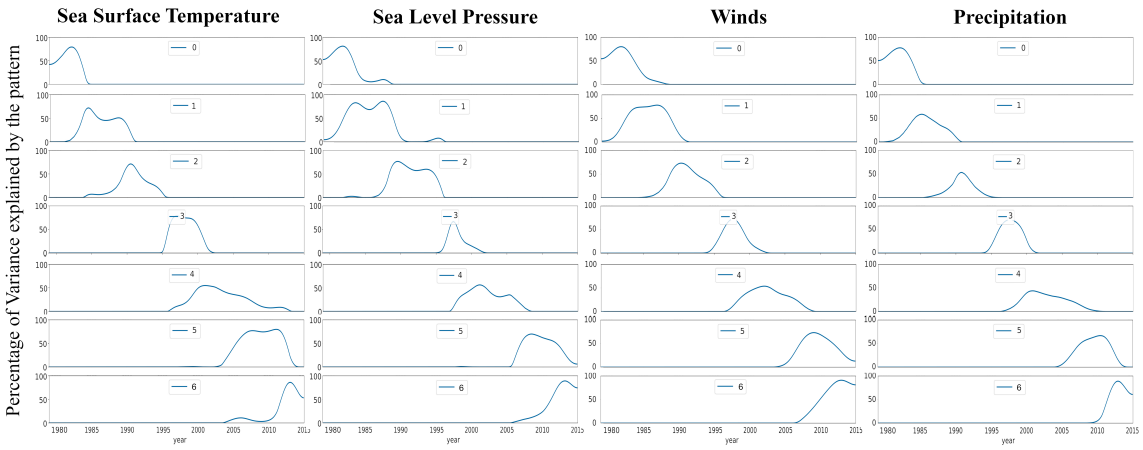

Figure 3.3: SOM patterns variance evolution in time. Each column represents one of the variables used to calculate the SOM: SST, pressure, winds and precipitation anomalies from left to right. Each row represents the variance evolution for the SOM pattern: pattern number zero, one, two, three, four, five and six from top to bottom. The curve from a specific pattern and variable represents when the observational data from this variable had its variance explained by this pattern.

From the 80's to the early 90's the SPI indicates positive/negative precipitation anomalies in NE/WAF, which can be explained by the SOM patterns transition from number 0 to 1 (Figure 3.2). Positive SST followed by a low pressure anomaly in the Tropical-South Atlantic starts moving towards the 20S latitude line, this pattern weakens the South Atlantic Subtropical High and creates a southward wind anomaly. As we can see in the SOM precipitation pattern 1 (Figure 3.2) a southward displacement of the decadal ITCZ brings positive precipitation anomalies to the South American coast. In the WAF coast near the Benguela region (10S-30S, 10E) a SST dipole shapes the eastern Atlantic part of this pattern. The negative SST anomaly north of the Benguela region prevents precipitation in the WAF region and creates the anti-correlation of highly positive/negative precipitation anomalies in NE/WAF.

The SST and pressure anomalies in the south Atlantic change from pattern 2 to 3 to 4. From 1995 to 2000 a fast development of positive pressure anomaly together with a negative SST anomaly forming in the South Atlantic was responsible for an intensification of the South Atlantic Subtropical High, shifting the direction of the previous wind anomaly. This brings negative precipitation anomalies in NE Brazil, reaching its highest intensity right before the year 2000. The south Atlantic WAF dipole (near the Benguela coast) also shifts from pattern 1 to 2 to 3, creating a negative pressure anomaly on the south west African coast, inducing and intensifying northward wind anomalies, flooding the WAF region (Figure 3.2 and 3.3 - transition between patterns 2, 3 and 4). The precipitation anti-correlation from these two regions is then maintained and it is illustrated by the SPI phase inversion with the NE drought reaching its lowest value in 1999, while WAF reaches its highest precipitation peak in 2004 (Figure 3.1).

From 2005 to 2015 the SPI and the SOM patterns evolve from positive to negative NE precipitation anomaly, while WAF SPI stabilizes (stays between the -1 and 1 interval). A positive southwest wind anomaly near the Equator induces a northerly displacement of the decadal ITCZ, flooding Northwest Africa. However, in the Benguela region the negative/positive pressure/SST anomaly from pattern 4 weakens, inducing drought along the WAF coast (Figure 3.2 - patterns 5 and 6). At that time, an entire sector in the South Atlantic (from 50W to 10E) is developing a positive pressure anomaly followed by a negative precipitation anomaly belt, dominating the South Atlantic reaching the NE coast from 2010 to 2015 and intensifying the positive pressure anomaly in the subtropical south

Atlantic (Figure 3.2 - patterns 3 to 4 to 5 to 6).

It should be noted that although the time period with respect to the SOM patterns evolution 4 to 5 to 6 show a positive precipitation anomaly in NE (peaking in 2009, Figure 3.1), a long term transition to a drier climate is seen in the final PPT pattern in Figure 3.2. In her study using an Integrated Drought Index (IDI), Cunha et al., (2019) found this, especially from 2011 to 2017, when droughts affected most parts of Brazil, which confirms the precipitation pattern 6 from Figure 3.2. By evaluating the frequency of drought occurrence, Cunha et al., (2019) identified 2014/2015 as an atypical year due to the higher occurrence of severe and moderate droughts from north to south Brazil.

3.1.2 1870-2019 Atlantic SST anomaly cycles.

Detrended tropical Atlantic SST anomalies from 1870 to 2019 were used to create another 800 grid neural network (Figure B.1). In the form of a directed graph, Figure 3.4 represents all the Atlantic SST anomaly decadal variability from 1870 to 2019. Figure 3.4 shows a 30 to 50 year periodicity. The beginning of each cycle is defined by positive SST anomalies in the north Atlantic and mainly negative SST anomalies in the south and subtropical south Atlantic (Central panel in Figure 3.4). These patterns evolve to a positive SST anomaly in the Subtropical South Atlantic (SSA - from 30°S to 60°S) and South Atlantic (SA - from 0° to 30°S), sometimes also developing a negative SST anomaly in the North Atlantic (NA - from 0° to 30°N). This Atlantic pattern evolution repeats itself with small changes during the studied interval (from 1870 to 2019) depicting a cyclicity from those decadal SST patterns in this 30 to 50 year scale.

Cycle 1 (1870-1940)

The SST anomaly begins in 1870 with the main pattern, central square in Figure 3.4. A latitudinal dipole with a positive SST anomaly in NA and mainly negative SST anomalies in SA and SSA. This pattern seems to appear from time to time, being considered the centerpiece of the SOM graph. The evolution of this cycle is a transition from the positive anomaly in the NA to the SSA in a seesaw movement (shown by the α arc in Figure 3.4).

Two small variations take place: one in the beginning of the 20th century when the positive SST anomaly migrates to the SA (from 0° to 30°S) and returns to the central pattern (from 1900 to 1905 shown by the β arc in Figure 3.4); and another one where a

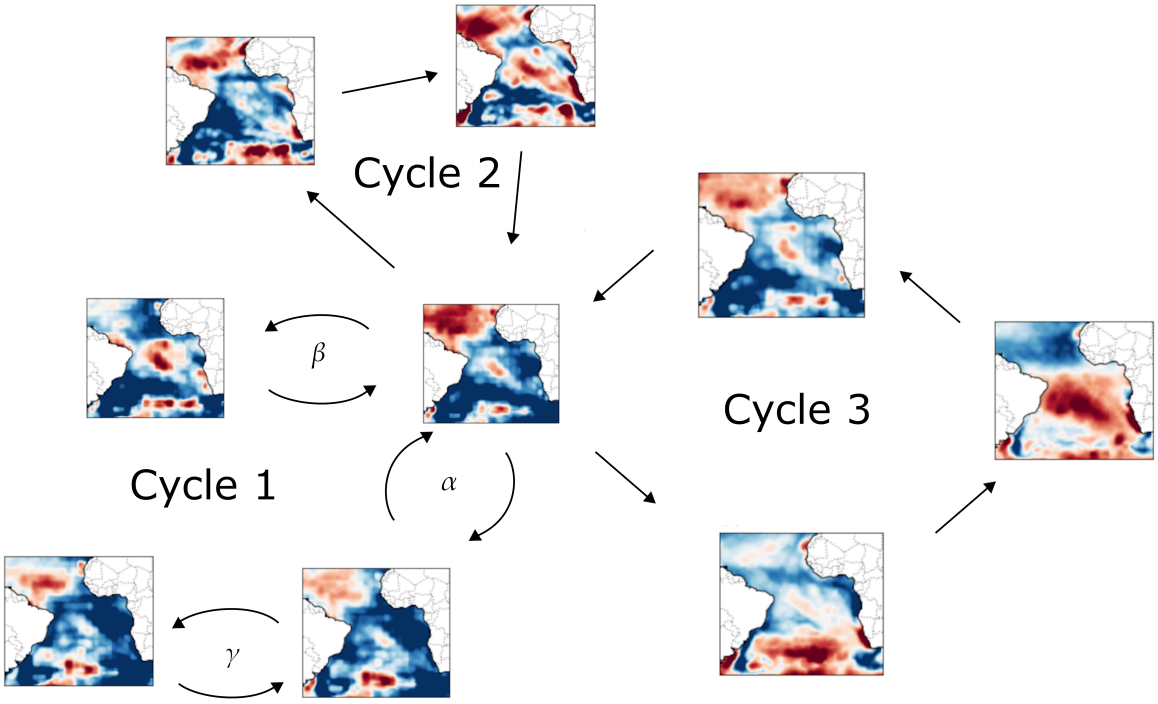

Figure 3.4: SST anomaly cycles identified by the SOM algorithm. Cycle number 1 (left) was identified as the evolution of SST anomaly from 1870 to 1940; cycle number 2 (top) was identified from 1940 to 1970; cycle number 3 (right) represents the SST anomaly in the 1970-2019 period. This Figure is a directed graph representation of the video contained in the Supplementary Materials.

small intensification from the NA SST positive anomaly takes place during a positive SST anomaly in SSA (happens twice, in 1910-1912 and 1920-1924 shown by the γ arc in Figure 3.4), soon returning to the main pattern.

Cycle 2 (1940-1970)

Cycle number 2 closely resembles cycle one, with one difference: this time the SSA develops a positive SST anomaly while the NA SST anomaly remains positive. The SSA positive anomaly is closer to Africa's coast (especially near the Benguela region - from the tip of West Africa coast to 12°S). The positive anomaly transitions north to the South Atlantic (peaking in 1960), returning to the main pattern in 1970.

Cycle 3 (1970-2019)

Cycle number 3 is the same pattern evolution seen in Figure 3.2, this time captured in the full SST anomaly time series. The pattern begins with a positive SST anomaly in SSA,

with a small negative SST anomaly in the Equator line. From 1980 to 2018 the positive SST anomaly slowly migrates into the NA while the regions around it assume a mainly negative SST anomaly, returning to the main pattern once again. As analyzed in section 3.1, this cycle was the main driver of the decadal anti-correlation seen between NE and WAF precipitation anomalies from 1979-2015 (Figure 3.1).

3.2 Model data Analysis

The Model data SST and PPT anomalies from the 21 different model runs were used to create two 20.000 neuron SOM (one for each variable). Using the DBI and Silhouette indexes for obtaining the optimal number of clusters, the feature space contains 300 SST clusters and 150 PPT clusters, representing all the possible system states of the tropical Atlantic region system in those runs, shown in Figures C.2 and C.3, respectively. Each model run time series evolution in the SOM feature space is depicted with the use of directed graphs. Each model run's variability was calculated using Shannon's Entropy.

3.2.1 Directed Graphs

The SOM is a feature space that captures the input data non linear patterns. The cluster patterns represent the possible states from the data (Figures C.2 and C.3). At each time step, the numerical model simulation is measured at one state. The full time series of each numerical model simulation is represented as a directed graph, red for the SST series and blue for the PPT model simulation (Figure 3.5 and 3.6). Each node represents a SOM cluster (a.k.a possible state of the system), which can be seen in Figures C.3 and C.2. The size of each node resembles the number of time steps the system was identified with that SOM cluster. Higher degree nodes are darker, in other words, if the system is found in a state, evolves and cycles back to that state its color gets darker.

To exemplify the model SOM directed graphs, Figure 3.5 shows in detail the evolution in time of the ECE PI run, projected in the possible states of the feature space. This figure is similar to the 20th century observations directed graph (Figure 3.4). They both depict the Atlantic ocean decadal SST cycles surrounding a main pattern (central panel from Figure 3.4 and model SOM cluster 66 in Figure 3.5). From these structures we can define and calculate macro-properties such as the entropy from the system in each time

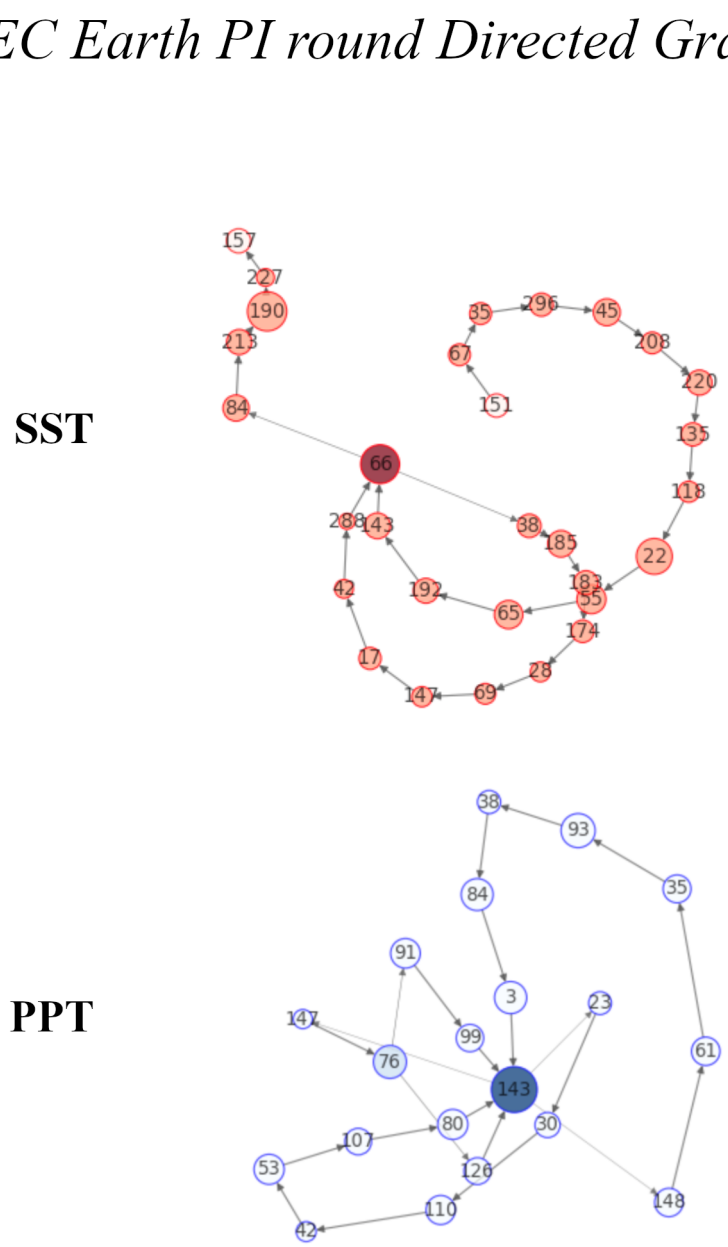

Figure 3.5: Directed graph from ECE - PI. The red graphs are for the SST system evolution and the blue graphs are for the PPT evolution. Each node represents a specific SOM Cluster, depicted in Figure C.2 and C.3

series.

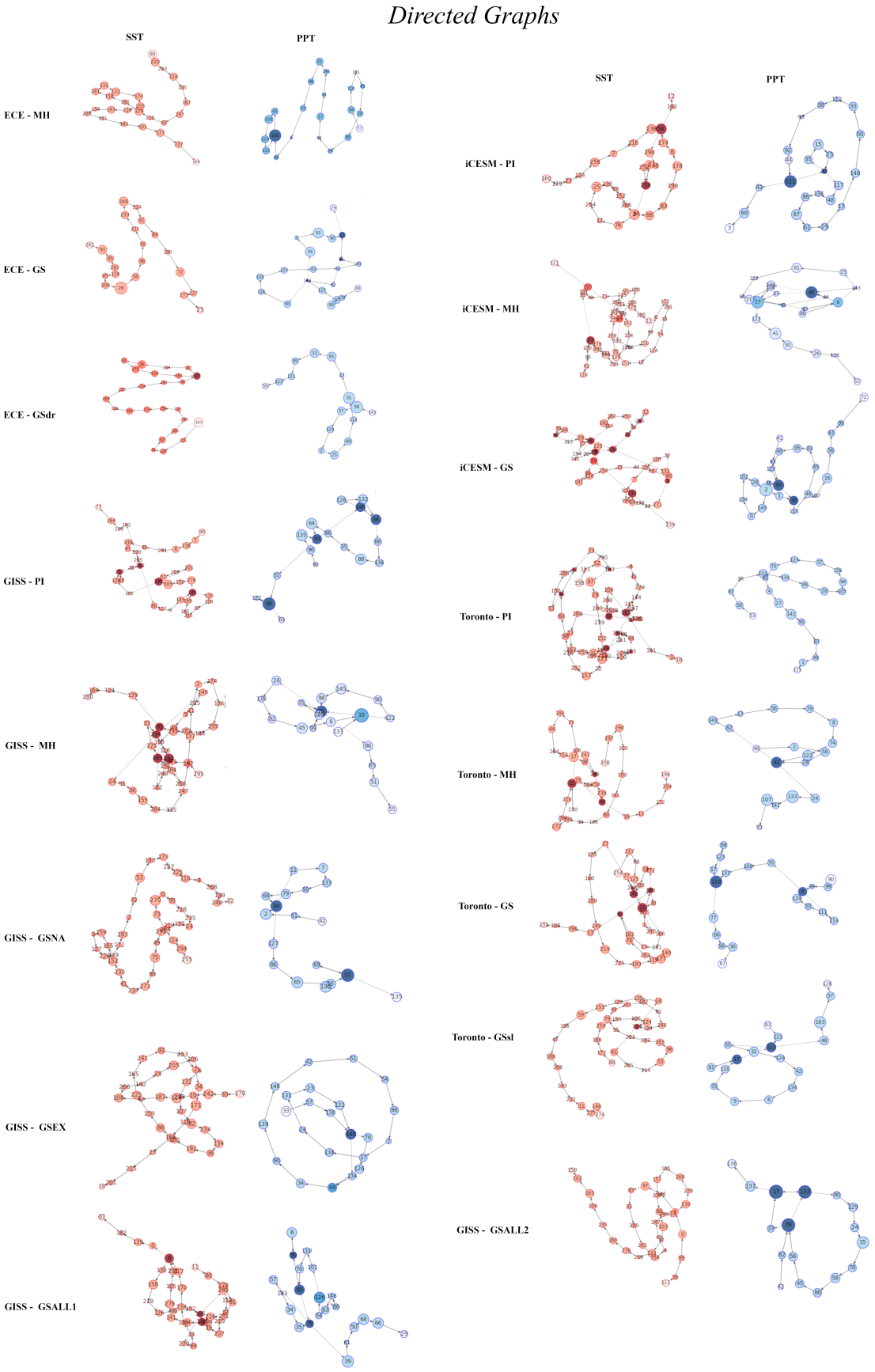

Figure 3.6: Directed graphs from: ECE - PI, MH, GS and GS with dust reduction runs (GSdr); GISS - PI, MH, GS with North Africa (NA), Extra-Tropical (EX) and full vegetation (ALL1 and ALL2) runs; iCESM - PI, MH and GS runs; and CESM Toronto - PI, MH, GS and GS with soil and lake input runs. The red graphs are for the SST system evolution and the blue graphs are for the PPT evolution. Each node represents a specific SOM Cluster, depicted in Figures C.2 and C.3

3.2.2 Shannon's Entropy

Many names can be used in reference to the randomness in a dataset. With climate data, variability can be seen as the lack of uniformity over multi-scales (Sang, 2013). Entropy measures a system's organization given its observable states. A variety of techniques can be used to apply entropy in climate studies (Mishra et al., 2009; Roushangar et al., 2020). In this study, Shannon's Entropy is used to measure randomness in a time series (Equation 2.1).

At any time step, the probability of our system (the simulated Atlantic Ocean variables) being found in each possible state (Px_k) can be calculated empirically from its model run's time series. If a specific model run has been in only one state during its whole time series, that system has probability equals one to be found in that specific state and zero on the others. Considering that $\sum_{k=1}^{N} P(x_k) = 1$, the entropy from that time series is the lowest possible, zero. Using the clusterized feature space of the SOM as the possible states of our system (x_k), at each time step our system is identified with a cluster in the feature space. Therefore, the entropy from each time series will reflect its variability in the feature space. Since we are using the same space for all the models and scenarios, we can compare their variability.

Figure 3.7 is the Entropy from each time series using different feature space clustering. The interval of clusters ('x' axis) was defined by the elbow location of the Silhouette and DBI indexes (Figure C.1).

Model variability analysis

The model runs' entropies were calculated from a set of clusterings to obtain a confidence interval that accounts for the uncertainty of the ideal number of clusters from the feature space. Table 3.1 shows the entropy mean and 95% confidence interval (two standard deviations) from Figure 3.7 for each model and scenario.

Table 3.1 - Entropy mean and Standard Deviation from the model runs.

Model	Scenario	H_{sst}	H_{ppt}
ECE	PI	3.32 ± 0.06	2.92 ± 0.09
ECE	MH	3.20 ± 0.06	2.94 ± 0.12

Continues in the next page...

ECE	GS	3.06 ± 0.09	2.91 ± 0.10
ECE	GSdr	3.10 ± 0.06	2.64 ± 0.13
GISS	PI	3.48 ± 0.06	2.84 ± 0.11
GISS	MH	3.56 ± 0.07	2.92 ± 0.08
GISS	GSALL1	3.47 ± 0.07	2.99 ± 0.08
GISS	GSALL2	3.42 ± 0.05	2.79 ± 0.08
GISS	GSEX	3.44 ± 0.09	3.06 ± 0.08
GISS	GSNA	3.61 ± 0.05	2.89 ± 0.07
iCESM	PI	3.30 ± 0.06	3.06 ± 0.11
iCESM	MH	3.67 ± 0.07	2.87 ± 0.17
iCESM	GS	3.58 ± 0.06	3.08 ± 0.10
CESM Toronto	PI	3.80 ± 0.04	3.06 ± 0.09
CESM Toronto	MH	3.40 ± 0.07	2.80 ± 0.10
CESM Toronto	GS	3.42 ± 0.09	2.97 ± 0.16
CESM Toronto	GSsl	3.38 ± 0.07	2.92 ± 0.08

ECE

The top row of Figure 3.7 shows the entropy values for the ECE model runs. ECE model runs presented the lowest SST variability and the lowest PPT variability from all model scenarios in the GSdr run. Lower variability implies a more organized system. Its decadal SST variability is larger in PI (3.32 ± 0.06, red curve), followed by MH (3.20 ± 0.06, blue curve) and lower in GS and GSdr (3.06 ± 0.09 and 3.10 ± 0.06, green and black curves respectively). This points to a dependency of the Atlantic Ocean decadal SST variability in both the mid-Holocene radiation parameters and the presence of vegetation in North Africa. For the decadal PPT all the different scenarios except the GSdr show similar variability (2.9 ± 0.1). The dust reduction from GSdr directly affects the simulation's cloud formation, which gets translated to a lower decadal PPT variability (2.64 ± 0.13, the lowest variability from all different models and scenarios).

GISS

The second row of Figure 3.7 shows the entropy values for the iCESM model runs. GISS presented the lowest SST variability amplitude between different scenarios. The entropies from different runs are all contained inside the uncertainties limits from one another. The decadal SST variability is larger in the MH and the GSNA runs (3.56 ± 0.07 and 3.61 $pm0.05$, blue and black curves). However, the uncertainties from the PI, GSEX,

GSALL1 and 2 upper limits overlap their variabilities (3.48 ± 0.06, 3.44 ± 0.09, 3.47 ± 0.07 and 3.42 ± 0.05, red, cyan, green and yellow curves respectively). The decadal PPT variability was higer in the GSEX run (3.06 ± 0.08), although comparable to the remaining runs due to their high standard deviations (PI - 2.84 ± 0.11, red curve; MH - 2.92 ± 0.08, blue curve; GSALL1 - 2.99 ± 0.08 green curve; and GSNA 2.89 ± 0.07, black curve). The lowest decadal PPT variability was GSALL2 run (2.79 ± 0.08), eventhough this run only differs of GSALL1 by their initial conditions their differences point to an internal variability issue, once again emphasizing the importance of a large ensemble for properly attributing model simulation uncertainties to different scenarios.

iCESM

The third row of Figure 3.7 shows the entropy values for the iCESM model runs. The decadal SST variability of the MH and the GS runs were similar within their uncertainties (3.67 ± 0.07 and 3.58 ± 0.06, blue and green curves respectively), while the PI run presented lower decadal SST variability (3.30 ± 0.06, red curve). While the decadal PPT variability is nearly the same for the GS and PI scenarios (3.08 ± 0.010 and 3.06 ± 0.11) and slightly lower in the MH run (2.87 ± 0.17), although its large standard deviation makes their 95% confidence interval overlap.

CESM Toronto

The bottom row of Figure 3.7 shows the entropy values for the CESM Toronto model runs. The decadal SST variability of the MH, GS and GSsl runs were equal inside their uncertainties (3.40 ± 0.07, 3.42 ± 0.09 and 3.38 ± 0.07, blue, green and black curves, respectively), while the PI scenario showed larger decadal SST variability with lower standard deviation (3.80 ± 0.04, red curve). The decadal PPT, on the other hand, shown similar variability in the PI, GS and GSsl runs (3.06 ± 0.09, 2.97 ± 0.16 and 2.92 ± 0.08, respectively), while the MH run appears with lower variability (2.80 ± 0.10), although its upper uncertainty boundary is comparable to the bottom uncertainty boundary of GSsl's variability (2.92 ± 0.08).

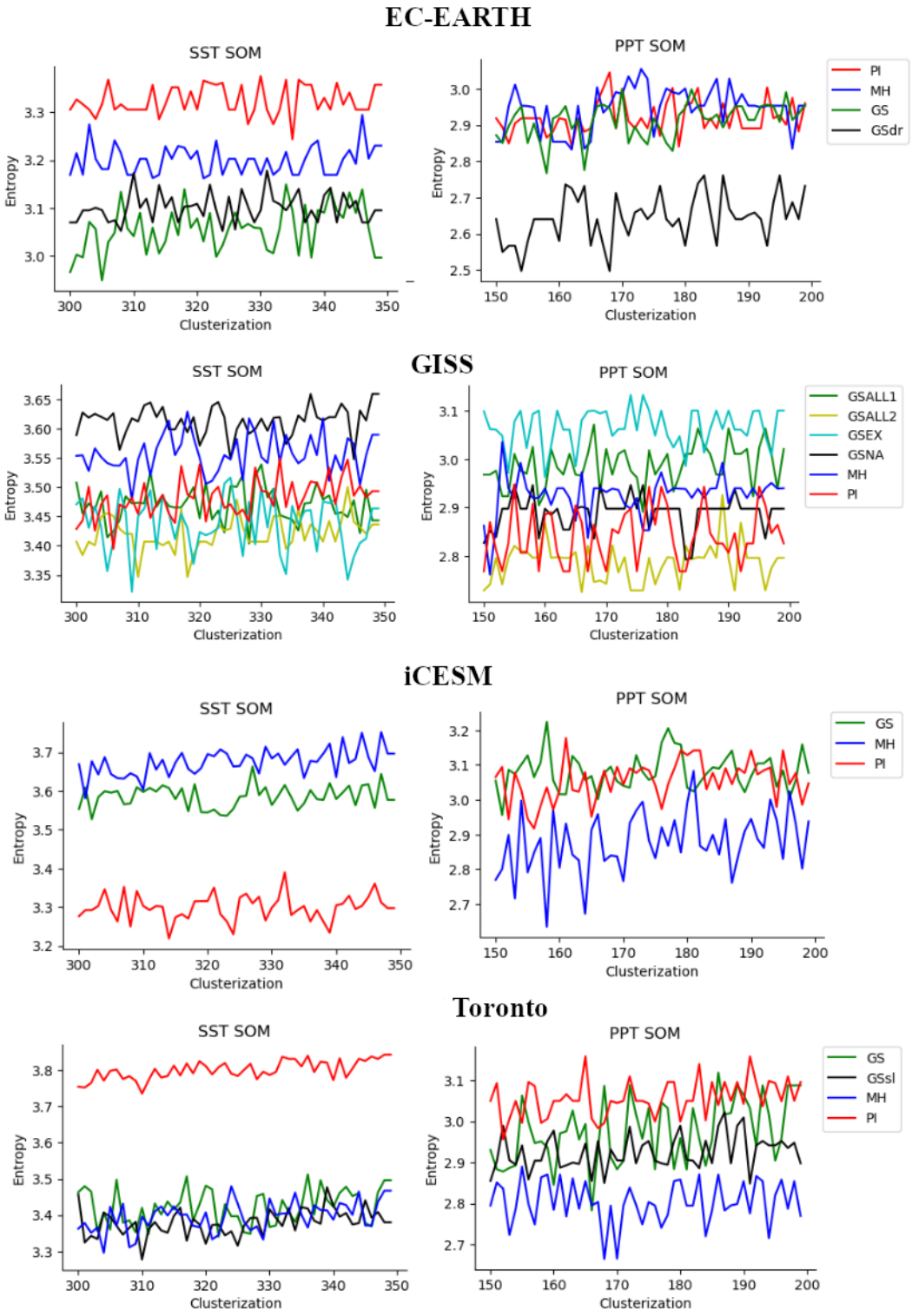

Figure 3.7: Entropy values for each model scenario with different feature space clusterings. The y axis is the Entropy value, calculated from equation 2.1. The x axis is the number of clusters (a.k.a possible states) inside the feature space. The Left Column are the Entropies from the SST patterns and the right column from the PPT patterns. From top to bottom the model runs are: EC-Earth - Pre-Industrial (PI, red), mid-Holocene (MH, blue), Green Sahara (GS, green) and Grenn Sahara with dust reduction (GSrd, black); GISS - Full vegetation (GSALL1, green; GSALL2 yellow), Extra-tropical vegetation (GSEX, cyan), North Africa vegetation only (GSNA, black), MH (blue) and PI (red); iCESM - GS (green), MH (blue) and PI (red); CESM Toronto - GS (green), GS with soil and lake inputs (GSsl, black), MH (blue) and PI (red)

Scenario variability analysis

Different models may have different parameterizations and physics. However, the Entropies were computed in a common feature space and all the model runs were simulated for 100 years, therefore, a comparison between different models is possible. Figure 3.8 shows the different models comparison within the similar scenarios.

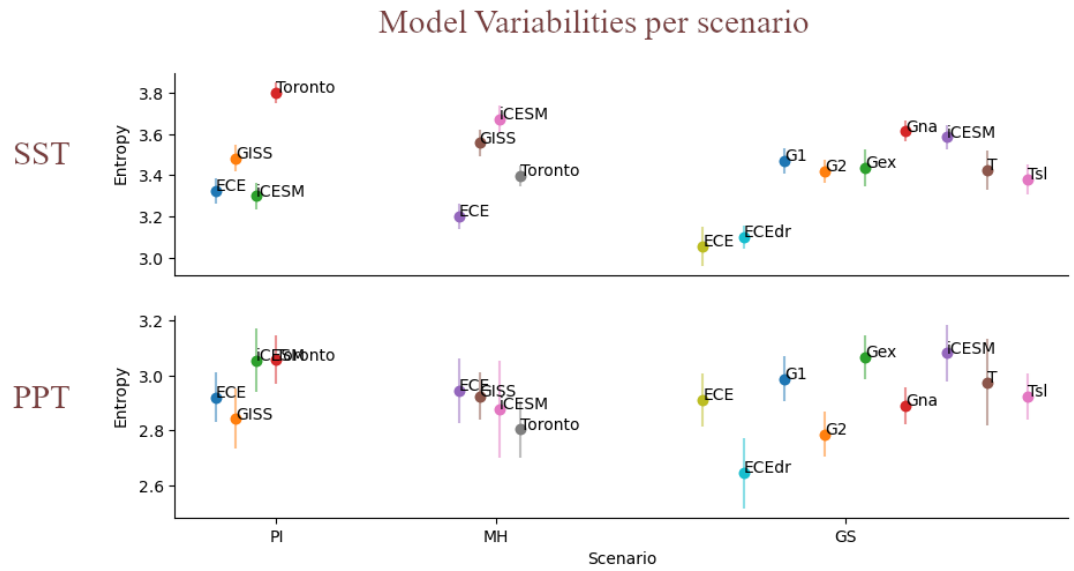

Figure 3.8: Entropy mean values and 95% confidence interval for each model scenario.

The largest entropy in the PI runs is seen in the CESM Toronto model (3.80 ± 0.04). This happens because the system evolved to a higher number of states in a shorter time. Larger entropy values implies a more chaotic system with larger variability. The different scenarios did not maintain the initial biases seen in the PI runs. There was no consensus over the growth or shrinkage of variability in any specific scenario among the different models. Also, some models showed a very entangled SST and PPT variability system. The CESM Toronto model, for example, showed changes in the decadal SST variability along with the PPT variability. On the other hand, ECE-Earth presented an almost constant PPT variability in all scenarios but GSdr (over 0.3 lower than the rest of the scenarios entropies, Table 3.1), but presented equivalent SST variabilities in the GS and GSdr scenarios. Also, the CESM models have shown diverging results. CESM-Toronto had its larger SST variability in the PI run, while iCESM had its lowest.

Chapter 4

Discussion

4.1 The tropical Atlantic modes

The SOM neuron grid does not form an orthonormal base. However, they can be seen as a set of vectors in which the input data is being projected. The SOM cluster maps are a projection from the data into a representative feature space, preserving the datasets' nonlinearity (Liu et al., 2015). However, it is no coincidence that the SOM patterns closely resemble modes of variability obtained with the definition of climate indices and the Principal Component Analysis (PCA) of the Atlantic decadal SST. When using the PCA, the linear modes obtained maximize the variance projected into an orthonormal base of the data (Preisendorfer e Mobley, 1988; Jolliffe, 2002; Haykin, 2009). In Oja's work (1982), it was proved that a single layer perceptron with a simple Hebbian updating rule, similar to the SOM algorithm (using cooperation and competition from the neurons weights in a non-supervised learning algorithm), naturally converges to the input data's first component (Oja, 1982; Haykin, 2009).

The SOM algorithm is able to extract the linear Atlantic modes. Instead of forcing orthonormality between the patterns, they force the patterns to closely correlate the input data features. Nevertheless, there are differences and similarities between the main Atlantic modes obtained via PCA and the ones shown here that are in our interest to discuss. The climate indices and principal components are useful in investigating high correlated events and local climate features, but it can be challenging to link multiple climate indices to larger scale phenomena (Deser et al., 2010). On a global scale, climate data is highly nonlinear and affected by a number of different ocean and atmosphere interactions and feedback, making a nonlinear approach such as the SOM methodology a more efficient

tool for research.

4.1.1 Atlantic Meridional Mode (AMM)

The AMM is the Atlantic ocean's SST first component in the PCA, it represents variations in the north–south gradient of the Atlantic (Tozuka et al., 2009; Deser et al., 2010; Brierley e Wainer, 2018). It can be defined and calculated as the difference of SST anomalies in two boxes, one in the North: 15–5N/50–20W, and the other one in the South: 5S–15S/20W–10E.

As we can see in Figures 3.2 and 3.4, most SST SOM patterns show a non neutral AMM, characterized by a dipole involving positive/negative SSTs in NA/SA (seen, for example, in Figure 3.2 - SOM pattern 1 and inverted in SOM pattern 6). These temperature gradients across the equator are correlated to the ITCZ decadal displacement into the warmer hemisphere and mainly to precipitation anomalies in the NE region (Deser et al., 2010).

4.1.2 Atlantic Niño

The Atlantic Niño can be defined as an equatorial Atlantic cold tongue which develops from the middle of SA all the way to the WAF coast region ($3N - 3S, 20 - 0W$). It is known to intensify atmospheric convection near the equator and keep the summer rain band from advancing into the WAF region, while increasing precipitation in interdecadal timescales at the Gulf of Guinea coast (Giannini et al., 2003; Deser et al., 2010; Wainer e Soares, 1997).

In the SOM patterns this mode appears as part of the dipole in the Benguela region (Figure 3.2), previously discussed in section 3.1.

4.1.3 South Atlantic Subtropical Dipole (SASD)

The SASD is a SST dipole in the SA (Northeast: 30–40S/10–30W, Southwest: 15–25S/0–20W) oriented in the northeast-southwest direction (Wainer et al., 2014). SST anomalies in this region have been linked to changes in the South Atlantic Subtropical High and the South Atlantic Convergence Zone (Garreaud et al., 2009; Marengo et al., 2012).

The SOM Patterns from Figures 3.2 and 3.4 can also be identified with the SASD behavior in the SSA region.

4.2 Methodology notes

There are a few relevant points that may be of interest to any future work using SOM.

4.2.1 Comparing observational with model data

The SOM methodology presented here could be used for detecting model biases. In theory, it is possible to compare observational data to numerical model outputs. However in practice, there are too many biases in numerical models output among themselves and when comparing historical simulation runs to the observational data. For example, observed SST trends have characterized recent warming in Indo-Pacific Warm Pool and slight cooling in the eastern Pacific, consistent with the strengthening of the Walker circulation cell (Coats e Karnauskas, 2017), while the Southern Ocean SST trends show different signs due to regional sea ice extent decrease or increase (Gorenstein et al., 2022b). In contrast, state-of-the-art coupled climate models generally project warming in the eastern equatorial Pacific, Walker circulation weakening, and Southern Ocean warming (Wills et al., 2022).

Also, timescales play an important role in the numerical climate model dynamics and the statistical bias correction of their outputs (Haerter et al., 2011; Döös et al., 2020). In this study, we did not compare the observational data SOM patterns to the model outputs SOM patterns, but, in theory, we could create a feature space trained with the decadal anomaly patterns of the Atlantic ocean climate variables in historical model runs and observational data together. In this scope, the observational data variability and the different model scenarios variability could be compared as their systems would be analyzed in the same feature space. However, instead of creating an homogeneous feature space, this approach would most likely result in a clear separation from model and observational data in the feature space, emphasizing the biases between observations and numerical models variability found in multi-decadal timescale (McGregor et al., 2018; Wills et al., 2022).

4.2.2 SOM neural grid dimension

As mentioned before, the multivariable SOM using observational data had a rectangular grid of 800 neurons (20x40, Figure A.2), while the full time dimension of the data was 1728 (4 variables with 432 monthly data each). The single variable SOM using 1870-2019 SST anomalies had also a 800 neuron grid (40x20, Figure B.1)), the time dimension had

1788 monthly SST anomalies. Also, both model data SOMs (PPT and SST) had 20 thousand neurons grids (200x100, not shown), the model data had a total of approximately 28 thousand months for each variable (counting all the models in all the different scenarios). Those dimensions were chosen after extensive research, trial and error. They were a consequence from the chosen metric, used in the neural network's learning and the posterior analysis (R-squared differences from images). Since we want our feature space to be able to capture and project the nuances and evolution of the data, the grid's dimension must be comparable to the input data's dimension we wish to reduce, otherwise the R-squared difference wont be positive when calculating the variance and correlation of each SOM neuron at each time step from the time series (Figure A.2). Therefore, the SOM neural grid should not apply a drastic dimensionality reduction by itself. It is the clustering of that grid (clusterization of the feature space) that will provide the ultimate dimensionality reduction for our problem.

Chapter 5

Summary and Conclusion

Dimensionality reduction

Using SOM as a non-linear pattern extraction tool to project the data into a feature space produced results which led to easier data interpretation. The 1979-2015 data had originally a time dimension of 432 months for each variable (SST, PPT, pressure and wind anomalies), which was reduced to 7 cluster patterns (Figure 3.2). In the 1870-2019 Atlantic decadal SST anomalies data the dimensionality reduction was even greater, from 1788 months to 9 patterns (Figure 3.4). In the model data SOM, the 28.000 months of SST and PPT anomalies were reduced to 300 and 150 cluster patterns (a.k.a system possible states), respectively.

1980-2015 Atlantic multi-variable Cycle

The 1979-2015 multivariable SOM (Figure 3.2) describe how SST, wind and pressure anomalies from the Atlantic Ocean control the anti-correlated precipitation between NE/WAF regions at decadal scale (Figure 3.1). The combined AMM, Benguela niño and SASD were responsible for the decadal precipitation in the tropical Atlantic region and its variability. The variance evolution of the variables (Figure 3.3) show that the SST anomaly from pattern 0, which peaked in 1980, seems to be responsible for the following peak in the pressure and wind patterns 1 in the SA, bringing rain to the NE coast in the 80's. The inversion from this pattern for positive/negative SST/pressure anomalies in SA seems to be responsible for the drought pattern in NE during the 90's. Patterns 1, 2 and 3 also show an interesting evolution on the WAF coast, in the Benguela region (from 10S to 30S, 10E). The positive/negative SST anomaly dipole reverses in the Benguela region from pattern 1 to 3, negative pressure anomalies develop high latitudinal wind anomalies

to the Equator, which persisted until 2005 (pattern 4), culminating in a highly positive precipitation pattern from 1995 to 2005 in WAF (over 2 SPI, Figure 3.1). Forming the anti-correlation from the NE and WAF SPI from 1980 to 2000's.

In the subsequent pattern evolution from 4 to 5 to 6: a SST positive anomaly forms north of the Equator. With a negative SST anomaly in SA and SSA. Simultaneously, the SST and pressure dipole in the Benguela region weakens, creating a positive pressure anomaly centered in the SSA, which dominates the whole basin. This increased the South Atlantic Subtropical High, shifting the decadal ITCZ belt position to the north and generating the drought scenario seen in Brazil during the 2011-2019 time period (Cunha et al., 2019).

The SST, pressure, wind and precipitation anomalies coupling was not forced. The SOM algorithms were calculated together as a multivariable neural network and individually, both approaches yielded the exact same results. Resulting in the same number of clusters, same patterns and variance evolution in time. Also, these 1979-2015 SOM patterns show a high level of anti-symmetry. SOM Cluster patterns 0 and 1 (Tropical Atlantic in 1980s) are almost completely anti-symmetrical to the SOM Cluster patterns 5 and 6 (Tropical Atlantic in 2010s). This points to a possible decadal variability envelope the Atlantic ocean oscillated in.

1870-2019 Atlantic SST anomaly cycles

Cycles 1, 2 and 3 from the 1870-2019 Atlantic decadal SST anomalies are very alike, they all contain the same evolution of positive SST anomaly, beginning in the north (0°-30N), passing to the subtropical south (30°S-60°S) and ending in SA region (0°-30°S). This suggests a repetition with small variations of this 30 to 40 year cycle. Although Atlantic decadal SST variability may not be fully represented by cycles 1,2 and 3 it is all the observational information at scientific disposal for now.

Each cycle may contain different precipitation patterns, as they are evolving from slightly different initial conditions. Climate signals from SST anomalies can differ in position and magnitude of their amplitude maximum, generating each time an unique climate response (Cai et al., 2020). Cycle 2 exhibits a positive SST anomaly in the West African coast, resembling patterns 0 and 1 in Figure 3.2, where the NE and WAF precipitation were highly anti-correlated (from 1980 to 1995 with positive SPI in NE and negative in

WAF). Although the α transition in cycle 1 does not show a positive SST anomaly pattern in WAF's coast. In other words, there may be other precipitation, wind and pressure anomaly patterns emerging from the Atlantic SST cycles that are not present in the 1979-2015 satellite data available.

Numerical models variability

The model data SOM has shown the methodology's true potential for data processing. Using 28 thousand montlhy tropical Atlantic SST and PPT anomalies, the SOM's feature space was able to discretize 17 climate simulations from 4 different numerical models (ECE, GISS, iCESM and CESM-Toronto) in eight different scenarios (PI, MH, GS, GSdr, GSNA, GSEX, GSALL and GSsl) into simple directed graphs (Figure 3.5 and 3.6. The directed graphs entropy was able to attribute the effect of different insolation, vegetation and dust emission scenarios to the tropical Atlantic's decadal variability (Figure 3.7). ECE GSdr scenario was the simulation run with the lowest variability, while the remaining ECE scenarios had equivalent PPT entropies (Table 3.1). This shows that the dust reduction was directly responsible for the model's PPT variability. While the vegetation inputs (GS and GSdr) had a more noticeable effect on the SST variability. The GISS model shown the lowest SST entropy amplitude between different scenarios. iCESM and CESM-Toronto have shown divergences in the SST entropies, indicating that the internal variability from the CESM models may provoke variability differences larger than the ones seen between different scenarios. Therefore, large ensembles are needed for more robust findings and uncertainty attributions for the numerical models' internal variability.

Next steps

The possible next steps for this study are: 1 - Improving the feature space clustering. Many of the Cluster patterns from the model SOM are alike (Figures C.2 and C.3), reducing the number of possible states in feature space will help elucidate and depict the tropical Atlantic cycles, specially in the model data SOM. 2 - Use a feature space from a SOM trained with observational data only to classify the possible states of a model run simulation. This approach can be used to find and shed light upon model and observation biases, as well as their causes. 2 - The development of a Recurrent Neural Network (RNN) or Koopman Neural network (Azencot et al., 2020). Designed to evolve dynamical systems,

these networks could be trained with observational and/or model data to make predictions about future states of the Atlantic Ocean influence region. Or even, reconstruct the precipitation patterns from before precipitation data started being collected. 3 - Assessment of other influences such as the Pacific Ocean tropical and extratropical teleconnections in NE and WAF regions. The Pacific ocean variability modes are a known source of influence in South American climate and the Atlantic Ocean (Cai et al., 2020). They may help explain the dynamic evolution of the Tropical Atlantic SOM patterns. 4 - Study and compare the usage of the SOM algorithm with a Deep clustering auto-encoder network, since the clustering of data is a highly desirable capability in a range of disciplines and sciences (Guo et al., 2017; Shah e Koltun, 2018; Bo et al., 2020; Chadebec et al., 2022b).

The codes developed by this study to create SOM feature spaces using climate data are available in a GitHub repository through the following link: https://github.com/IuriGorenstein/SOM.

Bibliography

Adams J., , 2017 climate indices, an open source Python library providing reference implementations of commonly used climate indices

Adler R., Huffman G., Chang A., Ferraro R., Xie P., Janowiak J., Rudolf B., Schneider U., Curtis S., Bolvin D., Gruber A., Susskind J., Arkin P., , 2003 The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present)

Ajjur S. B., Al-Ghamdi S. G., Global Hotspots for Future Absolute Temperature Extremes From CMIP6 Models, Earth and Space Science, 2021, vol. 8

Arias P., Bellouin N., Coppola E., Jones R., Krinner G., Marotzke J., Naik V., Palmer M., Plattner G.-K., Rogelj J., et al., Climate Change 2021: The Physical Science Basis. Contribution of Working Group14 I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Technical Summary, 2021a

Arias P. A., Bellouin N., Coppola E., Jones R. G., Krinner G., Marotzke J., Naik V., Palmer M. D., Plattner G.-K., Rogelj J., et al., Technical summary, Climate Change, 2021b, pp 33–144

Arkin P., Xie P., for Atmospheric Research Staff (Eds). Last modified 17 Apr 2020. N. C., , 2020 The Climate Data Guide: CMAP: CPC Merged Analysis of Precipitation

Azencot O., Erichson N. B., Lin V., Mahoney M., Forecasting Sequential Data Using Consistent Koopman Autoencoders. In Proceedings of the 37th International Conference on Machine Learning , vol. 119 of Proceedings of Machine Learning Research, PMLR, 2020, p. 475

Berger Milankovitch Theory and climate, AGU. Ressearch Letters, 1988

Berger A., Long-Term Variations of Daily Insolation and Quaternary Climatic Changes, Journal of the Atmospheric Sciences, 1978, vol. 35, p. 2362–2367

Bo D., Wang X., Shi C., Zhu M., Lu E., Cui P., Structural deep clustering network. In Proceedings of The Web Conference 2020 , 2020, p. 1400

Bonfils C., Santer B., Investigating the possibility of a human component in various pacific decadal oscillation indices, Clim Dyn, 2011, vol. 37, p. 1457–1468

Bova S., Rosenthal Y., Liu Z., Godad S. P., Yan M., Seasonal origin of the thermal maxima at the Holocene and the last interglacial, Nature, 2021, vol. 589, p. 548–553

Breiman L., Statistical modeling: The two cultures, Statistical science, 2001, vol. 16, p. 199

Brierley C., Wainer I., Inter-annual variability in the tropical Atlantic from the Last Glacial Maximum into future climate projections simulated by CMIP5/PMIP3, Climate of the Past, 2018, vol. 14, p. 1377

Brunton S. L., Budišić M., Kaiser E., Kutz J. N., Modern Koopman theory for dynamical systems, arXiv preprint arXiv:2102.12086, 2021

Bühler J. C., Axelsson J., Lechleitner F. A., Fohlmeister J., LeGrande A. N., Midhun M., Sjolte J., Werner M., Yoshimura K., Rehfeld K., Investigating stable oxygen and carbon isotopic variability in speleothem records over the last millennium using multiple isotope-enabled climate models, Climate of the Past, 2022, vol. 18, p. 1625

Burges C., , 2005 Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, chapter Geometric Methods for Feature Selection and Dimensional Reduction: A Guided Tour

Burkea K. D., Williams J. W., Chandler M. A., Haywood A. M., Lunt D. J., Otto-Bliesner B. L., Pliocene and Eocene provide best analogs for nearfuture climates, PNAS, 2018

Cai W., McPhaden M. J., Grimm A., Rodrigues R. R., Taschetto Andréa S.and Garreaud R. D., Dewitte B., Poveda G., Ham Y.-G., Santoso A., Ng B., Anderson W., Wang G.,

Geng T., Jo H.-S., Marengo J. A., Alves L. M., Osman M., Li S., Wu L., Karamperidou C., Takahashim K., Vera C., Climate impacts of the El Niño–Southern Oscillation on South America, Nat Rev Earth Environ, 2020, vol. 1, p. 215–231

Chadebec C., Vincent L. J., Allassonnière S., Pythae: Unifying Generative Autoencoders in Python – A Benchmarking Use Case, arXiv preprint arXiv:2206.08309, 2022a

Chadebec C., Vincent L. J., Allassonnière S., Pythae: Unifying Generative Autoencoders in Python – A Benchmarking Use Case, arXiv preprint arXiv:2206.08309, 2022b

Chang L.-C., Chang F.-J., Yang S.-N., Tsai F.-H., Chang T.-H., Herricks E. E., Self-organizing maps of typhoon tracks allow for flood forecasts up to two days in advance, Nature Communications, 2020, vol. 11, p. 1983

Coats S., Karnauskas K., Are simulated and observed twentieth century tropical Pacific sea surface temperature trends significant relative to internal variability?, Geophysical Research Letters, 2017, vol. 44, p. 9928

Colose C. M., LeGrande A. N., Vuille M., The influence of volcanic eruptions on the climate of tropical South America during the last millennium in an isotope-enabled general circulation model, Climate of the Past, 2016, vol. 12, p. 961

Costa M. S., Oliveira-Júnior F., Santos P., Filho W., Gois G., Blanco C., Teodoro P., Junior C., Santiago D., Souza E., Jardim A., Rainfall extremes and drought in Northeast Brazil and its relationship with El Niño–Southern Oscillation, Royal Meteorological Society, 2021, vol. 41, p. E2111

Cunha A. P. M. A., Leal M. Z. K. D., Costa L., Cuarta L. A., Marengo J. A., Tomasella J., Vieira R. M., Barbosa A. A., Cunningham C., Garcia J. V. C., Broedel E., Alvalá R., Ribeiro G., Extreme Drought Events over Brazil from 2011 to 2019, Atmosphere, 2019

da Silva L. I. L., , 2022 Speech at COP27

Davies D. L., Bouldin D. W., A cluster separation measure, IEEE transactions on pattern analysis and machine intelligence, 1979, pp 224–227

Deser C., Alexander M. A., Xie S.-P., Phillips A. S., Sea Surface Temperature Variability: Patterns and Mechanisms, Annual Review of Marine Science, 2010, vol. 2, p. 115

Deser C., Capotondi A., Saravanan R., Phillips A. S., Tropical Pacific and Atlantic climate variability in CCSM3, Journal of Climate, 2006, vol. 19, p. 2451

Deser C., Phillips A., Bourdette V., Teng H., Uncertainty in climate change projections: the role of internal variability, Climate dynamics, 2012, vol. 38, p. 527

Desgraupes B., Clustering indices, University of Paris Ouest-Lab Modal'X, 2013, vol. 1, p. 34

Dhrubajyoti S., Karnauskas K. B., Goodkin N. F., Tropical Pacific SST and ITCZ Biases in Climate Models: Double Trouble for Future Rainfall Projections?, AGU, Geophys. Res. Lett., 2019

Ding Z., Lu R., Wang L., Yu L., Liu X., Liu Y., Liu L., Bai M., Wang S., Early-Mid Holocene climatic changes inferred from colors of eolian deposits in the Mu Us Desert, Geoderma, 2021, vol. 401, p. 115172

Döös K., Brodeau L. P. L., Campino A. A., , 2020 Basic numerical methods in meteorology and oceanography

Döscher R., Acosta M., Alessandri A., Anthoni P., Arneth A., Arsouze T., Bergmann T., Bernadello R., Bousetta S., Caron L.-P., et al., The EC-earth3 Earth system model for the climate model intercomparison project 6, Geoscientific Model Development Discussions, 2021, vol. 1, p. 2021

Ferrero B., Tonelli M., Marcello F., Wainer I., Long-term regional dynamic sea level changes from CMIP6 projections, Advances in Atmospheric Sciences, 2021, vol. 38, p. 157

Galvan D., Effting L., Cremasco H., Conte-Junior C. A., The Spread of the COVID-19 Outbreak in Brazil: An Overview by Kohonen Self-Organizing Map Networks, Medicina, 2021, vol. 57

Garreaud R. D., Vuille M., Compagnucci R., Marengo J., Present-day South American climate, Palaeogeography, Palaeoclimatology, Palaeoecology, 2009, vol. 281

Giannini A., Saravanan R., Chang P., Oceanic forcing of Sahel rainfall on interannual to interdecadal time scales, Science, 2003, vol. 302, p. 1027

Gibson P. B., Perkins Kirkpatrick S. E., Uotila P., Pepler A. S., Alexander L. V., On the use of self organizing maps for studying climate extremes, Journal of Geophysical Research:Atmospheres, 2017, vol. 122, p. 3891

Gorenstein I., Prado L. F., Bianchini P. R., Wainer I., Griffiths M. L., Pausata F. S., Yokoyama E., A fully calibrated and updated mid-Holocene climate reconstruction for Eastern South America, Quaternary Science Reviews, 2022, vol. 292, p. 107646

Gorenstein I., Wainer I., Mata M. M., Tonelli M., Revisiting Antarctic sea-ice decadal variability since 1980, Polar Science, 2022, vol. 31, p. 100743

Government of Brazil, 2020 Fourth National Communication of Brazil, UNFCCC, 537 pp. https://unfccc.int/documents/267657

Griffiths M. L., Johnson K. R., Pausata F. S. R., White J. C., Henderson G. M., Wood C. T., Hongying Yang V. E., Conrad C., Sekhon N., End of Green Sahara amplified mid- to late Holocene megadroughts in mainland Southeast Asia, Nat Commun, 2020, vol. 11, p. 4204

Gu Q., Gervais M., Exploring North Atlantic and North Pacific Decadal Climate Prediction Using Self-Organizing Maps, Journal of Climate, 2021, vol. 34, p. 123–141

Guenang G. M., Kamga F. M., Computation of the Standardized Precipitation Index (SPI) and Its Use to Assess Drought Occurrences in Cameroon over Recent Decades, Journal of Apllied Meteorology and Climatology, 2014, vol. 53 (10): 2310–2324

Guo X., Liu X., Zhu E., Yin J., Deep clustering with convolutional autoencoders. In International conference on neural information processing , 2017, p. 373

Guttman N. B., COMPARING THE PALMER DROUGHT INDEX AND THE STANDARDIZED PRECIPITATION INDEX, JAWRA, 2007, vol. 34, p. 113

Haerter J., Hagemann S., Moseley C., Piani C., Climate model bias correction and the role of timescales, Hydrology and Earth System Sciences, 2011, vol. 15, p. 1065

Hagos S. M., Cook K. H., Influence of Surface Processes over Africa on the Atlantic Marine ITCZ and South American Precipitation, American Meteorological Society - Journal of Climate, 2005

Ham Y., Kug J., Park J., Two distinct roles of Atlantic SSTs in ENSO variability: North Tropical Atlantic SST and Atlantic Niño, Geophysical Research Letters, 2013

Harrison S. P., Bartlein P., , 2012 in Henderson-Sellers A., McGuffie K., eds, , The Future of the World's Climate (Second Edition) second edition edn, Elsevier Boston pp 403–436

Haykin S., Neural Networks and Machine Learning. Bookman, 2009

Hey A. J., Tansley S., Tolle K. M., et al., The fourth paradigm: data-intensive scientific discovery. vol. 1, Microsoft research Redmond, WA, 2009

Hounsou-Gbo G. A., Servain J., Araujo M., Caniaux G., Bourlès B., Fontenele D., Martins E. S. P., SST indexes in the Tropical South Atlantic for forecasting rainy seasons in Northeast Brazil, Atmosphere, 2019, vol. 10, p. 335

Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H., Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv preprint arXiv:1704.04861, 2017

Huo Y., Peltier W. R., Chandan D., Mid-Holocene monsoons in South and Southeast Asia: dynamically downscaled simulations and the influence of the Green Sahara, Clim. Past, 2021, vol. 17, p. 1645–1664

Ionita M., Scholz P., Chelcea S., Assessment of droughts in Romania using the Standardized Precipitation Index, Nat Hazards, 2016, vol. 81, p. 1483–1498

IPCC, 2020 AR5 Reference Regions

Jimenez-Muñoz J. C., Mattar C., Barichivich J., Santamaría-Artigas A., Takahashi K., Malhi Y., Sobrino J. A., van der Schrier G., Record-breaking warming and extreme drought in the Amazon rainforest during the course of El Niño 2015–2016, Nature, 2016

Jolliffe I., , 2002 Principal component analysis. New York: Springer-Verlag

Kalnay e. a., The NCEP/NCAR 40-year reanalysis project, Bull. Amer. Meteor. Soc., 1996, vol. 77, p. 437

Kaufman D., McKay N., Routson C., Erb M., Davis B., Heiri O., Jaccard S., Tierney J., Zhilich . . S., A global database of Holocene paleotemperature records, Scientific Data, 2020

Kay J. E., Deser C., Phillips A., Mai A., Hannay C., Strand G., Arblaster J. M., Bates S., Danabasoglu G., Edwards J., et al., The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability, Bulletin of the American Meteorological Society, 2015, vol. 96, p. 1333

Kohonen T., Self-organized formation of topologically correct feature maps, Biological Cybernetics, 1982, vol. 43, p. 59–69

Kohonen T., Self-Organizing Maps. Springer-Verlag Berlin Heidelberg, 1995

Kohonen T., Exploration of very large databases by self-organizing maps, Houston, 1997, vol. 1, p. PL1–PL6

LeGrande A. N., et al., Stability of ENSO and its tropical Pacific teleconnections over the Last Millennium, 2015

Lehner F., Deser C., Maher N., Marotzke J., Fischer E. M., Brunner L., Knutti R., Hawkins E., Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth System Dynamics, 2020, vol. 11, p. 491

Liu Y., Weisberg R. H., Mooers C. N., Performance evaluation of the self organizing map for feature extraction, Journal of Geophysical Research, 2015, vol. 111

Liu Z., Harrison S. P., Kutzbach J., Otto-Bliesner B., Global monsoons in the mid-Holocene and oceanic feedback, Climate Dynamics, 2002, vol. 22, p. 157–182

Liu Z., Harrison S. P., Kutzbach J., Otto-Bliesner B., Global monsoons in the mid-Holocene and oceanic feedback, Climate Dynamics, 2004, vol. 22, p. 157–182

McGregor S., Stuecker M. F., Kajtar J. B., England M. H., Collins M., Model tropical Atlantic biases underpin diminished Pacific decadal variability, Nature Climate Change, 2018, vol. 8, p. 493

McKee T. B., Drought monitoring with multiple time scales. In Proceedings of 9th Conference on Applied Climatology, Boston, 1995 , 1995

MacQueen J., et al., Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability , vol. 1, 1967, p. 281

Mantua N., Hare S., Zhang Y., Wallace J., Francis R., A Pacific interdecadal oscillation with impacts on salmon production, Bull Am Met Soc, 1997, vol. 58, p. 1069–1079

Marengo J. A., Liebmann B., A. M. G., Misra V., Dias P. L. S., Cavalcanti I. F. A., Carvalho L. M. V., Berbery E. H., Ambrizzi T., Vera C. S., Saulo A. C., Paegle J. N., Zipser E., Seth A., Alves L. M., Recent developments on the South American monsoon system, International Journal of Climatology, 2012, vol. 32, p. 1

Melin P., Monica J. C., Sanchez D., Castillo O., Analysis of Spatial Spread Relationships of Coronavirus (COVID-19) Pandemic in the World using Self Organizing Maps, Chaos, Solitons & Fractals, 2020, vol. 138, p. 109917

Mishra A. K., Özger M., Singh V. P., An entropy-based investigation into the variability of precipitation, Journal of Hydrology, 2009, vol. 370, p. 139

Nash D. J., Cort G. D., Chase B. M., Verschuren D., Nicholson S. E., Shanahan T. M., Asrat A., Lezine A.-M., Grab S. W., African hydroclimatic variability during the last 2000 years, Quaternary Science Reviews, 2016, vol. 154, 15 December 2016, Pages 1-22

NCAR, 2020 The Climate Data Guide: CRU TS Gridded precipitation and other meteorological variables since 1901

Nnamchi H. C., Li J., Influence of the South Atlantic Ocean Dipole on West African Summer Precipitation, American Meteorological Society - Journal of Climate, 2011

NOAA/OAR/ESRL, 2020 GHCN Gridded V2

Nobre P., Shukla J., Variations of sea surface temperature, wind stress, and rainfall over the tropical Atlantic and South America, J. Climate, 1996, vol. 9, p. 2464–79

Oja E., Simplified neuron model as a principal component analyzer, Journal of mathematical biology, 1982, vol. 15, p. 267

Pausata F. S. R., Zhang Q., Muschitiello F., Lu Z., Chafik L., Niedermeyer E. M., Stager J. C., Cobb K. M., Liu Z., Greening of the Sahara suppressed ENSO activity during the mid-Holocene, Nat Commun, 2017, vol. 8, p. 16020

Polo I., Rodríguez-Fonseca B., Losada T., García-Serrano J., Tropical Atlantic Variability Modes (1979–2002). Part I: Time-Evolving SST Modes Related to West African Rainfall, American Meteorological Society - Journal of Climate, 2008

Preisendorfer R. W., Mobley C. D., Principal component analysis in meteorology and oceanography, Developments in atmospheric science, 1988, vol. 17

Rayner N., Parker D., Horton E., , 2003 Hadley Centre for Climate Prediction and Research, Met Office, Bracknell, UK global analyses of sea surface temperature

Rice S. E., Patrick S., Index of State Weakness In the Developing World Report. Brookings Global Economy and Development, 2008

Richardson A., Risien C., Shillington F., Using self-organizing maps to identify patterns in satellite imagery, Progress in Oceanography, 2003, vol. 59, p. 223

Rojas O., YanYun L., Cumani R., Understanding the drought impact of El Niño on the global agricultural areas: an assessment using FAO's Agricultural Stress Index (ASI), Environment and Natural Resources Management Series, Climate Change, 2014

Roushangar K., Alizadeh F., Adamowski J., Saghebian S. M., Exploring the multiscale changeability of precipitation using the entropy concept and self-organizing maps, Journal of Water and Climate Change, 2020, vol. 11, p. 655

Saada N., Abu-Romman A., Multi-site Modeling and Simulation of the Standardized Precipitation Index (SPI) in Jordan, Journal of Hydrology: Regional Studies, 2017, vol. 14, p. 83

Samarasinghe S., Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. Auerbach publications, 2016

Sang Y.-F., Wavelet entropy-based investigation into the daily precipitation variability in the Yangtze River Delta, China, with rapid urbanizations, Theoretical and Applied Climatology, 2013, vol. 111, p. 361

Santos J. L., The Impact of El Niño - Southern Oscillation Events on South America, Advances in Geosciences, European Geosciences Union, 2006, 6, pp.221-225. ⟨hal-00296943⟩, 2006

Schmidt G. A., Kelley M., Nazarenko L., Ruedy R., Russell G. L., Aleinov I., Bauer M., Bauer S. E., Bhat M. K., Bleck R., et al., Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, Journal of Advances in Modeling Earth Systems, 2014, vol. 6, p. 141

Shah S. A., Koltun V., Deep continuous clustering, arXiv preprint arXiv:1803.01449, 2018

Smith Mayle F. E., Impact of mid- to late Holocene precipitation changes on vegetation across lowland tropical South America: a paleo-data synthesis, Quaternary Research, 2017, p. 1–22

Smith B., Wårlind D., Arneth A., Hickler T., Leadley P., Siltberg J., Zaehle S., Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, Biogeosciences, 2014, vol. 11, p. 2027

Sun Q., Miao C., Duan Q., Ashouri H., Sorooshian S., Hsu K., A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons, AGU - Reviews of Geophysics, 2017

Svoboda M., Hayes M., Wood D., Standardized Precipitation Index User Guide. World Meteorological Organizatio,(WMO-No. 1090), Geneva, 2012

Torres R. R., Benassi R. B., Martins F. B., Lapola D. M., Projected impacts of 1.5 and 2°C global warming on temperature and precipitation patterns in South America, International Journal of Climatology, 2021

Tozadore D. C., Ranieri C. M., Nardari G. V., Romero R. A., Guizilini V. C., Effects of emotion grouping for recognition in human-robot interactions. In 2018 7th Brazilian Conference on Intelligent Systems (BRACIS) , 2018, p. 438

Tozuka T., Yamagata T., et al., Interannual variability of the Guinea Dome and its possible link with the Atlantic Meridional Mode, Climate dynamics, 2009, vol. 33, p. 985

Van Der Maaten L., Postma E., Van den Herik J., et al., Dimensionality reduction: a comparative, J Mach Learn Res, 2009, vol. 10, p. 13

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I., Attention is all you need, Advances in neural information processing systems, 2017, vol. 30

Villamayor J., Influence of the Sea Surface Temperature Decadal Variability on Tropical Precipitation: West African and South American Monsoon, Universidad Complutense de Madrid, 2020, PhD dissertation

Wainer I., Prado L. F., Khodri M., Otto Bliesner B., Reconstruction of the South Atlantic Subtropical Dipole index for the past 12,000 years from surface temperature proxy, Scientific reports, 2014, vol. 4, p. 1

Wainer I., Soares J., North northeast Brazil rainfall and its decadal-scale relationship to wind stress and sea surface temperature, Geophysical Research Letters, 1997, vol. 24, p. 277

Wang F., Franco-Penya H.-H., Kelleher J. D., Pugh J., Ross R., An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity. In International Conference on Machine Learning and Data Mining in Pattern Recognition , 2017, p. 291

Wanner H., Beer J., Bütikofer J., Crowley T. J., Cubasch U., Flückiger J., Goosse H., Grosjean M., Joos F., Kaplan J. O., Küttel M., Müller S. A., Prentice I. C., Solomina O., Stocker T. F., Tarasov P., Wagner M., Widmann M., Mid- to Late Holocene climate change: an overview, Quaternary Science Reviews, 2008, vol. 27, p. 1791

Wills R. C., Dong Y., Proistosecu C., Armour K. C., Battisti D. S., Systematic Climate Model Biases in the Large-Scale Patterns of Recent Sea-Surface Temperature and Sea-Level Pressure Change, Geophysical Research Letters, 2022, vol. 49, p. e2022GL100011

Willshaw D., von der Malsburg C., How patterned neural connections can be set up by self-organization, Proceedings of the Royal Society of London Series B, 1976, vol. 194, p. 431–445

Wyser K., van Noije T., Yang S., von Hardenberg J., O'Donnell D., Döscher R., On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6, Geoscientific Model Development, 2020, vol. 13, p. 3465

Xie H., Ringler C., Zhu T., Waqas A., Droughts in Pakistan: a spatiotemporal variability analysis using the Standardized Precipitation Index, Water International, 2013, vol. 38, p. 620

Zhang M., Liu Y., Zhang J., Wen Q., AMOC and Climate Responses to Dust Reduction and Greening of the Sahara during the Mid-Holocene, Journal of Climate, 2021, vol. 34, p. 4893–4912

Zhang Y., Wallace J., Battisti D., ENSO-like interdecadal variability: 1900–93's, J Clim, 1997, vol. 10, p. 1004–1020

Zhao Y., Harrison S., Mid-Holocene monsoons: a multi-model analysis of the interhemispheric differences in the responses to orbital forcing and ocean feedbacks, Clim Dy, 2012, vol. 39, p. 1457–1487

Appendix

Apêndice A

SOM

A.1 Neuro Net Example

Networks acquire knowledge from the data in a process called learning. The SOM algorithm works with two different steps. The first one is a competition between neurons. To measure how connected two images are, the SOM calculates the correlation/distance between those images. At each training round, the correlation between every input data and each neuron from the grid is calculated. The best matching unit (bmu) from the grid and the input data has its weight increased, making that neuron grid point more similar to that monthly map (data input) (Kohonen, 1982, 1995; Haykin, 2009). The second step of the SOM algorithm is the cooperation step. All the neighbors from the first step winning neuron (bmu) also have their weight connections to that input data increased. This way the neural network is widely influenced by that input data with a gradual decrease of their weight connection with the input data as their distance to the bmu increases, creating a topologically separated space (Kohonen, 1997; Haykin, 2009). In the initial ordering phase all neurons are considered neighbors. After a minimum distance from the neurons are detected, a fine tune that considers only neurons close to each other as neighbors finishes the learning process.

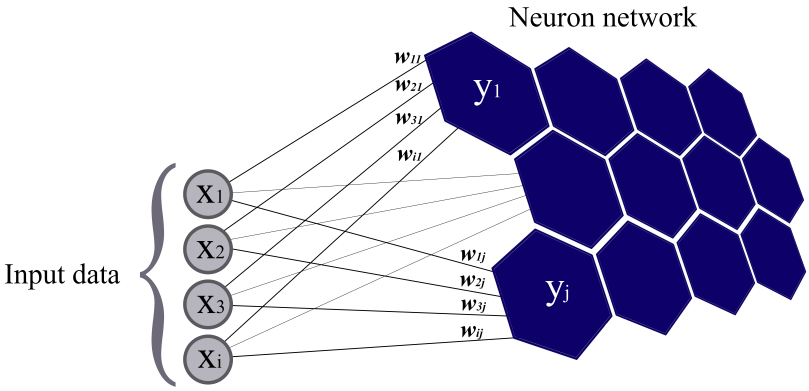

Figure A.1: Hexagonal Neuron Grid representation. For i in range of the input data time series, x_i are the input data. For j in range of the neuron grid size, y_j are each neuron from the grid. The arcs are the connection from the input data to the neuron with weight $w_{i,j}$.

A.2 Multivariable observational data SOM feature space encode and decode process

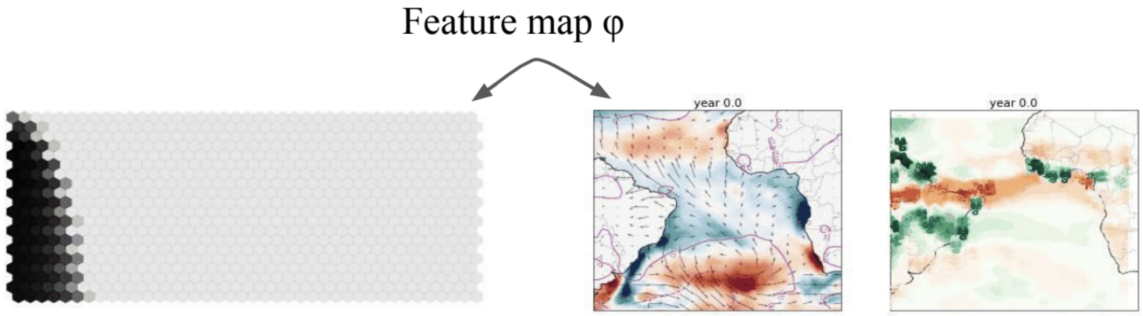

Figure A.2: To the left, hexagonal Neuron Grid being exposed to climate data from the Atlantic ocean on the right. The Feature map function Φ exemplifies that the data is being represented and transformed to the feature space and back to the input space. The Neuron Grid shadow is showing the positive correlated neurons.

Apêndice B

Observational data SOM

B.1 800 neuron grid (trained with 1870-2019 SST reanalysis data)

Figure B.1: 800 neural network grid from the Atlantic SST anomalies. Each neuron is represented by a SOM pattern. The feature space represents various features from the Atlantic SST anomalies.

Apêndice C

Model data SOM

C.1 Silhouette and DBI curves for model data SOM clusterization

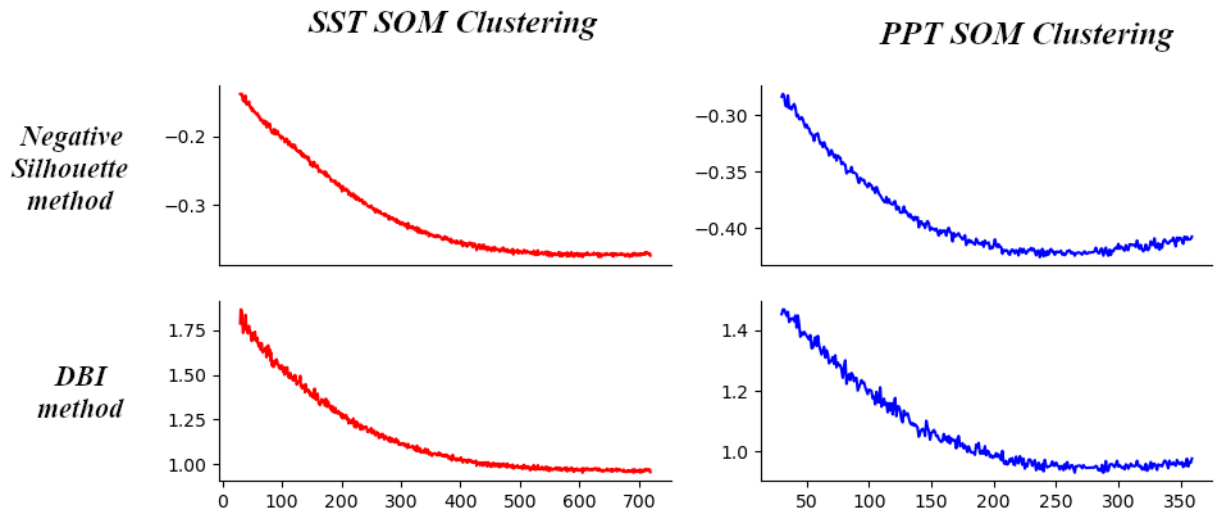

Figure C.1: Silhouette (top row) and Davies–Bouldin (bottom row) indexes for the Sea Surface temperature SOM (left column, red curves) and precipitation SOM (right column, blue curves) feature space clusterings. The 'x' axis indicates the number of clusters in which the feature space was separated.

C.2 SST anomaly model data SOM clusterization Patterns

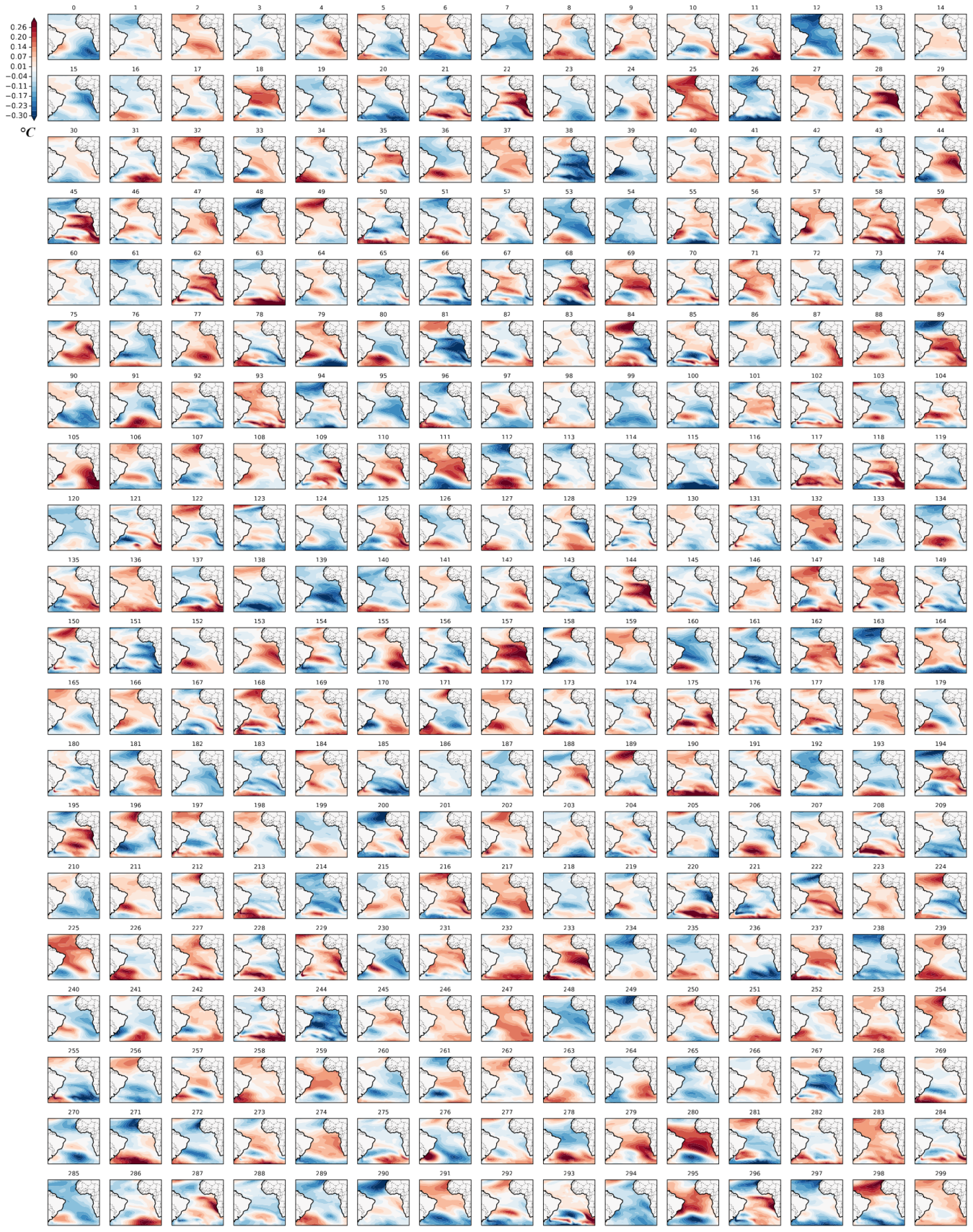

Figure C.2: SST model data SOM cluster patterns. From cluster 0 to cluster 299. Colorbar in Celsius: Blue indicating negative SST anomalies and red for positive.

C.3 PPT anomaly model data SOM clusterization Patterns

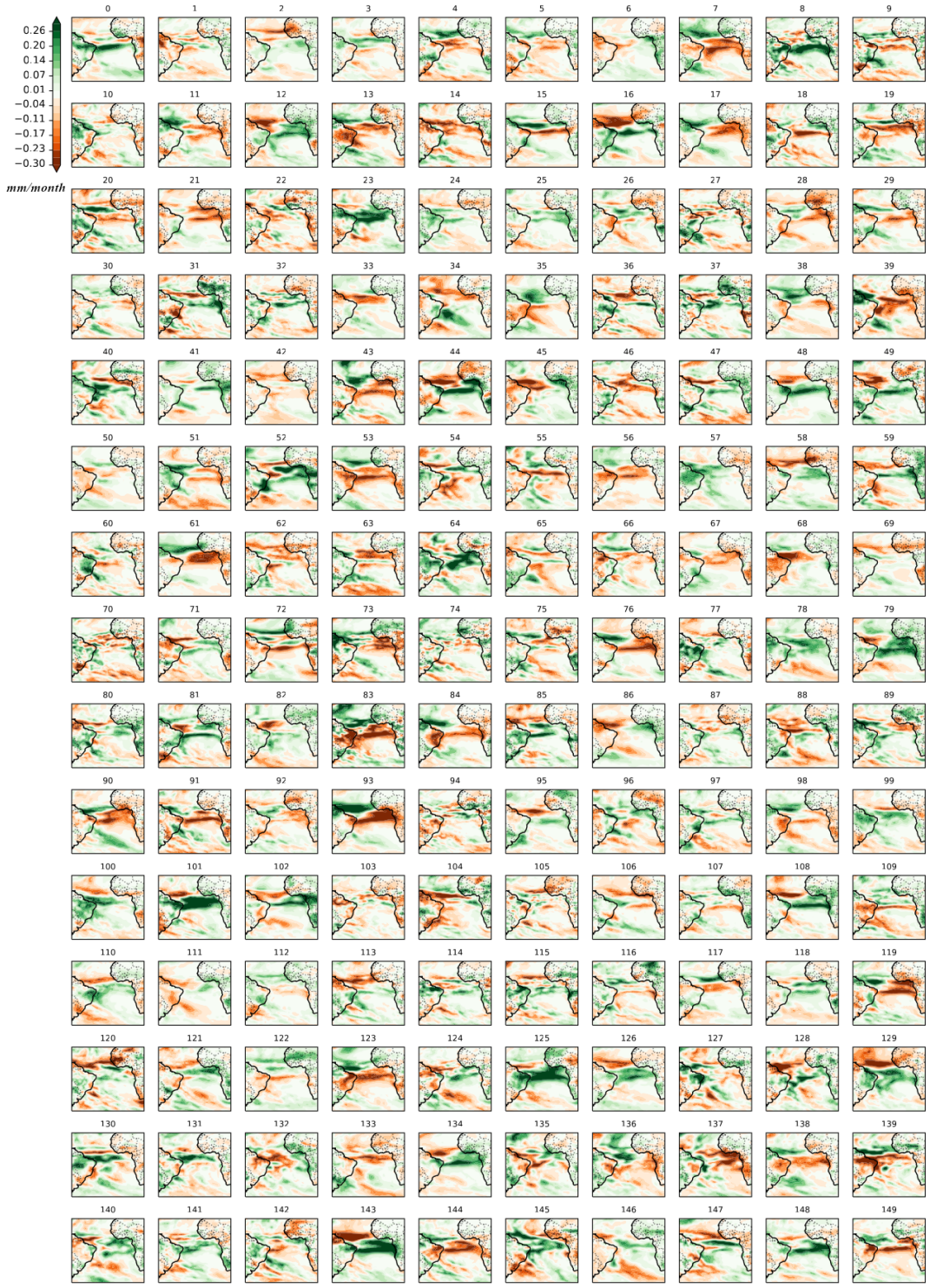

Figure C.3: Precipitation model data SOM cluster patterns. From cluster 0 to cluster 149. Colorbar in mm/month: Orange indicating negative precipitation anomalies and green for positive.