UNIVERSITÄT BAYREUTH

Long-term Precipitation and Vegetation Dynamics in the Caatinga (Brazil)

Master Thesis in Environmental Geography

Jannis Viola
Student number: 1540810
jannis.viola@uni-bayreuth.de

supervised by
Univ.-Prof. Dr. Cyrus Samimi[a]
Univ.-Prof. Dr. Bettina Engelbrecht[b]

[a] Working Group of Climatology, University of Bayreuth
[b] Working Group of Plant Ecology, University of Bayreuth

Bayreuth, 21st January, 2021

Contents

List of Figures

List of Tables

Abstract

The influence of drylands on global carbon variability is often underestimated. In semi-arid ecosystems, the seasonal distribution of precipitation is a strong governing factor of vegetation. The Caatinga seasonally dry tropical forest (SDTF) is situated at the intersection of several large-scale atmospheric systems, that lead to highly heterogeneous rainfall patterns in space and time. The exact areas of the influence zones are unknown, yet of critical importance for water availability throughout the region. The aim of this study is to examine the spatiotemporal dynamics of precipitation and vegetation and their long-term developments in the Caatinga. The analysis was based on remotely sensed fraction of absorbed photosynthetic active radiation (FPAR) and Gauge-based precipitation data. The spatial extent of the influence of the atmospheric systems, which source Caatinga precipitation was captured by "Partitioning around Medoids" (PAM) clustering. The temporal characteristics of the precipitation regimes were assessed with a wide range of metrics, including descriptive statistics, phenological parameters, time series structural patterns and long-term trends. Additionally, the strength of local precipitation-vegetation was captured by non-parametric Spearman correlation and lagged linear regressions. The study concluded that three major influence zones shape the dynamics of precipitation and vegetation in the Caatinga. It was shown that the precipitation regimes extensively shape Caatinga vegetation distribution and variability. FPAR forms distinct macroecological units that follow the dynamics of the precipitation regimes, independent of local geologic, edaphic and floristic settings or anthropogenic disturbances. Long-term changes in precipitation were negative and limited to a small proportion of the region. Trends in vegetation were not linked to precipitation and match previous research on atmospheric CO_2 fertilization effects in semi-arid ecosystems. The findings illustrate the links between seasonal rainfall distribution and Caatinga land surface phenology. This study highlights the importance of understanding the characteristics of local precipitation regimes for semi-arid ecosystems.

1 Introduction

Climate change is globally affecting the environment across all scales. The repercussions range from increasing temperatures, rising atmospheric CO_2 concentrations and UV intensity, to changes in the frequency and magnitude of rainfall events, as well as extreme events such as heatwaves, droughts and floods (Allen et al., 2014). The effects of droughts and temperature anomalies on plant mortality have already been observed globally (Allen et al., 2010) and changing average climate conditions are expected to affect the primary productivity of terrestrial vegetation and plant distribution patterns (De Keersmaecker et al., 2015). Besides the pressure on ecosystems, this may have major socio-economic implications, such as drought-related losses of harvests or increased storm hazards (Field et al., 2012).

Vegetation productivity is an essential variable in ecosystem functioning (Reid et al., 2005). Primary production provides the energetic basis and material substrate for nearly all heterotrophic organisms and is a crucial component of the many services that ecosystems provide (Sessa and Dolman, 2008; Hollmann et al., 2013). Temporal variability of primary production has important consequences for overall ecosystem functioning, and for the stability and value of production-derived ecosystem services (Haberl et al., 2014; Maurer et al., 2020).

Terrestrial primary production is a key regulator of the global carbon cycle (Melillo et al., 1993). Ecosystems can act as a carbon sink when their carbon fixation exceeds their carbon losses. Fixation takes place in photosynthetic processes of plants, while carbon losses occur in auto- and heterotrophic respiration or by natural or human disturbances (Kirschbaum et al., 2001). The global biomes differ in their contributions to global primary production and influence on the global carbon budget. While the mean annual sink is dominated by highly productive ecosystems (e.g. tropical forests), trends and interannual variability of the sink are dominated by semi-arid ecosystems. The carbon balance of semi-arid ecosystems is strongly associated with circulation-driven variability in precipitation and temperatures (Ahlström et al., 2015; Maurer et al., 2020). Therefore, continental-scale or even regional-scale changes in climate can impact the global carbon budget by shifting regional carbon balances from sinks to sources or opposite (Davidson et al., 2012; Poulter et al., 2014; Brienen et al., 2015). Quantifying the variability of primary production and the sensitivity to large-scale climatic variation is essential to predict interannual to decadal variations of the global carbon cycle (Ahlström et al., 2015) and for understanding how the global carbon cycle will respond to future climatic conditions (Maurer et al., 2020).

Semi-arid regions have been the subject of only few targeted studies that place their importance in a global context (Ahlström et al., 2015). Dryland biomes cover approximately 42 % of the global land surface (Sorensen, 2007) and support approximately 30% of the world's population (Group, 2011) with livestock and crops for global food production (Adeel et al., 2005). These biomes contain 7 of the 25 global biodiversity hotspots (Myers et al., 2000) and more than half of all bird and mammal species (Davies et al., 2012; Gudka et al., 2014). They are some of the most threatened, yet disregarded ecosystems (Janzen, 1988; Durant et al., 2012) and under pressure not only due to climate change but also human activity (Miles et al., 2006; Change, 2007). Their large proportion of global land surfaces and highly variable carbon stocks make dryland ecosystems pivotal for accounting of the global carbon balance, despite their low average carbon stock per area unit (Brandt et al., 2018). Drylands are characterized by a limited availability of water (Cherlet et al., 2018). These ecosystems can exhibit increased carbon uptake in anomalously wet years, which induce a rapid growth response of drought-adapted biota (Huxman et al., 2004; Chen et al., 2009). Due to the asymmetry in both the interannual distribution of rainfall and the response of the Gross Primary Productivity (GPP), this response was estimated to be as half as big as the contribution from the CO_2 fertilization effect during 1990-2013 (Haverd et al., 2017). Within drylands tropical dry forests act as an important carbon sink (Malhi, 2012). Observational data on African dryland vegetation carbon stocks (Zhang et al., 2018) suggests that models currently may underestimate the role of dryland savannahs as carbon sinks and sources (Poulter et al., 2014; Ahlström et al., 2015). Further research on the influence of semi-arid ecosystems on the global carbon budget is needed (Ahlström et al., 2015; Haverd et al., 2017; Maurer et al., 2020).

Among dry forests, the neotropic realm and their largest contributor, the Caatinga seasonally dry tropical forest (SDTF), are particularly underrepresented, although research interest and numbers of publications are rising (da Silva et al., 2017; de Queiroz et al., 2017). Caatinga is the name of the ecosystem and of a well defined social region. It is located in the north east of Brazil (S 3° - S 17°; W 35° - W 45°) and covers an area of approximately 913 000 km^2. Caatinga vegetation-drought dynamics are reported to have a significant influence on the global carbon budget (Pereira et al., 2014).
Home to approximately 30 million people, it is the most densely populated semiarid region in the world (Santos et al., 2014), while having the lowest development indicators of Brazil (Leal et al., 2005; da Silva and Barbosa, 2017), pushing the region beyond carrying capacity (da Silva et al., 2017). Subsistence farming, together with wood extraction, livestock grazing and rainfed agriculture is the prevalent source of income, making the local livelihood vulnerable to climate variability (Sun et al., 2007; Justino et al., 2013). Major crops are maize and beans (Marengo

et al., 2018), which have already been observed to be affected by changes in temperature and precipitation regimes in the tropics, and it is expected that these impacts will further increase (Borrell et al., 2020). The climatic changes projected to occur in the future may cause significant impacts in South America in the present century. These impacts will be more intense in tropical regions specifically the Brazilian Amazon and Northeast Brazil (NEB) (Baettig et al., 2007; Metz et al., n.d.; Althoff et al., 2016). The Caatinga is expected to be most vulnerable due to the high variation in water availability and the current economic situation, potentially leading to a social crisis (Barbosa et al., 2006; Marengo et al., 2014).

Caatinga vegetation is extremely heteorogenous and ranges from open cactus shrub on rocky outcrops to semi-decidous forests and forms the largest neotropical SDTF and woodland (SDTFW, *sensu* Gaussen (1973); de Queiroz et al. (2017)). The vegetation is dominated by small-leaved, thorny trees as well as many succulents and therophitic herbs. Most of the woody species are deciduous (de Queiroz et al., 2017) and a huge variety of anatomical, morphological and ecophysiological features related to water use efficiency and safe growth of the plants exist (Machado et al., 1997; Mansur and Barbosa, 2000; da Silva et al., 2004; Souza et al., 2010; Figueiredo et al., 2012; de Lima et al., 2012). Especially different forms of deciduousness, thought to have evolved as an adaptation to seasonal drought in the tropics (Axelrod, 1966), are a key feature of Caatinga woody vegetation (de Queiroz et al., 2017; Mooney et al., 1995; Eva et al., 2004; Bowman and Prior, 2005). They are strongly favoured in regions, where the metabolic cost of capturing nutrients is small and there is a distinct growing season (Givnish, 2002; Ruggiero et al., 2002). In contrast, evergreen species are expected in areas with infertile soils, uniform distribution of annual rainfall and climatic unpredictability (Bowman and Prior, 2005; Ouédraogo et al., 2016).

An important aspect of the Caatinga is the very heterogeneous climatic setting and locally different sources of precipitation. The region is positioned at the intersection of large-scale atmospheric systems that affect the area with varying intensities at different times throughout the year (Pezzi and Cavalcanti, 2001; Marengo et al., 2018; Gomes et al., 2019). Three major influence zones have been suggested, which develop distinct precipitation regimes. These systems divide the region into a northern, a central-eastern and a south-western area (Brahmananda Rao et al., 1993; de Andrade et al., 2017).

The high socio-economic vulnerability of the Caatinga, the complex climatological setting and the valuable wildlife with high rates of endemism (Leal et al., 2005; Queiroz et al., 2015; de Queiroz et al., 2017) demand a better understanding of the precipitation regimes and a spatially explicit delineation of their major influence zones. Besides the strong governing influence of mean annual precipitation (MAP) on Caatinga productivity and diversity (Rito et al., 2017), other aspects of rainfall dynamics can have far reaching implications for Caatinga vegeta-

tion. Seasonality, or the severity of the dry season and the resulting drought stress can influence ecosystem productivity, land surface phenology, tree cover and trait distribution in dry tropical forests (Aragão et al., 2007; Chen et al., 2009; Ouédraogo et al., 2016; Zhang et al., 2018). Commonly measured plant traits, utilised for predicting plant growth, net primary production (NPP) dynamics or the response of ecosystems to climate change (Díaz and Cabido, 1997; Díaz et al., 2016; Reich, 2012; Suding et al., 2008) often do not align well with gradients of water availability (Wright et al., 2004). These weak trait-precipitation relationships were attributed to an improper selection of traits related to hydraulic functioning (Griffin-Nolan et al., 2018). On the other hand, Schwartz et al. (2020) reported that, although ecologists have long recognized correlations between water availability and species distributions (Gentry, 1988), they usually only characterized rainfall regimes with parameters such as MAP or simple dry season metrics. The incomplete consideration of the wide range of environmental and especially precipitation characteristics, sometimes referred to as "multi-dimensional climate space" (Schwartz et al., 2020), might add to the weak trait-precipitation correlation reported by Wright et al. (2004); Griffin-Nolan et al. (2018). This gap calls for spatially explicit characterizations of precipitation dynamics and an evaluation of local precipitation regimes beyond MAP. To understand and anticipate local patterns of water availability throughout the year, the temporal characteristics of the individual regimes and their long-term developments have to be assessed. In addition, the effects on land surface phenology, as well as the strength of the vegetation-precipitation coupling could yield important information for policy makers, infrastructure planning, conservation efforts and the design of ecological studies.

This study aimed to explicitly delineate the influence zones of the precipitation sourcing atmospheric systems of the Caatinga and their temporal characteristics. Subsequently, their effects on local vegetation dynamics were determined by analyzing remote sensing data .
The following questions where addressed:

1. Precipitation
    a) Where are the true influence zones of the different precipitation sourcing systems?
    b) What are the spatiotemporal characteristics of the precipitation regimes?
    c) Are there long-term changes in the amount of precipitation?
2. Vegetation
    a) How strong is the link between observed precipitation patterns and vegetation dynamics?
    b) Is it possible to identify macroecological units of Caatinga vegetation that follow the precipitation regimes?
    c) Are there long-term changes in vegetation productivity?

To facilitate this, precipitation gauge data was clustered. As a proxy for vegetation, the fraction of absorbed photosynthetic active radiation (FPAR) was used. Different temporal parameters, assessing long-term conditions, variability, seasonal and phenological patterns, time series behaviour and trends where derived for both precipitation and vegetation data. The results were divided into subgroups based on the preceding clustering and compared statistically.

2 Materials and Methods

The following sections will present the study area and a review of the atmospheric systems influencing Caatinga precipitation (2.1 & 2.1). Subsequently, the acquisition of the data and preprocessing (2.2), clustering (2.3), the spatiotemporal analyses (2.4), the correlation analyses (2.5) and the statistical tests (2.6) will be presented.

2.1 Study Area

The extent of forest cover in dryland biomes is disputed (Bastin et al., 2017; Brandt et al., 2020), as is the exact extent of the Caatinga (da Silva et al., 2017). This study follows the definition of the "Instituto Brasileiro de Geografia e Estatística" (IBGE, Brazilian Institute of Geography and Statistics) and the Ministério do Meio Ambiente (MMA, Ministry of the Environment), who include the Campo Maior area in the north-west and exclude the São Francisco-Gurgéia along the south-western border. The Caatinga is situated in the federal states Ceará, Rio Grande do Norte, Piauí, Paraíba, Pernambuco, Alagoas, Sergipe and Bahia.

Geology and Soils

Around 70 % of Caatinga area is based on crystalline basements and 30% to sedimentary basins that form flattened surfaces between 300 m and 500 m above sea level, which are interspersed with some isolated mountain ranges and high-altitude plateaus up to 1000 m, most prominently the Chapada Diamantina and the Chapada do Araripe (Sampaio, 1995). The geological settings lead to distinct floras with different plant assemblages, diversity and life-form spectra (de Queiroz et al., 2017). The region has different soil types, ranging from rocky, or sandy areas to relatively fertile and profound areas, especially along rivers (da Silva et al., 2017). Caatinga soils are generally not very fertile and are low in soil organic carbon (Schulz et al., 2016). Due to the prevailing crystalline geological basements, water reserves originate from rainfall that is concentrated over a few months of the year, along with intermittent surface runoff, amplifying the importance of precipitation regimes for water availability (de Andrade et al., 2017; Farrick and Branfireun, 2015). In many areas, soil erosion and salinisation lead to degraded soils (Santos et al., 2014). Several studies have documented that over 90 % of the region has a moderate to high risk of desertification due to intense land use (Vieira et al., 2015).

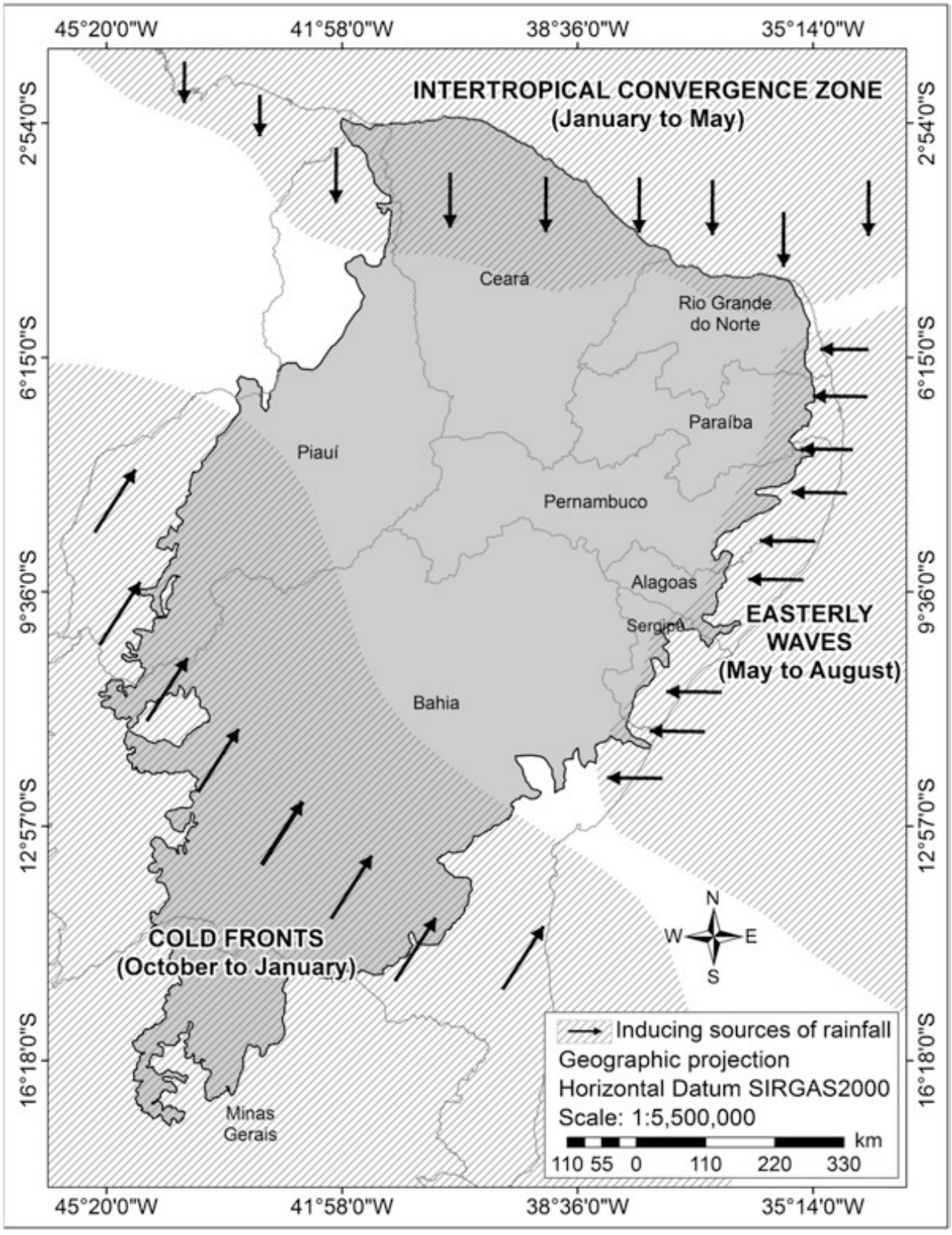

Figure 2.1: Study area and political boundaries of the federal states; shaded areas: predominant precipitation sources and approximate spatial extents of the precipitation regimes (de Andrade et al., 2017) (note: different borders in the North- and Southwest compared to this study)

Regional Climate Systems

To understand the complex precipitation and vegetation dynamics of the Caatinga, it is crucial to comprehend the influences of the different large-scale atmospheric systems. Fig. 2.1 shows a schematic representation of the Caatinga and the influence zones of the prevailing precipitation sourcing systems. The northern area can be attributed to the innertropical convergence zone (ITCZ) (Santos e Silva et al., 2014). The wet season starts around January and is caused by frontal systems (Ferreira and da Silva Mello, 2005) and the begin of the southward migration of the ITCZ (Torres and Ferreira, 2011; Hastenrath, 2012; Guerreiro et al., 2013). The ITCZ location and its seasonal latitudinal migration is mainly controlled by a delayed feedback of tropical Atlantic and Pacific sea surface tempperature (SST) variability (Hastenrath and Greischar, 1993; Moncunill, 2006; Hastenrath, 2012; Rodrigues and McPhaden, 2014; Marengo et al., 2018). When Atlantic SST favour northeast trade winds accompanied by a weakening of the southeast trade winds, the ITCZ migrates to the Southern Hemisphere. This results in increased precipitation, especially during March and April (Hastenrath, 2012). The precipitation of these months can account for over 50% of the total annual precipitation in the northern Caatinga (de Andrade et al., 2016). The end of the wet season is determined by the return of the ITCZ to the Northern Hemisphere (Guerreiro et al., 2013). The ITCZ control on precipitation leads to a highly predictable seasonal climate in the northern Caatinga, which allows the forecast of large climate events with a leading time of a few months (Hastenrath, 1990; Hastenrath and Greischar, 1993; Hastenrath, 2012; Marengo et al., 2013, 2016, 2018; Giannini et al., 2004; Nobre et al., 2006; Rodrigues and McPhaden, 2014; Hounsou-Gbo et al., 2016). Although the ITCZ mechanism is long known, its influence zone is unclear (Sun et al., 2007).
Precipitation of the eastern region of the Caatinga is induced by a combination of local convection, the sea breeze and easterly wave disturbances (EWD). EWDs are disturbances that move westward with the trade winds that are associated with the subtropical ridges (Gomes et al., 2015). They account for at least 60% of the total rainfall over the east coast of Northeast Brazil (NEB) (Gomes et al., 2019). The southeasterly winds blow perpendicular to the coast and are modulated in strength and direction by the subtropical high in the South Atlantic (Kousky, 1979; Brahmananda Rao et al., 1993; Kayano, 2003; Torres and Ferreira, 2011). The South Atlantic subtropical Anticyclone (SASA) and its latitudinal position in the South Atlantic Basin (SAB) is responsible for changes in sea level pressures, wind speeds and air temperatures in northern Brazil (Gilliland and Keim, 2018). The position of the SASA and the South Atlantic Ridge is important for the low-level water vapour transport towards NEB (Brahmananda Rao et al., 1993; De Lima Moscati and Gan, 2007; Gomes et al., 2015). Their position influences the occurrence and propagation speed of EWDs (Kayano, 2003). EWDs reach the coast east of the center of the Caatinga. They can either move inwards and cross the Caatinga or get deflected along the north coast and contribute to local rainfall along their path (Gomes et al.,

2019). While the isolated effect of EWDs is weak for the whole Caatinga (Kayano, 2003), it is the main source of rainfall variability along the east coast and in the interior NEB region (De Lima Moscati and Gan, 2007). Only 14 % of the reported EWDs move across the region towards the Amazon (Gomes et al., 2019).
The wet season in the Southwest is induced by northward incursions of cold fronts (CF) from the South Atlantic Ocean that move northwestward from the southern Brazilian coast (Rao et al., 1996; Marengo et al., 2018). Frontal systems interact with convective activity from the Amazon Basin or from the northern coast of Brazil (Kousky, 1979; Torres and Ferreira, 2011) and cause a precipitation maximum during November & December. CFs are linked to the northward displacement of the South Atlantic Convergence Zone (SACZ). They are additionally modulated by the upper-level cyclonic vortex situated over the Atlantic Ocean, intense convection over the eastern Amazon, and an eastward shift in the position of the Bolivian high. At low levels the flow is directed from the Amazon towards the South of NEB and there is an eastward shift of the SASA with a weakening of the southeast trade winds over the Northeastern coast (Chaves and Cavalcanti, 2001; De Lima Moscati and Gan, 2007; Gilliland and Keim, 2018).
The influence of several oscillating atmospheric systems is well established for the Caatinga, especially on the northern and southern areas. Previous research demonstrated that rainfall anomalies in NEB are linked to several modes of natural variability, including El Niño/Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), the North Atlantic Oscillation (NAO), and the Atlantic Multidecadal Oscillation (AMO). While being a major driver of precipitation variability, ENSO is only responsible for a fraction of droughts, which can occur during both phases and are additionally governed by other remote processes (Wang and Fiedler, 2006; Rodrigues et al., 2011; Hastenrath, 2012; Rodrigues and McPhaden, 2014; Amorim et al., 2014; Barbosa and Kumar, 2016).

2.2 Data and Processing

The precipitation data used is the "GPCC Full Data Monthly Product Version 2018 Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data" dataset at 0.25° degree resolution (Meyer-Christoffer et al., 2018) of the Global Precipitation Climatology Center (GPCC). It was obtained as a netcdf file from http://gpcc.dwd.de. The GPCC data is an influencial dataset, used for other climatologies like CHELSA (Karger et al., 2017) or for benchmarking other precipitation products (Prakash et al., 2015). It is a gridded global dataset provided in world geodetic system (WGS) 84 with a time coverage from 01.01.1981 to 01.12.2016 and includes 6 variables (excluding dimension variables): precipitation in mm/month per grid, number of gauges per grid, number of infilled gauges per grid, interpolation error without infilled stations per grid, interpolation error with infilled stations per grid and difference of new minus

old interpolation method per grid. All analyses are based on precipitation.

As a proxy for vegetation productivity, the fraction of absorbed photosynthetically active radiation (FPAR) was used. The energy required by terrestrial vegetation to produce organic materials from mineral components is provided by solar radiation in the spectral range 400-700 nm, known as photosynthetically active radiation (PAR). The proportion of PAR that is effectively absorbed by plants is called FPAR. It can be used to assess the primary productivity of canopies, the associated fixation of atmospheric CO_2 and the energy balance of the surface and was defined as an essential climate variable (ECV) (, GCOS). Monitoring FAPAR can provide information on the amount and health cycle of vegetation (Sessa and Dolman, 2008; GCOS, 2011; Hollmann et al., 2013). Compared to the famous normalized difference vegetation index (NDVI), FPAR has the advantage that it is a real-world biophysical unit, that can be verified *insitu*.
The FPAR data used is the "NOAA Climate Data Record (CDR) of Leaf Area Index (LAI) and Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Version 4" at 0.05° resolution (Claverie et al., 2016) of the National Oceanic and Atmospheric Administration (NOAA). It was obtained as netcdf files from https://www.ncei.noaa.gov using GNU wget 1.20 FTP-server downloading software (Foundation, 2010). It is a gridded global dataset provided in WGS84, with a daily coverage from 01.01.1981 ongoing and includes 3 variables (excluding dimension variables): leaf area index, fraction of absorbed photosynthetic active radiation and quality control bit flags. All analyses are based on FPAR. While the authors reported saturation effects for broadleaf biomes with high leaf area index (LAI) (>4.5) and FPAR (>0.8) values, this can be neglected for most of the area. The whole Caatinga was categorized as one land cover class during calculation (Claverie et al., 2016). Time series with <10% missing data were included (Vasseur and Yodzis, 2004). To match the monthly temporal resolution of the precipitation dataset and to handle missing values, maximum value composites were generated. The maximum value of each pixel of each month was chosen to represent that time step (Barbosa et al., 2019). Spatially, the FPAR data was aggregated to 0.25° resolution using an average.

The extent and Caatinga borders are taken from the shapefile of the "Biomes 2004" map (IBGE, 2004) and was obtained from https://www.ibge.gov.br. It is a vector dataset provided in 1:5000000 scale and originally distributed using the "Sistema de Referencia Geocéntrico para las Américas geocentric reference system" (SIRGAS) and was reprojected to WGS84 prior application.

To ensure compatibility and to facilitate analysis, all data was preprocessed using the raster package and R Studio (R Core Team, 2019; Hijmans, 2019). The datasets were temporally har-

monized to match the 35 year period from 1982 to 2016, the maximum possible overlap. The data was spatially restricted to the official Caatinga borders and big water surfaces were masked out, resulting in 1049 remaining pixels per dataset that each yield a time series of 420 monthly time steps. If possible, the color scales of the plots were generated using the "viridis" package (Garnier, 2018), to ensure readability when printing in greyscale and for the colourblind. Due to the mainly diagnostic nature of this study, all data presented here is unfiltered and not detrended (as in Giannini et al. (2004)). This allows a better quantification of the direct impacts of precipitation variability over the Caatinga vegetation response by retaining the real amount of rainfall which is directly related to the vegetation or damages to local economy (Lucena et al., 2011).

2.3 Clustering

Clustering is a common unsupervised machine learning task, in which the data set has to be automatically partitioned into "clusters". The k-medoids clustering algorithm "Partitioning Around Medoids" (PAM) (Kaufman and Rousseeuw, 1990) was chosen, because it outperforms traditional k-means in terms of execution time, overlapping clusters and reduced noise (Arora et al., 2016). It allows clustering using a wide range of distance metrics and has been used in a huge variety of fields, ranging from environmental or climatological studies to facial recognition (Bracken et al., 2015; O'Donnell et al., 2017; Molaei et al., 2019), as well as explicitly for regionalisation and precipitation regimes (Belagoune et al., 2017; Ma et al., 2020; Saunders et al., 2020).

PAM takes a dissimilarity matrix $\mathbf{D}$ as input and produces a set of cluster centers or "medoids", which are themselves elements in the set being clustered. This makes them more robust representations of the cluster centers compared to k-means approaches. This is particularly important since rainfall probability distributions are seldom gaussian (Öztürk, 1981; Thompson, 1984; Fatichi et al., 2012) and in the presence of skewed distributions or extreme values, or if many elements do not clearly belong to any cluster (Van der Laan et al., 2003; Arora et al., 2016; Schubert and Rousseeuw, 2018).

The medoids identify the clusters and are selected as follows: K denotes the number of clusters and $\mathbf{M} = (M_1, \ldots, M_K)$ is any size K collection of the p elements $\mathbf{x}_j$. With $\mathbf{M}$, the dissimilarity $d(\mathbf{x}_j, M_k)$ of each element and each member of $\mathbf{M}$ is calculated. For each element $\mathbf{x}_j$, the minimum and minimizer is denoted by $\min_{k=1,\ldots,K} d(\mathbf{x}_j, M_k) = d_1(\mathbf{x}_j, \mathbf{M})$ and $\min^{-1}_{k=1,\ldots,K} d(\mathrm{x}_j, M_k) = l_1(\mathrm{x}_j, \mathrm{M})$. PAM selects the medoids $\mathbf{M}^*$ by minimizing the sum of such distances $\sum_j d_1(\mathbf{x}_j, M)$. Each medoid M_k^* identifies a cluster, defined as the elements which are closer to this medoid than to any other. Each element gets their respective cluster membership assigned.

Using the R package "cluster" (Maechler et al., 2019) PAM was applied to raw, scaled and centred precipitation data, with $k = 3$ according to the 3 suggested rainfall sources of the Caatinga (de Andrade et al., 2017) (see sec. 2.1 & fig. 2.1). The results were congruent, so only the raw-based version was used for further analyses. To assess clustering quality and estimate the optimal number of clusters as detectable in the data, the "NbClust" and "factoextra" packages (Charrad et al., 2014; Kassambara and Mundt, 2020) were used. Within cluster sum of squares (Scree- or Elbowplot), average silhouette method (Kaufman and Rousseeuw, 1990) and gap statistic (Tibshirani et al., 2001) were applied. Clustering quality results can be found in the appendix 5.2.

2.4 Spatiotemporal Analyses

To characterize the precipitation and vegetation dynamics, a suite of spatial and temporal parameters was calculated. The metrics measure long-term averaged conditions, strength, seasonal dynamics and reliability, as well as time series behaviour and intrinsic predictability from an information theory point of view.

2.4.1 Baseline Conditions

First, as a measure of overall water availability, mean annual precipitation (MAP) was calculated for all precipitation time series as the cumulative sum of averaged monthly values ($\bar{x}$):

$$\mathrm{MAP} = \Sigma\left[\bar{\mathrm{x}}_{January}, \bar{\mathrm{x}}_{February}, \cdots, \bar{\mathrm{x}}_{December}\right] \tag{2.1}$$

As a first inventory and to capture the typical vegetation conditions, the mean annual maximum FPAR (MAMF) was used, as it captures the peak of the season and represents the underlying vegetation or land cover type more accurately than a cumulative sum (Ceccherini et al., 2013). It is calculated using the arithmetic mean of all annual maximum FPAR values:

$$\mathrm{MAMF} = \frac{1}{n}\sum_{i=1}^{35} max_i = \frac{max_{1982} + max_{1983} + \cdots + max_{2016}}{n}. \tag{2.2}$$

Additionally, the total mean FPAR (MF), the mean annual FPAR range (MAR) were calculated. Plots can be found in the supplements: fig. 5.6.

2.4.2 Variability

In order to capture long-term variability of the time series, the coefficient of variation (CV) was calculated. The CV has been used to quantify the reliability of rainfall regimes (Fatichi et al., 2012) or vegetation variability (Schucknecht et al., 2013). The CV, which in the context of precipitation can be understood as the reliability of the rainfall amount or "relative seasonality" (Schwartz et al., 2020), was chosen to facilitate cross-regional comparisons, because it is unbiased by the absolute local mean values unlike e.g. standard deviation (Oliver, 1980) and was calculated for both datasets. CV is the ratio of the standard deviation (σ) to the mean ($\bar{x}$), expressed as percentage:

$$\mathrm{CV} = 100 \times (\sigma/\bar{\mathrm{x}}) \tag{2.3}$$

Relative seasonality has been used to explain species distributions and spatial patterns in life history strategies (Schwartz et al., 2020). The CV of monthly precipitation as measure of relative seasonality (Abel et al., 2020) is widely used in biogeography research (e.g. (Beaumont et al., 2005; Parmentier et al., 2007; Rubio de Casas et al., 2017)) and is included in the famous BIOLCIM dataset (Booth et al., 2014). Due to high correlations with other parameters, the CV of monthly values has been moved to the appendix (fig. 5.8) and will not be further discussed.

2.4.3 Rainfall Concentration, Seasonality and Timing

Rainfall concentration and dry season length can be strong factors governing vegetation structure and dynamics in the dry tropics (Engelbrecht et al., 2007; Bowman and Prior, 2005). However, the explanatory power of dry season length can be limited. The length of time during which precipitation falls below a threshold may not necessarily translate to water availability or the severity of water stress during that period (Schwartz et al., 2020). On a monthly time scale, it did not prove to be a sensible choice (see supplements; fig.5.9). Alternatively, to assess seasonal concentration of precipitation and to measure dry season intensity, the precipitation concentration index (PCI) was calculated for each precipitation time series. The PCI is a dimensionless measure of average annual rainfall concentration and was originally introduced by Oliver (1980). It is calculated based on monthly values and their CV as shown in the following equation:

$$\mathrm{PCI} = \frac{100}{12} \times \left[1 + \left(\frac{cv}{100}\right)^2\right] \tag{2.4}$$

The theoretical limits of the PCI are 8.3 when the rainfall in each month of the year is the same (CV=0) and 100 when the entire yearly rainfall occurs in one month (CV=$100\sqrt{11}$). PCI values double when the same monthly increment occurs in half of the designated period. A figure of the theoretical PCI scaling can be found in the appendix (fig. 5.10). Later Michiels et al. (1992) expanded the PCI to differentiate between seasonal concentration (PCI1) and temporal

concentration (PCI2). The authors showed that averaging across all respective months of a time series and calculating the PCI captures the average seasonal concentration at this station (PCI1). Calculating the PCI for each individual year and averaging across the interim results captures the temporal concentration of the typical season (PCI2). Irregular or changing timing of the rainy season during the observation period leads to a broad average rainfall distribution and low PCI1. The PCI2 captures the rainfall concentration of the typical individual season more closely, independent of its timing. Thus, the interplay and differences between PCI1 and PCI2 not allow only for a characterization of the average annual seasonality from uniform to strongly seasonal concentration, but give additional insights into the regularity of seasonal timing and concentration. This was defined as a third parameter PCIΔ, by substracting PCI1 from the PCI2 values.
Due to its calculation, the PCI heavily decreases for distributions without months without values. This nonlinear behavior makes it less intuitive to interpret seasonal concentration of vegetation that forms a steady cover, so this metric was omitted for FPAR.

Seasonality, Phenology

The timing of the season is an important parameter for the assessment of regional precipitation or phenological patterns in ecohydrological studies (Marengo et al., 2001; Liebmann et al., 2007; Feng et al., 2013). The onset and cessation of the wet or dry season are common measures of timing, but measures related to the date of peak rainfall or productivity occurrence are also applied (Feng et al., 2013; Schwartz et al., 2020; de Jesus et al., 2020). To capture temporal differences in timing of the seasons, the typical peak and onset month was determined for each pixel and both datasets. For the onset month, the mean annual distribution was calculated and its total mean was set as threshold to decide between wet and dry season. To anticipate seasonal patterns that span from one year to the other, each year was replicated so that at least one full season was created. Then, the start month of the longest contiguous sequence of values above the mean was determined. For the peak of the season, the month of the maximum value of each year was determined and discrete monthly increments ($Jan = "1", Feb = "2", \cdots, Dec = "12"$) were assigned. In order to account for the circular nature of discrete monthly numbers ($|1_{Jan} - 2_{Feb}| = 1$, $|1_{Jan} - 12_{Dec}| = 1$), a circular approach had to be used. The monthly increments were transformed to radians before calculating the median value of all yearly maximum months $\tilde{m}_{seas}$ using the circular package (Agostinelli and Lund, 2017). The circular median was chosen to avoid fractioning the monthly increments.

2.4.4 Time Series Predictability and Structural Complexity

Information theory quantifiers (ITQ) were applied to capture differences in time series behavior, deal with nonlinearity and as a measure of intrinsic predictability of the systems at hand (Sippel

et al., 2016; Pennekamp et al., 2019). ITQs capture the structure and dynamics of series and have been successfully applied in many fields, ranging from medical research, characterisation of precipitation regimes, model benchmarking and to assessment of intrinsic predictability of different systems (Ouyang et al., 2013; Zhang et al., 2019; Liu et al., 2019; Sippel et al., 2016; Pennekamp et al., 2019). Using permutation entropy (PE) and Martín-Platino-Rosso statistical complexity (MPR) allows to classify time series behaviour in terms of structure and predictability (Sippel et al., 2016; Ribeiro et al., 2017; Pennekamp et al., 2019).They are metrics based on the space of probability density functions (PDFs) for datasets. This makes it possible to compare different sets and to classify them according to the properties of their underlying processes, i.e. stochastic vs. deterministic. ITQs can separate time series that are difficult to distinguish by conventional analysis. They can discern stochastic processes like correlated noise (Vasseur and Yodzis, 2004), noisy series with underlying mechanistic forcings like seasonal cycles, and deterministic chaotic maps, which do not contain any noise (Rosso et al., 2007). The position within causal information planes (entropy-complexity plane) is an especially fast tool to discern time series structural behavior (Rosso et al., 2012).
Ordinal pattern statistics and the seminal works of Bandt and Pompe (2002) with incorporated weight information (Fadlallah et al., 2013) are the basis for the ITQs. The embedding dimension $D = 4$ time steps and a time lag $\tau = 1$ was used, to match the time series length requirements, following the Lehmer-permutation scheme (Olivares et al., 2012). ITQs capture global properties of the data and relate to the information content and the complexity of the data.
The permutation Shannon entropy (PE) is a measure of the information content of a time series and can be used to quantify uncertainity, disorder, state-space volume and lack of information (Shannon, 1948; Brissaud, 2005). The obtained parameters PE ($\mathcal{H}[P]$) & MPR ($\mathcal{C}[P]$) are robust against noise and nonstationarity, when the transformations do not happen within the width of the embedding dimension (Lopez-Ruiz et al., 1995; Rosso et al., 2007, 2012; Sippel et al., 2016). Based on a probability distribution with N possible states $P = \{p_i; i = 1, \ldots, N\}$ with $\sum_{i=1}^{N} p_i = 1$, the Shannon information measure (Entropy) reads:

$$\mathrm{S}[P] = -\sum_{i=1}^{N} p_i \ln [p_i] \tag{2.5}$$

The Shannon entropy is minimal when all but one of the p_i 's vanish, and maximal when all $p_i's$ are equal, i.e. for the uniform distribution $P_e = \left\{p_i = \frac{1}{N}, \forall i = 1, \ldots, N\right\}$. In this case, $\mathrm{S}\left[P_e\right] = \mathrm{S}_{\max} = \ln \mathrm{N}$. However, these two situations are extreme cases and unlikely to occur in any natural phenomenon considered here. In the following normalized Shannon entropy, $0 \leq \mathcal{H} \leq 1$, was used:

$$\mathcal{H}[P] = \mathrm{S}[P]/\mathrm{S}_{\max}. \tag{2.6}$$

Contrary to information content, there is no universally accepted definition of complexity. The focus of this thesis is on describing the complexity of time series and not the complexity of the underlying systems. Statistical complexity (MPR,$\mathcal{C}[P]$) is low for ordered settings like simple oscillations or trends (minimal PE) and for unordered, uncorrelated white noise (maximal PE). Between the two extremes, data is more difficult to characterize and has higher complexity. $\mathcal{C}[P]$ captures structures in the data that relate to organization, correlational structures, memory and other properties (Feldman et al., 2008). This is is achieved by measuring a disequilibrium, which is the distance from the equilibrium (uniform) distribution of the accessible states of a system (Lamberti et al., 2004). The Jensen-Shannon divergence quantifies differences between probability distributions or symbolic sequences (Grosse et al., 2002) and has been used in a wide range of biodiversity research, including genetics (Kartal et al., 2020), gut microbiomes (Goedert et al., 2014), and mangrove forests (Jia et al., 2019). Based on the seminal notion advanced by Lopez-Ruiz et al. (1995), this statistical complexity measure (Martin et al., 2003) is based on the normalized Shannon entropy (PE) and the disequilibrium $Q_J\left[P, P_d\right]$ defined in terms of the Jensen-Shannon divergence $\mathcal{J}\left[P, P_e\right]$:

$$\mathcal{Q}_J\left[P, P_e\right] = Q_0 \mathcal{J}\left[P, P_e\right] = Q_0 \left\{ \mathrm{S}\left[\left(P + P_e\right)/2\right] - \mathrm{S}[P]/2 - \mathrm{S}\left[P_e\right]/2 \right\} \tag{2.7}$$

where Q_0 is a normalization constant chosen such that $0 \leq \mathcal{Q}_J \leq 1$:

$$Q_0 = -2 \left\{ \frac{N+1}{N} \ln(N+1) - \ln(2N) + \ln N \right\}^{-1}. \tag{2.8}$$

It is defined via the functional product form:

$$\mathcal{C}[P] = \mathcal{Q}_J\left[P, P_e\right] \cdot \mathcal{H}[P]. \tag{2.9}$$

$\mathcal{C}[P]$ or "MPR" was used for plotting and classifying the time series in the entropy-complexity plane (Rosso et al., 2012) (see fig. 5.14), but lead to spatial patterns very similar to PE and was moved to the appendix (see 5.13).

2.4.5 Trends

In order to understand long-term dynamics of precipitation and vegetation in the Caatinga, especially with changing long-term conditions, trend analyses were applied. Trend analyses are often performed using a pixelwise linear model. They provide the slope coefficient of an ordinary least squares regression (OLS) between the values over time and a linear time-series(Fensholt et al., 2015). However, the outcome of a regression analysis is influenced by several factors related to the data (data choice, data period and quality), the choice of vegetation metric and the nonlinearity of processes in relation to the predicting variable used as the trend indicator.

Since environmental time series often do not meet parametric assumptions of normality and homoscedasticity (Hirsch and Slack, 1984), median trend (Theil-Sen)(Sen, 1968) procedure has also been applied to detect trends in vegetation (Fensholt et al., 2012, 2013). The Theil-Sen (TS) estimator is a rank-based test based on nonparametric statistics that calculates the slope and intercept of all pairs of observations of the time series, which makes it effective for noisy series. Due to its use of the median, up to approximately 29% of the samples can be unrelated noise without impacting the statistic (Hoaglin et al., 2000). TS is proven to be robust against seasonality, non-normality, heteroscedasticity, temporal autocorrelation (intra- and inter annual scale), as well as censored data (Hirsch and Slack, 1984; Van Belle and Hughes, 1984; Akritas et al., 1995; Alcaraz-Segura et al., 2010). It has been recommended for analysing time series of vegetation (De Beurs and Henebry, 2005). It is based on the simple linear regression model

$$Y = \alpha_{reg} + \beta X + \varepsilon \tag{2.10}$$

with the slope β, the y-intercept α_{reg} and the error term ε. The Theil-Sen estimator determines the slope of the regression line via the median of the slopes of all lines that can be drawn through the data points:

$$\mathrm{TS}(x, y) = \mathrm{median}_{k,l \in \{1,\ldots,n\}} \left(\frac{y_l - y_k}{x_l - x_k} \right). \tag{2.11}$$

It was calculated for each pixel time series of both data sets with a significance level α set to 0.05.

2.5 Precipitation-Vegetation Coupling

The strength of the correlation between precipitation and vegetation time series was assessed using pixelwise correlation and lagged regression analyses.
As a measure of temporal correlation of the two sets of time series, the Spearman correlation coefficient ρ_t is defined as the Pearson correlation coefficient between the rank variables. Spearman rank correlation is a non-parametric test, making it robust against outliers, noise and non-normality of the data (Hollander et al., 2013). It is widely used in environmental research and for precipitation time series (Ceccherini et al., 2013; Zhang et al., 2018). The implementation of Team (2019) was used for calculation. For a sample of size n, the n raw scores X_i, Y_i are converted to ranks $\mathrm{rg}_{X_i}, \mathrm{rg}_{Y_i}$, internally and r_s is computed as

$$r_s = \rho_{\mathrm{rg}_X, \mathrm{rg}_Y} = \frac{\mathrm{cov}\left(\mathrm{rg}_X, \mathrm{rg}_Y\right)}{\sigma_{\mathrm{rg}_X} \sigma_{\mathrm{rg}_Y}} \tag{2.12}$$

where ρ denotes the usual Pearson correlation coefficient, but applied to the rank variables, $\mathrm{cov}\,(\mathrm{rg}_X, \mathrm{rg}_Y)$ is the covariance of the rank variables, σ_{rg_X} and σ_{rg_Y} are the standard deviations of the rank variables.
Lagged linear regression was applied as a second metric and to incorporate the signal of the preceding season. For each time series pair, FPAR values of a given month are regressed on precipitation values of the concurrent and past months. The coefficient of determination R_t^2 denotes the percentage of explained variability of the FPAR data around its mean. It is defined as the models R^2, adjusted for the number of predictors. Based on literature review (Gessner et al., 2013; Barbosa and Kumar, 2016) and preliminary tests, the precipitation of the concurrent and 3 preceding months were chosen as predictors. With the y-intercept α_{reg}, and the slopes $\beta\,(X_t, X_{t-1}, \cdots, X_{t-3})$, as well as the error term ε, the equation reads:

$$Y_t = \alpha_{reg} + \beta X_t + \beta X_{t-1} + \beta X_{t-2} + \beta X_{t-3} + \varepsilon. \tag{2.13}$$

The lagged regression analyses are of exploratory nature and individual choice of lag lengths or assessing potential bias and other model diagnostics was beyond the scope of this work.

2.6 Statistical Analyses

In order to support and verify the results of the clustering and to test for differences in spatial distribution, the results of the temporal analyses were compared statistically. All parameters were grouped according to their precipitation cluster membership and the subgroups were compared. After tests for independence of the data (Strasser and Weber, 1999), the nonparametric Wilcoxon-Mann-Whitney Test(Bauer, 1972; Hollander et al., 1999) was applied. The implementation in the "coin" package (Hothorn et al., 2008) was used, which can handle ties. The Wilcoxon-Mann-Whitney Test tests for the equality of the distributions of a numeric response variable in two or more independent groups against shift alternatives. Testing for dependent data via the nonparametric Wilcoxon rank-sum Test (Bauer, 1972; Hollander et al., 1999) included in the "ggsignif" package (Ahlmann-Eltze, 2019) yielded similar results and was moved to the annex (fig. 5.19). Comparing subgroup medians via Kruskall-Multi-Comparison tests (Kruskal and Wallis, 1952; Siegel, 1988) using the "pgirmess" package (Giraudoux, 2018) also yielded similar results and was not further pursued.

For circular data, such as the seasonal timing and phenological parameters, the Watson-Wheeler test of homogeneity of means (Batschelet, 1981) was applied. The implementation of the "circular" package (Agostinelli and Lund, 2017) was used. The Watson-Wheeler (also known as Mardia-Watson-Wheeler, or uniform score, or Watson-Williams) test is a non-parametric test to compare the mean or variance of two or several samples and can be applied for circular data.

3 Results

Clustering precipitation data with k=3 clusters resulted in a northern cluster k_3N (cluster 1), covering 35.56 % of the area, an eastern cluster k_3E (cluster 2), covering 21.45 % which protrudes towards the center of the region at around S 7° and the biggest cluster in the South k_3S (cluster 3) with 43 %. A map of the clusters and the precipitation distributions of each cluster are shown in figure 3.1. Clustering quality indices can be found in the appendix (fig. 5.2).

The spatial distribution of MAP (see fig 3.2 (a)) shows an area of increased rainfall with values of up to 1612 mm/yr^{-1} in the Northwest (S 4°; W 43°). Other areas of increased precipitation are located along the north-eastern border, or uplands such as the Chapada do Araripe in the northern center (S 7°; E 39°) and the Chapada Diamantina (S 14°; W 42°, see also the elevation map in the appendix, fig. 5.1). East and west of the Chadapa do Araripe, "valleys", i.e. areas of decreased MAP are visible, indicting two drier belts protruding to the northern coast. The center of the Caatinga (ranging between S 9°-S 11° and W 39°-W 42°) exhibits the lowest MAP, with the minimum at 355 mm/yr^{-1}.

The overall variability of precipitation exhibits the highest values of up to 174 % in the center of the area. The Northeast, the northern uplands and the eastern border exhibit the lowest values, with the eastern area leading with CV_p of values that reach 79 %. The "valley" feature is visible again, exhibiting increased CV_p in two belts from the center of the region to the northern coast. k_3N shows the narrowest distribution and an average CV_p of 130 %. k_3S has the highest average CV_p at 134 % and intermediate spread. k_3E reaches the lowest values with an average of 119 % and the widest range.

The average annual rainfall concentration ranges from a strongly seasonal distribution in the north with PCI1 values up to 21, to uniform rainfall distribution in the south east (<9), close to the absolute possible minimum. The area of very low PCI1 values reaches westwards up to the Chapada Diamantina ($\tilde{\text{S}}$ 13°; W 42°) and stretches along the eastern border into the north. The southern peak and Southwest exhibit intermediate PCI1 values that match the central northern values. The center of the Caatinga is the area of the highest mismatch of seasonal timing, captured by the difference between PCI1 and PCI2, ΔPCI. The PCI1 values of this area are intermediate, but the PCI2 values are very high due to very strong seasonal concentration in individual years. The ΔPCI also exhibits high values east and west of the Chapada do Araripe, indicating unstable seasonal timing and concentration. The north-western edge of the Caatinga shows almost no difference between averaged seasonal pattern and individual seasonal concentration, indicating seasonal distribution and temporal stability. k_3N exhibits the highest average PCI1 (17.12) and narrowest distribution, followed by k_3S ($\bar{x}$PCI1: 14.13) and k_3E ($\bar{x}$PCI1:

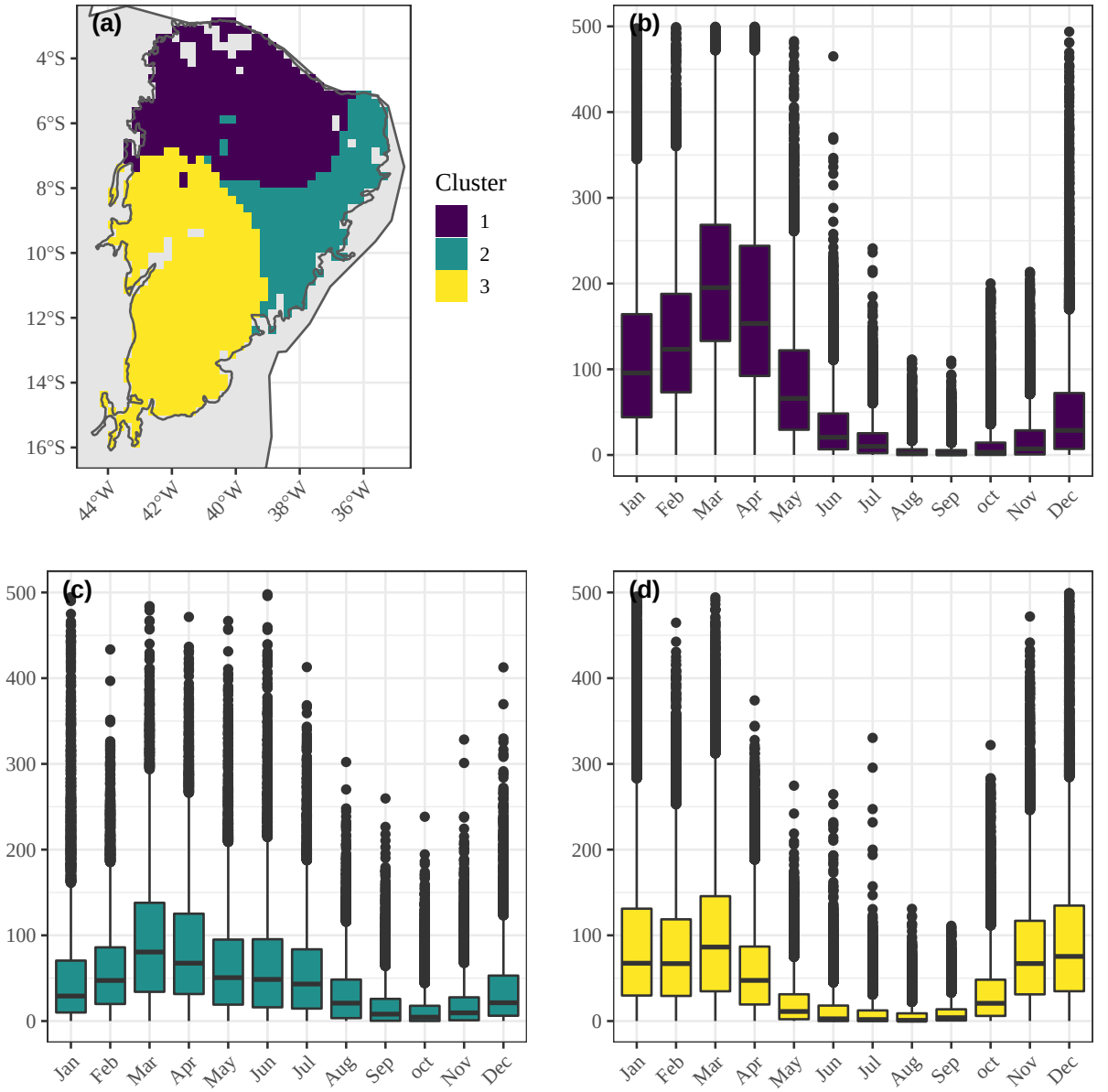

Figure 3.1: (a) Spatial representation of the precipitation clustering, Cluster 1 = k_3N (purple), Cluster 2 = k_3E (turquoise), Cluster 3 = k_3S (yellow); (b) - (d) Precipitation distributions of the clusters 1982-2016. Precipitation values of each month of the respective cluster are represented as boxplots. Boxes range from 25th to 75th percentiles, thick black lines are the median, whiskers stretch $1.5 * IQR$ (interquartile range). *Outliers have been truncated at* 500 mm *for plotting.*

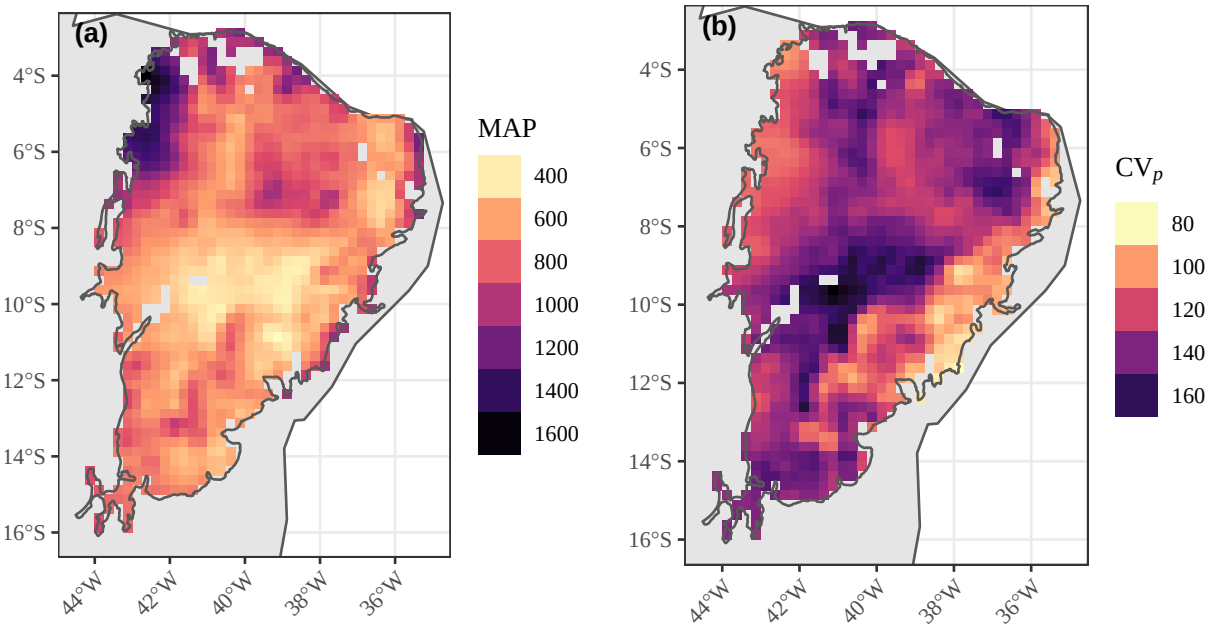

Figure 3.2: Spatial distribution of (a) MAP (mm^{-yr}) and (b) CV_p (%) based on the years 1982-2016.

12.35). For ΔPCI, k_3N exhibits the narrowest range with an average difference of 4.52, k_3E the widest range with an intermediate difference of 7 PCIΔ and k_3S the highest average difference of 8.38 PCIΔ.

Precipitation onset and peak maps (fig. 3.4) show the median onset and maximum precipitations months. They form 3 patches, a clear north-south divide that has a small southward notch north of the Chapada Diamantina (S̃ 13°; W 42°) and a detached area along the eastern border of the Caatinga. The wet season of the northern region mostly starts in January and peaks during March, the wet season onset of the eastern patch is around March, with a peak during austral winter. The southern area starts mainly in November and peaks subsequently in January.

The precipitation time series behaviour and predictability exhibits a clear north-south divide (S 7°; W 44° - S 5°; W 38°). The minimum 0.71 PEp is in the very north, indicating higher predictability and the presence of structure, such as long-term correlations and cycles (seasons) in the time series. High values are prevalent throughout most of the rest of region, up to 0.97 PE_p, indicating low predictability and stochastic time series, close to (white) noise behaviour. The very southern tip shows again slightly decreased PEp values, indicating the area of structural (cyclical) influences by the SACZ. MPRp closely mirrors the spatial distribution of PEp and can be found in the appendix 5.13.

All observed precipitation trends are negative. The strongest precipitation trends are in the

south east (S 12°, W 38°) with values up to −10 mm per decade. Merely 16.49 % of the region exhibits significant changes in precipitation ($p < 0.05$). k_3N exhibits the smallest rates of change and k_2E the strongest and most variable trends.

For vegetation, the spatial distribution of MAMF values loosely resembles MAP with low values in the dry core area and in the Northeast. Generally, there is a north-west to south-east gradient of declining values. The north-western area has the highest FPAR peak (excluding the total maximum in the Chapada Diamantina, MAMF 0.87 %). The lowest MAMF values seem to follow the distribution of the highest PCIΔ values (fig. 3.3 (c)).
Vegetation variability CV_f is lower than precipitation CV_p and seemingly linked to topography. It is the lowest in the north-western patch, along the northern border and well defined uplands. The core area of the Caatinga and the areas north-east and north-west of the Chapada do Araripe uplands show the highest CV_f with values above 40 %. k_3N has an average CV_f of 27.5 % and the biggest range, while k_3E has the highest average CV_f of 29.9 % and the lowest range. k_3S has the lowest average FPAR variability of 25.8 %.
Vegetation seasonality (fig. 3.7) exhibits similar, but slightly fuzzier patterns compared to precipitation with a 1-3 month delay (see fig. 3.4). The predominant onset month in the North is February, followed by peak values around May and June. The eastern patch along the border is again visible, but bigger compared to the precipitation maps with season onset during austral fall and peak during austral winter. The southern tip and western central parts exhibit season onset between November and January and a peak mainly during austral fall and even earlier for the most southern tip. "Valley" like patterns, that were already visible in the spatial distributions of MAP, CV_p (fig. 3.2 (a) & (b)), PCI2, PCIΔ (3.3 (b) & (c)) are again visible east and west of the northern uplands and connected the southern and central part of Caatinga that peaks during austral fall.

48.23 % of the Caatinga exhibits significant trends in FPAR between 1982 and 2016 (fig. 3.8 (b)). In contrast to trends in precipitation (fig. 3.5 (b)), 46.61% of the total area trend positively with a maximum up to $4.7\,\%/dec^{-1}$ and only 1.62 % show a negative trend. Only 7.72 % of the total area is trending in both precipitation and FPAR. The region around the northern uplands forms the largest coherent patch. Coherent patches are also located in the north western corner and the southern tip of the region. The fragmented negative trends are mainly concentrated in the central part. k_3N and k_3E exhibit trends of $\tilde{2}$ % per decade, k_3S $\tilde{1}.5$ % per decade. For the time period 1982-2010 Caatinga FPAR exhibited an average increase of 10.31 % compared to the baseline mean 1982-2010. For the reference period 1982-2016 the changes amount to an increase of 12.37 %.

The vast majority of time series pairs is correlated positively (99%, $p < 0.05$, fig. 3.9 (a)). The areas of the highest ρ_t are along the south western border, the western central area an the northern hinterland. The area north west of San Salvador and parts of the Chapada Diamantina are not significantly correlated, as well as the north-western edge of the region, a small patch area at the Rio São Francisco dam and a patch near Juazeiro do Norte (S 7.5°; W 39.3°). The few negatively correlated pixels are restricted to areas of strong human influence, the north-western edge or next to the non-significant patches.
The coefficients of determination of the lagged regressions results in a north-west to south-east gradient (fig. 3.9 (b)). The closest link is in the Northwest, where precipitation of the same and the three preceding months explain over 40% of the FPAR variability ($p < 0.05$). Only very low adj. R_t^2 are achieved along the eastern border and in the south-east . The area around San Salvador and the North of the state Bahia does not show any significant relationship.

Most cluster-based subgroups exhibited highly significant differences between precipitation- and vegetation-based parameters (see 3.1 & 3.2). For K_3E and k_3S, MAP exhibited only weakly significant differences. Between them, PE_p and R_t^2 did not exhibit significant differences ($p < 0.05$). Vegetation trends of k_3N and k_3E exhibited similar distributions and were not statistically discernible ($p < 0.05$).

Table 3.1: Group comparisons of the temporal parameters, split according to precipitation cluster memberships. Wilcoxon-Mann-Whitney Test; significance levels: $***$ <0.001, $**$ <0.01, $*$ <0.05, $-$ >0.05; exact p.values can be found in the supplements (see 5.1)

	k_3N-k_3E	k_3N-k_3S	k_3E-k_3S
MAP	$***$	$***$	$*$
CV_p	$***$	$***$	$***$
PCI1	$***$	$***$	$***$
PCI1	$***$	$***$	$***$
PCIΔ	$***$	$***$	$***$
TS_p	$***$	$***$	$***$
PE_p	$***$	$***$	$-$
MAMF	$***$	$***$	$***$
CV_f	$***$	$***$	$***$
TS_f	$-$	$***$	$***$
PE_f	$***$	$***$	$***$
ρ_t	$***$	$*$	$***$
R_t^2	$***$	$***$	$-$

Table 3.2: Group comparisons of the temporal parameters, split according to precipitation cluster memberships. Watson-Wheeler test of homogeneity of means for circular data; significance levels: $***$ <0.001, $**$ <0.01, $*$ <0.05, $-$ >0.05; exact p.values can be found in the supplements (see 5.1)

	k_3N-k_3E	k_3N-k_3S	k_3E-k_3S
ONS_p	$***$	$***$	$***$
$PEAK_p$	$***$	$***$	$***$
ONS_f	$***$	$***$	$***$
$PEAK_f$	$***$	$***$	$***$

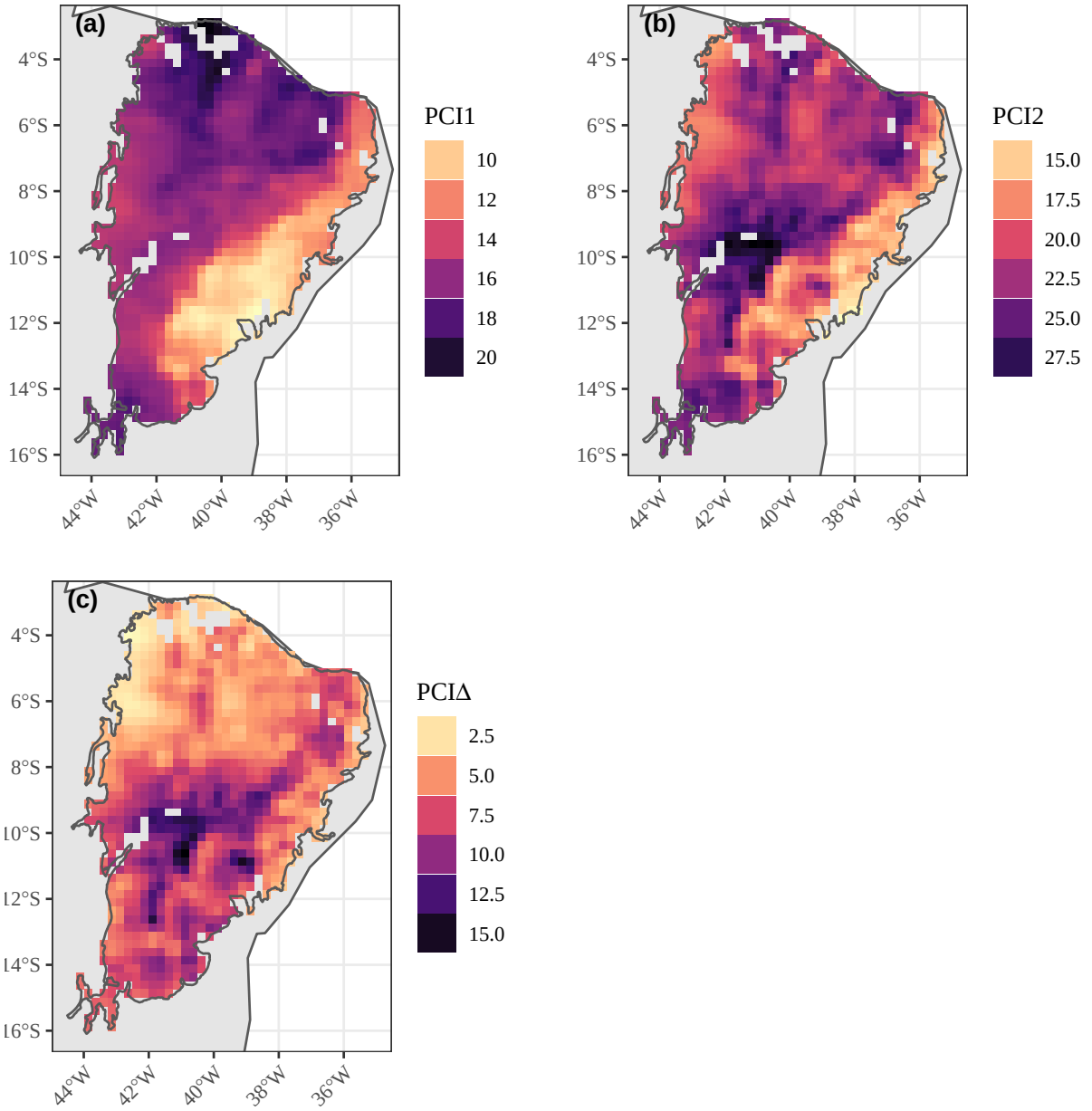

Figure 3.3: Spatial distribution of (a) PCI1, (b) PCI2, and (c) PCIΔ values based on the years 1982-2016. *Note: color scales are individual for each panel.*

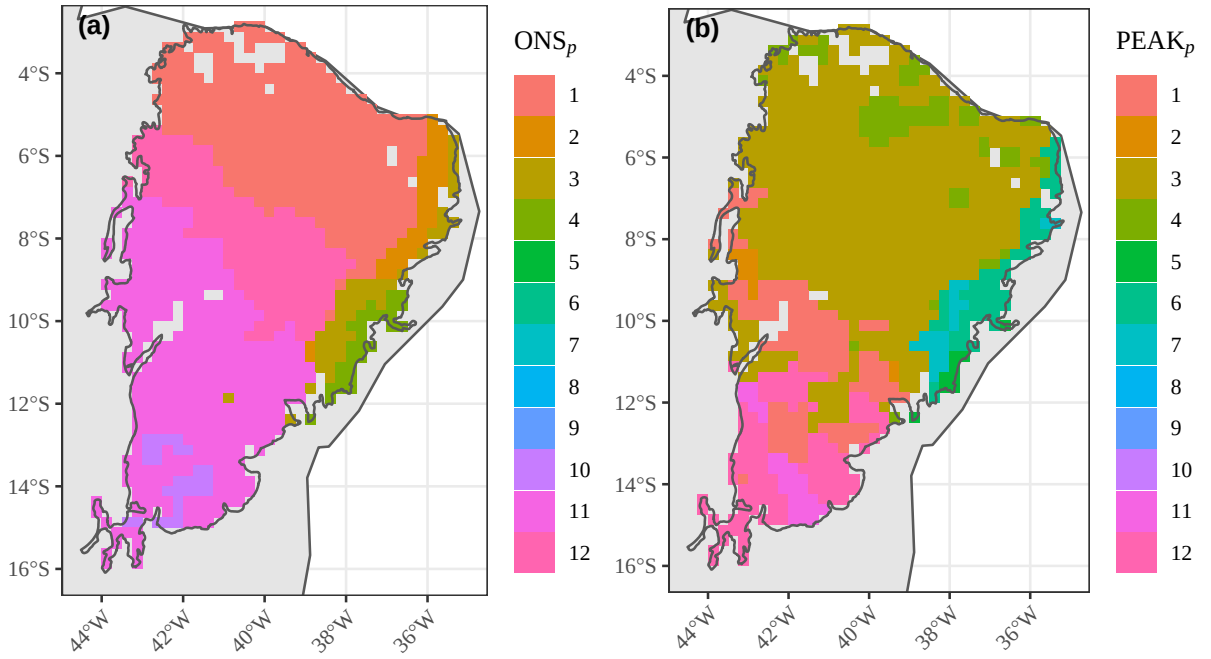

Figure 3.4: Spatial distribution of precipitation season parameters. Median a) wet season onset and b) precipitation maximum month based on the years 1982-2016. *Monthly values are coded from 1="January" to 12="December" as circular colors.*

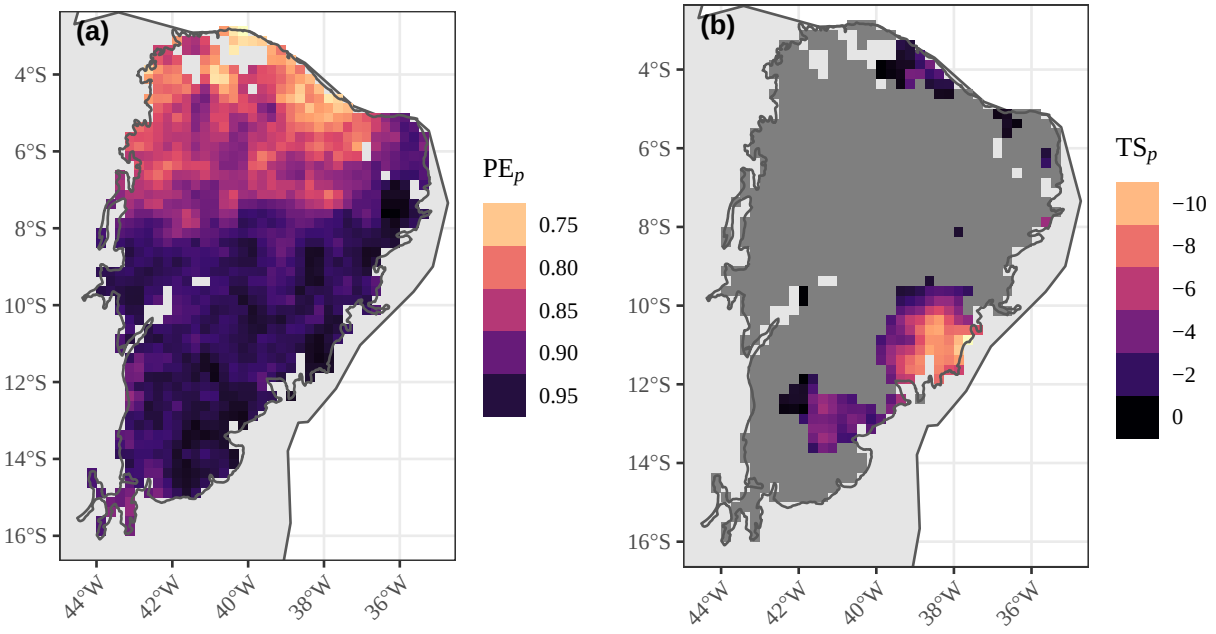

Figure 3.5: Spatial distribution of a) PE$_p$ ($\mathcal{H}$, *disorder*) and b) significant changes (TS$_p$ in mm^{-dec} ($p < 0.05$)) based on the years 1982-2016.

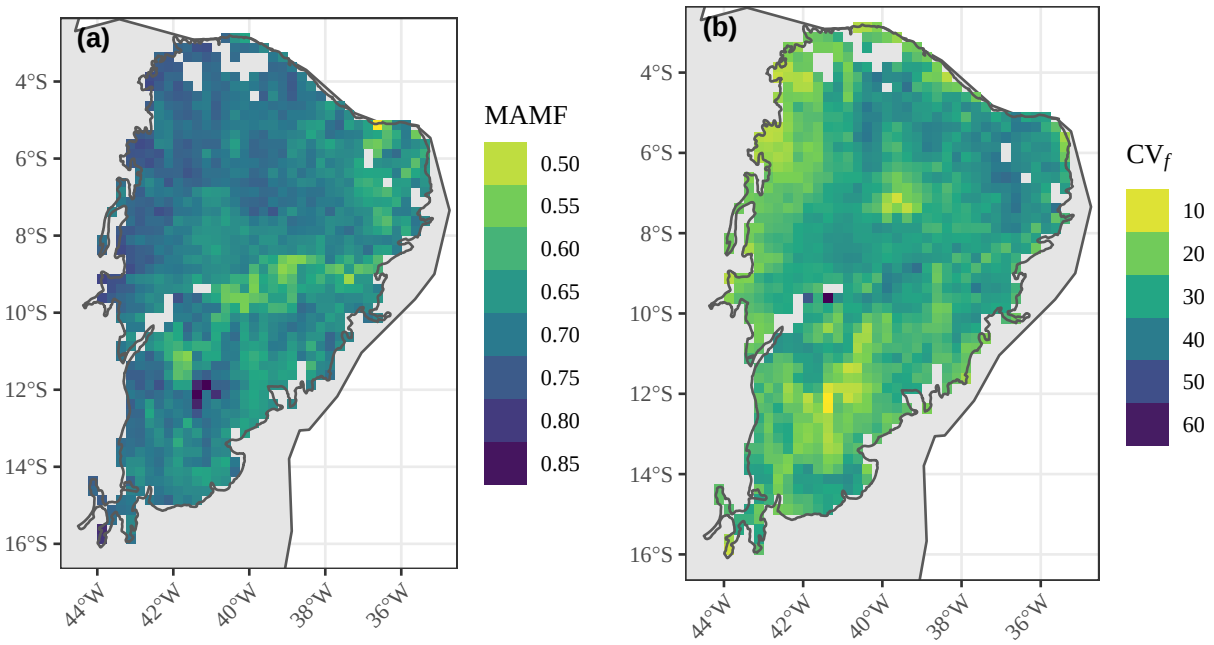

Figure 3.6: Spatial distribution of a) MAMF (%) and b) CV$_f$ (%) based on the years 1982-2016.

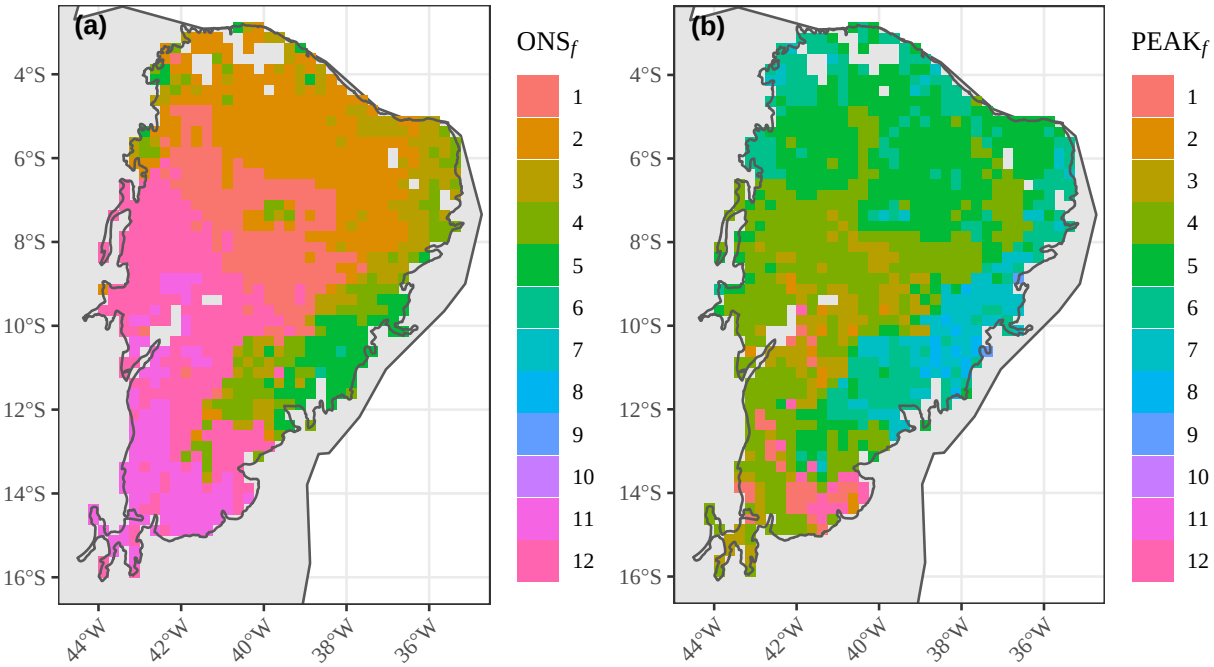

Figure 3.7: Spatial distribution of FPAR phenologies. Median a) season onset and b) maximum months based on the years 1982-2016. *Monthly values coded from 1="January" to 12="December" as circular colors.*

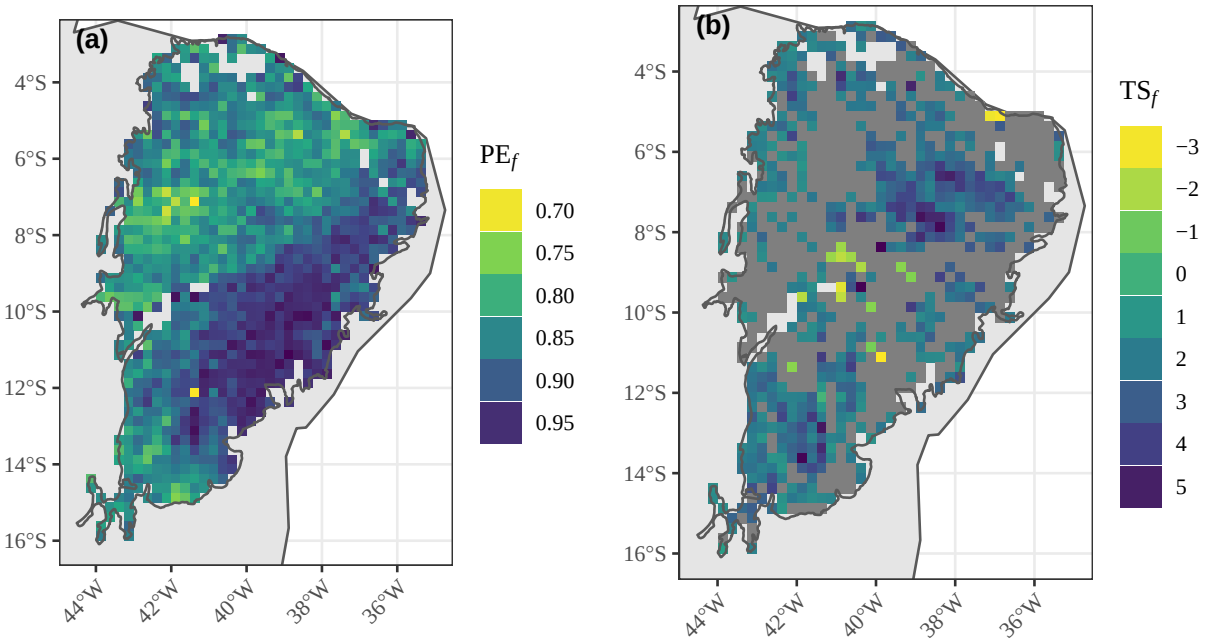

Figure 3.8: Spatial distribution of a) PE_f and b) significant changes per decade (TS_f in $\%^{-dec}$ ($p < 0.05$)) based on the years 1982-2016.

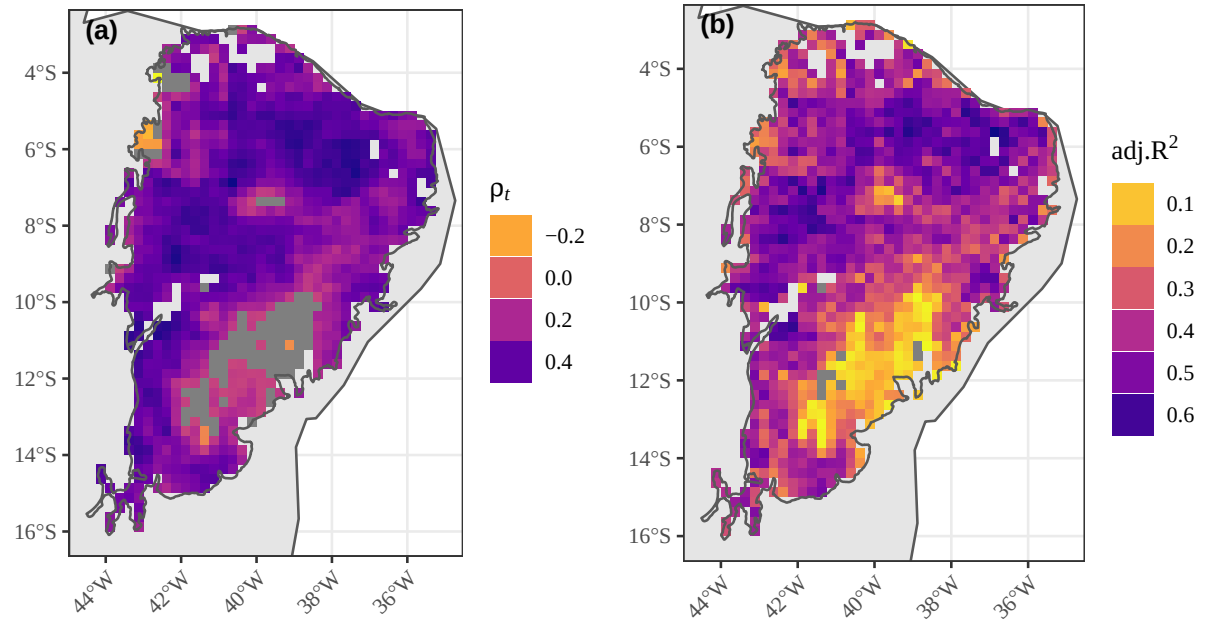

Figure 3.9: Spatial distribution of the strength of precipitation-vegetation coupling. (a) ρ_t ($p < 0.05$) and b) R_t^2 ($p < 0.05$) based on the years 1982-2016.

4 Discussion

In the following sections, the methods and results of this study will be critically reflected and interpreted in regard to the research questions posed in section 1. The first part (4.1) will evaluate the results of the analyses of the precipitation data. In section 4.1.1, the clustering results and the spatial extent of the influence zones will be discussed. Section 4.1.2 aims to define the individual precipitation regimes of the clusters and to evaluate the applied metrics. Subsequently, long-term precipitation trends will be discussed in regard to their causes and future perspectives (sec. 4.1.3). The second part (4.2), includes the results of the vegetation analyses. First, the patterns of precipitation-vegetation correlation will interpreted and a possible explanation for the spatial patterns will be presented in section 4.2.1. Section 4.2.2 will address the formation of macroecological vegetation units based on the precipitation regimes. Lastly, long-term changes in FPAR will be discussed in reference to possible drivers in section 4.2.3.

4.1 Precipitation

4.1.1 Spatial Extent of the Precipitation Regimes

The first aim of this study was to regionalize the Caatinga by comparing spatial patterns obtained by clustering GPCC precipitation data to theoretical influence zones of large scale atmospheric systems derived from literature review. While there is a vast body of literature on the large scale atmospheric systems sourcing Caatinga precipitation, this was the first study to spatially explicitly delineate their influence areas by clustering precipitation data. Although clustering quality indices suggested optimal numbers of clusters (k) ranging from 2 to 4 (see 5.2), the choice of $k = 3$ proved to be reasonable. The clustering yielded spatial patterns that resemble the theoretical influence zones depicted by de Andrade et al. (2017) and in section 2.1. They resembled northern cluster k_3N (ITCZ influence), a central-eastern cluster k_3E (EWD influence) and a south-western cluster k_3S (CF influence). Splitting the temporal parameters according to their pixels' cluster membership resulted in three distinct precipitation regimes and statistically different subgroups for all time series characteristics besides MAP and PE_p (see tables 3.1 & 3.2). While the results allowed for attributing the areas to their main influence system, wide ranges of some parameters implied high intracluster heterogeneity, either from overlapping influence areas or clustering mismatches. The spatial extent of k_3N was smaller than expected, while k_3E seemed to be the most heterogeneous cluster with the biggest ranges in the individual parameters, or with the most incorrectly admitted areas. A strong argument for the underestimation of the northern influence zone are the precipitation peaks in March and April, that are

prevalent in the seasonal profiles of clusters k_3E and k_3S (fig. 3.1 (c) & (d)). These two months followed the pattern of the ITCZ seasonal distribution and brought more precipitation than the typical levels of the two other clusters.
Large areas exhibited $PEAK_p$ that matches the ITCZ time signature, while the spatial distribution of parameters such as seasonal concentration (PCI1) and predictability (PE_p) spatial distribution indicated a smaller northern zone. The contributions of those two months are higher than the local baseline precipitation patterns. However, their influence on the local precipitation regimes and vegetation patterns might be negligible - due to the short water retention potential of the majority of the soils in the Caatinga (see sec. 2.1), individual months of strong precipitation might surpass a threshold and no longer shape or benefit local vegetation productivity (Heisler-White et al., 2008; Ross et al., 2012; Zhang et al., 2013). With the data at hand, it is not possible to discern whether precipitation contributions to the seasonal profiles stem from truly mismatched areas, or areas of overlapped and shared influences. PAM clustering showed meaningful and fast results for the core influence areas, while some ambiguity remained, especially in bordering zones, where the influences overlapped. A possible solution could be complex network analyses (Ciemer, 2019) or regionalization based on fuzzy clustering approaches (Satyanarayana and Srinivas, 2011; Dikbas et al., 2012) that allow for partial and mixed cluster memberships. Also, finer temporal scales could allow tracking or classification of rainfall events according to their sourcing system and to derive a ratio of their contributions.

4.1.2 Temporal Characteristics of the Precipitation Regimes

The analyses of seasonal dynamics in terms of timing and seasonal precipitation distributions revealed three main areas with distinct seasonal dynamics. The influence of the ITCZ system caused strong seasonal climate with "well defined" seasonality for k_3N and a strong temporal concentration of the precipitation in a few months. The erratic nature of the EWD and CF systems shaped the annual seasonal precipitation distribution of k_3E and k_3S and lead to weak seasonal precipitation patterns. Onset and peak patterns of the precipitation season (see 3.4) were in line with findings of Liebmann and Mechoso (2011); Hastenrath (2012); Gomes et al. (2019). Seasonal concentration patterns (PCI1 & PCI2, see fig.3.3) also supported the matching to the respective sourcing systems.
The timing dependent parameters ONS_p & $PEAK_p$ proved to be valuable to discriminate between the dominating sourcing systems. They clearly captured the strong differences in seasonal timing within the Caatinga, which are even today not often acknowledged in studies (e.g. de Medeiros et al. (2020)). Failing to acknowledge the different seasonal timings potentially weakens the results of e.g. precipitation-teleconnection correlations when cumulative wet season precipitation values are used, that effectively aggregate over the wrong time period for parts of the Caatinga. Additionally, timing dependent parameter distributions within the cluster

subgroups revealed ambiguous cluster memberships and mismatches. Some pixels exhibited seasonal timing patterns that clearly belonged to another precipitation sourcing system (see also supplementary materials, fig. 5.11).
The PCI proved to be a useful parameter in assessing seasonal dynamics, however interpretation of its results was less straightforward. Due to its nonlinear scaling (Oliver (1980); Michiels et al. (1992), see 2.4.3 & appendix 5.10), the PCI signal is smaller for distributions with precipitation in every month, regardless of the total amplitude. Hence, the differences between weakly seasonal distributions are less pronounced and the application of the PCI can only be recommended for regions that are at least semi-arid.

Precipitation variability was assessed in terms of variation in quantity and timing of the seasons. The interior region of the Caatinga, originally attributed to different clusters, (around S 10°; W 41°) exhibited the highest CV and PCIΔ values. High PCI2 values indicated a very concentrated season in individual years, caused by individual events, such as passing of EWDs or CFs (Gomes et al., 2019; Chaves and Cavalcanti, 2001), which contrasted the broad average distribution captured by PCI1. The resulting high PCIΔ values indicated a very high interannual variability of the timing of the precipitation. This area was characterized by very concentrated, short wet seasons that occured very irregularly throughout the year. Additionally, in this area, the lowest baseline precipitation met the highest precipitation variability, in both amount and timing. This pattern could be interpreted as a "dead spot" of the three sourcing systems. The area is south of a reliable ITCZ contribution to rainfall (Hastenrath, 2012), and only 14 % of the reported EWDs move across the region (Gomes et al., 2019). Additionally, the uplands of South Brazil, as well as local topography such as the Chapada Diamantina shield this area from advancing CFs (Chaves and Cavalcanti, 2001).

To complement the traditional descriptive statistics and to distinguish the different precipitation regimes from a structural point of view, the time series dynamics were assessed using information theory quantifiers. The spatial distribution of PE_p mapped out a clear northwest-southeast division that closely resembles the extent of the northern cluster k_3N, while clusters k_3E and k_3S did not exhibit significant differences. The lower PE values of the northern cluster k_3N indicated cyclic (i.e. seasonal) components and autocorrelation processes that are governed by underlying deterministic mechanisms. This supported the image of relatively stable timing and high predictability of the ITCZ controlled precipitation, as reported by Hastenrath (2012); Marengo et al. (2016, 2018); Giannini et al. (2004); Nobre et al. (2006); Rodrigues and McPhaden (2014); Hounsou-Gbo et al. (2016)). The other clusters exhibited behaviour that is closer to stochastic, k-noise series (i.e. pink to white noise) (Vasseur and Yodzis, 2004; Sippel et al., 2016) without strong long-term autocorrelations or cyclic components. This causes low predictability

and rainfalls that are still difficult to forecast (Torres and Ferreira, 2011; Pennekamp et al., 2019).

Since CFs and EWDs are reported to occur several times per month (Nieto-Ferreira et al., 2011; Gomes et al., 2019), the coarse temporal scale of this study might have prohibited further differentiation between the two systems in terms of PE_p. The use of different embedding dimensions (Cuesta-Frau et al., 2019), lagged permutations (Zunino et al., 2010) or of a third ITQ could help to better distinguish regional structures and differing dynamical behavior (Rosso et al., 2007; Sippel et al., 2016). A possible addition is Fisher-Information (Frieden, 2004), which is asymetric, i.e. sensitive to local changes in the pattern distribution and could therefore differentiate between systems with equal PE_p but different permutation pattern distributions (Olivares et al., 2012). However, the similar PE values of k_3E and k_3S could also be related to the structural similarity of the EWD and CF sourcing systems. While Yamazaki and Rao (1977) first reported EWDs as the source of the east coast wet season, Kousky (1979, 1980) stated far reaching CFs or their remnants that occur along the coast as the source of rainfall in eastern Northeast Brazil. Gomes et al. (2019) recently reported that the main system associated with EWD formation are CFs (72% of all captured EWDs) and that the most intense CFs tend to move towards lower latitudes, disturbing the trade winds exactly during the eastern rainy season (austral autumn and winter). Therefore, EWD and CF systems would share a common origin, explaining the structural similarity of their precipitation time series, while the temporal offset of the seasons could be explained by the time the turbulences need to reach low latitudes and to trigger EWD occurrences.

4.1.3 Long-term changes in precipitation

Long-term precipitation trends exhibited highly localized and negative patterns. While a tendency towards wetter conditions in South America has been reported (Magrin et al., 2007), also in the the semiarid regions of Brazil (Guerreiro et al., 2013), the results of this study (see section 3) added to reports of drier conditions for Northeast Brazil (Haylock et al., 2006; Marengo et al., 2010; Barbosa and Kumar, 2016). The findings did not match expected drying trends for NEB, that are caused by an earlier and stronger northward shift of the ITCZ (Nobre and Shukla, 1996; Allen et al., 2014), but supported another mechanism.
Cluster k_3N exhibited almost no trends, which is in line with findings of Lucena et al. (2011), who reported a lack of precipitation trends in NNEB despite changing ENSO, PDO and NAO dynamics. They observed that ENSO-related low-frequency variations in the tropical oceans occurred together with symmetric long-term changes in PDO/NAO during the second half of the twentieth century. This symmetrical, in-phase (i.e., inverse impacts), long-term climate behaviour could have inhibited their possible indirect mutual impact on NNEB seasonal rainfall. The resulting precipitation trends would cancel each other out. In contrast, k_3E and k_3S ex-

hibited clear drying trends (fig. 3.5 (b)), which is in line with findings by Dunn et al. (2020). A possible explanation for the spatial distribution of the detected trends lies in the decreasing moisture transport by the eastern EWD source system. Comparing fig. 3.5 (b) and fig. 4.1 showed that the spatial distribution of the significant trends and typical EWD storm tracks align well. The patterns revealed that even some areas that were associated with other precipitation systems, obtained a considerable amount of precipitation from passing EWDs. Negatively trending areas along the northern coast, that had not been attributed to k_3E could be explained as the areas of the ITCZ cluster, that lie along the northern EWD route and were additionally fed by local EWD-topography interaction, along the uplands of the western border of Ceara. The trends of k_3E could be attributed to the main route of EWDs hitting the eastern coast. The areas east of the Chapada Diamantina and originally attributed to k_3S seem to have obtained an important but decreasing share of rainfall from interaction of westward propagating EWDs with local convection and topography. This seems to be a relatively new development, as Brahmananda Rao et al. (1993) reported no trends in precipitation for ENEB during the years 1964-1983. While the isolated effect of EWDs is weak for the whole Caatinga (Kayano, 2003), it is the main source of rainfall variability along the west coast and in the interior NEB region (De Lima Moscati and Gan, 2007), and only 14 % of the reported EWDs move across the region towards the Amazon (Gomes et al., 2019). A westward expansion, or a latitudinal shift of the SASA, accompanied by a weakening of the easterly tradewinds (pathway of EWDs), could lead to decreased EWD activity and moisture transport into the exact areas that are trending in fig. 3.5 (b). Additionally, the location of the SACZ is an important source of precipitation over SNEB and is linked to the displacement of the SASA and weaker trade winds (Chaves and Cavalcanti, 2001), which also leads to decreasing precipitation contributions in the southern Caatinga. A poleward shift of the SACZ has been reported for the period 1979-2014 (Zilli et al., 2019). Based on different emission scenarios, climate change predictions and the resulting atmospheric warming, it is expected that the position of the SASA will continue to shift poleward and intensify in the twenty-first century (Nunez et al., 2009; Soares and Marengo, 2009; Marengo et al., 2010; Seth et al., 2010; Gilliland and Keim, 2018). Therefore, the interior of the Caatinga, the South and especially its eastern border will be the most affected areas, due to fewer EWD moisture transports. This could also offer a mechanistic explanation for Marengo et al. (2018), who reported that areas most affected by the 2010-2016 drought were areas in the northern part of the state Bahia, which is particularly important for agricultural production (k_3E, see fig 3.1 (a)).

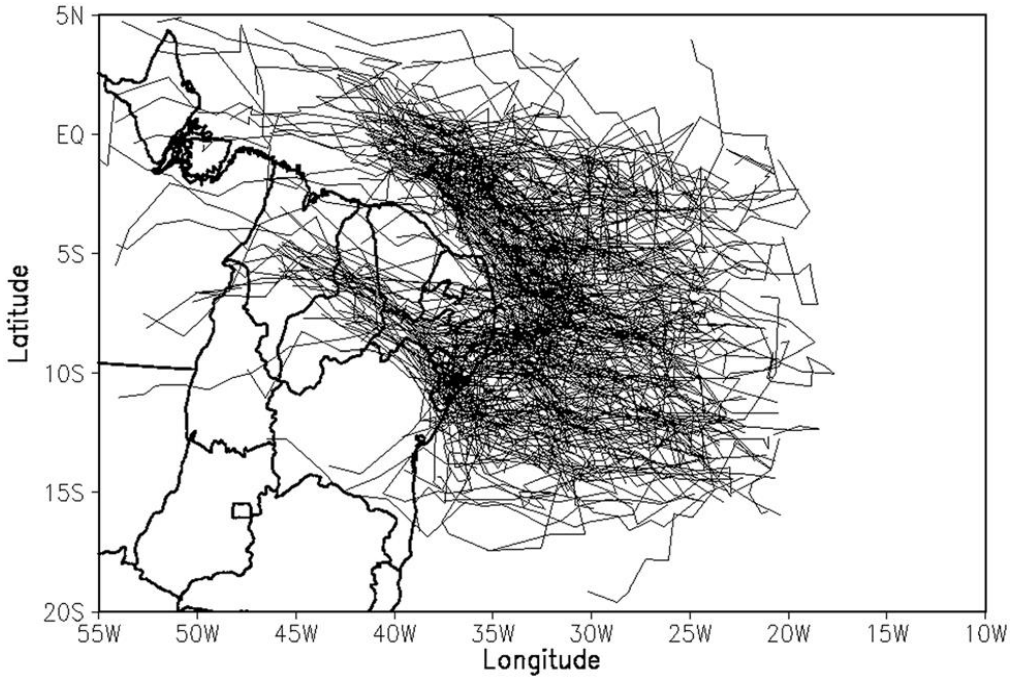

Figure 4.1: Typical EWD routes (thin lines) as captured by cyclonic vorticity center tracks on the ERAI data (1989-2009) for all EWDs detected at 850 hPa (Gomes et al., 2019). Black lines indicate the coastline and federal states.

4.2 Vegetation

4.2.1 Strength of the Precipitation-Vegetation Link

A strong correlation between precipitation and vegetation was expected for the Caatinga SDTF (Barbosa and Kumar, 2016; Barbosa et al., 2019; de Jesus et al., 2020). To verify this, non-parametric spearman correlation and lagged linear models were applied for each precipitation-vegetation time series pair. Low to intermediate correlations were found, with hotspots in the central northern and western areas. The patterns of increased ρ_t & R_t^2 (fig. 3.9) matched deforestation and total human impact factors as reported by da Silva and Barbosa (2017). The uplands exhibited weak or no significant correlation, as well as huge areas in the area of the state Bahia, probably due to strong agricultural activities in these areas (Marengo et al., 2018). The results of the lagged regression were similarly heterogeneous and explained 1 % to 66 % variability in vegetation (R_t^2). Again, the highest values were in the central North and the West of the region. While Rito et al. (2017) reported that precipitation is a much stronger control on Caatinga vegetation than human acute and chronic disturbance, the areas of increased ρ_t & R_t^2 lined up with findings of Antongiovanni et al. (2018) (see fig. 4.2). The authors reported more Caatinga remnants and the highest remaining Caatinga land cover in these areas. In areas

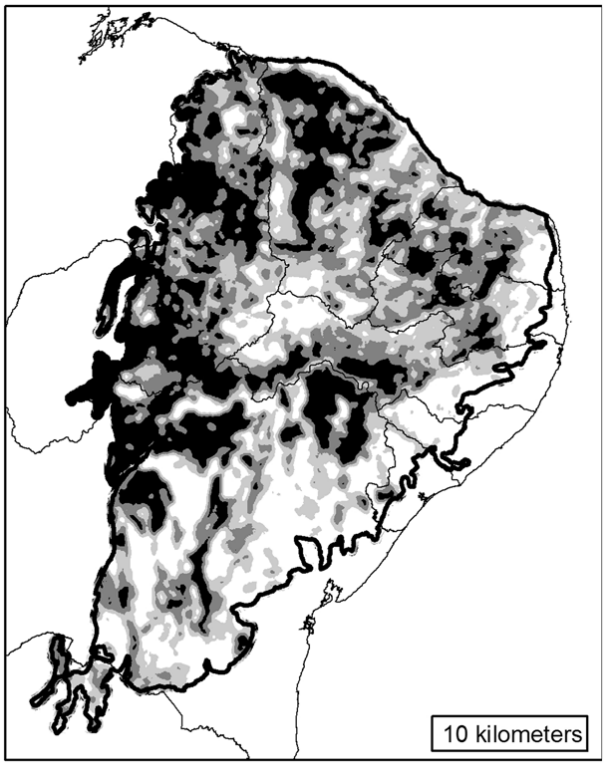

Figure 4.2: Fragmentation patterns of the Caatinga SDTF at 10 km resolution (Antongiovanni et al., 2018). Dark areas indicate more forest cover and less fragmented Caatinga remnants. Thin black lines are coast lines and borders of the federal states, the thick black line is the border of the Caatinga (*sensu* IBGE (2004)).

without strong orographic precipitation effects, the strength of the link between Caatinga vegetation and precipitation could therefore act as a proxy for human disturbance. However, the strength of the correlations should be used from an exploratory, landscape scale point of view due to the coarse spatial resolution, mixed pixel effects (Foody, 2004; Chen et al., 2018) and issues regarding the linear model. The effects of agriculture can only be identified if the area is big enough to clearly change the signature of the pixel (720 km^2 to 770 km^2 in the Caatinga). Pixels which sharply stood out from their surroundings, for example in CV_f and PE_f, were only observed along Rio São Francisco dam and in the Chapada Diamantina. Besides better spatiotemporal resolution, generalized linear regression to adjust for the bounded response of the FPAR variable, moving window analyses (e.g. Abel et al. (2020)) or analyses of spatial patterns of model performance (e.g. bias) could provide further insights.

4.2.2 Macroecological Vegetation Units

The strong effect of precipitation on SDTF species diversity, distribution, structure and productivity (Ruggiero et al., 2002; Rito et al., 2017) and the correlation between precipitation and Caatinga vegetation (Barbosa and Kumar, 2016; Barbosa et al., 2019; de Jesus et al., 2020) lead to the hypothesis of "Caatinga regimes" - Regional, landscape scale areas of distinct Caatinga vegetation dynamics that match the precipitation regimes presented in sections 3 & 4. They were presumed to share dynamics and patterns with their precipitation regime, independent of other local factors such as topographic, geologic, edaphic or floristic settings. Splitting the spatiotemporal parameters of FPAR into groups according to the precipitation clusters k_3N,k_3E and K_3S yielded significantly different subgroups for all tested vegetation-based parameters (see section 3) besides trends (TS_f). The results supported the hypotheses of subregional similarity of vegetation dynamics, even across ecoregions, as presented in da Silva et al. (2017). Due to the coarse spatial resolution of the vegetation data at hand, final conclusions on species or community level are not possible. However, the results allowed for interpretations on a regional, landscape scale. The findings heavily imply varying gradients of species distributions, community structure and functional traits related to drought sensitivity according to the influence zones defined in sections 3, 4.1.1 & 4.1.2.

4.2.3 Long-term Changes in Vegetation Productivity

Vegetation exhibited significant changes in more than half of the region. In contrast to the strong link of precipitation and vegetation dynamics, trends of precipitation and vegetation were not related and overlapped in merely $<8\,\%$ of the area. The majority of detected FPAR trends were positive, indicating that for the years 1982-2016, the Caatinga acted as carbon sink in terms of aboveground primary productivity as captured by FPAR. Given the extraordinary dry conditions during the 1980s in the Caatinga (de Medeiros et al., 2020), regeneration of vegetation conditions could play a role in the observed positive trends, as rebound effects of vegetation after the droughts during the 1970s and 1980s have been discussed as a source of the greening of the Sahel (Olsson et al., 2005). However, especially in the light of new droughts after 2010 (Marengo et al., 2018), other factors seem to be more important. Trends in primary productivity have been explained by structural changes in vegetation composition, or CO_2 fertilization (Brandt et al., 2019; Abel et al., 2020). Due to CO_2 fertilization, the increase in water use efficiency of photosynthesis with rising atmospheric CO_2 levels has long been anticipated to lead to increased foliage cover in warm and dry environments (Farquhar, 1997). This effect has been globally confirmed in numerous studies, using satellite and *in situ* observations and facilitates changes towards woodier and more densely vegetated landscapes (Buitenwerf et al., 2012; Donohue et al., 2009; Morgan et al., 2007; Lu et al., 2016). For areas where water is the limiting factor for vegetation growth, Donohue et al. (2013) reported an CO_2 induced increase

in green foliage cover of around 11 %. Adjusted to their time period (1982-2010), the results of this study are well in line with an average increase of 10.31% (see supplemenary materials, fig. 5.16), explaining most of the observed positive trends, especially in the absence of precipitation trends. Changes in vegetation type may also affect the changes. In semi-arid grasslands and woodlands, annual and perennial species differ in their response (e.g. for NDVI: Dall'Olmo and Karnieli (2002)). Irregular precipitation regimes and changing timing are reported to favour woody vegetation (Zhang et al., 2018). Newingham et al. (2013) reported no increase in dryland perennial plant biomass after 10 yr of CO_2 fertilization, which would additionally hint towards an increase of woody biomass and not an increase of perennials in the positively trending areas in the Caatinga. With the data at hand, it is unfortunately not possible to ultimately decipher if the observed changes result from CO_2 fertilization or structural changes in vegetation composition.

5 Conclusion

The major influence zones of large-scale atmospheric systems in the Caatinga were delineated for the first time, using spatially explicit PAM clustering of gauge-based precipitation data. While the eastern and southern clusters exhibited similarities, all obtained zones possessed distinct precipitation regimes and long-term dynamics. Their characteristics were found to govern Caatinga SDTF dynamics. FPAR data could be split into three significantly different marcoecological units (termed *Caatinga regimes*) according to the precipitation clusters. These were superimposed on the vegetation, across local geologic, edaphic, floristic settings or anthropogenic disturbance regime. Precipitation and vegetation did not exhibit common long-term trends. Vegetation exhibited spatially heterogeneous linkage to precipitation, with a higher correlation in areas that are reported to be less impacted by anthropogenic disturbance (Antongiovanni et al., 2018).

In conclusion, the findings showed that three distinct macroecological units exist within the Caatinga STDF, which are governed by and follow the dynamics of the prevailing precipitation regimes. With an expected increase in interannual variability of rainfall (Kharin et al., 2007; Field et al., 2012), increasingly unpredictable seasonality and an altered number of heavy rainfall and drought events (Smith, 2011; Fischer et al., 2013), a better spatiotemporal characterisation of precipitation regimes is (Zhang et al., 2018). The nuanced control of precipitation regimes on dryland vegetation dynamics has only recently come into the focus of environmental research and current models possibly underestimate the crucial role of dryland ecosystems for the global carbon balance (Poulter et al., 2014; Ahlström et al., 2015; Ouédraogo et al., 2016; Biederman et al., 2017; Brandt et al., 2018). Understanding relationships between precipitation regimes and land surface phenology in seasonally dry tropical forests is of fundamental significance for modelling tropical vegetation dynamics, quantifying their carbon budget and predicting their future response to climate changes.

Bibliography

Abel, C., Horion, S., Tagesson, T., De Keersmaecker, W., Seddon, A. W., Abdi, A. M. and Fensholt, R. (2020), 'The human–environment nexus and vegetation–rainfall sensitivity in tropical drylands', *Nature Sustainability* pp. 1–8.

Adeel, Z., Safriel, U., Niemeijer, D., White, R., De Kalbermatten, G., Glantz, M., Salem, B., Scholes, B., Niamir-Fuller, M., Ehui, S. et al. (2005), 'Ecosystems and human well-being: desertification synthesis. a report of the millennium ecosystem assessment', *World Resources Institute. Washington DC* .

Agostinelli, C. and Lund, U. (2017), *R package* `circular`*: Circular Statistics (version 0.4-93)*, CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA.
URL: *https://r-forge.r-project.org/projects/circular/*

Ahlmann-Eltze, C. (2019), *ggsignif: Significance Brackets for 'ggplot2'*. R package version 0.6.0.
URL: *https://CRAN.R-project.org/package=ggsignif*

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., Kato, E., Poulter, B., Sitch, S., Stocker, B. D., Viovy, N., Wang, Y. P., Wiltshire, A., Zaehle, S. and Zeng, N. (2015), 'The dominant role of semi-arid ecosystems in the trend and variability of the land co2 sink', *Science* **348**(6237), 895–899.

Akritas, M. G., Murphy, S. A. and Lavalley, M. P. (1995), 'The theil-sen estimator with doubly censored data and applications to astronomy', *Journal of the American Statistical Association* **90**(429), 170–177.

Alcaraz-Segura, D., Chuvieco, E., Epstein, H. E., Kasischke, E. S. and Trishchenko, A. (2010), 'Debating the greening vs. browning of the north american boreal forest: differences between satellite datasets', *Global Change Biology* **16**(2), 760–770.

Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. T. et al. (2010), 'A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests', *Forest ecology and management* **259**(4), 660–684.

Allen, S. K., Plattner, G.-K., Nauels, A., Xia, Y. and Stocker, T. F. (2014), 'Climate change 2013: The physical science basis. an overview of the working group 1 contribution to the fifth assessment report of the intergovernmental panel on climate change (ipcc).', *Eguga* p. 3544.

Althoff, T. D., Menezes, R. S. C., de Carvalho, A. L., de Siqueira Pinto, A., Santiago, G. A. C. F., Ometto, J. P. H. B., von Randow, C. and Sampaio, E. V. d. S. B. (2016), 'Climate change impacts on the sustainability of the firewood harvest and vegetation and soil carbon stocks in a tropical dry forest in santa teresinha municipality, northeast brazil', *Forest Ecology and Management* **360**, 367–375.

Amorim, A. C. B., Chaves, R. R. and Santos e Silva, C. M. (2014), 'Influence of the tropical atlantic ocean's sea surface temperature in the eastern northeast brazil precipitation'.

Antongiovanni, M., Venticinque, E. M. and Fonseca, C. R. (2018), 'Fragmentation patterns of the caatinga drylands', *Landscape Ecology* **33**(8), 1353–1367.

Aragão, L. E. O., Malhi, Y., Roman-Cuesta, R. M., Saatchi, S., Anderson, L. O. and Shimabukuro, Y. E. (2007), 'Spatial patterns and fire response of recent amazonian droughts', *Geophysical Research Letters* **34**(7).

Arora, P., Varshney, S. et al. (2016), 'Analysis of k-means and k-medoids algorithm for big data', *Procedia Computer Science* **78**, 507–512.

Axelrod, D. I. (1966), 'Origin of deciduous and evergreen habits in temperate forests', *Evolution* pp. 1–15.

Baettig, M. B., Wild, M. and Imboden, D. M. (2007), 'A climate change index: Where climate change may be most prominent in the 21st century', *Geophysical Research Letters* **34**(1).

Bandt, C. and Pompe, B. (2002), 'Permutation entropy: A natural complexity measure for time series', *Phys. Rev. Lett.* **88**, 174102.

Barbosa, H. A., Kumar, T. L., Paredes, F., Elliott, S. and Ayuga, J. (2019), 'Assessment of caatinga response to drought using meteosat-seviri normalized difference vegetation index (2008–2016)', *ISPRS journal of photogrammetry and remote sensing* **148**, 235–252.

Barbosa, H., Huete, A. and Baethgen, W. (2006), 'A 20-year study of ndvi variability over the northeast region of brazil', *Journal of arid environments* **67**(2), 288–307.

Barbosa, H. and Kumar, T. L. (2016), 'Influence of rainfall variability on the vegetation dynamics over northeastern brazil', *Journal of Arid Environments* **124**, 377–387.

Bastin, J.-F., Berrahmouni, N., Grainger, A., Maniatis, D., Mollicone, D., Moore, R., Patriarca, C., Picard, N., Sparrow, B., Abraham, E. M. et al. (2017), 'The extent of forest in dryland biomes', *Science* **356**(6338), 635–638.

Batschelet, E. (1981), 'Circular statistics in biology.', *Academic Press, 111 Fifth Ave., New York, Ny 10003, 1981, 388* .

Bauer, D. F. (1972), 'Constructing confidence sets using rank statistics', *Journal of the American Statistical Association* **67**(339), 687–690.

Beaumont, L. J., Hughes, L. and Poulsen, M. (2005), 'Predicting species distributions: use of climatic parameters in bioclim and its impact on predictions of species' current and future distributions', *Ecological modelling* **186**(2), 251–270.

Belagoune, F., Boutoutaou, D. and Bellout, M. (2017), Spatio-temporal analysis of extreme rainfall and flooding in the semi-arid region of algeria and its application to the regionalization of flood flow estimates, *in* 'AIP Conference Proceedings', Vol. 1814, AIP Publishing LLC, p. 020027.

Biederman, J. A., Scott, R. L., Bell, T. W., Bowling, D. R., Dore, S., Garatuza-Payan, J., Kolb, T. E., Krishnan, P., Krofcheck, D. J., Litvak, M. E. et al. (2017), 'Co 2 exchange and evapotranspiration across dryland ecosystems of southwestern north america', *Global Change Biology* **23**(10), 4204–4221.

Booth, T. H., Nix, H. A., Busby, J. R. and Hutchinson, M. F. (2014), 'Bioclim: the first species distribution modelling package, its early applications and relevance to most current maxent studies', *Diversity and Distributions* **20**(1), 1–9.

Borrell, J., Dodsworth, S., Forest, F., Pérez-Escobar, O. A., Lee, M., Mattana, E., Stevenson, P., Howes, M.-J., Pritchard, H. W., Ballesteros, D. et al. (2020), 'The climatic challenge: Which plants will people use in the next century?', *Environmental and Experimental Botany* **170**, 103872.

Bowman, D. and Prior, L. D. (2005), 'Why do evergreen trees dominate the australian seasonal tropics?', *Australian Journal of Botany* **53**(5), 379–399.

Bracken, C., Rajagopalan, B., Alexander, M. and Gangopadhyay, S. (2015), 'Spatial variability of seasonal extreme precipitation in the western united states', *Journal of Geophysical Research: Atmospheres* **120**(10), 4522–4533.

Brahmananda Rao, V., De Lima, M. C. and Franchito, S. (1993), 'Seasonal and interannual variations of rainfall over eastern northeast brazil', *Journal of Climate* **6**(9), 1754–1763.

Brandt, M., Hiernaux, P., Rasmussen, K., Tucker, C. J., Wigneron, J.-P., Diouf, A. A., Herrmann, S. M., Zhang, W., Kergoat, L., Mbow, C. et al. (2019), 'Changes in rainfall distribution promote woody foliage production in the sahel', *Communications biology* **2**(1), 1–10.

Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L. V., Hiernaux, P., Diouf, A. A. et al. (2020), 'An unexpectedly large count of trees in the west african sahara and sahel', *Nature* pp. 1–5.

Brandt, M., Wigneron, J.-P., Chave, J., Tagesson, T., Penuelas, J., Ciais, P., Rasmussen, K., Tian, F., Mbow, C., Al-Yaari, A. et al. (2018), 'Satellite passive microwaves reveal recent climate-induced carbon losses in african drylands', *Nature ecology & evolution* **2**(5), 827–835.

Brienen, R. J., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., Lopez-Gonzalez, G., Monteagudo-Mendoza, A., Malhi, Y., Lewis, S. L. et al. (2015), 'Long-term decline of the amazon carbon sink', *Nature* **519**(7543), 344–348.

Brissaud, J.-B. (2005), 'The meanings of entropy', *Entropy* **7**(1), 6896.

Buitenwerf, R., Bond, W., Stevens, N. and Trollope, W. (2012), 'Increased tree densities in s outh a frican savannas:¿ 50 years of data suggests co 2 as a driver', *Global Change Biology* **18**(2), 675–684.

Ceccherini, G., Gobron, N., Migliavacca, M. and Robustelli, M. (2013), Long-term measurements of plant phenology over europe derived from seawifs and meris, *in* 'ESA Living Planet Symposium', Vol. 722 of *ESA Special Publication*, p. 71.

Change, I. P. O. C. (2007), 'Climate change 2007: synthesis report', *Geneva: IPCC* .

Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A. (2014), 'NbClust: An R package for determining the relevant number of clusters in a data set', *Journal of Statistical Software* **61**(6), 1–36.
URL: *http://www.jstatsoft.org/v61/i06/*

Chaves, R. R. and Cavalcanti, I. F. A. (2001), 'Atmospheric circulation features associated with rainfall variability over southern northeast brazil', *Monthly Weather Review* **129**(10), 2614–2626.

Chen, S., Lin, G., Huang, J. and Jenerette, G. D. (2009), 'Dependence of carbon sequestration on the differential responses of ecosystem photosynthesis and respiration to rain pulses in a semiarid steppe', *Global Change Biology* **15**(10), 2450–2461.

Chen, X., Wang, D., Chen, J., Wang, C. and Shen, M. (2018), 'The mixed pixel effect in land surface phenology: A simulation study', *Remote Sensing of Environment* **211**, 338–344.

Cherlet, M., Hutchinson, C., Reynolds, J., Hill, J., Sommer, S. and von Maltitz, G. (2018), 'World atlas of desertification (p. 295). luxembourg: Publication office of the european union'.

Ciemer, C. (2019), Complex systems analysis of changing rainfall regimes in South America and their implications for the Amazon rainforest, PhD thesis, Humboldt University of Berlin.

Claverie, M., Matthews, J., Vermote, E. and Justice, C. (2016), 'A 30+ year avhrr lai and fapar climate data record: Algorithm description and validation', *Remote Sensing* **8**(3), 263.

Cuesta-Frau, D., Murillo-Escobar, J. P., Orrego, D. A. and Delgado-Trejos, E. (2019), 'Embedded dimension and time series length. practical influence on permutation entropy and its applications', *Entropy* **21**(4), 385.

da Silva, E., Nogueira, R., de Azevedo Neto, A., de Brito, J., Cabral, E. et al. (2004), 'Ecophysiological aspects of ten species occurring in "caatinga" dry forest in cabaceiras county, paraiba state, brazil.', *Iheringia, Série Botânica* **59**(2), 201–205.

da Silva, J. M. C. and Barbosa, L. C. F. (2017), Impact of human activities on the caatinga, *in* 'Caatinga', Springer, pp. 359–368.

da Silva, J. M. C., Barbosa, L. C. F., Leal, I. R. and Tabarelli, M. (2017), The caatinga: understanding the challenges, *in* 'Caatinga', Springer, pp. 3–19.

Dall'Olmo, G. and Karnieli, A. (2002), 'Monitoring phenological cycles of desert ecosystems using ndvi and lst data derived from noaa-avhrr imagery', *International Journal of Remote Sensing* **23**(19), 4055–4071.

Davidson, E. A., de Araújo, A. C., Artaxo, P., Balch, J. K., Brown, I. F., Bustamante, M. M., Coe, M. T., DeFries, R. S., Keller, M., Longo, M. et al. (2012), 'The amazon basin in transition', *Nature* **481**(7381), 321–328.

Davies, J., Poulsen, L., Schulte-Herbrüggen, B., Mackinnon, K., Crawhall, N., Henwood, W. D., Dudley, N., Smith, J., Gudka, M. et al. (2012), 'Conserving dryland biodiversity', *IUCN, UNEP-WCMC, UNCCD, Nairobi, Bonn* .

de Andrade, E. M., do Nascimento Aquino, D., Chaves, L. C. G. and Lopes, F. B. (2017), Water as capital and its uses in the caatinga, *in* 'Caatinga', Springer, pp. 281–302.

de Andrade, E. M., Sena, M. G. T., da Silva, A. G. R., Pereira, F. J. S. and Lopes, F. B. (2016), 'Uncertainties of the rainfall regime in a tropical semi-arid region: the case of the state of ceará', *Revista Agroambiente Online* **10**(2), 88–95.

De Beurs, K. M. and Henebry, G. M. (2005), 'Land surface phenology and temperature variation in the international geosphere–biosphere program high-latitude transects', *Global Change Biology* **11**(5), 779–790.

de Jesus, J. B., Kuplich, T. M., de Carvalho Barreto, Í. D., da Rosa, C. N. and Hillebrand, F. L. (2020), 'Temporal and phenological profiles of open and dense caatinga using remote sensing: response to precipitation and its irregularities'.

De Keersmaecker, W., Lhermitte, S., Tits, L., Honnay, O., Somers, B. and Coppin, P. (2015), 'A model quantifying global vegetation resistance and resilience to short-term climate anomalies and their relationship with vegetation cover', *Global Ecology and Biogeography* **24**(5), 539–548.

de Lima, A. L. A., Sampaio, E. V. d. S. B., de Castro, C. C., Rodal, M. J. N., Antonino, A. C. D. and de Melo, A. L. (2012), 'Do the phenology and functional stem attributes of woody species allow for the identification of functional groups in the semiarid region of brazil?', *Trees* **26**(5), 1605–1616.

De Lima Moscati, M. C. and Gan, M. A. (2007), 'Rainfall variability in the rainy season of semiarid zone of northeast brazil (neb) and its relation to wind regime', *International Journal of Climatology: A Journal of the Royal Meteorological Society* **27**(4), 493–512.

de Medeiros, F. J., de Oliveira, C. P., e Silva, C. M. S. and de Araújo, J. M. (2020), 'Numerical simulation of the circulation and tropical teleconnection mechanisms of a severe drought event (2012–2016) in northeastern brazil', *Climate Dynamics* pp. 1–15.

de Queiroz, L. P., Cardoso, D., Fernandes, M. F. and Moro, M. F. (2017), Diversity and evolution of flowering plants of the caatinga domain, *in* 'Caatinga', Springer, pp. 23–63.

Díaz, S. and Cabido, M. (1997), 'Plant functional types and ecosystem function in relation to global change', *Journal of vegetation science* **8**(4), 463–474.

Díaz, S., Kattge, J., Cornelissen, J. H., Wright, I. J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Prentice, I. C. et al. (2016), 'The global spectrum of plant form and function', *Nature* **529**(7585), 167–171.

Dikbas, F., Firat, M., Koc, A. C. and Gungor, M. (2012), 'Classification of precipitation series using fuzzy cluster method', *International journal of climatology* **32**(10), 1596–1603.

Donohue, R. J., McVICAR, T. R. and Roderick, M. L. (2009), 'Climate-related trends in australian vegetation cover as inferred from satellite observations, 1981–2006', *Global Change Biology* **15**(4), 1025–1039.

Donohue, R. J., Roderick, M. L., McVicar, T. R. and Farquhar, G. D. (2013), 'Impact of co2 fertilization on maximum foliage cover across the globe's warm, arid environments', *Geophysical Research Letters* **40**(12), 3031–3035.

Dunn, R. J., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., Lippmann, T., Allan, R., Aguilar, E., Barry, A. A. et al. (2020), 'Development of an updated global land in situ-based data set of temperature and precipitation extremes: Hadex3', *Journal of Geophysical Research: Atmospheres* **125**(16), e2019JD032263.

Durant, S., Pettorelli, N., Bashir, S., Woodroffe, R., Wacher, T., De Ornellas, P., Ransom, C., Abáigar, T., Abdelgadir, M., El Alqamy, H. et al. (2012), 'Forgotten biodiversity in desert ecosystems', *Science* **336**(6087), 1379–1380.

Engelbrecht, B. M., Comita, L. S., Condit, R., Kursar, T. A., Tyree, M. T., Turner, B. L. and Hubbell, S. P. (2007), 'Drought sensitivity shapes species distribution patterns in tropical forests', *Nature* **447**(7140), 80–82.

Eva, H. D., Belward, A. S., De Miranda, E. E., Di Bella, C. M., Gond, V., Huber, O., Jones, S., Sgrenzaroli, M. and Fritz, S. (2004), 'A land cover map of south america', *Global Change Biology* **10**(5), 731–744.

Fadlallah, B., Chen, B., Keil, A. and Príncipe, J. (2013), 'Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information', *Phys. Rev. E* **87**, 022911.
URL: *https://link.aps.org/doi/10.1103/PhysRevE.87.022911*

Farquhar, G. D. (1997), 'Carbon dioxide and vegetation', *Science* **278**(5342), 1411–1411.

Farrick, K. K. and Branfireun, B. A. (2015), 'Flowpaths, source water contributions and water residence times in a mexican tropical dry forest catchment', *Journal of Hydrology* **529**, 854–865.

Fatichi, S., Ivanov, V. Y. and Caporali, E. (2012), 'Investigating interannual variability of precipitation at the global scale: Is there a connection with seasonality?', *Journal of climate* **25**(16), 5512–5523.

Feldman, D. P., McTague, C. S. and Crutchfield, J. P. (2008), 'The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing', *Chaos: An Interdisciplinary Journal of Nonlinear Science* **18**(4), 043106.

Feng, X., Porporato, A. and Rodriguez-Iturbe, I. (2013), 'Changes in rainfall seasonality in the tropics', *Nature Climate Change* **3**(9), 811–815.

Fensholt, R., Horion, S., Tagesson, T., Ehammer, A., Grogan, K., Tian, F., Huber, S., Verbesselt, J., Prince, S. D., Tucker, C. J. et al. (2015), Assessment of vegetation trends in drylands from time series of earth observation data, *in* 'Remote Sensing Time Series', Springer, pp. 159–182.

Fensholt, R., Langanke, T., Rasmussen, K., Reenberg, A., Prince, S. D., Tucker, C., Scholes, R. J., Le, Q. B., Bondeau, A., Eastman, R. et al. (2012), 'Greenness in semi-arid areas across the globe 1981–2007–an earth observing satellite based analysis of trends and drivers', *Remote sensing of environment* **121**, 144–158.

Fensholt, R., Rasmussen, K., Kaspersen, P., Huber, S., Horion, S. and Swinnen, E. (2013), 'Assessing land degradation/recovery in the african sahel from long-term earth observation based primary productivity and precipitation relationships', *Remote Sensing* **5**(2), 664–686.

Ferreira, A. G. and da Silva Mello, N. G. (2005), 'Principais sistemas atmosféricos atuantes sobre a região nordeste do brasil e a influência dos oceanos pacífico e atlântico no clima da região', *Revista Brasileira de Climatologia* **1**(1).

Field, C. B., Barros, V., Stocker, T. F. and Dahe, Q. (2012), *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*, Cambridge University Press.

Figueiredo, K. V., Oliveira, M. T., Oliveira, A. F. M., Silva, G. C. and Santos, M. G. (2012), 'Epicuticular-wax removal influences gas exchange and water relations in the leaves of an exotic and native species from a brazilian semiarid region under induced drought stress', *Australian Journal of Botany* **60**(8), 685–692.

Fischer, E. M., Beyerle, U. and Knutti, R. (2013), 'Robust spatially aggregated projections of climate extremes', *Nature Climate Change* **3**(12), 1033–1038.

Foody, G. M. (2004), Sub-pixel methods in remote sensing, *in* 'Remote sensing image analysis: Including the spatial domain', Springer, pp. 37–49.

Foundation, F. S. (2010), 'GNU wget', http://www.gnu.org/software/wget/.
URL: *http://www.gnu.org/software/wget/*

Frieden, B. (2004), 'Science from fisher information: A unification–cambridge univ', *Press.2004* .

Garnier, S. (2018), *viridis: Default Color Maps from 'matplotlib'*. R package version 0.5.1.
URL: *https://CRAN.R-project.org/package=viridis*

Gaussen, H. (1973), 'International classification and mapping of vegetation', *Ecology and Conservation* **6**.

GCOS (2011), Systematic observation requirements for satellite-based products for climate. 2011 update supplemetnatl details to the satellite 39 based component og the implementation plan for the global observing system for climate in support of the unfccc (2010 update), *in* 'Tech. rep', World Meteorological Organisation (WMO).

(GCOS), G. C. O. S. (2003), 'Second report on the adequacy of the global observing systems for climate in support of the unfccc'.

Gentry, A. H. (1988), 'Changes in plant community diversity and floristic composition on environmental and geographical gradients', *Annals of the Missouri botanical garden* pp. 1–34.

Gessner, U., Naeimi, V., Klein, I., Kuenzer, C., Klein, D. and Dech, S. (2013), 'The relationship between precipitation anomalies and satellite-derived vegetation activity in central asia', *Global and Planetary Change* **110**, 74–87.

Giannini, A., Saravanan, R. and Chang, P. (2004), 'The preconditioning role of tropical atlantic variability in the development of the enso teleconnection: Implications for the prediction of nordeste rainfall', *Climate Dynamics* **22**(8), 839–855.

Gilliland, J. M. and Keim, B. D. (2018), 'Position of the south atlantic anticyclone and its impact on surface conditions across brazil', *Journal of Applied Meteorology and Climatology* **57**(3), 535–553.

Giraudoux, P. (2018), *pgirmess: Spatial Analysis and Data Mining for Field Ecologists*. R package version 1.6.9.
URL: *https://CRAN.R-project.org/package=pgirmess*

Givnish, T. J. (2002), 'Adaptive significance of evergreen vs. deciduous leaves: solving the triple paradox', *Silva fennica* **36**(3), 703–743.

Goedert, J. J., Hua, X., Yu, G. and Shi, J. (2014), 'Diversity and composition of the adult fecal microbiome associated with history of cesarean birth or appendectomy: analysis of the american gut project', *EBioMedicine* **1**(2-3), 167–172.

Gomes, H. B., Ambrizzi, T., da Silva, B. F. P., Hodges, K., Dias, P. L. S., Herdies, D. L., Silva, M. C. L. and Gomes, H. B. (2019), 'Climatology of easterly wave disturbances over the tropical south atlantic', *Climate Dynamics* **53**(3-4), 1393–1411.

Gomes, H. B., Ambrizzi, T., Herdies, D. L., Hodges, K. and Pontes da Silva, B. F. (2015), 'Easterly wave disturbances over northeast brazil: an observational analysis', *Advances in Meteorology* **2015**.

Griffin-Nolan, R. J., Bushey, J. A., Carroll, C. J., Challis, A., Chieppa, J., Garbowski, M., Hoffman, A. M., Post, A. K., Slette, I. J., Spitzer, D. et al. (2018), 'Trait selection and community weighting are key to understanding ecosystem responses to changing precipitation regimes', *Functional Ecology* **32**(7), 1746–1756.

Grosse, I., Bernaola-Galván, P., Carpena, P., Román-Roldán, R., Oliver, J. and Stanley, H. E. (2002), 'Analysis of symbolic sequences using the jensen-shannon divergence', *Physical Review E* **65**(4), 041905.

Group, U. N. E. M. (2011), 'Global drylands: a un system-wide response'.

Gudka, M., Davies, J., Poulsen, L., Schulte-Herbrüggen, B., MacKinnon, K., Crawhall, N., Henwood, W. D., Dudley, N. and Smith, J. (2014), 'Conserving dryland biodiversity: a future vision of sustainable dryland development', *Biodiversity* **15**(2-3), 143–147.

Guerreiro, M. J. S., Maia de Andrade, E., Abreu, I. and Lajinha, T. (2013), 'Long-term variation of precipitation indices in ceará state, northeast brazil', *International Journal of Climatology* **33**(14), 2929–2939.

Haberl, H., Erb, K.-H. and Krausmann, F. (2014), 'Human appropriation of net primary production: patterns, trends, and planetary boundaries', *Annual Review of Environment and Resources* **39**, 363–391.

Hastenrath, S. (1990), 'Prediction of northeast brazil rainfall anomalies', *Journal of Climate* **3**(8), 893–904.

Hastenrath, S. (2012), 'Exploring the climate problems of brazil's nordeste: a review', *Climatic Change* **112**(2), 243–251.

Hastenrath, S. and Greischar, L. (1993), 'Circulation mechanisms related to northeast brazil rainfall anomalies', *Journal of Geophysical Research: Atmospheres* **98**(D3), 5093–5102.

Haverd, V., Ahlström, A., Smith, B. and Canadell, J. G. (2017), 'Carbon cycle responses of semi-arid ecosystems to positive asymmetry in rainfall', *Global change biology* **23**(2), 793–800.

Haylock, M. R., Peterson, T. C., Alves, L. M., Ambrizzi, T., Anunciação, Y., Baez, J., Barros, V., Berlato, M., Bidegain, M., Coronel, G. et al. (2006), 'Trends in total and extreme south american rainfall in 1960–2000 and links with sea surface temperature', *Journal of climate* **19**(8), 1490–1512.

Heisler-White, J. L., Knapp, A. K. and Kelly, E. F. (2008), 'Increasing precipitation event size increases aboveground net primary productivity in a semi-arid grassland', *Oecologia* **158**(1), 129–140.

Hijmans, R. J. (2019), *raster: Geographic Data Analysis and Modeling*. R package version 3.0-7. **URL:** *https://CRAN.R-project.org/package=raster*

Hirsch, R. M. and Slack, J. R. (1984), 'A nonparametric trend test for seasonal data with serial dependence', *Water Resources Research* **20**(6), 727–732.

Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (2000), *Understanding robust and exploratory data analysis*, Wiley New York.

Hollander, M., Wolfe, D. A. and Chicken, E. (1999), 'Nonparametric statistical methods john wiley & sons', *New York* .

Hollander, M., Wolfe, D. A. and Chicken, E. (2013), *Nonparametric statistical methods*, Vol. 751, John Wiley & Sons.

Hollmann, R., Merchant, C. J., Saunders, R., Downy, C., Buchwitz, M., Cazenave, A., Chuvieco, E., Defourny, P., de Leeuw, G., Forsberg, R. et al. (2013), 'The esa climate change initiative: Satellite data records for essential climate variables', *Bulletin of the American Meteorological Society* **94**(10), 1541–1552.

Hothorn, T., Hornik, K., van de Wiel, M. A. and Zeileis, A. (2008), 'Implementing a class of permutation tests: The coin package', *Journal of Statistical Software* **28**(8), 1–23.

Hounsou-Gbo, G. A., Servain, J., Araujo, M., Martins, E. S., Bourlès, B. and Caniaux, G. (2016), 'Oceanic indices for forecasting seasonal rainfall over the northern part of brazilian northeast', *American Journal of Climate Change* **5**(02), 261–274.

Huxman, T. E., Smith, M. D., Fay, P. A., Knapp, A. K., Shaw, M. R., Loik, M. E., Smith, S. D., Tissue, D. T., Zak, J. C., Weltzin, J. F. et al. (2004), 'Convergence across biomes to a common rain-use efficiency', *Nature* **429**(6992), 651–654.

IBGE, B. (2004), 'Instituto brasileiro de geografia e estatística', *Mapa do Biomas do Brasil* **2004**.

Janzen, D. H. (1988), 'Tropical dry forests', *Biodiversity* pp. 130–137.

Jarvis, A., Reuter, H., Nelson, A. and Guevara, E. (2008), 'Cgiar-csi srtm 90m database', *Hole-filled SRTM for the globe Version* **4**.

Jia, M., Wang, Z., Wang, C., Mao, D. and Zhang, Y. (2019), 'A new vegetation index to detect periodically submerged mangrove forest using single-tide sentinel-2 imagery', *Remote Sensing* **11**(17), 2043.

Justino, F., Oliveira, E. C., Rodrigues, R. d., Gonçalves, P. H. L., Souza, P. J. O. P., Stordal, F., Marengo, J., da Silva, T. G., Delgado, R. C., da Silva Lindemann, D. et al. (2013), 'Mean and interannual variability of maize and soybean in brazil under global warming conditions', *American Journal of Climate Change* **2013**.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P. and Kessler, M. (2017), 'Climatologies at high resolution for the earth's land surface areas', *Scientific data* **4**, 170122.

Kartal, Ö., Schmid, M. W. and Grossniklaus, U. (2020), 'Cell type-specific genome scans of dna methylation divergence indicate an important role for transposable elements', *Genome biology* **21**(1), 1–29.

Kassambara, A. and Mundt, F. (2020), *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.
URL: *https://CRAN.R-project.org/package=factoextra*

Kaufman, L. and Rousseeuw, P. J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis.*, John Wiley.

Kayano, M. (2003), 'Low-level high-frequency modes in the tropical atlantic and their relation to precipitation in the equatorial south america', *Meteorology and Atmospheric Physics* **83**(3-4), 263–276.

Kharin, V. V., Zwiers, F. W., Zhang, X. and Hegerl, G. C. (2007), 'Changes in temperature and precipitation extremes in the ipcc ensemble of global coupled model simulations', *Journal of Climate* **20**(8), 1419–1444.

Kirschbaum, M., Eamus, D., Gifford, R., Roxburgh, S. and Sands, P. (2001), Definitions of some ecological terms commonly used in carbon accounting, *in* 'Net Ecosystem Exchange Workshop', pp. 18–20.

Kousky, V. E. (1979), 'Frontal influences on northeast brazil', *Monthly Weather Review* **107**(9), 1140–1153.

Kousky, V. E. (1980), 'Diurnal rainfall variation in northeast brazil', *Monthly Weather Review* **108**(4), 488–498.

Kruskal, W. H. and Wallis, W. A. (1952), 'Use of ranks in one-criterion variance analysis', *Journal of the American Statistical Association* **47**(260), 583–621.

Lamberti, P., Martin, M., Plastino, A. and Rosso, O. (2004), 'Intensive entropic non-triviality measure', *Physica A: Statistical Mechanics and its Applications* **334**(1-2), 119–131.

Leal, I. R., Da Silva, J. M. C., Tabarelli, M. and Lacher Jr, T. E. (2005), 'Changing the course of biodiversity conservation in the caatinga of northeastern brazil', *Conservation Biology* **19**(3), 701–706.

Liebmann, B., Camargo, S. J., Seth, A., Marengo, J. A., Carvalho, L. M., Allured, D., Fu, R. and Vera, C. S. (2007), 'Onset and end of the rainy season in south america in observations and the echam 4.5 atmospheric general circulation model', *Journal of Climate* **20**(10), 2037–2050.

Liebmann, B. and Mechoso, C. R. (2011), The south american monsoon system, *in* 'The Global Monsoon System: Research and Forecast', World Scientific, pp. 137–157.

Liu, B. Y., Zhu, Q., Riley, W. J., Zhao, L., Ma, H., Van Gordon, M. and Larsen, L. (2019), 'Using information theory to evaluate directional precipitation interactions over the west sahel region in observations and models', *Journal of Geophysical Research: Atmospheres* **124**(3), 1463–1473.

Lopez-Ruiz, R., Mancini, H. L. and Calbet, X. (1995), 'A statistical measure of complexity', *Physics letters A* **209**(5-6), 321–326.

Lu, X., Wang, L. and McCabe, M. F. (2016), 'Elevated co 2 as a driver of global dryland greening', *Scientific reports* **6**, 20716.

Lucena, D. B., Servain, J. and Gomes Filho, M. F. (2011), 'Rainfall response in northeast brazil from ocean climate variability during the second half of the twentieth century', *Journal of Climate* **24**(23), 6174–6184.

Ma, Y., Lu, M., Bracken, C. and Chen, H. (2020), 'Spatially coherent clusters of summer precipitation extremes in the tibetan plateau: Where is the moisture from?', *Atmospheric Research* **237**, 104841.

Machado, I. C., Barros, L. M. and Sampaio, E. V. (1997), 'Phenology of caatinga species at serra talhada, pe, northeastern brazil', *Biotropica* **29**(1), 57–68.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K. (2019), *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.0 — For new features, see the 'Changelog' file (in the package source).

Magrin, G., Travasso, M., Baethgen, W., Grondona, M., Giménez, A., Cunha, G., Castaño, J. and Rodriguez, G. (2007), Past and future changes in climate and their impacts on annual crops yield in south east south america, *in* 'Proc. TGICA Workshop', Vol. 1, Ipcc wg i, Nadi, Fiji.

Malhi, Y. (2012), 'The productivity, metabolism and carbon cycle of tropical forest vegetation', *Journal of Ecology* **100**(1), 65–75.

Mansur, R. and Barbosa, D. (2000), 'Physiological behavior in young plants of four trees species of caatinga submitted the two cycles of water stress', *Phyton* **68**, 97–106.

Marengo, J. A., Alves, L. M., Alvala, R., Cunha, A. P., Brito, S. and Moraes, O. L. (2018), 'Climatic characteristics of the 2010-2016 drought in the semiarid northeast brazil region', *Anais da Academia Brasileira de Ciências* **90**(2), 1973–1985.

Marengo, J. A., Alves, L. M., Soares, W. R., Rodriguez, D. A., Camargo, H., Riveros, M. P. and Pabló, A. D. (2013), 'Two contrasting severe seasonal extremes in tropical south america in 2012: flood in amazonia and drought in northeast brazil', *Journal of climate* **26**(22), 9137–9154.

Marengo, J. A., Ambrizzi, T., Da Rocha, R. P., Alves, L. M., Cuadra, S. V., Valverde, M. C., Torres, R. R., Santos, D. C. and Ferraz, S. E. (2010), 'Future change of climate in south america in the late twenty-first century: intercomparison of scenarios from three regional climate models', *Climate Dynamics* **35**(6), 1073–1097.

Marengo, J. A., Liebmann, B., Kousky, V. E., Filizola, N. P. and Wainer, I. C. (2001), 'Onset and end of the rainy season in the brazilian amazon basin', *Journal of Climate* **14**(5), 833–852.

Marengo, J., Chou, S., Torres, R., Giarolla, A., Alves, L. and Lyra, A. (2014), 'Climate change in central and south america: recent trends, future projections, and impacts on regional agriculture'.

Marengo, J., Espinoza, J., Ronchail, J. and Alves, L. (2016), 'Tropical south america east of the andes, in state of the climate in 2015', *Bulletin of the American Meteorological Society* **97**(8), S184–s185.

Martin, M., Plastino, A. and Rosso, O. (2003), 'Statistical complexity and disequilibrium', *Physics Letters A* **311**(2), 126–132.

Maurer, G. E., Hallmark, A. J., Brown, R. F., Sala, O. E. and Collins, S. L. (2020), 'Sensitivity of primary production to precipitation across the united states', *Ecology Letters* **23**(3), 527–536.

Melillo, J. M., McGuire, A. D., Kicklighter, D. W., Moore, B., Vorosmarty, C. J. and Schloss, A. L. (1993), 'Global climate change and terrestrial net primary production', *Nature* **363**(6426), 234–240.

Metz, B. et al. (n.d.), 'Intergovernmental panel on climate change (ipcc), 2007. climate change 2007: Mitigation', *Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* .

Meyer-Christoffer, A., Becker, A., Finger, P., Schneider, U. and Ziese, M. (2018), 'Gpcc climatology version 2018 at 0.25: Monthly land-surface precipitation climatology for every month and the total year from rain-gauges built on gts based and historical data', *Global Precipitation Climatology Centre (GPCC, http://gpcc. dwd. de/) at Deutscher Wetterdienst* .

Michiels, P., Gabriels, D. and Hartmann, R. (1992), 'Using the seasonal and temporal precipitation concentration index for characterizing the monthly rainfall distribution in spain', *Catena* **19**(1), 43–58.

Miles, L., Newton, A. C., DeFries, R. S., Ravilious, C., May, I., Blyth, S., Kapos, V. and Gordon, J. E. (2006), 'A global overview of the conservation status of tropical dry forests', *Journal of biogeography* **33**(3), 491–505.

Molaei, A. M., Zakeri, B. and Andargoli, S. M. H. (2019), 'Two-dimensional doa estimation for multi-path environments by accurate separation of signals using k-medoids clustering', *IET Communications* **13**(9), 1141–1147.

Moncunill, D. F. (2006), The rainfall trend over ceará and its implications, *in* '8ª Conferência Internacional de Meteorologia e Oceanografia do Hemisfério Sul, Foz do Iguaçu', pp. 315–323.

Mooney, H. A., Medina, E., Gentry, A. and Bullock, S. (1995), *Diversity and floristic composition of neotropical dry forestsSeasonally dry tropical forests*, Cambridge University Press.

Morgan, J. A., Milchunas, D. G., LeCain, D. R., West, M. and Mosier, A. R. (2007), 'Carbon dioxide enrichment alters plant community structure and accelerates shrub growth in the shortgrass steppe', *Proceedings of the National Academy of Sciences* **104**(37), 14724–14729.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. and Kent, J. (2000), 'Biodiversity hotspots for conservation priorities', *Nature* **403**(6772), 853–858.

Newingham, B. A., Vanier, C. H., Charlet, T. N., Ogle, K., Smith, S. D. and Nowak, R. S. (2013), 'No cumulative effect of 10 years of elevated co$_2$ on perennial plant biomass components in the mojave desert', *Global Change Biology* **19**(7), 2168–2181.

Nieto-Ferreira, R., Rickenbach, T. M. and Wright, E. A. (2011), 'The role of cold fronts in the onset of the monsoon season in the south atlantic convergence zone', *Quarterly Journal of the Royal Meteorological Society* **137**(657), 908–922.

Nobre, P., Marengo, J., Cavalcanti, I. F. d. A., Obregon, G., Barros, V., Camilloni, I., Campos, N. and Ferreira, A. G. (2006), 'Seasonal-to-decadal predictability and prediction of south american climate', *Journal of climate* **19**(23), 5988–6004.

Nobre, P. and Shukla, J. (1996), 'Variations of sea surface temperature, wind stress, and rainfall over the tropical atlantic and south america', *Journal of Climate* **9**(10), 2464–2479.

Nunez, M. N., Solman, S. A. and Cabré, M. F. (2009), 'Regional climate change experiments over southern south america. ii: Climate change scenarios in the late twenty-first century', *Climate Dynamics* **32**(7-8), 1081–1095.

O'Donnell, J. L., Kelly, R. P., Shelton, A. O., Samhouri, J. F., Lowell, N. C. and Williams, G. D. (2017), 'Spatial distribution of environmental dna in a nearshore marine habitat', *PeerJ* **5**, e3044.

Olivares, F., Plastino, A. and Rosso, O. A. (2012), 'Ambiguities in bandt-pompe's methodology for local entropic quantifiers', *Physica A: Statistical Mechanics and its Applications* **391**(8), 2518–2526.

Oliver, J. E. (1980), 'Monthly precipitation distribution: A comparative index', *The Professional Geographer* **32**(3), 300–309.

Olsson, L., Eklundh, L. and Ardö, J. (2005), 'A recent greening of the sahel–trends, patterns and potential causes', *Journal of Arid Environments* **63**(3), 556–566.

Ouédraogo, D.-Y., Fayolle, A., Gourlet-Fleury, S., Mortier, F., Freycon, V., Fauvet, N., Rabaud, S., Cornu, G., Bénédet, F., Gillet, J.-F. et al. (2016), 'The determinants of tropical forest deciduousness: disentangling the effects of rainfall and geology in central africa', *Journal of Ecology* **104**(4), 924–935.

Ouyang, G., Li, J., Liu, X. and Li, X. (2013), 'Dynamic characteristics of absence eeg recordings with multiscale permutation entropy analysis', *Epilepsy research* **104**(3), 246–252.

Öztürk, A. (1981), 'On the study of a probability distribution for precipitation totals', *Journal of Applied Meteorology* **20**(12), 1499–1505.

Parmentier, I., Malhi, Y., Senterre, B., Whittaker, R. J., ATDN, Alonso, A., Balinga, M. P., Bakayoko, A., Bongers, F., Chatelain, C. et al. (2007), 'The odd man out? might climate explain the lower tree α-diversity of african rain forests relative to amazonian rain forests?', *Journal of Ecology* **95**(5), 1058–1071.

Pennekamp, F., Iles, A. C., Garland, J., Brennan, G., Brose, U., Gaedke, U., Jacob, U., Kratina, P., Matthews, B., Munch, S., Novak, M., Palamara, G. M., Rall, B. C., Rosenbaum, B., Tabi, A., Ward, C., Williams, R., Ye, H. and Petchey, O. L. (2019), 'The intrinsic predictability of ecological time series and its potential to guide forecasting', *Ecological Monographs* **89**(2), e01359.

Pereira, M. P. S., Justino, F., Malhado, A. C. M., Barbosa, H. and Marengo, J. (2014), 'The influence of oceanic basins on drought and ecosystem dynamics in northeast brazil', *Environmental Research Letters* **9**(12), 124013.

Pezzi, L. and Cavalcanti, I. (2001), 'The relative importance of enso and tropical atlantic sea surface temperature anomalies for seasonal precipitation over south america: a numerical study', *Climate Dynamics* **17**(2–3), 205–212.

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y. et al. (2014), 'Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle', *Nature* **509**(7502), 600–603.

Prakash, S., Gairola, R. and Mitra, A. (2015), 'Comparison of large-scale global land precipitation from multisatellite and reanalysis products with gauge-based gpcc data sets', *Theoretical and Applied Climatology* **121**(1-2), 303–317.

Queiroz, R. T., Moro, M. F. and Loiola, M. I. B. (2015), 'Evaluating the relative importance of woody versus non-woody plants for alpha-diversity in a semiarid ecosystem in brazil', *Plant Ecology and Evolution* **148**(3), 361–376.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org*

Rao, V. B., Cavalcanti, I. F. and Hada, K. (1996), 'Annual variation of rainfall over brazil and water vapor characteristics over south america', *Journal of Geophysical Research: Atmospheres* **101**(D21), 26539–26551.

Reich, P. B. (2012), 'Key canopy traits drive forest productivity', *Proceedings of the Royal Society B: Biological Sciences* **279**(1736), 2128–2134.

Reid, W. V., Mooney, H. A., Cropper, A., Capistrano, D., Carpenter, S. R., Chopra, K., Dasgupta, P., Dietz, T., Duraiappah, A. K., Hassan, R. et al. (2005), *Ecosystems and human well-being-Synthesis: A report of the Millennium Ecosystem Assessment*, Island Press.

Ribeiro, H. V., Jauregui, M., Zunino, L. and Lenzi, E. K. (2017), 'Characterizing time series via complexity-entropy curves', *Physical Review E* **95**(6), 062106.

Rito, K. F., Arroyo-Rodríguez, V., Queiroz, R. T., Leal, I. R. and Tabarelli, M. (2017), 'Precipitation mediates the effect of human disturbance on the brazilian caatinga vegetation', *Journal of Ecology* **105**(3), 828–838.

Rodrigues, R. R., Haarsma, R. J., Campos, E. J. and Ambrizzi, T. (2011), 'The impacts of inter–el niño variability on the tropical atlantic and northeast brazil climate', *Journal of Climate* **24**(13), 3402–3422.

Rodrigues, R. R. and McPhaden, M. J. (2014), 'Why did the 2011–2012 la niña cause a severe drought in the brazilian northeast?', *Geophysical Research Letters* **41**(3), 1012–1018.

Ross, I., Misson, L., Rambal, S., Arneth, A., Scott, R., Carrara, A., Cescatti, A. and Genesio, L. (2012), 'How do variations in the temporal distribution of rainfall events affect ecosystem fluxes in seasonally water-limited northern hemisphere shrublands and forests', *Biogeosciences* **9**, 1007–1024.

Rosso, O. A., Carpi, L. C., Saco, P. M., Ravetti, M. G., Plastino, A. and Larrondo, H. A. (2012), 'Causality and the entropy–complexity plane: Robustness and missing ordinal patterns', *Physica A: Statistical Mechanics and its Applications* **391**(1-2), 42–55.

Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A. and Fuentes, M. A. (2007), 'Distinguishing noise from chaos', *Phys. Rev. Lett.* **99**, 154102.

Rubio de Casas, R., Willis, C. G., Pearse, W. D., Baskin, C. C., Baskin, J. M. and Cavender-Bares, J. (2017), 'Global biogeography of seed dormancy is determined by seasonality and seed size: a case study in the legumes', *New Phytologist* **214**(4), 1527–1536.

Ruggiero, P. G. C., Batalha, M. A., Pivello, V. R. and Meirelles, S. T. (2002), 'Soil-vegetation relationships in cerrado (brazilian savanna) and semideciduous forest, southeastern brazil', *Plant Ecology* **160**(1), 1–16.

Sampaio, E. (1995), 'Overview of the brazilian caatinga', *Seasonally dry tropical forests* pp. 35–63.

Santos e Silva, C. M., Silva, A., Oliveira, P. and Lima, K. (2014), 'Dynamical downscaling of the precipitation in northeast brazil with a regional climate model during contrasting years', *Atmospheric Science Letters* **15**(1), 50–57.

Santos, M. G., Oliveira, M. T., Figueiredo, K. V., Falcao, H. M., Arruda, E. C., Almeida-Cortez, J., Sampaio, E. V., Ometto, J. P., Menezes, R. S., Oliveira, A. F. et al. (2014), 'Caatinga, the brazilian dry tropical forest: can it tolerate climate changes?', *Theoretical and Experimental Plant Physiology* **26**(1), 83–99.

Satyanarayana, P. and Srinivas, V. (2011), 'Regionalization of precipitation in data sparse areas using large scale atmospheric variables–a fuzzy clustering approach', *Journal of Hydrology* **405**(3-4), 462–473.

Saunders, K., Stephenson, A. and Karoly, D. (2020), 'A regionalisation approach for rainfall based on extremal dependence', *Extremes* pp. 1–26.

Schubert, E. and Rousseeuw, P. J. (2018), 'Faster k-medoids clustering: Improving the pam, clara, and clarans algorithms', *CoRR* **abs/1810.05691**.

Schucknecht, A., Erasmi, S., Niemeyer, I. and Matschullat, J. (2013), 'Assessing vegetation variability and trends in north-eastern brazil using avhrr and modis ndvi time series', *European Journal of Remote Sensing* **46**(1), 40–59.

Schulz, K., Voigt, K., Beusch, C., Almeida-Cortez, J. S., Kowarik, I., Walz, A. and Cierjacks, A. (2016), 'Grazing deteriorates the soil carbon stocks of caatinga forest ecosystems in brazil', *Forest Ecology and Management* **367**, 62–70.

Schwartz, N. B., Lintner, B. R., Feng, X. and Powers, J. S. (2020), 'Beyond map: A guide to dimensions of rainfall variability for tropical ecology', *Biotropica* **52**(6), 1319–1332.

Sen, P. K. (1968), 'Estimates of the regression coefficient based on kendall's tau', *Journal of the American Statistical Association* **63**(324), 1379–1389.

Sessa, R. and Dolman, H. (2008), 'Terrestrial essential climate variables for climate change assessment, mitigation and adaptation (gtos 52)'.

Seth, A., Rojas, M. and Rauscher, S. A. (2010), 'Cmip3 projected changes in the annual cycle of the south american monsoon', *Climatic Change* **98**(3-4), 331–357.

Shannon, C. E. (1948), 'A mathematical theory of communication', *Bell System Technical Journal* **27**(3), 379–423.

Siegel, S. (1988), 'Castellan. nonparametric statistics for the social sciences'.

Sippel, S., Lange, H., Mahecha, M. D., Hauhs, M., Bodesheim, P., Kaminski, T., Gans, F. and Rosso, O. A. (2016), 'Diagnosing the dynamics of observed and simulated ecosystem gross primary productivity with time causal information theory quantifiers', *Plos One* **11**(10), 1–29.

Smith, M. D. (2011), 'The ecological role of climate extremes: current understanding and future prospects', *Journal of Ecology* **99**(3), 651–655.

Soares, W. R. and Marengo, J. A. (2009), 'Assessments of moisture fluxes east of the andes in south america in a global warming scenario', *International Journal of Climatology: A Journal of the Royal Meteorological Society* **29**(10), 1395–1414.

Sorensen, L. (2007), 'A spatial analysis approach to the global delineation of dryland areas of relevance to the cbd programme of work on dry and sub-humid lands', *UK, Cambridge* .

Souza, B. D., Meiado, M. V., Rodrigues, B. M. and Santos, M. G. (2010), 'Water relations and chlorophyll fluorescence responses of two leguminous trees from the caatinga to different watering regimes', *Acta Physiologiae Plantarum* **32**(2), 235–244.

Strasser, H. and Weber, C. (1999), 'On the asymptotic theory of permutation statistics'.

Suding, K. N., Lavorel, S., Chapin Iii, F., Cornelissen, J. H., DIAz, S., Garnier, E., Goldberg, D., Hooper, D. U., Jackson, S. T. and Navas, M.-L. (2008), 'Scaling environmental change through the community-level: a trait-based response-and-effect framework for plants', *Global Change Biology* **14**(5), 1125–1140.

Sun, L., Li, H., Ward, M. N. and Moncunill, D. F. (2007), 'Climate variability and corn yields in semiarid ceará, brazil', *Journal of Applied Meteorology and Climatology* **46**(2), 226–240.

Team, R. C. (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Thompson, C. S. (1984), 'Homogeneity analysis of rainfall series: An application of the use of a realistic rainfall model', *Journal of Climatology* **4**(6), 609–619.

Tibshirani, R., Guenther, W. and Hastie., T. (2001), 'Estimating the number of clusters in a data set via the gap statistic', *Journal of the Royal Statistical Society Series B* .

Torres, R. R. and Ferreira, N. J. (2011), 'Case studies of easterly wave disturbances over northeast brazil using the eta model', *Weather and Forecasting* **26**(2), 225–235.

Van Belle, G. and Hughes, J. P. (1984), 'Nonparametric tests for trend in water quality', *Water resources research* **20**(1), 127–136.

Van der Laan, M., Pollard, K. and Bryan, J. (2003), 'A new partitioning around medoids algorithm', *Journal of Statistical Computation and Simulation* **73**(8), 575–584.

Vasseur, D. A. and Yodzis, P. (2004), 'The color of environmental noise', *Ecology* **85**(4), 1146–1152.

Vieira, R. d. S. P., Tomasella, J., Alvalá, R., Sestini, M., Affonso, A., Rodriguez, D., Barbosa, A., Cunha, A., Valles, G., Crepani, E. et al. (2015), 'Identifying areas susceptible to desertification in the brazilian northeast', *Solid Earth* **6**(1), 347–360.

Wang, C. and Fiedler, P. C. (2006), 'Enso variability and the eastern tropical pacific: A review', *Progress in Oceanography* **69**(2-4), 239–266.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H., Diemer, M. et al. (2004), 'The worldwide leaf economics spectrum', *Nature* **428**(6985), 821–827.

Yamazaki, Y. and Rao, V. B. (1977), 'Tropical cloudiness over the south atlantic ocean', *Journal of the Meteorological Society of Japan. Ser. II* **55**(2), 205–207.

Zhang, L., Li, H., Liu, D., Fu, Q., Li, M., Faiz, M. A., Khan, M. I. and Li, T. (2019), 'Identification and application of the most suitable entropy model for precipitation complexity measurement', *Atmospheric research* **221**, 88–97.

Zhang, W., Brandt, M., Tong, X., Tian, Q. and Fensholt, R. (2018), 'Impacts of the seasonal distribution of rainfall on vegetation productivity across the sahel', *Biogeosciences* **15**(1), 319.

Zhang, Y., Susan Moran, M., Nearing, M. A., Ponce Campos, G. E., Huete, A. R., Buda, A. R., Bosch, D. D., Gunter, S. A., Kitchen, S. G., Henry McNab, W. et al. (2013), 'Extreme precipitation patterns and reductions of terrestrial ecosystem production across biomes', *Journal of Geophysical Research: Biogeosciences* **118**(1), 148–157.

Zilli, M. T., Carvalho, L. M. and Lintner, B. R. (2019), 'The poleward shift of south atlantic convergence zone in recent decades', *Climate Dynamics* **52**(5-6), 2545–2563.

Zunino, L., Soriano, M. C., Fischer, I., Rosso, O. A. and Mirasso, C. R. (2010), 'Permutation-information-theory approach to unveil delay dynamics from time-series analysis', *Physical Review E* **82**(4), 046212.

Supplements

Table 5.1: Cluster comparisons with Wilcoxon-Mann-Whitney Test, exact p.values

	1-2	1-3	2-3
MAP	0.00e+00	0.00e+00	1.96e-02
CV_p	4.36e-09	1.63e-05	6.60e-17
PCI1	0.00e+00	0.00e+00	3.27e-18
PCI2	1.22e-12	1.49e-06	3.91e-22
PCIΔ	6.12e-37	3.17e-101	1.54e-11
TS_p	7.16e-11	8.88e-07	1.35e-08
PE_p	1.90e-87	3.33e-123	3.29e-01
MAMF	0.00e+00	9.29e-12	8.18e-16
CV_f	1.74e-04	2.97e-05	0.00e+00
TS_f	7.04e-01	1.59e-06	4.80e-05
PE_f	2.44e-52	1.45e-15	0.00e+00
ρ_t	4.40e-11	1.15e-02	4.32e-06
R_t^2	3.40e-14	2.19e-13	2.35e-01

Table 5.2: Cluster comparisons with Watson-Wheeler Test for circular data, exact p.values

	k_3N-k_3E	k_3N-k_3S	k_3E-k_3S
ONS_p	1.39e-37	1.39e-123	1.84e-82
$PEAK_p$	5.66e-29	1.72e-102	2.66e-59
ONS_f	4.49e-09	1.42e-40	5.26e-35
$PEAK_f$	1.03e-26	1.03e-66	1.15e-20

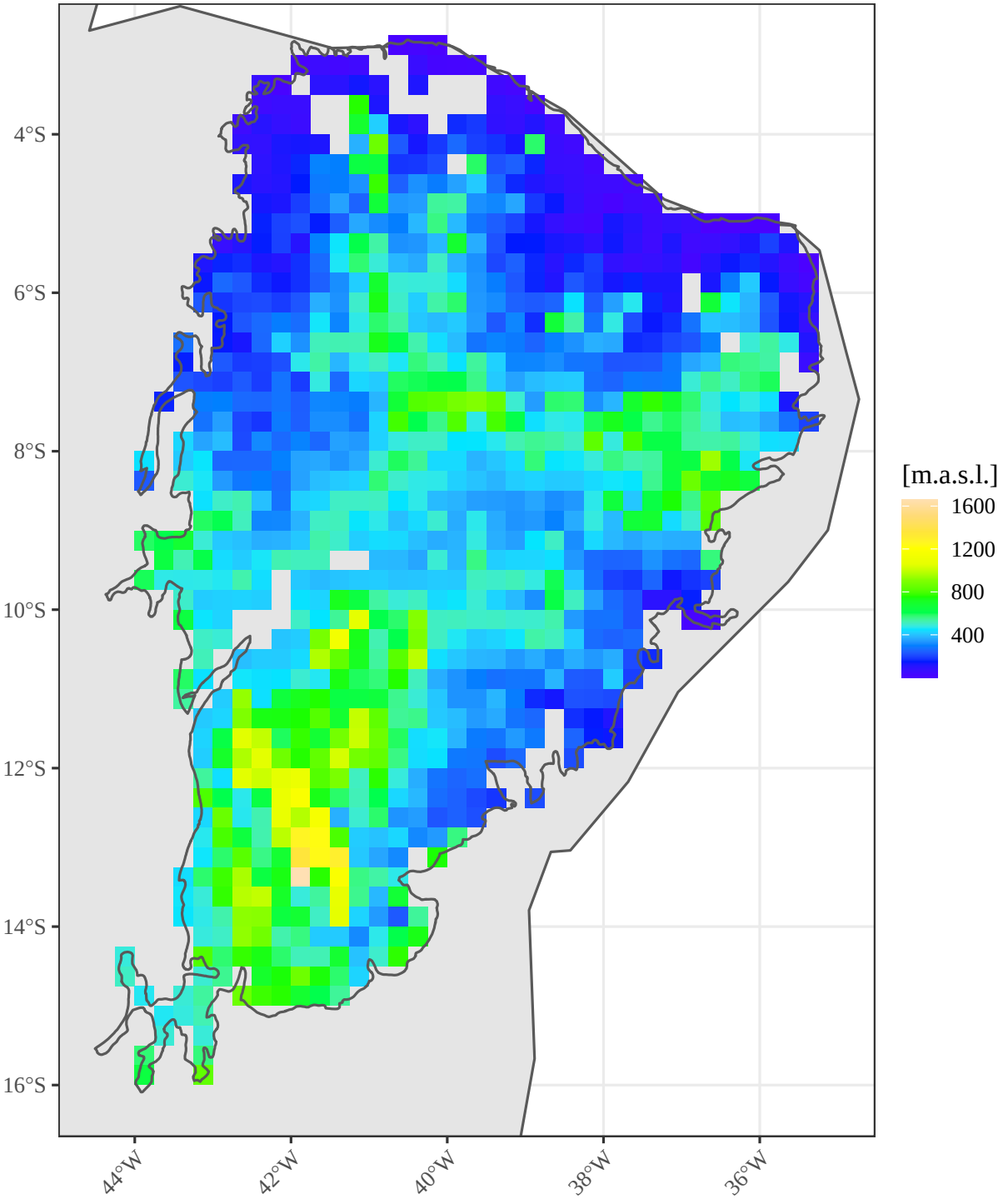

Figure 5.1: SRTM 90m altitude (Jarvis et al., 2008) aggregated to 0.25° resolution

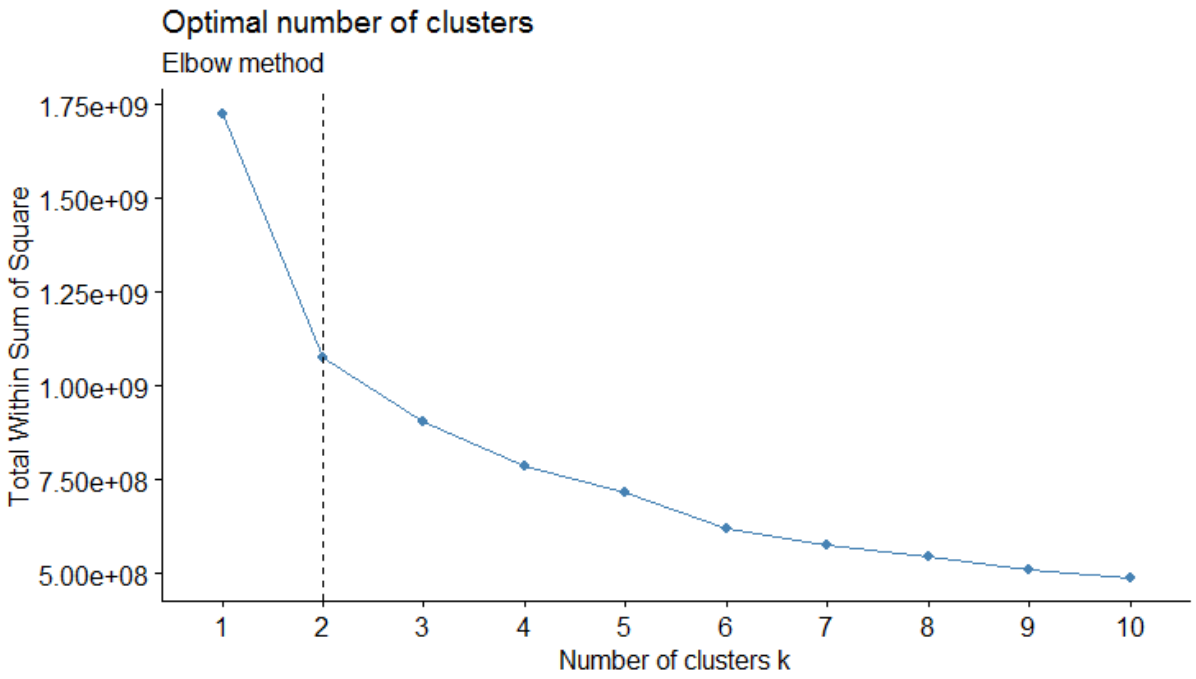

Figure 5.2: Screeplot plot.

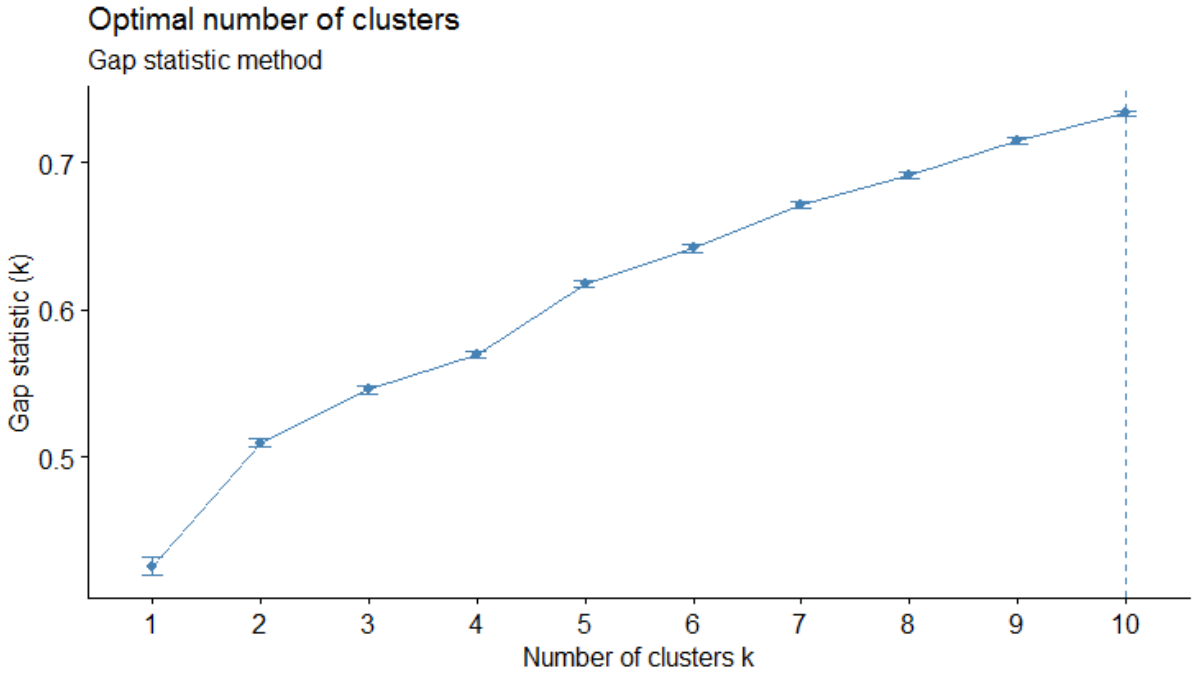

Figure 5.3: Gap Statistic (Tibshirani et al., 2001).

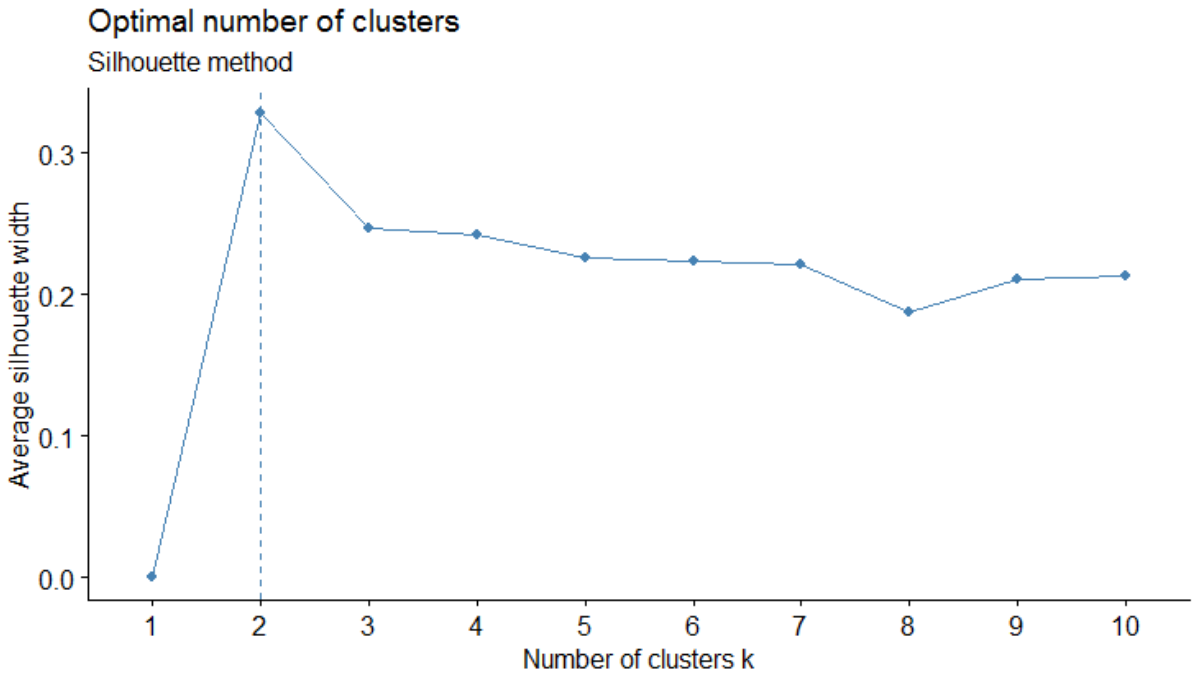

Figure 5.4: Silhuette plot.

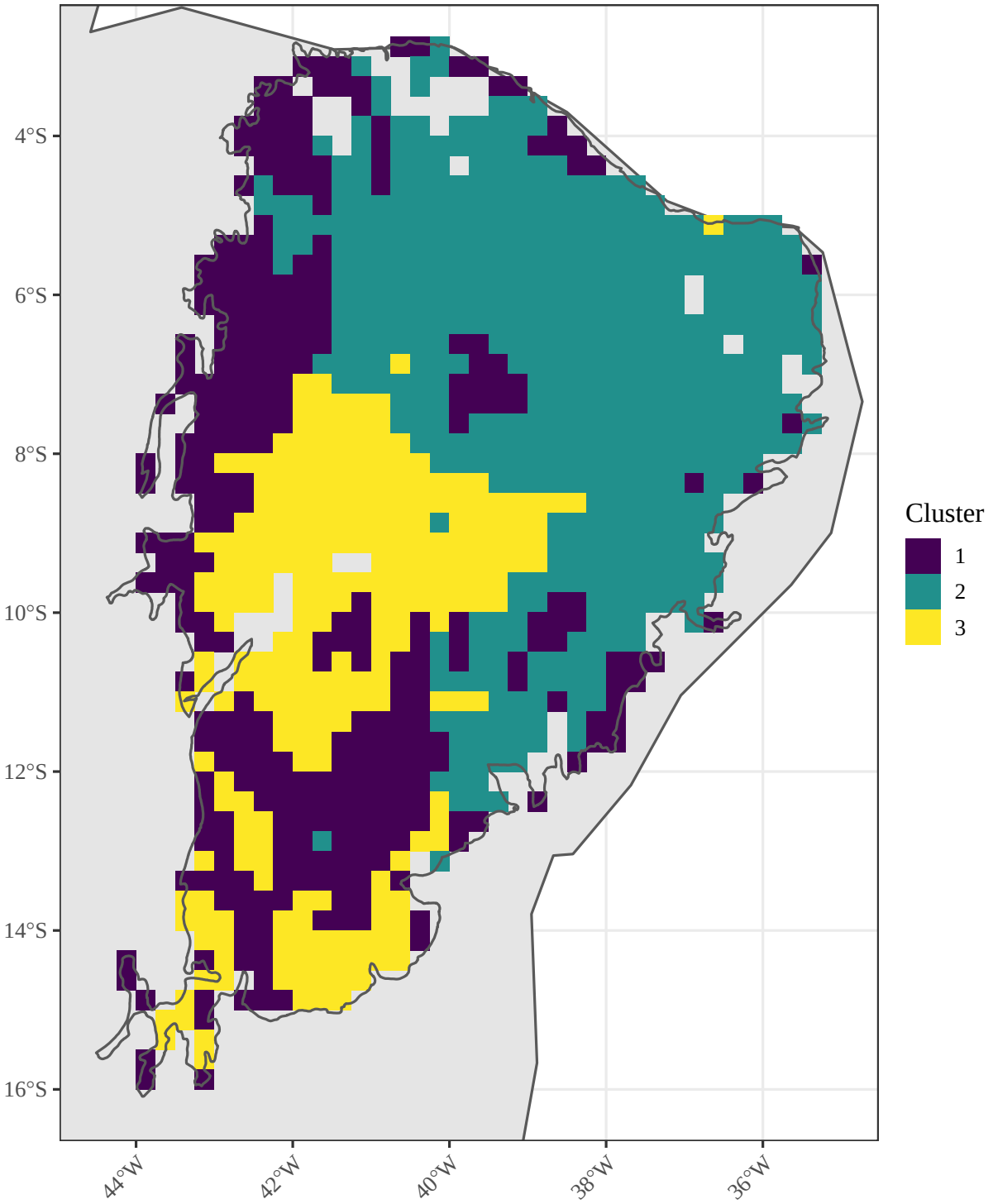

Figure 5.5: Spatial representation of the results of clustering FPAR data with $k = 3$.

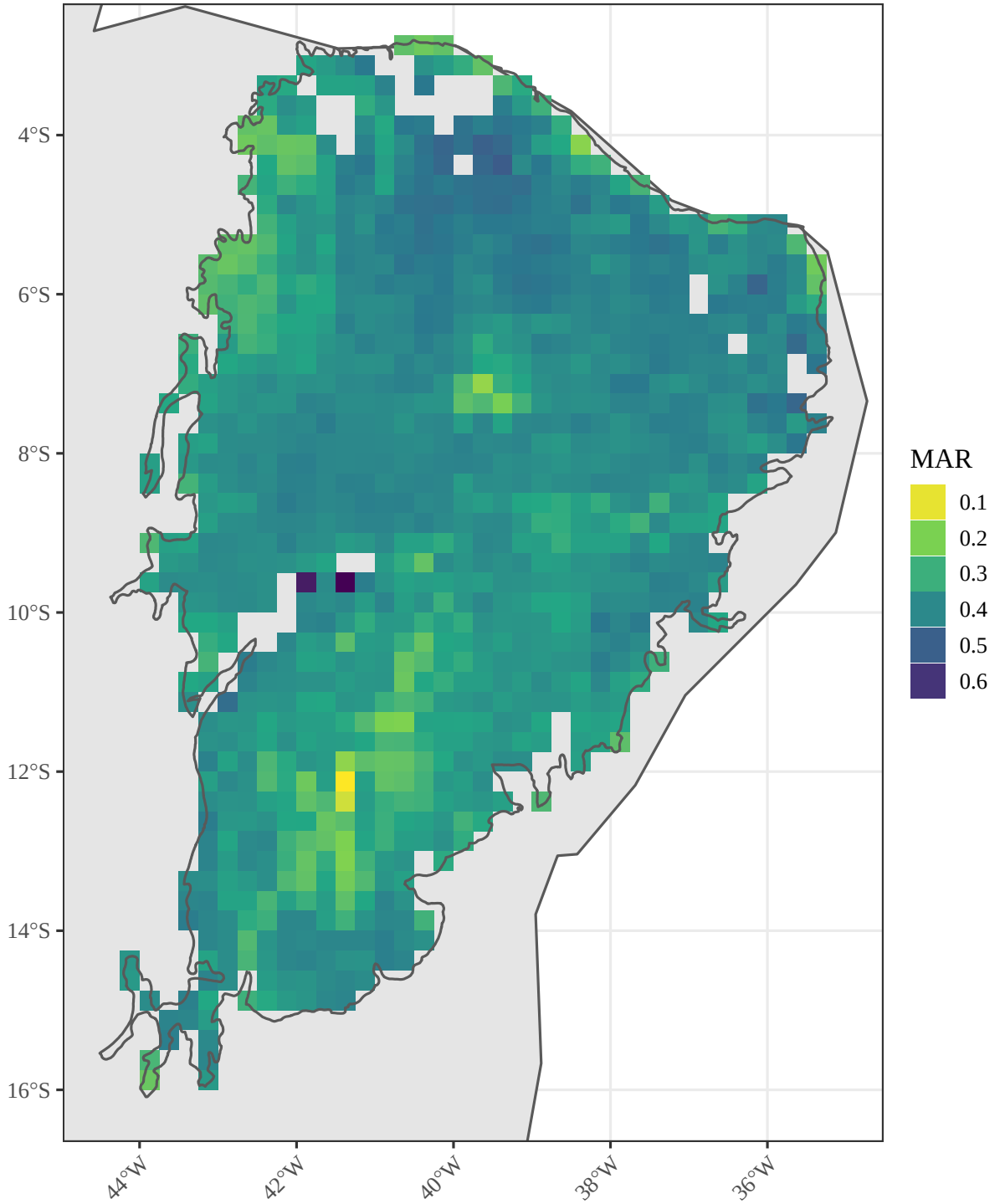

Figure 5.6: Mean annual range (MAR) of FPAR data 1982-2016.

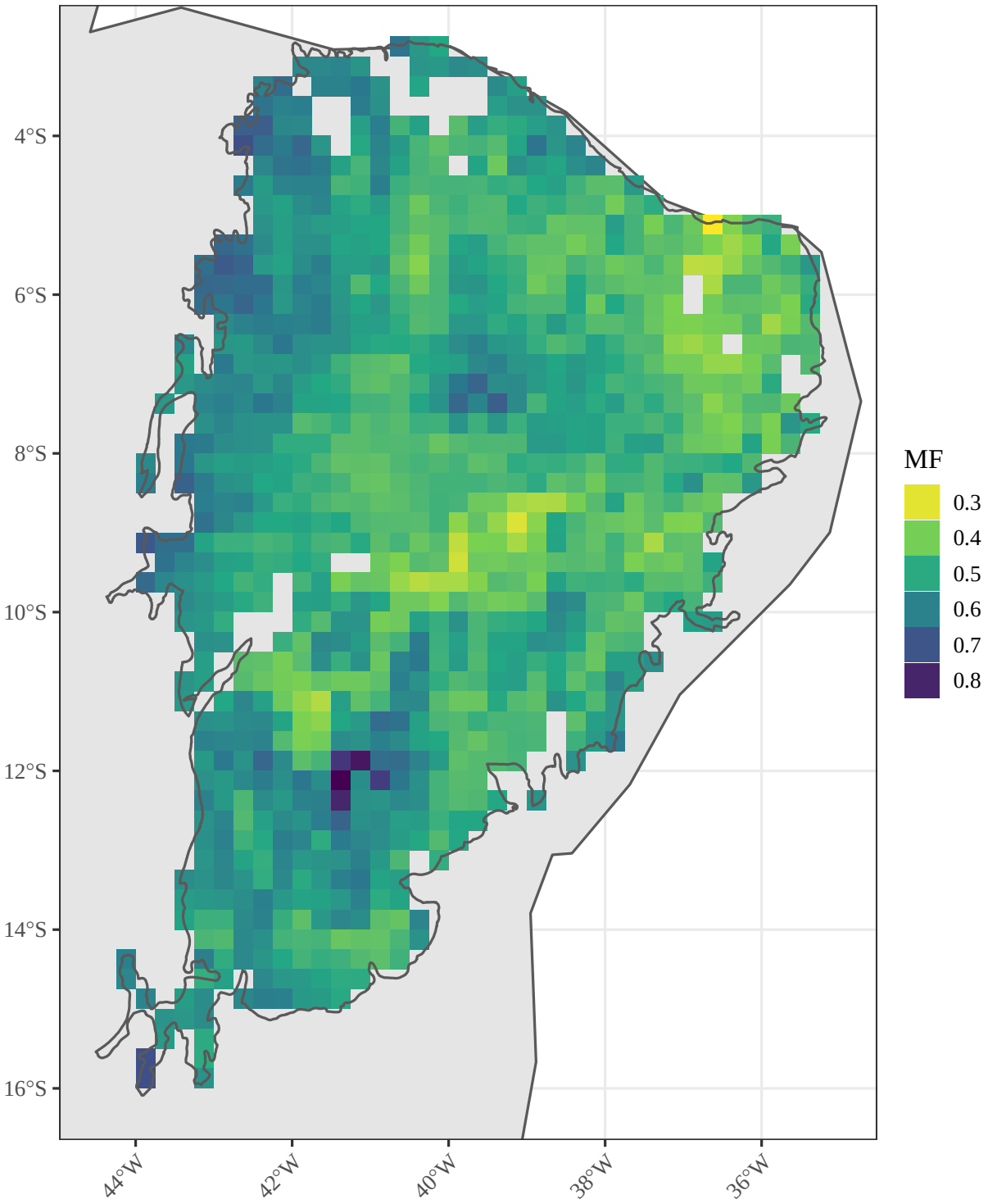

Figure 5.7: Mean (MF) of FPAR data 1982-2016.

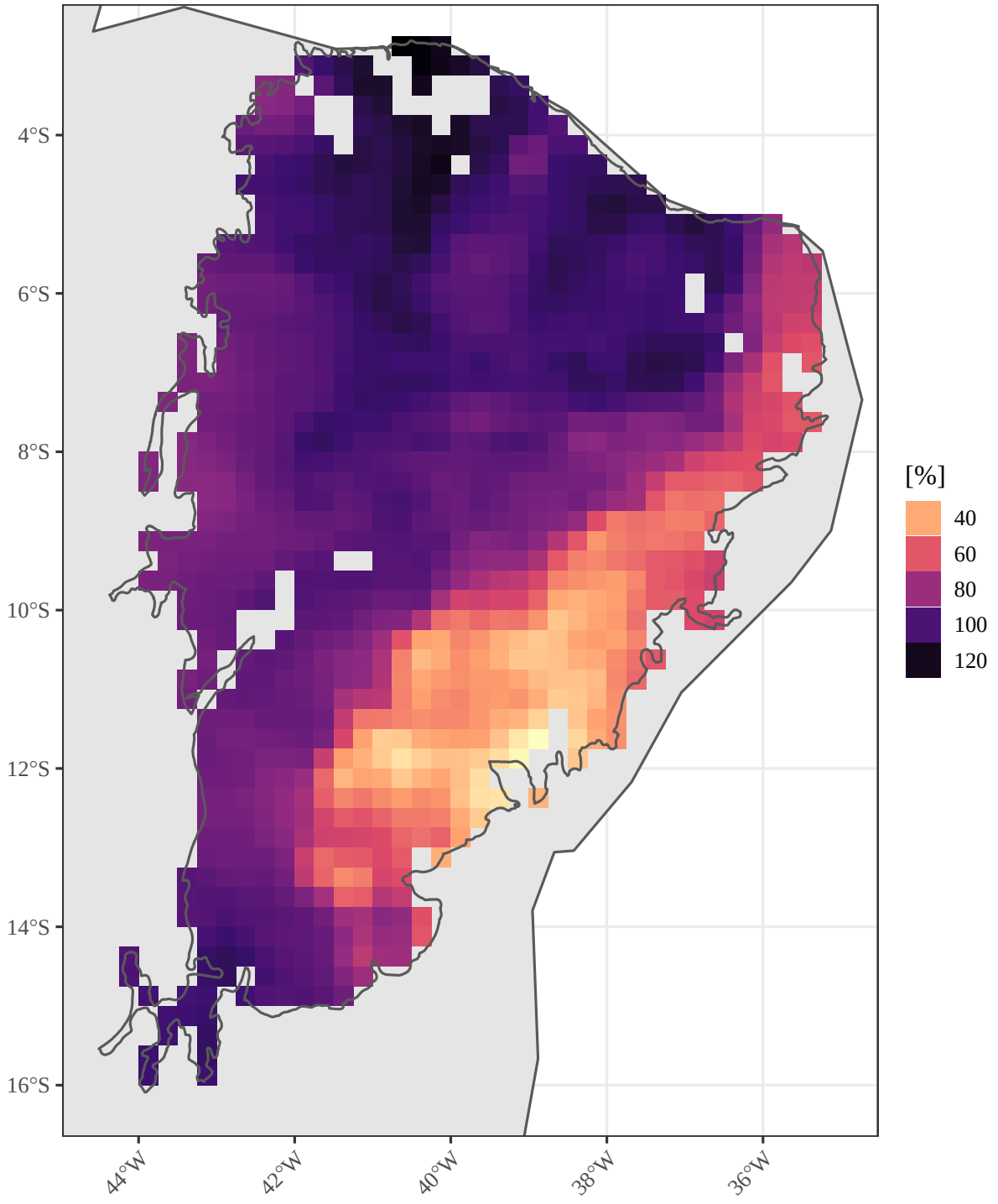

Figure 5.8: Precipitation relative seasonality (CV of monthly values of the average year) 1982-2016.

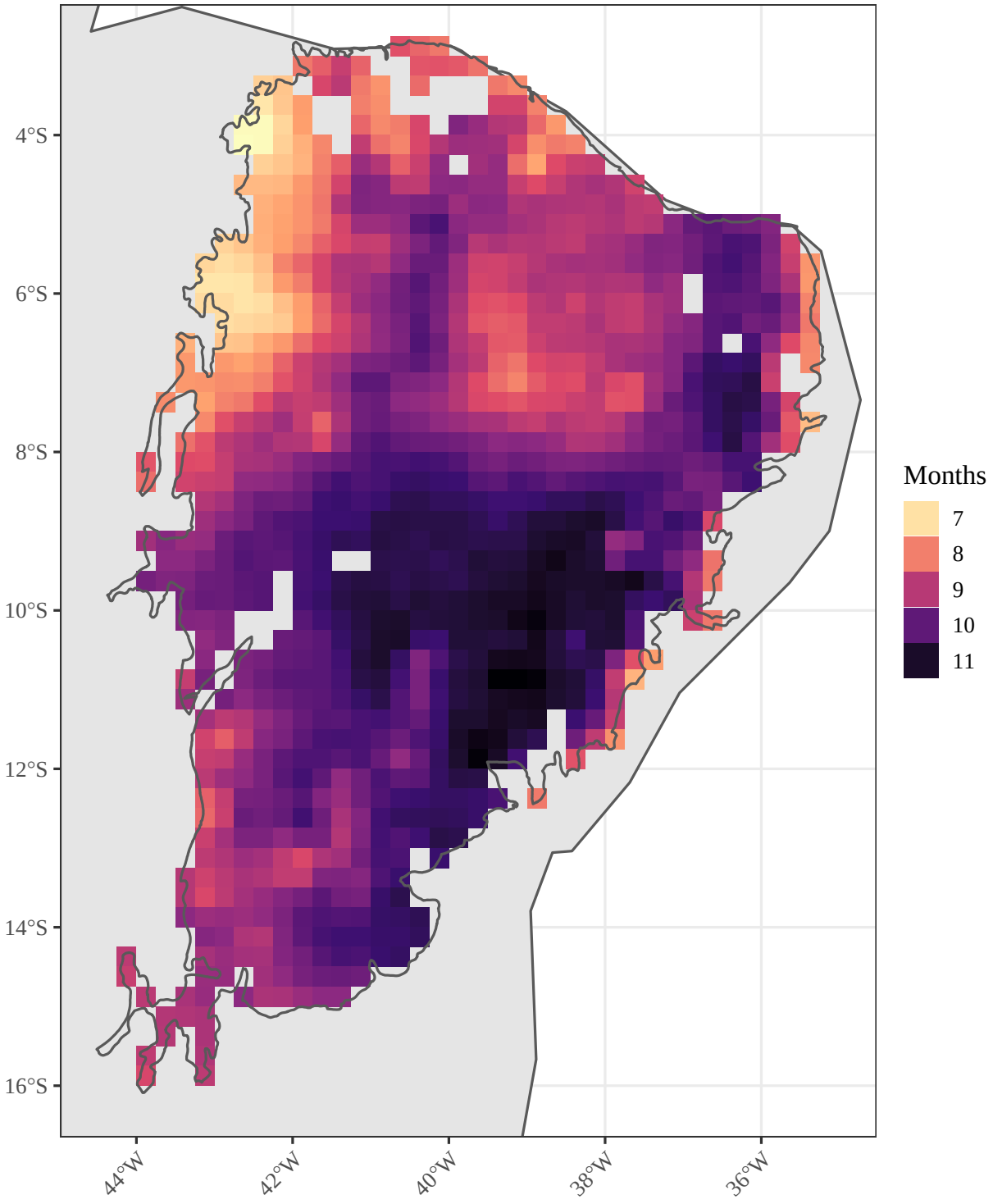

Figure 5.9: Mean dry season length 1982-2016 in months. $Dryseason = months < 100mm$ *sensu* (Aragão et al., 2007; Schwartz et al., 2020).

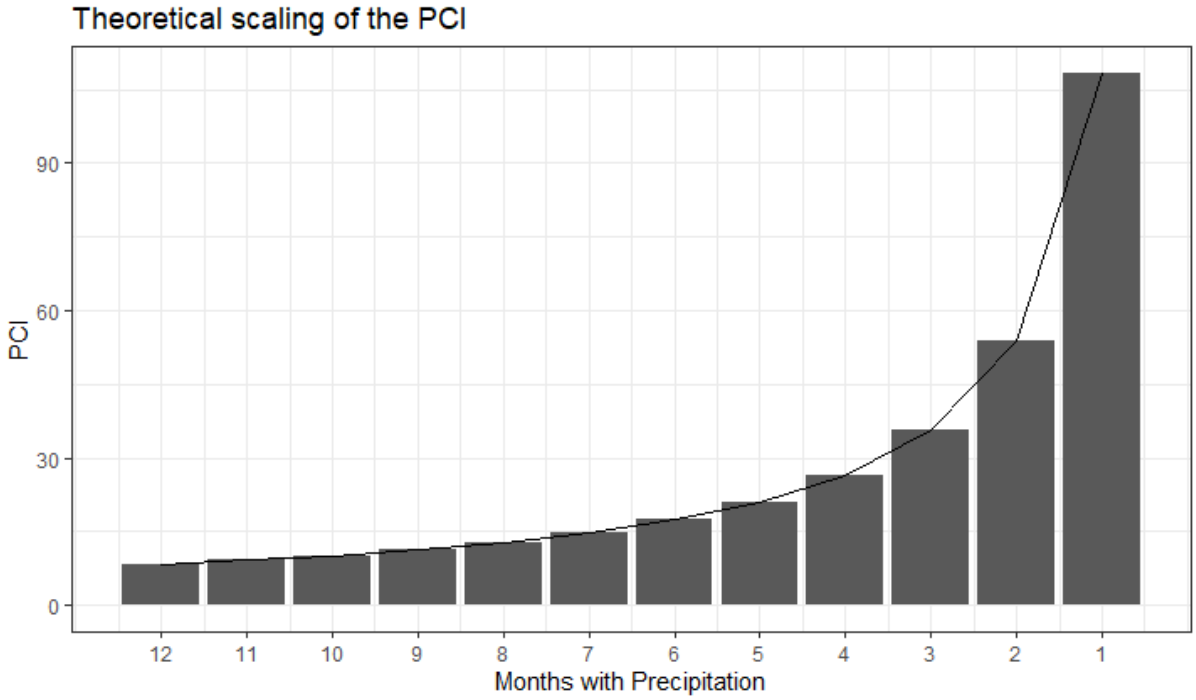

Figure 5.10: Theoretical scaling of the PCI. Low sensitivity for distributions without zero precipitation months (Oliver, 1980; Michiels et al., 1992).

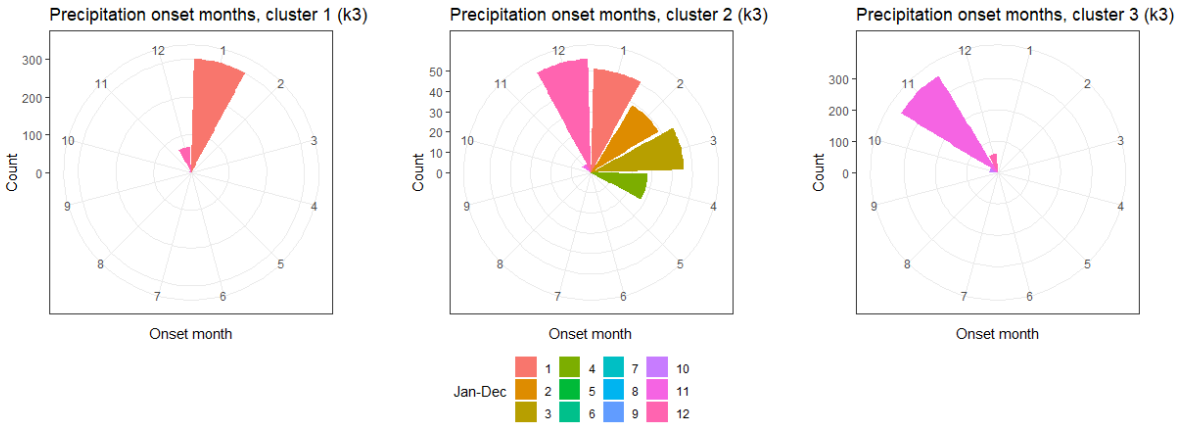

Figure 5.11: Rose diagrams of ONS_p distribution of the precipitation clusters 1982-2016.

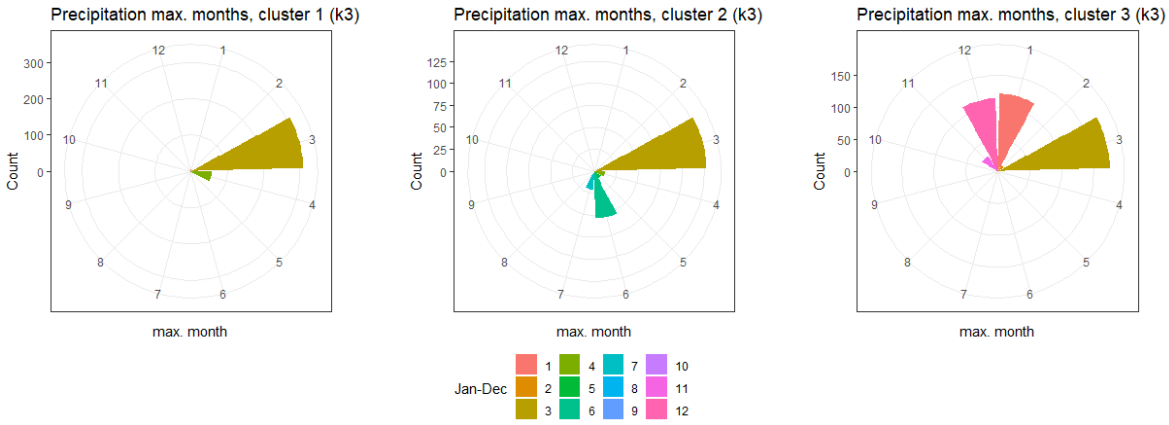

Figure 5.12: Rose diagrams of $PEAK_p$ distribution of the precipitation clusters 1982-2016.

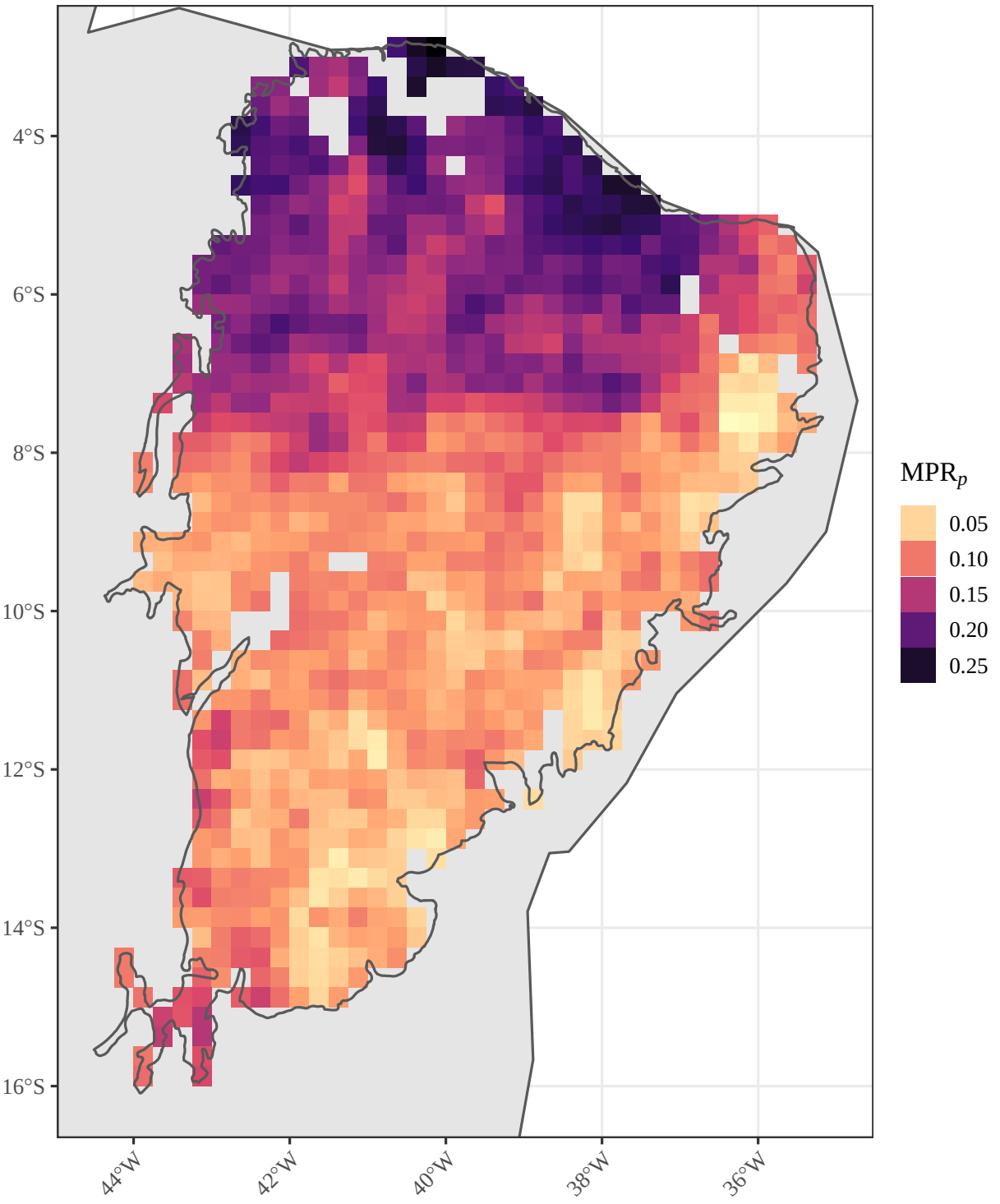

Figure 5.13: Spatial distribution of MPR$_p$. Low values indicate less structured/ordered, less complex time series.

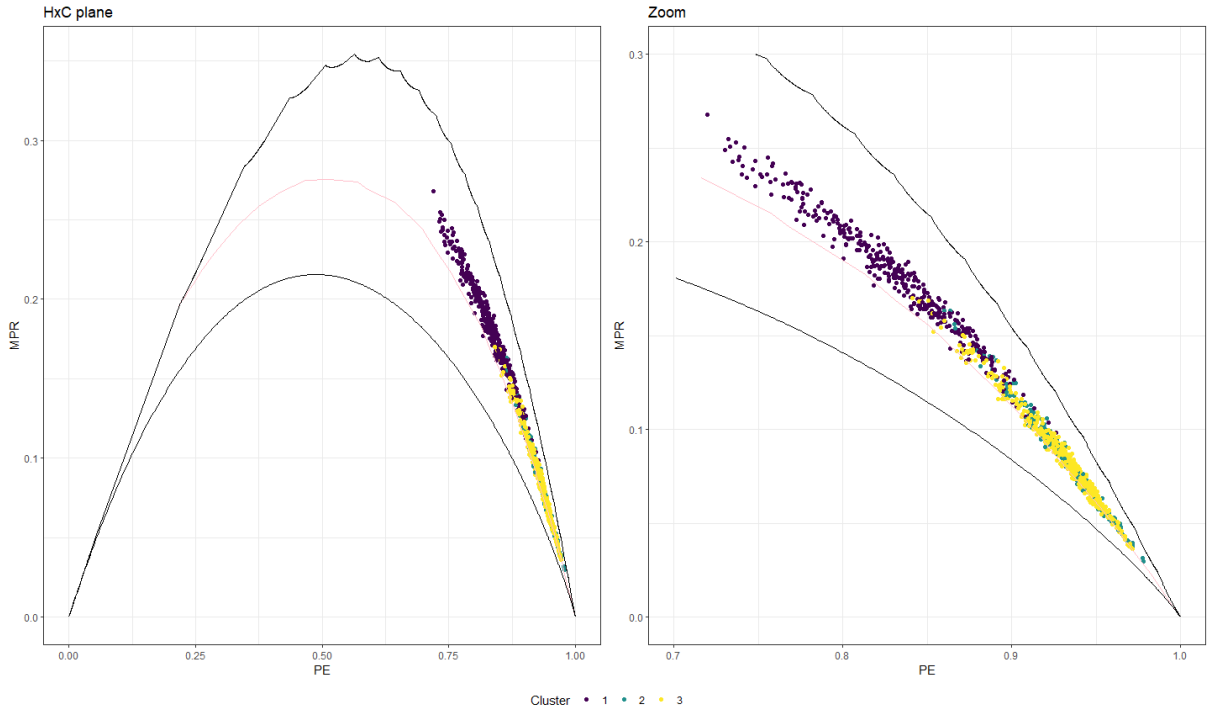

Figure 5.14: Representation of the precipitation time series in the Entropy-Complexity plane (HxC) (Rosso et al., 2012). Each dot represents one time series, colors according to precipitation cluster memberships. Black lines are the theoretical limits, the pink line represents k-noise ("colored" noise, $1/f^{\beta}$ exponents from 0-4.

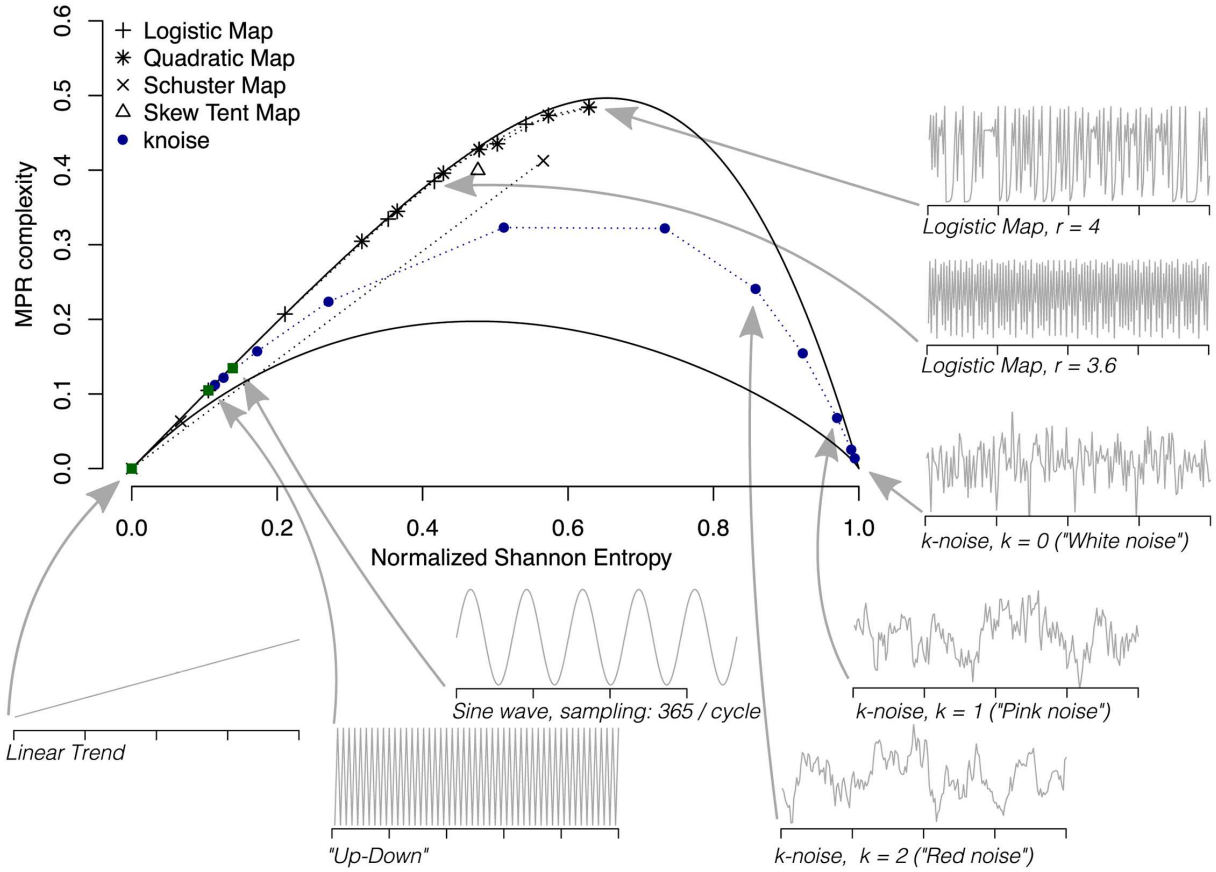

Figure 5.15: Schematic representation of the classification of time series in the Entropy-Complexity plane (HxC) (Rosso et al., 2012), taken from Sippel et al. (2016).

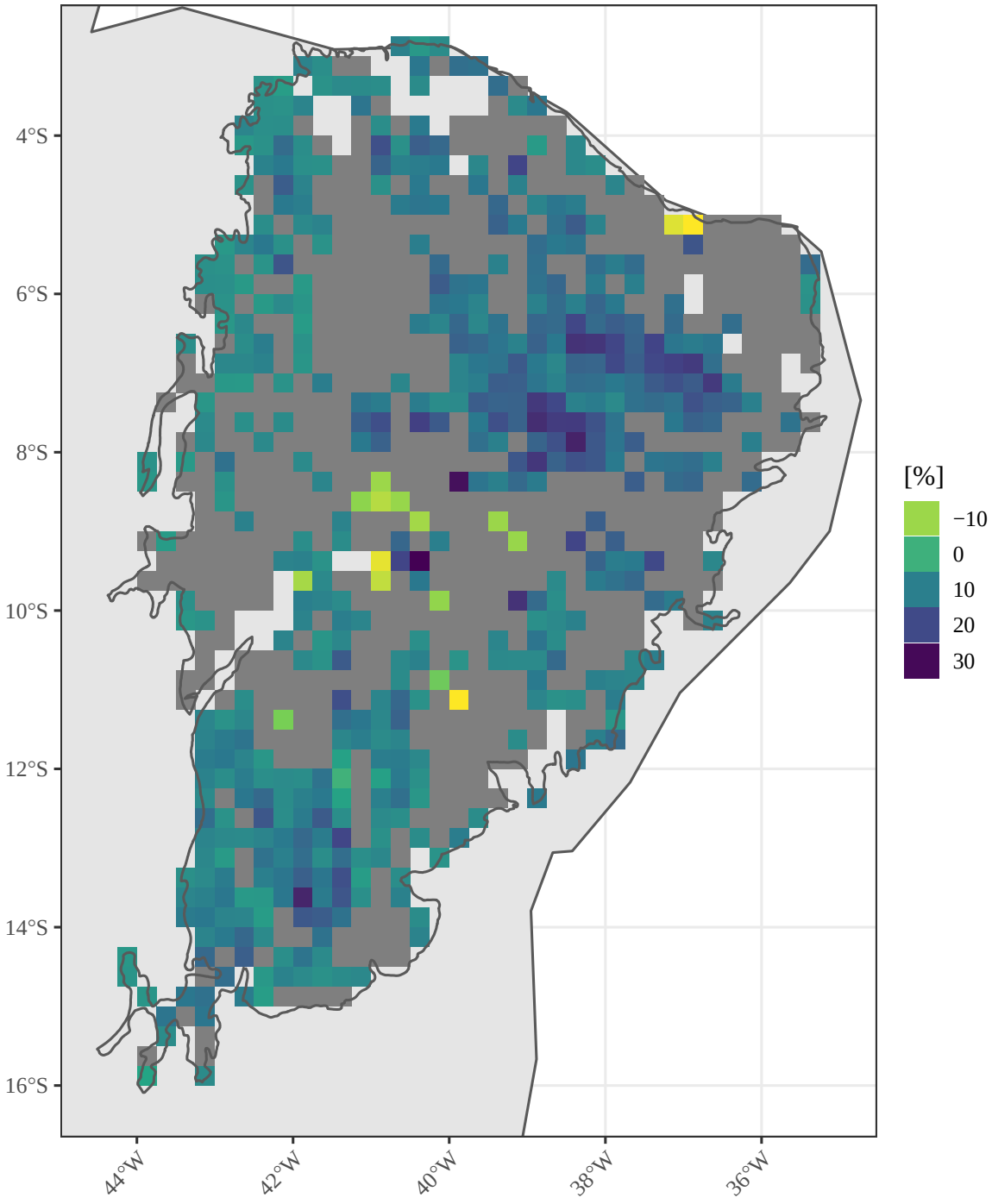

Figure 5.16: FPAR trends compared to baseline average 1982-2010. Comparison to Donohue et al. (2013).

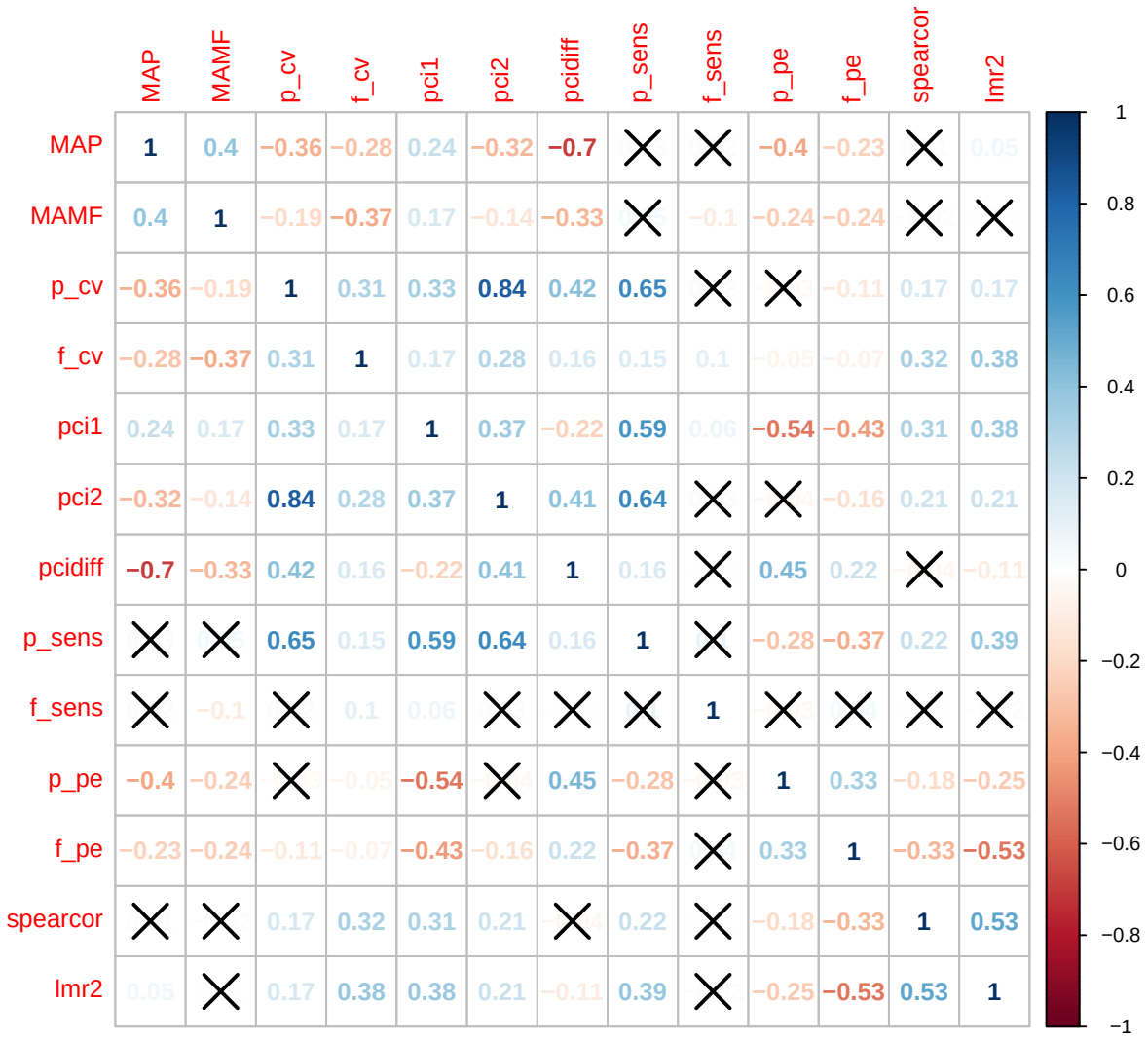

Figure 5.17: Spatial spearman rank correlation ρ_S of the parameters ($p < 0.05$).

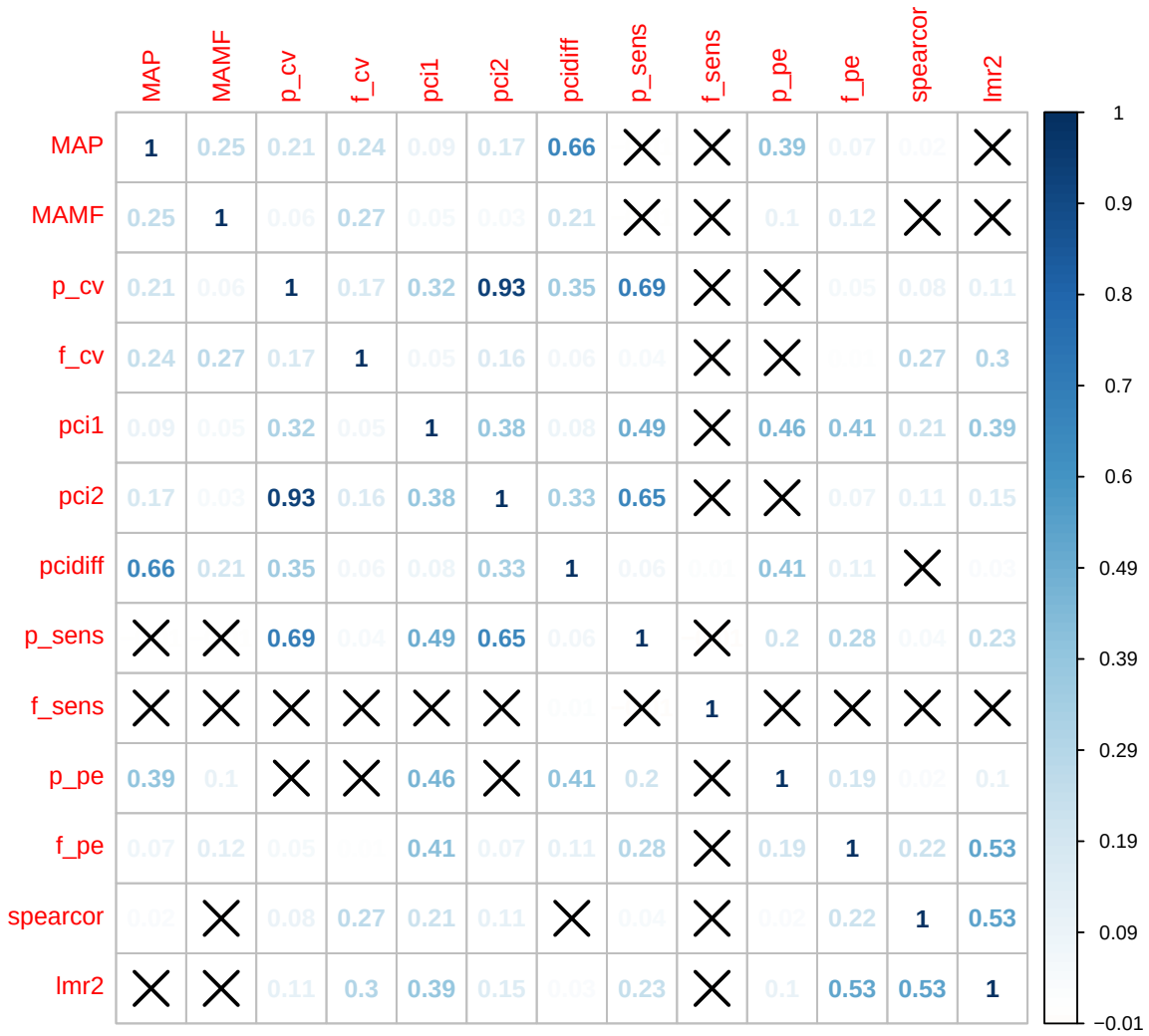

Figure 5.18: Coefficients of determination for spatial regressions $adj.R^2$ of the parameters ($p < 0.05$).

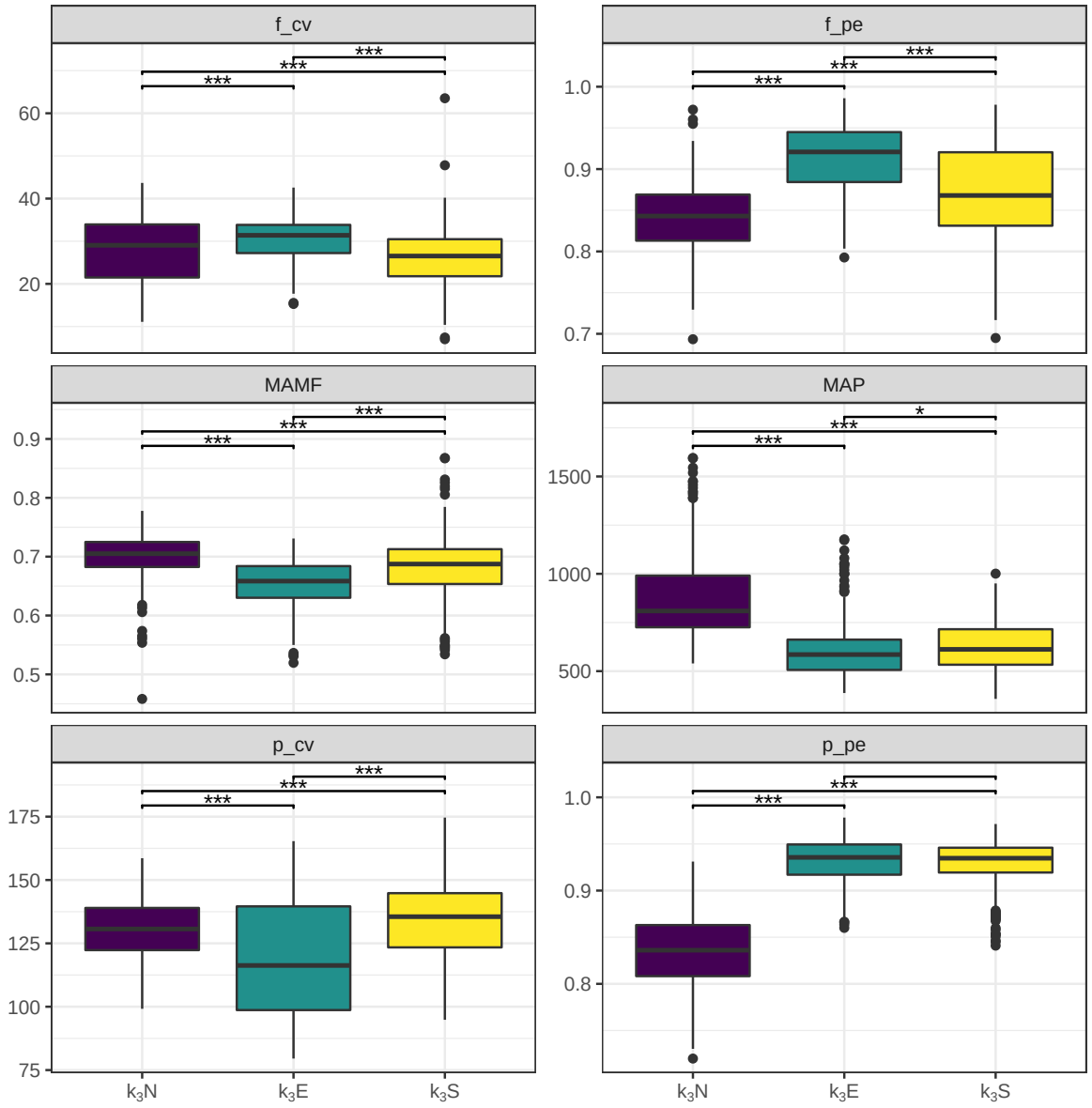

Figure 5.19: Boxplots of the spatiotemporal parameters (rowwise: CV_f & PE_f, MAMF & MAP, CV_p & PE_p). Subgroups according to precipitation cluster memberships. Testing for dependent data via the nonparametric Wilcoxon rank-sum Test (Bauer, 1972; Hollander et al., 1999); significance levels: $***$ <0.001, $**$ <0.01, $*$ <0.05, $-$ >0.05.

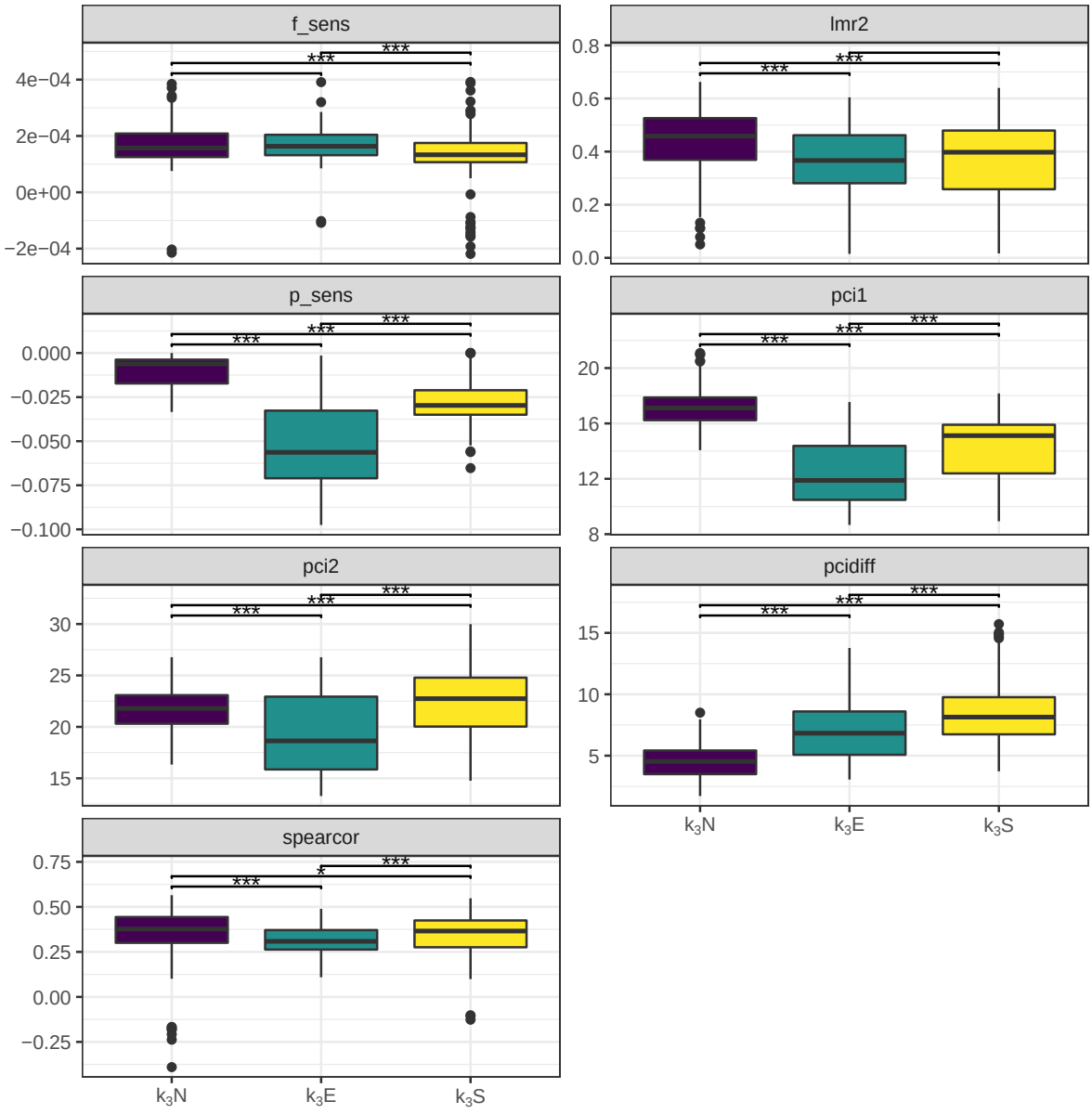

Figure 5.20: Boxplots of the spatiotemporal parameters (rowwise: TS_f & R_t^2, TS_p & PCI1, PCI2 & PCIΔ, ρ_t). Subgroups according to precipitation cluster memberships. Testing for dependent data via the nonparametric Wilcoxon rank-sum Test (Bauer, 1972; Hollander et al., 1999); significance levels: $***$ <0.001, $**$ <0.01, $*$ <0.05, $-$ >0.05.

List of Abbreviations

Symbol	Description
SDTF	Seasonally Dry Tropical Forest
FPAR	Fraction of Absorbed Photonsynthetic Active Radiation
PAM	Partitiong around Medoids Clustering (Kaufman and Rousseeuw, 1990)
TS	Theil-Sen Estimator (Sen, 1968)
GPP	Gross Primary Productivity
MAP	Mean Annual Precipitation
NEB	Northeast Brazil
ITCZ	Inner Tropical Convergence Zone
SST	Sea Surface Temperature
EWD	Easterly Wave Disturbances
SASA	South Atlantic Subtropical Anticyclone
CF	Cold Fronts
SACZ	South American Convergence Zone
SNEB	south Northeast Brazil
ENEB	east Northeast Brazil
NNEB	north Northeast Brazil
ENSO	El Niño/Southern Oscillation
PDO	Pacific Decadal Oscillation
NAO	North Atlantic Oscillation
GPCC	Global Precipitation Climatology Centre
WGS	World Geodetic System
PAR	Photosynthetic Active Radiation
CO_2	Carbondioxide
NOAA	National Oceaninc and Atmospheric Administration
CDR	Climate Data Record
LAI	Leaf Area Index
FTP	File Transfer Protocol
IBGE	Instituto Brasileiro de Geografia e Estatística
MMA	Ministério do Meio Ambiente
SIRGAS	Sistema de Referencia Geocéntrico para las Américas geocentric reference system

MAMF	Mean Annual Maximum FPAR
MF	Mean FPAR
MAR	Mean Annual Range FPAR
CV	Coefficient of Variation
PCI	Precipitation Concentration Index
ITQ	Information Theory Quantifier
PE	Permutation (Shannon) Entropy
MPR	Martín-Platino-Rosso statistical complexity
PE_p	Permutation Entropy of Precipitation
MPR_p	Statistical complexity of Precipitation
PE_f	Permutation Entropy of Vegetation
MPR_f	Statistical complexity of Vegetation
PDF	Probability Density Function
MAX_p	Precipitation Season Median Maximum Month
ONS_p	Precipitation Season Median Onset Month
MAX_f	Vegetation Season Median Maximum Month
ONS_f	Vegetation Season Median Onset Month
IQR	Interquartile Range
ρ	Correlation Coefficient
R^2	Coefficient of Determination
hPA	Hectopascal

Acknowledgements

First, i want to thank my supervisor Prof. Dr. Cyrus Samimi for his ideas and for giving me the opportunity to develop the topic quite freely. Then, i want to thank Prof. Dr. Bettina Engelbrecht who helped me a lot to structure my thoughts and who made the connection to Brazil. The time at UFPE Recife and Catimbau National Park was an incredible experience where I made friends that are still in my life today. I also want to thank the working group of climatology, especially Harald, Liu and Isabelle, for revising my script and many useful comments during the process. At last, i want to thank my friends and family for many revisions, great discussions, moral support and a lot of patience.

Declaration of Authorship

I hereby declare that the thesis titled "Long-term precipitation and vegetation dynamics in the Caatinga (Brazil)" is my own unaided work. All direct or indirect sources used are acknowledged as references. This thesis was not previously presented to another examination board and has not been published.

Hiermit erkläre ich, dass ich die vorgelegte Arbeit „Long-term precipitation and vegetation dynamics in the Caatinga (Brazil)" eigenständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen und Hilfsmittel benutzt habe. Weiter wurde diese Arbeit bisher nicht veröffentlicht oder zur Erlangung eines akademischen Grades eingereicht.

Bayreuth, Jannis Viola