



A new approach to detect extreme events: a case study using remotely-sensed precipitation time-series data

Philipe Riskalla Leal ^{a,*}, Ricardo José de Paula Souza e Guimarães ^b, Fábio Dall Cortivo ^a, Rayana Santos Araújo Palharini ^a, Milton Kampel ^a

^a National Institute for Space Research (INPE, Instituto Nacional de Pesquisas Espaciais), Avenida dos Astronautas 1758, São José dos Campos, São Paulo, 12227-010, Brazil

^b Evandro Chagas Institute (IEC), Rodovia BR-316 km 7 s/n – Levilândia, Ananindeua, Pará, 67030-000, Brazil

ARTICLE INFO

Keywords:

Extreme event detection
Precipitation time-series analysis
Brazilian Amazon region
Climate change

ABSTRACT

Detecting and predicting extreme events are of major importance for socioeconomic, healthcare and ecological purposes. This study proposes an alternative model to detect extreme events based on analyses of probability distribution functionffns

s ($f_{(X)}$), called Optimum Probability Distribution Function Searcher Model (Opt.PDF-model). The Opt.PDF-model involves the optimization of a fitness function between an histogram and a set of theoretical $f_{(X)}$, and the subsequent evaluation of the Probability Point Function (PPF) of the fittest theoretical ($f_{(X)}$) to assess threshold values for the classification of extreme events. Any occurrence in the dataset with a PPF value equal to or greater than 90% was considered an extreme event candidate. A satellite-derived precipitation time-series (*Climate Hazards Group InfraRed Precipitation with Station data*) was used to calibrate and validate the proposed model, with data on accumulated precipitation from more than 30 years (Jan.1981 to Dec.2018) of the Brazilian Amazon region. The proposed method was pairwise cross-validated with two other extreme event models based on Gamma and Gaussian distributions, as applied by the European Drought Observatory of the European Environment Agency. Additionally, all three extreme event classification models were cross-validated relative to the El Niño Southern Oscillation (ENSO). By means of the Opt.PDF-model, it was possible to evidence two positive temporal trends for the area of study: one for more intense precipitation events, and another for less intense events. The pairwise cross-validation analysis returned specific Kappa's similarity indices, and the highest similarity was observed between the Gamma and the Opt.PDF models (48% for PPF_(97.7%)). This analysis indicated that extreme event detection models are highly sensitive to distribution family priors and to threshold definitions. The proposed approach and the results obtained here are potentially useful for climate change warnings, and can be extended to other scientific areas that involve time-series analyses.

1. Introduction

Extreme events are generally easy to recognize but difficult to define (Diaz and Murnane, 2008). This perception is due to the variability in the definitions of extreme event, and in the context in which they are applied (Broska et al., 2020; Harris, 1996; Makkonen et al., 2013). Extreme events can be defined as events that would normally be as uncommon as, or rarer than, a certain percentile of a probability distribution function ($f_{(X)}$) estimated from observations expressed as departures from daily or monthly means (NASEM, 2016). They may also be

defined as individual local weather events whose intensity exceeds critical levels on a continuous scale (IPCC, 2021, 2008). Specifically for this essay, the definition of Kantz et al. (2005) is adopted, whereby extreme events are those events that are rare, occur irregularly, exhibit observable phenomena that assume extreme values, and are inherent to the system under study and not due to external forces.

Increases in frequency and in intensity of extreme events due to global warming are expected for the near future (Chen et al., 2013; Feng et al., 2014; IPCC, 2021). More recently, many extreme precipitation events have occurred around the world, with serious impacts on human

* Corresponding author.

E-mail addresses: leal.philipe@gmail.com (P. Riskalla Leal), ricardojpsg@gmail.com (R.J. de Paula Souza e Guimarães), fd cortivo@gmail.com (F. Dall Cortivo), rayana.palharini@gmail.com (R. Santos Araújo Palharini), milton.kampel@inpe.br (M. Kampel).

society and the ecological environment (Wu and Lau, 2016). Therefore, understanding extreme precipitation events is important not only because of their considerable social effects, but also due to changes in their characteristics and patterns over the last decades (Feng et al., 2014; Song et al., 2019).

A usual procedure to detect extreme events in climatological studies is based on analyses of time-series anomaly data (Akbari et al., 2017; Mann and Park, 1999). The anomaly time-series is defined in relation to a specified climatology, derived from a given temporal period; this period is evaluated over a long interval to provide sufficient data for statistical analyses, often around 30 years or more (Hansen et al., 2012). The climatological period should also be fixed to prevent potential changes in the nature of the distribution of anomalies (Hansen et al., 2012; NCAR Community, 2004; Thornton and Pi, 2004). If the available anomaly extraction techniques are themselves based on a decomposition (i.e., subtraction of a trend/seasonality pattern) of the time-series data, they may also be called detrending techniques in time-series analyses (Brownlee, 2018). The use of anomaly time-series facilitates analytical interpretation, and mitigates certain statistical constraints, such as variance inflation, when dealing with sparse matrices (Gilbert et al., 1992; Press et al., 1992). Nevertheless, this approach lacks analytical power because it neglects possible seasonal and intra-seasonal trends over the time-series (Greene, 2000; Herranz et al., 2017; Taylor and Letham, 2017). In addition, the respective decomposition technique may be overly sensitive to the presence of outliers in the time-series (Trenberth et al., 2003).

Another approach to detect extreme events in time-series is to consider a previously defined $f_{(X)}$ and indirectly estimate its hyper parameters from the time-series (Ozcan et al., 2013; Rajeevan et al., 2008). This technique lessens the burden of finding the best distribution family for the data, but may lack fitness precisely because of this assumption. Furthermore, by assuming a single $f_{(X)}$ structure over a given time-series data, one presupposes stationarity, which implies that there is no seasonal, trend or other time-cyclical-dependent effect in the time-series (Xavier et al., 2019). Therefore, effects such as seasonal variations, consequently, variations in one's cumulative distribution function (CDF) and threshold definitions are neglected (Camuffo et al., 2018).

Predefined thresholds to map precipitation events and detect the frequency of extreme event is also a common practice (Dereczynski et al., 2009; Goswami et al., 2006; Rajeevan et al., 2008; Rao et al., 2020). Nevertheless, fixing threshold values to define extreme events might be inappropriate in regions where spatial variability is relevant, resulting in a poorly defined ($f_{(X)}$) (Haylock and Nicholls, 2000; Manton et al., 2001). Therefore, this approach is restricted to areas where seasonal mean climatology and daily variability are well characterized and considered homogeneous (Goswami et al., 2006).

Due to the variety of possible techniques available to detect extreme events (Dereczynski et al., 2009; Le Guer and El Omari, 2012), the obtained results remain subject to substantial uncertainty. Therefore, a continued research effort is required to increase the reliability of such attributions, particularly for classes of events whose definitions are still poorly understood (NASEM, 2016).

Satellite-based studies are emerging as useful tools to assess regional and global extremes, particularly over the oceans and other poorly instrumented regions (Lockhoff et al., 2014). The Global Precipitation Measurement (GPM) mission and other capabilities will be beneficial in the coming years and decades (Hou et al., 2014). As satellite-based records lengthen in time, they can become viable for trend detection and attribution studies of extreme events (NASEM, 2016).

Given the importance of studying extreme events (Diaz and Murane, 2008; Ghil et al., 2011; Tawn et al., 2018), and based on the climate change scenario, this essay proposes an alternative method for detecting extreme events in time-series datasets based on the analysis of the empirical probability distribution function. The proposed model involves a fitness function between an empirical data-derived histogram

and a set of theoretical $f_{(X)}$. The respective point probability function ($PPF_{(x)}$) of the fittest theoretical $f_{(X)}$ is then applied to the derivation of threshold values in order to classify the time-series in potential extreme event. This algorithm has been implemented according to a hierarchical architecture based on an object-oriented programming paradigm (Barth et al., 2014; Hemmendinger, 2019).

This essay is structured in three parts. First, a conceptual review of the probability distribution function is presented. In the subsequent section, the proposed model is described detailing the schematics of its algorithm implementation. Lastly, the proposed model is applied to a precipitation time-series derived from a remote sensing database, and the results are pairwise cross-validated with two other extreme event classification approaches (i.e., models) applied by the European Drought Observatory of the European Environment Agency.

2. Conceptual review

To better understand the main features of the algorithm development and its results, some basic statistical definitions of probability functions are presented in this section.

2.1. Definitions

Let X be a random variable with a given probability distribution function ($f_{(X)}$). A probability density function (PDF) is a type of $f_{(X)}$ applied to continuous variables. It states that for each possible value of a random continuous variable there is a probability associated with it (Bussab and Morettin, 2010). A PDF graph is the limiting case of an histogram as the amount of data increases and the class intervals decrease in size, and the histogram can be defined as the empirical probability distribution function ($EPDF_{(x)}$) of X , in which the number of bins sets the probability distribution (Butterfield, 2014).

The CDF of a random variable X (F_X) may be expressed for $X \in \mathbb{R}$, with $F_{(X)} \rightarrow [0 : 1]$ according to equation (1):

$$F_X(x) = \mathbb{P}(X \leq x) \quad 1$$

The $f_X(x)$ of a random variable X is defined as the derivative of its CDF (equation (2)):

$$f_X(x) = F'_{(x)} = \frac{d(F_{(x)})}{dx} \quad 2$$

The point probability function ($PPF_{(x)}$) of a random variable X ($PPF_{(x)}$) is defined as the inverse of its CDF (equation (3)). PPF is also called quantile probability function in statistics literature (Wasserman, 2009), though here the PPF nomenclature is adopted.

$$PPF_{(x)} = F^{-1}_{(x)} \quad 3$$

Statistical moments represent a set of specific quantitative measurements of the shape of a distribution function. Examples are the mean, variance, skewness, and kurtosis of a random variable X . Whenever possible, the statistical moments were used for fitting the ($f_{(X)}$) families over the dataset. In the case of the Gamma family distribution, the iteration fitting method was applied to reduce bias in the estimation of the Gamma's parameters (Chandler and Scott, 2011).

3. Material and methods

3.1. Area of study

This study considers the state of Pará's daily mean precipitation time-series for analysis and validation by the models. Pará is one of the nine states that compose Brazil's Legal Amazon region, with an area of 1.2 million km², representing 14.7% of the entire Brazilian territory, and nearly 23% of the Brazilian Amazon region (IBGE, 2017).

Pará is located in the eastern part of the Amazon region, and west

from the semi-arid northeastern region of Brazil (see Fig. 1) comprising 144 municipalities, subdivided into six meso-regions (IBGE, 2017). According to the Brazilian census of 2010, its population was approximately 8,602,865 (IBGE, 2017).

Pará is under a tropical climate, which in turn is subdivided into three categories: "Tropical zone without a dry season" (Af), "Tropical zone of monsoon" (Am), and "Tropical zone with dry winters" (Aw) (Alvares et al., 2014). According to these authors, Pará's territory is majorly of the monsoon type, with a small portion of tropical zone without dry a season in the central and western areas. The annual mean temperature is 27 °C, with maximum and minimum of 33 °C and 20 °C (INMET, 2018). The annual mean precipitation is high, ranging from 1,800 mm up to 13,200 mm per year (Lima et al., 2010).

In the Amazon region, precipitation exhibits marked inter-annual variability, part of which is attributed to sea surface temperature (SST) variability and to climate events such as El Niño-Southern Oscillation – ENSO (Riveros et al., 2013). Other phenomena, such as squall lines (Alcântara, 2011; Alcântara and Souza, 2016; Cohen et al., 1989),

easterly waves (Merritt, 1964), an upper-level cyclonic vortex (Coutinho et al., 2010; KOUSKY and GAN, 1981; Souza et al., 2015) and the meridional SST gradient in the Tropical Atlantic (Aceituno, 1998; Marengo et al., 2001; Riveros et al., 2013) also play an important role in the region's precipitation system (Moraes and Francisco Filho, 2018).

The Amazon region has been deemed vulnerable to extreme drought events on account of global warming (Cox et al., 2004; Li et al., 2006), especially due to its importance in agricultural activities in the region (Moraes et al., 2006) and its importance to the Amazon weather and climate (Santosdos Oliveira, 2017; Souza et al., 2016). Notably, there is evidence of increased precipitation intensity over the past decades for the Amazon region (Riveros et al., 2013; Sterenberg, 2012), though that increase has not changed the total yearly amount precipitated (Aguilar et al., 2005). Meanwhile, it is important to account for changes in climate and precipitation regimes - since there are evidences that suggest a potential increase in precipitation intensity in the region over the following years (Marengo et al., 2010).

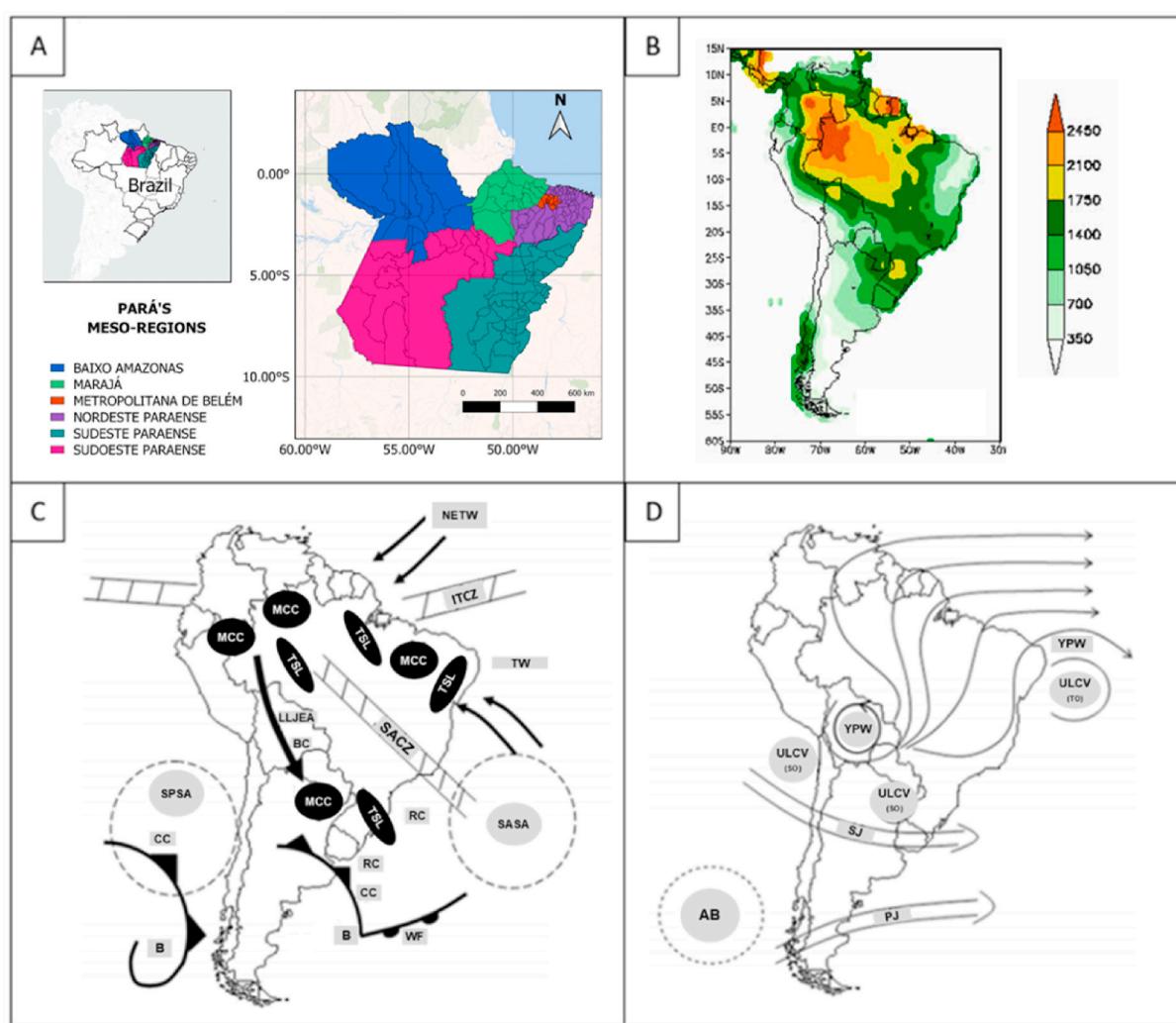


Fig. 1. Area of study, with Pará's meso-regions. (A) state of Pará in Brazil; (B) and (C) adapted from Shi et al. (2000), demonstrating the spatial distribution of the temporal mean precipitation over South America; (C) and (D) adapted from Reboita et al. (2010), depicting the main synoptic systems in South America.

Low troposphere wind system: NETW: NE Trade Winds; TW: Trade Winds; SASA: South Atlantic Subtropical Anticyclone; SPSA: South Pacific Subtropical Anticyclone; MCC: Mesoscale Convective Complex; CC: Cold Front; WF: Warm Front; LLJEA: Low-Level Jet East of the Andes; TSL: Tropical Squall Lines; SACZ: South Atlantic Convergence Zone; ITCZ: Intertropical Convergence Zone; SJ: Subtropical Jet; PJ: Polar Jet. See further description in Reboita et al. (2010).

High troposphere wind system:

AB: Bolivian High; AB: Atmospheric Blocking; SJ: Subtropical Jet; PJ: Polar Jet; ULCV_(SO): Upper Levels Cyclonic Vortices of Subtropical Origin; ULCV_(TO): Upper Levels Cyclonic Vortices of Tropical Origin. YPW: Yearly Periodic Wind systems - they disappear during winter (June-September).

3.2. The proposed Model's conceptual algorithm

The proposed model, named Optimum Probability Distribution Function Searcher Model (Opt.PDF-model), is a variation of the algorithm applied by the European Drought Observatory. It involves the optimization of a fitness function between a histogram and a set of theoretical $f_{(X)}$, and the subsequent evaluation of the Probability Point Function (PPF) of the fittest theoretical $f_{(X)}$ to assess threshold values for the classification of extreme events. The thresholds are expressed as a set of percentiles. The Opt.PDF-model is applicable to cross-dataset analysis and inter-comparisons between different studies and datasets. The algorithm architecture of the proposed model consists of a Manager class (*M-class*), two independent classes (*C1-class*, *C2-class*) and a third type, the Extreme Event Classifier (*E.E.-Classifier*).

M-class is responsible for filtering the set of available $f_{(X)}$ families

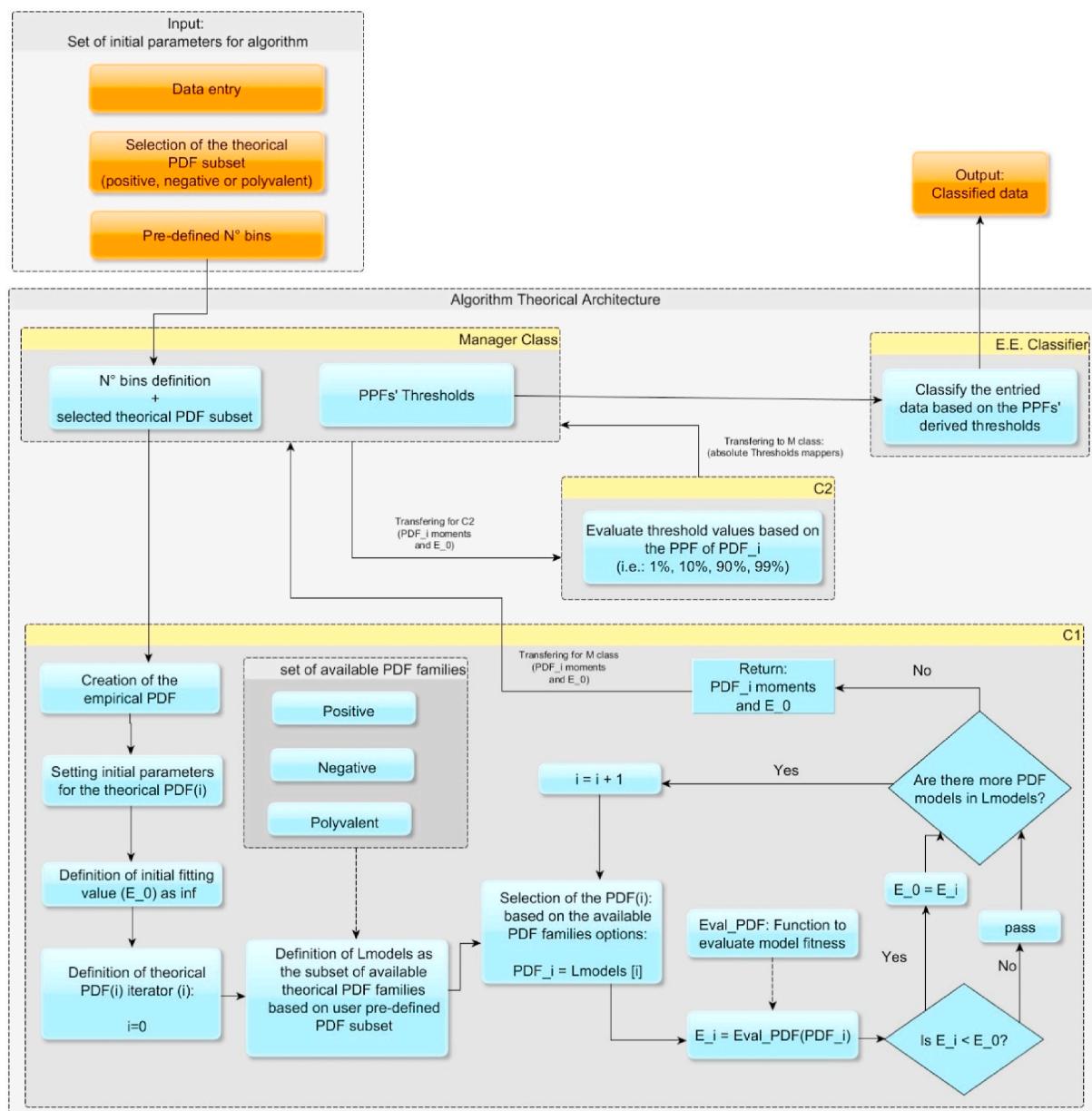


Fig. 2. Conceptual structure of the Optimum Probability Distribution Function Searcher Model (Opt.PDF-model). Arrows indicate the connection direction between given classes of the algorithm. Continuous lines represent a direct link between processes. Dashed lines represent entries that are not directly related to the user's initial entry of parameters. Yellow boxes represent the classes of the algorithm. Boxes with grey headline are sets of attributes/parameters related to the algorithm but not directly provided or manageable by the user. Blue boxes represent processes and definitions. Orange boxes represent input/output (IO) parameters. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

The C2-class then instantiates the theoretical PPF from the best fitted $f_{(X)}$, which is used for the acquisition of the absolute thresholds (in units of data) for later classification of the dataset by the E.E.-Classifier. These thresholds are then passed back to the M-class, which relays them to the E.E.-Classifier in the final step of the algorithm (the classification). A conceptual framework of the algorithm is presented in Fig. 2.

The evaluation of the theoretical $f_{(X)}$ fitness is performed in three steps: (i) C1 retrieves the best parameters of the iterated $f_{(X)}$ family from the provided dataset (i.e., the statistical moments of the training time-series). (ii) C1 fits the theoretical $f_{(X)}$ to the given data using the estimated parameters. It is worth of mentioning that the fitting algorithm is dependent on the $f_{(X)}$ family being evaluated in each iteration. Description of the methods are presented in Python's SciPy library documentation (Virtanen et al., 2020). (iii) C1 retrieves the residual term of the theoretical $f_{(X)}$ (that is now fitted to the dataset) from the EPDF of the provided dataset. The difference between the EPDF and the theoretical $f_{(X)}$ is evaluated by the root mean square error (RMSE) (equation (4)). This residual is called $E_{(i)}$ (error of the instance_(i)). From the tested subset of available theoretical $f_{(X)}$, the one with highest adjustment (lowest residual) is selected as the best choice to evaluate the thresholds that will be applied in the extreme event detection step.

$$RMSE = \sqrt{\frac{\sum_i^n (Y_{(i)} - \widehat{Y}_{(i)})^2}{N}} \quad 4$$

where $Y_{(i)}$ is the true value; $\widehat{Y}_{(i)}$ is the predicted value; N is the number of samples.

The user-defined number of bins is an important step in evaluating the model's fitness. Since selecting the number of bins for best fitting an empirical histogram into one's data is not straightforward (Wasserman, 2009), the proposed model's algorithm was developed in such a way as to allow the user to predefine this attribute based on prior knowledge of the data. If no bin is defined by the user, an automatic function to select the best number of bins is activated. This function applies the Freedman Diaconis Estimator (Freedman and Diaconis, 1981) and the Sturges algorithms (Sturges, 1926) to evaluate the best number of bins, and returns the maximum value (i.e., the best number of bins to be used) from the two methods.

The algorithm architecture was structured to allow possible grouping, slicing and aggregation operations over a given time-series. Therefore, the structure allows the algorithm to evaluate local and time-specific extreme event thresholds, allowing better management over time effects (Camuffo et al., 2018) and spatial effects (Tawn et al., 2018) of one's dataset.

3.3. Model validation

A subset of the Climate Hazards Group Infrared Precipitation with Station data (CHIRPS) (Funk et al., 2014) was used to evaluate precipitation climatological patterns in the area of study. The dataset consists of spatially gridded data with a spatial resolution of $\sim 5.6 \times 5.6 \text{ km}^2$, encompassing nearly 30 years of quasi-global rainfall data (50°S - 50°N), from 1981/01/01 to 2018/12/31. CHIRPS provides gauge-satellite precipitation estimates that have low latency, high resolution, low bias, and long period of record (Funk et al., 2015). CHIRPS precipitation data have been previously validated with rain gauges in different areas of Brazil (Nogueira et al., 2018; Paredes-Trejo et al., 2017), achieving good results in different seasons, and good correlation with observations from land stations.

CHIRPS precipitation data was retrieved from the Google Earth Engine (GEE) platform. The data was spatially aggregated into municipality-specific accumulated means for the state of Pará. GEE is a global platform for Earth science data and analysis powered by Google's cloud infrastructure, combining a catalog of satellite imagery and

geospatial datasets with global-scale analysis capabilities (Gorelick et al., 2017).

The daily mean precipitation dataset of Pará's municipalities was applied in the Opt.PDF-model validation. The Opt.PDF-model was compared with two established extreme event detection approaches applied by the European Drought Observatory, a service run by the EC's Joint Research Centre from the European Environment Agency (EEA; EDO, 2019).

The first model applies a Gaussian (Normal) distribution function over the anomaly time-series. Here anomaly is considered as the time-series detrended by its climatological average, derived according to equation (5).

$$Anomaly_{(t)} = \frac{[Precipitation_{(t)} - \overline{Precipitation}]}{Var(Precipitation_{(t)})} \quad 5$$

where t is the day of the time-series; $Precipitation_{(t)}$ is the daily accumulated precipitation averaged for a given region of interest (i.e., state of Pará) of a given day in the time-series; $Var(Precipitation_{(t)})$ is the variance of the daily accumulated precipitation time-series; $\overline{Precipitation}$ is the sample average of $Precipitation_{(t)}$. This extreme event detection approach is hereafter named Gauss-model.

The second model applies a Gamma PDF on the original time-series (no detrending technique is applied to the time-series prior to the distribution fitting). Hereafter, this model is hereafter named Gamma-model.

For each tested model (Opt.PDF-model, Gaussian and Gamma), the respective PPFs were estimated according to the European Drought Observatory percentage threshold definitions (2.30%, 6.70%, 15.90%, 84.10%, 93.30% and 97.70%). These percentage thresholds are considered comparable in time and space, and they can be evaluated at different time scales to monitor droughts and wetness in the environment (Vicente-Serrano et al., 2010). The results of such analysis can be understood as a set of three classified precipitation time-series (one for each tested model).

The three classified precipitation time-series were later grouped by month and specific PPF, and each monthly-accumulated occurrence was verified (the monthly-accumulated sum). To prevent eventual confusion between terminologies (frequency, amplitude, intensity), the "monthly-accumulated summed occurrences" is now abbreviated as MASO. As each extreme event classification model decomposes the time-series into specific PPFs, MASO is also PPF-specific ($MASO_{(PPF=i)}$).

The $MASO_{(PPF=i)}$ of each time-series were later cross-validated between each unique pairs of extreme event classification models. The cross-tabulation technique performs a pairwise comparison of all the classes (in the present case, the classes derived from the PPFs) of each treatment (i.e., of each tested model) (Bussab and Morettin, 2010). The Kappa similarity index (SI) was used for this pairwise comparisons; a SI equal to one implies that all events detected by a given classifier (model) "A" were also detected by a classifier (model) "B"; a SI of zero implies that none of the events detected by classifier "A" were detected by classifier "B" (Cohen, 1960). The SI can be estimated according to equations (6) and (7).

$$k = \frac{p_0 - p_e}{1 - p_e} \quad 6$$

$$p_e = \frac{\sum_k (n_{(k, c=1)} * n_{(k, c=2)})}{N^2} \quad 7$$

where p_0 stands for the proportion of units in which the classifiers agree; p_e stands for the proportion of units for which agreement is expected by chance; n is the number of times classifier (c) predicted category k; N is the number of observations; k is the number of classes (in this case, given that it is a pairwise test, $k = 2$).

In addition, the three-classified time-series were later segmented

according to El Niño Southern Oscillation (ENSO) variability periods (La Niña and El Niño). ENSO is recognized as an important regulatory driver for extreme precipitation (Cai et al., 2014; Grimm and Tedeschi, 2009) and as a dominant inter-annual mode of coupled atmosphere-ocean variability in several regions of the world (Trenberth and Stepaniak, 2001), including South America (Grimm, 2003; Penalba and Rivera, 2016). Therefore, ENSO events were used here as climatological indicators allowing further assessment of potential differences in classification between the models tested.

4. Results

According to the Opt.PDF-model, the $f_{(X)}$ family that best fitted the precipitation dataset was the Generalized Half Logistic (Genhalflogistic) with the following hyper-priors: 4.39×10^{-11} for c ; -1.76×10^{-9} for location, and 5.88 for scale. A general PDF of the Genhalflogistic family is described in equations (8), (9), and (10). The algorithm for estimating the hyper-parameters for this PDF family can be found elsewhere (Kumar et al., 2015).

$$f_X(x, c, loc, scale) = \frac{a}{b * scale} \quad 8$$

$$a = 2 * \left(1 - c * \left[\frac{(x - loc)}{scale}\right]\right)^{\left(\frac{1}{c-1}\right)} \quad 9$$

$$b = \left[1 + \left(1 - c * \left[\frac{(x - loc)}{scale}\right]\right)^{\left(\frac{1}{c}\right)}\right]^2 \quad 10$$

Results from the Opt.PDF-model suggest that the frequency of intra-

annual precipitation events is PPF-specific (Fig. 3); therefore, precipitation events of different intensities tend to occur in different frequencies over the year; furthermore, the amplitude of the intra-annual variability is also dependent on the precipitation intensity (see oscillation of the medians and the box-plot quantiles over the months for different PPFs in Fig. 3). With regard to the precipitation data of Pará's municipalities, the events from the caudal portions of the PDF (precipitations events of $PPF_{(0\%)}^*$ and $PPF_{(\geq 90\%)}^*$) are more intra-annually variable than those centered more around the median of the PDF (precipitation events whose intensity ranges between $PPF = 1\%$ and $PPF \leq 70\%$). The results imply that ordinary rains (precipitation events whose PPF is centered between 1% and 70%) present no clear intra-annual trend, no frequency variation over the year; while odd (rarer) precipitation events (precipitation events of $PPF_{(class)} = 0\%$ and $PPF_{(class)} \geq 90\%$) present higher intra-annual frequency variability.

Results from the MASO analyses of the daily accumulated mean precipitation data of the municipalities are presented in Fig. 4. The Gamma-model (blue line) resulted in higher MASO for low quantiles ($PPF_{(0\%)}$). For quantile 6.7%, Gamma-model classified more events than any other model considered here (higher MASO values), implying that the model is more sensitive to seasonal and intra-seasonal outliers in the respective PPF threshold. Based on Gamma-model's $MASO_{(PPF = 0\%)}$ and $MASO_{(PPF = 6.7\%)}$ temporal trend, results suggest that the frequency of those events is increasing over the years (positive trend). This increase of less intense precipitations implies that precipitation events are occurring more often over the years, which can be a consequence of climate change (Trenberth, 2013; UN, 2007). As droughts are naturally associated with sustained periods of no or very low precipitation (Makukumbura et al., 2019), this increase in less intense precipitations suggests a potential drought scenario trend for the studied area. This agrees with previous studies that pointed towards a drought tendency in

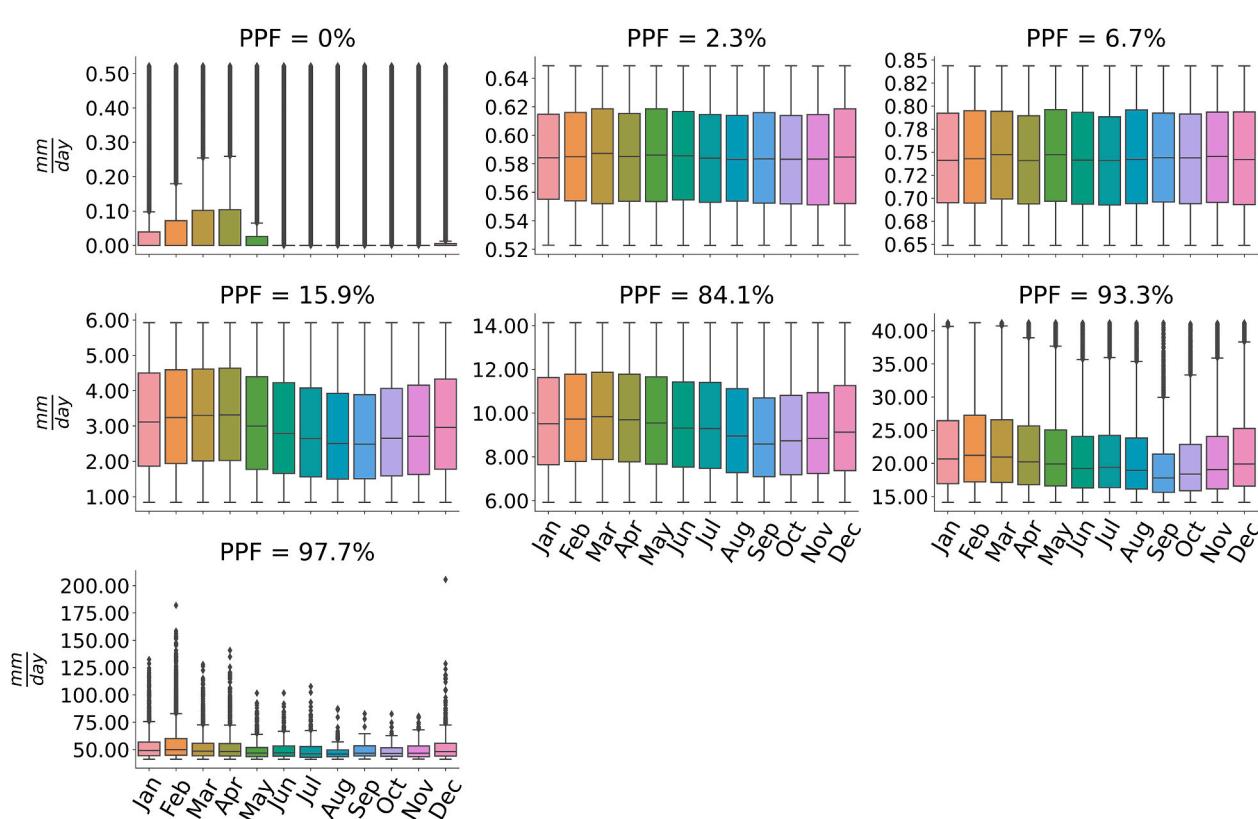


Fig. 3. State of Pará monthly precipitation distribution derived from Opt.PDF-model for eight different PPF thresholds (0%, 1%, 10%, 20%, 50%, 70%, 90%, 99%). Data distribution is presented in a box-plot structures showing the median, the confidence interval and the respective outlier precipitation events for each specific month and PPF threshold of the time-series. The red star indicates the mean of each month. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

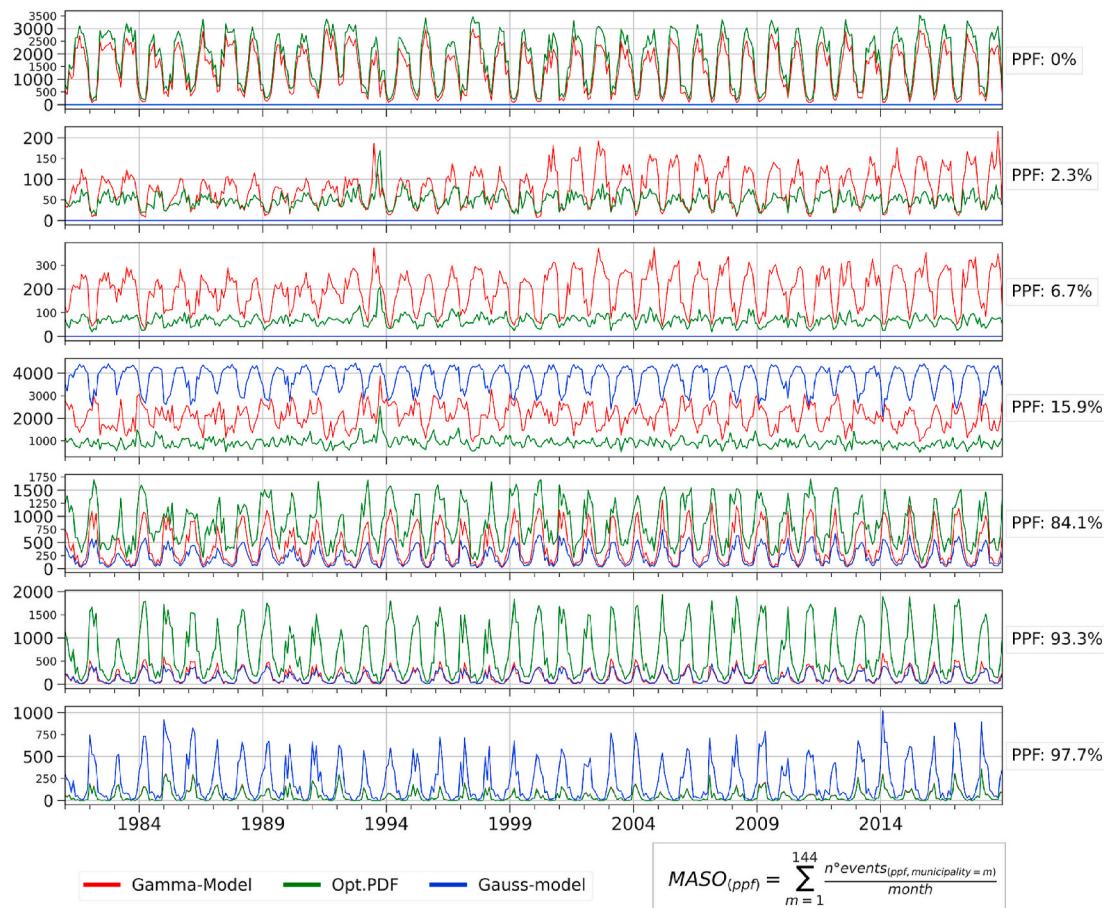


Fig. 4. Gamma distributed (Gamma-model) detections of monthly-accumulated precipitation events in the state of Pará from the original precipitation data, Gaussian distributed (Gauss-model) precipitation anomalies, and original precipitation data based on the Opt.PDF-model approach (Opt.PDF), for different PPF classes (on the right).

the Amazon region (Cox et al., 2004; Li et al., 2006). These potential changes could result in stiff socioeconomic impacts in the region, triggering public health problems, such as increase in cases of parasites, waterborne and vector-associated diseases (Diaz and Murnane, 2008; Ghil et al., 2011; Marcheggiani et al., 2010; SantosdosOliveira, 2017).

For PPFs larger than 84.1%, the Gamma-model provided lower MASO values (with frequency values reaching below 500 occurrences per month when considering daily precipitation events in every municipality of the state of Pará) than all other tested models. For the PPF_(97.7%), the Gamma-model's classification of extreme events was similar to that of the Opt.PDF-model.

Unlike the Gamma-model and the Opt.PDF-model, the Gauss-model was applicable only to PPFs higher than 6.7%, since no precipitation events with low-intensity values were detected in the time-series derived from the model. These findings agree with the PDF shape of the Gauss-model (Bussab and Morettin, 2010). Therefore, this PDF family structure may not be well-suited for lower values of precipitation anomaly data. For PPFs larger than 6.7%, the Gauss-model showed no trend variability in the MASOs time-series, except for PPF 97.7%. Gauss-model's MASO_(PPF = 97.7%) exhibited a positive trend over time, indicating an increase in the frequency of extreme precipitation events over the study period. According to the Gauss-model, precipitation events presented a well-structured annual cycle, especially for the PPF_(15.9%). In case of more intense precipitation (PPFs >15.9%), the response of the Gauss-model varied. For PPF_(93.3%), it resulted in lower MASOs than the Opt.PDF-model, and similar results were obtained with the Gamma-model. For PPF_(97.7%), the Gauss-model resulted in higher MASO, nearly three times higher when compared with the Opt.

PDF-model and the Gamma-model, reaching frequency values near 700 occurrences per month when taking into account daily precipitation events of every municipality. These later results imply that the Gauss-model is more sensitive to high-intensity extreme events than the other tested models. The Opt.PDF-model and Gamma-model resulted in near identical values for MASO_(PPF=97.7%), implying in similar extreme precipitation detection capability, suggesting that both models are less sensitive to highly positive extreme precipitation events and, therefore, lead to lower false positive results in extreme event detection (Bussab and Morettin, 2010).

According to the linear angular coefficient of the temporal trend of the monthly extreme events accumulated occurrence time-series (Fig. 4), the temporal frequency of the intense precipitation events (PPF_(97.7%)) are expected to increase between 0.11 (Opt.PDF-model) and 0.94 event per year (Table 1). This is disturbing news, since intense precipitation is usually associated with short-duration extreme events (Barbero et al., 2019) that significantly impact the population and infrastructure when flooding, erosion and landslides are present (Lenderink and Van Meijgaard, 2008). Furthermore, for high PPF classes (84.1% and 93.3%), all respective angular coefficients denoted an increase over time. These results are in agreement with global climate models (IPCC, 2021), which point towards a worldwide increase in the frequency of intense precipitation in the future (Myhre et al., 2019), with variations induced by an increase in the atmospheric moisture (Barbero et al., 2017) and convective potential energy availabilities (Xiao et al., 2016). Numerous observations across the globe (Guerreiro et al., 2018) and specifically in South America (Marengo et al., 2010) validate these prospects. In this essay, a positive temporal trend was also

Table 1

Relationship between the linear angular coefficient of the temporal trend of the monthly time-series of the occurrence of extreme events from each of the Extreme Event models and respective PPF classes.

Models/PPFs	0%	15.90%	2.30%	6.70%	84.10%	93.30%	97.70%
Gamma	-3.56	-1.24	0.92	1.03	1.95	0.74	0.12
Gaussian	0.00	-2.72	0.00	0.00	1.09	0.65	0.94
Opt.PDF	-1.40	-2.02	0.03	-0.06	0.39	2.92	0.11

observed for low-intense precipitation events ($PPF = 2.3\%$ and 6.7%), pointing towards a drought-trending scenario. Results in agreement with previous studies (Cann et al., 2013), which provides evidence for an increase in the frequency of heavy precipitations in the near future, even for regions with a drought tendency.

The similarity indexes (SI) indicate that the different classifiers show dissimilar results for each PPF class (Table 2). The highest SI (48%) was observed between the Gamma-model and the Opt.PDF-model for PPF 97.7% (intense precipitation MASO). Lower SI were found for other PPF classes and model comparisons (Table 3). Gamma-model and Gaussian-model were similar globally, with an SI of 54%, while Gauss-model and Opt.PDF-model were dissimilar (SI = 0%).

After segmenting the three extreme event time-series (Gaussian, Gamma and Opt.PDF-models) by the ENSO phases (La Niña and El Niño), each $MASO_{(PPF=i)}$ monthly frequency was evaluated. These frequencies were then applied to a cross-tabulation analysis for further assessment of the classifications of the tested models (Fig. 5). From left to right, the Gamma-model returned $MASO_{(PPF=i)}$ intensities between the Opt.PDF-model and the Gauss-model. $MASO_{(PPF=0\%)}$ and $MASO_{(PPF=15.9\%)}$ intensities increased, indicating that less intense precipitations tend to predominate on the area of study. The Opt.PDF-model returned lower values for $MASO_{(PPF=15.9\%)}$, but higher values for $MASO_{(PPF=0\%)}$, also reinforcing that less intense precipitations tend to predominate in the state of Pará. In addition, the Opt.PDF-model returned higher MASOs for more intense precipitation (PPFs 84.1% and 93.9%) compared to the Gamma-model and the Gauss-model. Opt.PDF-model's $MASO_{(PPF=97.7\%)}$ intensity showed similar values when compared with $MASO_{(PPF=97.7\%)}$ from the Gamma-model. Gauss-model resulted in lower MASO values, except for PPF 15.9% and 97.7%, in which the Gauss-model displays higher values than all other models. These findings reinforce the importance of carefully selecting a family of $f(x)$ for extreme event detection analysis. Since each family generates a specific set of thresholds for data classification, evaluating them properly is of great importance for managers, policy-makers and stakeholders with regard to actions to mitigate or adapt to climate changes.

5. Discussion

This study proposes and evaluates a new model for time-series detection of extreme events. Its findings reinforce that extreme event detection models based on the definition of probability thresholds are sensitive to $f(x)$ priors. Imposing a specific distribution family on extreme event analysis can result in biases in one's threshold definition, potentially compromising the extreme event detection and trend

Table 2

Cross-tabulation of the similarity index (SI) analysis between all classes and models. SI equals to one (1.0) when both models identically detect all events in a series, and zero (0.0) otherwise.

PPF	Gamma x Opt.PDF	Gaussian x Opt.PDF	Gaussian x Gamma
0%	0.0	0.0	0.0
2.3%	0.1	0.0	0.0
6.7%	0.0	0.0	0.0
15.9%	0.0	0.0	0.0
84.1%	0.0	0.0	0.0
93.3%	0.0	0.0	0.1
97.7%	0.48	0.0	0.0

Table 3

Kappa similarity index scores from the models pairwise comparisons.

Pair-model comparisons	Similarity Index
Opt.PDF vs. Gaussian	0.20
Gamma vs. Gaussian	0.54
Opt.PDF vs. Gamma	0.25

analysis. If thresholds are expressed in absolute terms, the number of extreme events may change with climate cycles, whereas relative-based threshold definitions (percentiles or standard deviation) keep distribution frequency constant (Camuffo et al., 2018).

Regarding ENSO variability, all tested models performed similarly (similar mean, median and standard deviations), regardless of which threshold was evaluated. These findings are consistent with the relatively weak relationships reported for the influence of ENSO on extreme precipitation events (Ashcroft et al., 2019; Pepler et al., 2014). Since there is significant variability between precipitation datasets (Palharini et al., 2020), future studies may evaluate the ENSO effects on the proposed model with different precipitation datasets.

It is important to highlight here that the proposed model, at least at its current version, is sensitive to event dependence (commonly seen in autocorrelated time-series). This is due to the fact that the current model applies the log-likelihood optimization (i.e., maximization) approach during the PDFs fitting; as the loglikelihood optimization approaches most often assume independence between events (Storvik, 2011), and given the (non)stationarity and the autocorrelation patterns commonly observed in time-series analyses (Dickey and Fuller, 2012; Xavier et al., 2019), for the current model's version, one should apply it on time-series that are desirably stationary and composed (resultant) of independent events (i.e., non-autocorrelated) (Xavier et al., 2019). A solution for this log-likelihood issue is to consider pre-processing routines into one's time-series analyses (Brownlee, 2018) prior to applying the proposed model. For instance, by applying detrending and/or deseasoning approaches (e.g., auto-regressive models) (Dickey and Fuller, 2012; Herranz et al., 2017), one can achieve resultant time-series that are, then, non-autocorrelated (Trenberth et al., 2003), and thus, directly applicable to the proposed model (Camuffo et al., 2018). These techniques were not applied in this study given to two main reasons: a) these pre-processing routines are not applied by the algorithms from the European Drought Observatory of the European Environment Agency (EEA; EDO, 2019), thus, invalidating any cross-validation analysis in this study; b) these preprocessing routines are considered beyond the proposal of the current study (and respective model); therefore, a subject for future studies. Despite the fact that the proposed model is sensitive to these time-series priors, the model's algorithm was developed in such a way to allow continuous integration with other modules and statistical routines; therefore, the proposed model not only supports custom parametrization, but also supports third-party pre- and post-processing routines. The Opt.PDF-model algorithmic architecture allows interoperability between processing routines, and upgradability (with new functionalities and time-series processing routines). Signal decomposition routines as the empirical orthogonal functions (EOF) and the Fourier decomposition, for instance, are potential future candidates (Brunton and Kutz, 2019; Elmaghriby et al., 2016; Monahan et al., 2009) to be integrated into the proposed model.

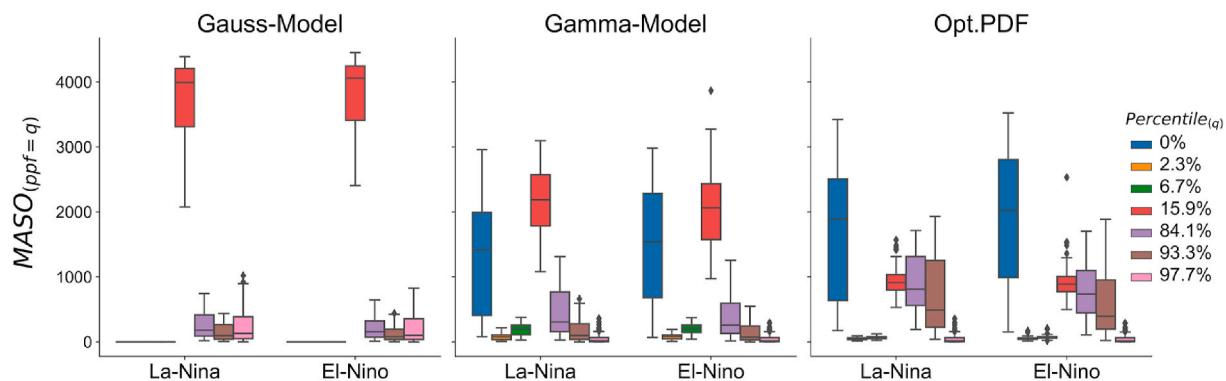


Fig. 5. Comparison between the monthly-accumulated occurrences of extreme events for each classifier (Gauss-Model, Gamma-Model, and Opt.PDF-model) based on the ENSO phases.

Legend: data distribution is presented in a box-plot structure. The red star icon represents the mean of the observed distribution in each class. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

This study may be of assistance for future research on time-series and for climate change analyses. Future studies may investigate more thoroughly the climate trends observed here in the area under consideration, as well as expand the proposed methodology to other geographical regions of interest.

6. Conclusions

This essay proposes a new extreme event classification algorithm (namely, the Opt.PDF-model) that involves a fitness function between an empirical PDF derived from a user-defined dataset and a set of theoretical $f_{(X)}$, with a subsequent classificatory step in which the percentile threshold values are evaluated.

The Opt.PDF-model was tested using the CHIRPS daily precipitation dataset for the eastern Amazon region in northern Brazil. High- and low-intensity precipitation events showed positive temporal trends, indicating that each accumulated frequency is expected to increase. This could be interpreted as a change in the precipitation regime over the years, whereby a decrease or an increase in precipitation is to be expected in the future. Since the intensity of heavy extreme precipitation events is expected to increase with global warming, these results may be considered as an early warning of climate change-induced impacts, with potentially relevant socioeconomic consequences that managers, policy makers and stakeholders must consider (IPCC, 2021).

The proposed Opt.PDF-model was cross-validated with two other standard extreme event detection models used by the European Drought Observatory: one that applies a Gaussian (Normal) distribution function over the anomaly time-series (here named Gauss-model), and another that applies the Gamma distribution function (here named Gamma-model). Inter-model comparisons denoted statistical disparity. Gauss-model was the most sensitive of the three models for low ($PPF_{(15.9\%)}^{}$) and intense ($PPF_{(97.7\%)}^{}$) precipitation events, with nearly twice the number of detections per month than the Gamma-model, and nearly four times the amount of the Opt.PDF-model. The pattern was inverted for other intensities of precipitation. For intermediate-high ($PPF_{(84.1\%)}^{}$) and high ($PPF_{(93.3\%)}^{}$) precipitations, the Opt.PDF-model returned the highest detection rates, nearly 4 times higher than those of the other two models. For less intense precipitation events ($PPF_{(2.3\%)}^{}$ and $PPF_{(6.7\%)}^{}$), the Gamma-model was more sensitive compared to the other ones.

As presented in this work, the application of different methods to detect extreme events can generate variations in the frequency estimation of those events. By imposing a specific family distribution over a given time-series, the results can be biased, varying according to the model used. With the proposed Opt.PDF-model, this constraint is mitigated, allowing a more general approach to detecting and classifying extreme events. The proposed model is still sensitive to autocorrelation and trend effects that may be present in one's time-series; thus, one

should consider applying detrending/deseasoning and/or other statistical decomposition strategies prior to applying the proposed model. Future studies may further upgrade the model proposed here by incorporating new subroutines to one's time-series analysis. Future studies may also test the current model by applying it on new and different datasets, with different temporal attributes (trend, seasonality, anomaly, noise), for different time aggregations (seasonal, annual, biannual etc.), and/or even between different time-frames (different periods of a single time-series). It is the authors' desire that the proposed model be further developed so to encompass different optimization strategies and pre-processing routines in the future; it would then allow the user to properly select which type of extreme-event detection strategy should be applied into one's specific dataset.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are grateful to the Climate Hazards Center, the University of Santa Barbara and the United States Geological Survey for the CHIRPS data, and to Google for the development of the Google Earth Engine platform. This study is a contribution from the Monitoring Oceans from Space Lab (MOceanS) of the Brazilian National Institute for Space Research (INPE) to the Open Network for Water-Related Diseases (ONWARD), a project funded by the UKRI-GCRF Network Award (EP/T003820/1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rsase.2021.100618>.

Funding

Philippe Riskalla Leal was supported by Brazil's Coordination for the Improvement of Higher Education Personnel (CAPES, Finance Code 001), and partially supported by the National Council for Scientific and Technological Development, CNPq (grant #313588/2019-8) under program 2019–2023 (no. 4444327/2019-5) of the National Institute for Space Research, INPE; Ricardo José de Paula Souza e Guimarães was partially supported by the National Council for Scientific and Technological Development (CNPq, grant #313588/2019-8); Fábio Dall'Corntivo was supported by the National Council for Scientific and

Technological Development (CNPq, grant 301042/2020-9).

Author statement

Conceptualization, PRL, MK, RJPSG, RSAP and FDC; methodology, PLR, RJPSG, MK and FDC; formal analysis, PRL and RSAP; resources, MK and RJPSG; writing—original draft preparation, PRL; writing—review and editing, PRL, MK, RJPSG, RSAP, and FDC; supervision, MK and RJPSG.

Ethical statement

The authors declare that all ethical practices have been followed in relation to the development, writing, and publication of the present work.

Availability of data and materials

Data availability is made by the respective data sources. The data processing scripts can be made available, on demand of possible interested parties.

References

- Aceituno, P., 1998. On the functioning of the southern oscillation in the south American sector. Part I: surface climate 116, 505–524.
- Aguilar, E., Peterson, T.C., Obando, P.R., Frutos, R., Retana, J.A., Solera, M., Soley, J., García, I.G., Araujo, R.M., Santos, A.R., Valle, V.E., Brunet, M., Aguilar, L., Álvarez, L., Bautista, M., Castaño, C., Herrera, L., Ruano, E., Sinay, J.J., Sánchez, E., Oviedo, G.I.H., Obed, F., Salgado, J.E., Vázquez, J.L., Baca, M., Gutiérrez, M., Centella, C., Espinosa, J., Martínez, D., Olmedo, B., Espinoza, C.E.O., Núñez, R., Haylock, M., Benavides, H., Mayorga, R., 2005. Changes in precipitation and temperature extremes in Central America and northern South America, 1961–2003. *J. Geophys. Res. Atmos.* 110, 1–15. <https://doi.org/10.1029/2005JD006119>.
- Alkbari, E., Alavipanah, S.K., Jeihouni, M., Hajeb, M., Haase, D., Alavipanah, S., 2017. A review of ocean/sea subsurface water temperature studies from remote sensing and non-remote sensing methods. *Water (Switzerland)* 9, 25. <https://doi.org/10.3390/w9120936>.
- Alcântara, C.R., 2011. Linha de instabilidade da Amazônia: estudo de caso e importância das características do perfil do vento na sua formação e desenvolvimento. *Cienc. Nat.* 3, 197–226. <https://doi.org/10.5902/2179460X9370>.
- Alcântara, C.R., Souza, E.P. De, 2016. Caracterização de perfis de vento no ambiente de formação das linhas de instabilidade amazônicas Wind profiles characterization in the training environment the lines of Amazon instability. <https://doi.org/10.5902/2179-460X18337>.
- Alvares, C.A., Stape, J.L., Sentelhas, P.C., José Leonardo, de Moraes Gonçalves, Sparovek, G., 2014. Köppen's climate classification map for Brazil. *Meteorol. Z.* 22, 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- Ashcroft, L., Karoly, D.J., Dowdy, A.J., 2019. Historical extreme rainfall events in southeastern Australia. *Weather and Climate Extremes* 25, 100210. <https://doi.org/10.1016/j.wace.2019.100210>.
- Barbero, R., Fowler, H.J., Lenderink, G., Blenkinsop, S., 2017. Is the intensification of precipitation extremes with global warming better detected at hourly than daily resolutions? *Geophys. Res. Lett.* 44, 974–983. <https://doi.org/10.1002/2016GL071917>.
- Barbero, R., Fowler, H.J., Blenkinsop, S., Westra, S., Moron, V., Lewis, E., Chan, S., Lenderink, G., Kendon, E., Guerreiro, S., Li, X.F., Villalobos, R., Ali, H., Mishra, V., 2019. A synthesis of hourly and daily precipitation extremes in different climatic regions. *Weather and Climate Extremes* 26, 100219. <https://doi.org/10.1016/j.wace.2019.100219>.
- Barth, T.J., Griebel, M., Keyes, D.E., Nieminen, R.M., Roose, D., Schlick, T., Langtangen, H.P., 2014. A Primer on Scientific Programming with Python Fourth Edition 6 Texts in Computational Science and Engineering, fourth ed. SPRINGER. <https://doi.org/10.1007/978-3-642-54959-5>.
- Broska, L.H., Paganietz, W.R., Vögele, S., 2020. Extreme events defined—a conceptual discussion applying a complex systems approach. *Futures* 115, 102490. <https://doi.org/10.1016/j.futures.2019.102490>.
- Brownlee, J., 2018. Introduction to Time Series Forecasting with Python: How to Prepare Data and Develop Models to Predict the Future. V1.6.
- Brunton, S.L., Kutz, J.N., 2019. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control, first ed. Brunton and Kutz.
- Bussab, W. de O., Morettin, P.A., 2010. Estatística Básica, sixth ed.
- Butterfield, J., 2014. Collins English Dictionary : Complete and Unabridged, 12th. HarperCollins.
- Cai, W., Borlace, S., Lengaigne, M., Van Renssch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M.J., Wu, L., England, M.H., Wang, G., Guilyardi, E., Jin, F.F., 2014. Increasing frequency of extreme El Niño events due to greenhouse warming. *Nat. Clim. Change* 4, 111–116. <https://doi.org/10.1038/nclimate2100>.
- Camuffo, D., della Valle, A., Becherini, F., 2018. A critical analysis of the definitions of climate and hydrological extreme events. *Quat. Int.* <https://doi.org/10.1016/j.quaint.2018.10.008>, 0–1.
- Cann, K.F., Thomas, D.R.H., Salmon, R.L., Wyn-Jones, A.P., Kay, D., 2013. Extreme weather-related weather events and waterborne disease. *Epidemiol. Infect.* 141, 671–686. <https://doi.org/10.1017/S0950268812001653>.
- Chandler, R.E., Scott, E.M., 2011. Statistical Methods for Trend Detection and Analysis in the Environmental Sciences. John Wiley & Sons, Ltd. <https://doi.org/10.1002/978111991571.fmatter>.
- Chen, J., Wei, F., Zheng, C., Wu, Y., Adriano, D.C., 2013. Background concentrations of elements in soils of China. *J. Chem. Inf. Model.* 53, 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. <https://doi.org/10.1177/001316446002000104>.
- Cohen, J.C.P., Silva Dias, M.A.F., Nobre, C.A., 1989. Aspectos Climáticos das Linhas de Instabilidade na Amazônia. CLIMANALÍSE - Boletim de Monitoramento e Análise Climática 4, 34–40.
- Community, N.C.A.R., 2004. Community Climate System Model: Version 3.
- Coutinho, M.D.L., Gan, M.A., Rao, V.B., 2010. Método objetivo de identificação dos vórtices ciclónicos de altos níveis na região Tropical Sul: validação. *Revista Brasileira de Meteorologia* 25, 311–323. <https://doi.org/10.1590/s0102-77862010000300003>.
- Cox, P.M., Betts, R.A., Collins, M., Harris, P.P., Huntingford, C., Jones, C.D., 2004. Amazonian forest dieback under climate-carbon cycle projections for the 21st century. *Theor. Appl. Climatol.* 78, 137–156. <https://doi.org/10.1007/s00704-004-0049-4>.
- Dereczynski, C.P., E. J.S. de O., Machado, C.O., 2009. CLIMATOLOGIA DA PRECIPITAÇÃO NO MUNICÍPIO DO RIO DE JANEIRO Universidade Federal do Rio de Janeiro - Instituto de Geociências – Departamento de Meteorologia , Centro Federal de Educação Tecnológica Celso Suckow da Fonseca do Rio de Janeiro , Recebido Jane. *Revista Brasileira de Meteorologia* 24, 24–38.
- Diaz, H.F., Murnane, R.J., 2008. Preface: the significance of weather and climate extremes to society: an introduction. *Climate Extremes and Society* 9780521870. <https://doi.org/10.1017/CBO9780511535840.002> xiii–xvi.
- Dickey, D.A., Fuller, W.A., 2012. Distribution of the Estimators for Autoregressive time series with a Unit root. *J. Am. Stat. Assoc.* 37–41.
- EEA: EDO, 2019. Meteorological and Hydrological Droughts.
- Elmaghraby, E.K., Khadra, S.A.A., Eissa, H.S., 2016. Using the fast fourier transform technique for climate time series decomposition. *Arab J.Nucl. Sci. Appl.* 49, 78–85.
- Feng, S., Hu, Q., Huang, W., Ho, C.H., Li, R., Tang, Z., 2014. Projected climate regime shift under future global warming from multi-model, multi-scenario CMIP5 simulations. *Global Planet. Change* 112, 41–52. <https://doi.org/10.1016/j.gloplacha.2013.11.002>.
- Freedman, D., Diaconis, P., 1981. On the histogram as a density estimator: L 2 Theory. *Wahrscheinlichkeitstheorie* 57, 453–476. <https://doi.org/10.1007/BF00920714>.
- Funk, C.C., Peterson, P.J., Landsfeld, M.F., Pedreros, D.H., Verdin, J.P., Rowland, J.D., Romero, B.E., Husak, G.J., Michaelsen, J.C., Verdin, A.P., 2014. A Quasi-Global Precipitation Time Series for Drought Monitoring: U.S. Geological Survey Data Series 832. <https://doi.org/10.3133/ds832>. Usgs 4.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes. *Scientific Data* 2 2. <https://doi.org/10.1038/sdata.2015.66> 2015.
- Ghil, M., Yiou, P., Hallegatte, S., Malamud, B.D., Naveau, P., Soloviev, A., Friederichs, P., Keilis-Borok, V., Kondrashov, D., Kossobokov, V., Mestre, O., Nicolis, C., Rust, H.W., Shebalin, P., Vrac, M., Witt, A., Zaliapin, I., 2011. Extreme events: Dynamics, statistics and prediction. *Nonlinear Process Geophys.* 18, 295–350. <https://doi.org/10.5194/npg-18-295-2011>.
- Gilbert, J.R., Moler, C., Schreiber, R., 1992. Sparse matrices in MATLAB: Design and implementation. *SIAM J. Matrix Anal. Appl.* 13, 333–356.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Rem. Sens. Environ.* 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Goswami, B.N., Venugopal, V., Sengupta, D., Madhusoodanan, M.S., Xavier, P.K., 2006. Increasing trend of extreme rain events over India in a warming environment. *Science*. <https://doi.org/10.1126/science.1132027>.
- Greene, W., 2000. Econometric Analysis, fifth ed. Prentice Hall.
- Grimm, A.M., 2003. How do La Niña events disturb the summer monsoon system in Brazil? *Climate Dynamics* on line 1–2.
- Grimm, A.M., Tedeschi, R.G., 2009. ENSO and extreme rainfall events in South America. *J. Clim.* 22, 1589–1609. <https://doi.org/10.1175/2008JCLI2429.1>.
- Guerreiro, S.B., Fowler, H.J., Barbero, R., Westra, S., Lenderink, G., Blenkinsop, S., Lewis, E., Li, X.F., 2018. Detection of continental-scale intensification of hourly rainfall extremes. *Nat. Clim. Change* 8, 803–807. <https://doi.org/10.1038/s41558-018-0245-3>.
- Hansen, J., Sato, M., Ruedy, R., 2012. Perception of Climate Change. In: Proceedings of the National Academy of Sciences of the United States of America, 109. <https://doi.org/10.1073/pnas.1205276109>.
- Hansen, J., Sato, M., Ruedy, R., 2012. Perception of climate change. *Proc. Natl. Acad. Sci. U. S. A.* 109, 367–370. <https://doi.org/10.1073/pnas.1205276109>.
- Harris, R.I., 1996. Gumbel re-visited - a new look at extreme value statistics applied to wind speeds. *J. Wind Eng. Ind. Aerod.* 59, 1–22. [https://doi.org/10.1016/0167-6105\(95\)00029-1](https://doi.org/10.1016/0167-6105(95)00029-1).
- Haylock, M., Nicholls, N., 2000. Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998. *Int. J. Climatol.* 20, 1533–1541. [https://doi.org/10.1002/1097-0088\(20001115\)20:13<1533::AID-JOC586>3.0.CO;2-J](https://doi.org/10.1002/1097-0088(20001115)20:13<1533::AID-JOC586>3.0.CO;2-J).

- Hemmendinger, D., 2019. Object-Oriented Programming. In: Britannica Academic. Encyclopædia Britannica.
- Herranz, E., Gentle, J., Wang, G., 2017. Unit roots in time series with changepoints. *Int. J. Stat. Probab.* 6, 127. <https://doi.org/10.5539/ijsp.v6n6p127>.
- Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K., Iguchi, T., 2014. The global precipitation measurement mission. *Bull. Am. Meteorol. Soc.* 95, 701–722. <https://doi.org/10.1175/BAMS-D-13-00164.1>.
- IBGE, 2017. Conheça Cidades e Estados do Brasil: Brasil/Pará [WWW Document]. URL <https://cidades.ibge.gov.br/> (accessed 07.18.2018).
- INMET, 2018. Normal Climatológica Do Brasil 1981-2010.
- IPCC, 2008. Climate Change 2007: Synthesis Report, Intergovernmental Panel on Climate Change [Core Writing Team IPCC. <https://doi.org/10.1256/004316502320517344>.
- IPCC, 2021. Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
- Kantz, H., Altmann, E.G., Hallerberg, S., Holstein, D., Riegert, A., 2005. Dynamical Interpretation of Extreme Events: Predictability and Predictions. In: Albereio, S., Jentsch, V., Kantz, H. (Eds.), Extreme Events in Nature and Society. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 69–93. https://doi.org/10.1007/3-540-28611-X_4.
- Kousky, V.E., Gan, M.A., 1981. Upper tropospheric cyclonic vortices in the tropical South Atlantic. *Tellus* 33, 538–551. <https://doi.org/10.1111/j.2153-3490.1981.tb01780.x>.
- Kumar, D., Jain, N., Gupta, S., 2015. The type I generalized half-logistic distribution based on upper record values. *Journal of Probability and Statistics* 2015. <https://doi.org/10.1155/2015/393608>.
- Le Guer, Y., El Omari, K., 2012. Chaotic advection for thermal mixing. *Adv. Appl. Mech.* 45, 189–237. <https://doi.org/10.1016/B978-0-12-380876-9.00005-7>.
- Lenderink, G., Van Meijgaard, E., 2008. Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nat. Geosci.* 1, 511–514. <https://doi.org/10.1038/ngeo262>.
- Li, W., Fu, R., Dickinson, R.E., 2006. Rainfall and its seasonality over the Amazon in the 21st century as assessed by the coupled models for the IPCC AR4. *J. Geophys. Res. Atmos.* 111, 1–14. <https://doi.org/10.1029/2005JD006355>.
- Lima, A.M.M. de, Cruz, F.M., Cavalcante, L.M., Chaves, L.M. de L., Junior, M.I., Santos, V.J.C., 2010. A Gestão da Oferta Hídrica no Estado do Pará e seus Aspectos Condicionantes. *Revista Brasileira de Recursos Hídricos* 15, 69–83.
- Lockhoff, M., Zolina, O., Simmer, C., Schulz, J., 2014. Evaluation of satellite-retrieved extreme precipitation over Europe using gauge observations. *J. Clim.* 27, 607–623. <https://doi.org/10.1175/JCLI-D-13-00194.1>.
- Madakumbura, G.D., Kim, H., Utsumi, N., Shioigama, H., Fischer, E.M., Seland, Ø., Scinocca, J.F., Mitchell, D.M., Hirabayashi, Y., Oki, T., 2019. Event-to-event intensification of the hydrologic cycle from 1.5 °C to a 2 °C warmer world. *Sci. Rep.* 9, 1–7. <https://doi.org/10.1038/s41598-019-39936-2>.
- Makkonen, L., Pajari, M., Tikannäki, M., 2013. Closure to “Problems in the extreme value analysis” (Struct. Safety 2008;30:405–419). *Struct. Saf.* 40, 65–67. <https://doi.org/10.1016/j.strusafe.2012.09.007>.
- Mann, M.E., Park, J., 1999. Oscillatory spatiotemporal signal detection in climate studies: a Multiple-taper spectral Domain approach. *Adv. Geophys.* 41, 1–131. [https://doi.org/10.1016/S0065-2687\(08\)60026-6](https://doi.org/10.1016/S0065-2687(08)60026-6).
- Manton, M.J., Haylock, M.R., Hennessy, K.J., Nicholls, N., Chambers, L.E., Collins, D.A., Daw, G., Finet, A., Gunawan, D., Inape, K., Isobe, H., Kestin, T.S., Lefale, P., Leyu, C.H., Lwin, T., Maitrepierre, L., Uprasitwong, N., Page, C.M., Pahalad, J., Plummer, N., Salinger, M.J., Suppiah, R., Tran, V.L., Trewin, B., Tibig, I., Yee, D., 2001. Trends in extreme daily rainfall and temperature in southeast Asia and the South Pacific : 1961 – 1998. *Int. J. Climatol.* 21, 269–284. <https://doi.org/10.1002/joc.610>.
- Marchegiani, S., Puccinelli, C., Ciadamidaro, S., Bella, V., Della, Carere, M., Blasi, M.F., Pacini, N., Funari, E., Mancini, L., 2010. Risks of water-borne disease outbreaks after extreme events. *Toxicol. Environ. Chem.* 92, 593–599. <https://doi.org/10.1080/02772240903252140>.
- Marengo, J.A., Liebmann, B., Kousky, V.E., Filizola, N.P., Wainer, I.C., 2001. Onset and End of the rainy season in the Brazilian amazon basin. *J. Clim.* 14, 833–852.
- Marengo, J.A., Rusticucci, M., Penalba, O., Renom, M., 2010. An intercomparison of observed and simulated extreme rainfall and temperature events during the last half of the twentieth century: Part 2: historical trends. *Climatic Change* 98, 509–529. <https://doi.org/10.1007/s10584-009-9743-7>.
- Merritt, E.S., 1964. Easterly waves and Perturbations, a reappraisal. *J. Appl. Meteorol.* 3, 367–382. [https://doi.org/10.1175/1520-0450\(1964\)003<0367:EWAPAR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1964)003<0367:EWAPAR>2.0.CO;2).
- Monahan, A.H., Fyfe, J.C., Ambaum, M.H.P., Stephenson, D.B., North, G.R., 2009. Empirical orthogonal functions: the Medium is the Message. *J. Clim.* 22, 6501–6514. <https://doi.org/10.1175/2009JCLI3062.1>.
- Moraes, D., Francisco Filho, M., 2018. Contribuição das chuvas do período da tarde em belém E possíveis relações com a normal climatológica. *Revista Brasileira de Climatologia* 23, 17–32.
- Moraes, B.C. de, Costa, J.M.N. da, Costa, A.C.L. da, Costa, M.H., 2006. Variação espacial e temporal da precipitação no Estado do Pará. Universidade Federal de Viçosa, Brasil. <https://doi.org/10.1590/s0044-59672005000200010>.
- Myhre, G., Alterskjaer, K., Stjern, C.W., Hodnebrog, Ø., Marelle, L., Samset, B.H., Sillmann, J., Schaller, N., Fischer, E., Schulz, M., Stohl, A., 2019. Frequency of extreme precipitation increases extensively with event rareness under global warming. *Sci. Rep.* 9, 16063 <https://doi.org/10.1038/s41598-019-52277-4>.
- NASEM, 2016. Attribution of extreme weather events in the context of climate change. In: Attribution of Extreme Weather Events in the Context of Climate Change. The National Academies Press, Washington, DC. <https://doi.org/10.17226/21852>.
- Nogueira, S.M.C., Moreira, M.A., Volpato, M.M.L., 2018. Evaluating precipitation estimates from Eta, TRMM and CHIRPS data in the south-southeast region of Minas Gerais state-Brazil. *Rem. Sens.* 10, 1–16. <https://doi.org/10.3390/rs10020313>.
- Ozcan, O., Bookhagen, B., Musaoglu, N., 2013. Analyzing spatiotemporal patterns of extreme precipitation events in Southeastern Anatolia. *Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci.-ISPRS Archives* 40, 195–200. <https://doi.org/10.5194/isprsarchives-XL-7-W2-195-2013>.
- Palharini, R.S.A., Vila, D.A., Rodrigues, D.T., Quispe, D.P., Palharini, R.C., de Siqueira, R.A., de Sousa Afonso, J.M., 2020. Assessment of the extreme precipitation by satellite estimates over South America. *Rem. Sens.* 12, 1–23. <https://doi.org/10.3390/rs12132085>.
- Paredes-Trejo, F.J., Barbosa, H.A., Lakshmi Kumar, T.V., 2017. Validating CHIRPS-based satellite precipitation estimates in Northeast Brazil. *J. Arid Environ.* 139, 26–40. <https://doi.org/10.1016/j.jaridenv.2016.12.009>.
- Penalba, O.C., Rivera, J.A., 2016. Precipitation response to El Niño/La Niña events in Southern South America – emphasis in regional drought occurrences. *Adv. Geosci.* 62, 1–14. <https://doi.org/10.5194/adgeo-42-1-2016>.
- Pepler, A., Coutts-Smith, A., Timbal, B., 2014. The role of East Coast Lows on rainfall patterns and inter-annual variability across the East Coast of Australia. *Int. J. Climatol.* 34, 1011–1021. <https://doi.org/10.1002/joc.3741>.
- Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1992. Sparse Linear Systems. In: Numerical Recipes in FORTRAN: the Art of Scientific Computing. Cambridge University Press, England, pp. 63–82.
- Rajeevan, M., Bhate, J., Jaswal, A.K., 2008. Analysis of variability and trends of extreme rainfall events over India using 104 years of gridded daily rainfall data. *Geophys. Res. Lett.* 35, 1–6. <https://doi.org/10.1029/2008GL035143>.
- Rao, G.V., Reddy, K.V., Srinivasan, R., Sridhar, V., 2020. Spatio-temporal analysis of rainfall extremes in the flood-prone Nagavali and Vamsadhara Basins in eastern India. *Weather and Climate Extremes*, 100265. <https://doi.org/10.1016/j.wace.2020.100265>.
- Reboita, M.S., Gan, M.A., Rocha, R.P. da, Ambrizzi, T., 2010. Regimes de precipitação na América do Sul: uma revisão bibliográfica. *Revista Brasileira de Meteorologia* 25, 185–204. <https://doi.org/10.1590/s0102-77862010000200004>.
- Riveros, M.P., Camargo, H., Soares, W.R., Alves, L.M., Pabló, A.D., Rodriguez, D.A., Marengo, J.A., 2013. Two Contrasting severe seasonal extremes in tropical south America in 2012: flood in Amazonia and drought in northeast Brazil. *J. Clim.* 26, 9137–9154. <https://doi.org/10.1175/jcli-d-12-00642.1>.
- Santos, C.A.C. dos, Oliveira, V.G. de, 2017. Trends in Extreme Climate Indices for Pará State, Brazil. *Revista Brasileira de Meteorologia*. <https://doi.org/10.1590/0102-778632120150053>.
- Shi, W., Higgins, R.W., Yarosh, E., Kousky, V.E., 2000. The Annual Cycle and Variability of Precipitation in Brazil. NCEP/Climate Prediction Center.
- Song, X., Zhang, J., Zou, X., Zhang, C., AghaKouchak, A., Kong, F., 2019. Changes in precipitation extremes in the Beijing metropolitan area during 1960–2012. *Atmos. Res.* 222, 134–153. <https://doi.org/10.1016/j.atmosres.2019.02.006>.
- Souza, E.B. de, Kayano, M.T., Tota, J., Pezzi, L., Fisch, G., Nobre, C., 2015. On the influences of the El Niño, La Niña and Atlantic Dipole Paterni on the Amazonian rainfall during 1960–1998. *Acta Amazonica* 30, 305–318. <https://doi.org/10.1590/1809-43922000302318>.
- Souza, E.B. de, Carmo, A.M.C., Moares, B.C., Nacif, A., Ferreira, D.B. da S., Rocha, E.J.P., Souza, P.J.D.O.P., 2016. Sazonalidade Da precipitação sobre a Amazônia legal Brasileira: Clima Atual E Projeções Futuras Usando O Modelo Regcm4 (seasonal precipitation over the Brazilian legal amazon: climate current c future projections using Regcm4 model). *Revista Brasileira de Climatologia* 18. <https://doi.org/10.5380/abclima.v18i0.43711>.
- Sterenberg, F., 2012. With flood and drought, scientists are puzzled by extreme weather in Brazil's northern regions. GLOBO.
- Storvik, G., 2011. Numerical Optimization of Likelihoods: Additional Literature for STK210.
- Sturges, H.A., 1926. The choice of a class interval. *J. Am. Stat. Assoc.* 21, 65–66. <https://doi.org/10.1080/01621459.1926.10502161>.
- Tawn, J., Shooter, R., Towe, R., Lamb, R., 2018. Modelling spatial extreme events with environmental applications. *Spatial Statistics* 28, 39–58. <https://doi.org/10.1016/j.spasta.2018.04.007>.
- Taylor, S.J., Letham, B., 2017. Forecasting at scale. *PeerJ* 1–25. <https://doi.org/10.7287/peerj.preprints.3190v2>.
- Thornton, M., Pi, O.W., 2004. Temperature anomaly GIS Demonstration project. *Change* 1–17.
- Trenberth, K.E., 2013. Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. CAMBRIDGE UNIVERSITY PRESS, Cambridge, United Kingdom. <https://doi.org/10.5194/cp-6-379-2010>.
- Trenberth, K.E., Stepaniak, D.P., 2001. Indices of El Niño evolution. *J. Clim.* 14, 1697–1701.
- Trenberth, K.E., Dai, A., Rasmusson, R.M., Parsons, D.B., The Mendeley Support Team, Markoff, M.S., Cullen, A.C., Hamlet, A.F., Lettenmaier, D.P., Alexander, L.v., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Klein Tank, a.M.G., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, a., Rupa Kumar, K., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D.B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M., Vazquez-Aguirre, J.L., 2003. Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.* 84, D05109 <https://doi.org/10.1029/2005JD006290>.

- UN, 2007. Climate Change: Impacts, Vulnerabilities and Adaptation in Developing Countries., United Nations Framework Convention on Climate Change. <https://doi.org/10.1029/2005JD006289>.
- Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *J. Clim.* 23, 1696–1718. <https://doi.org/10.1175/2009JCLI2909.1>.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A.R.~J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, Ilhan, Feng, Y., Moore, E.W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.~A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., Contributors, S.1.0, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods.* <https://doi.org/10.1038/s41592-019-0686-2>.
- Wasserman, L., 2009. All of Statistics: A Concise Course in Statistical Inference, Technometrics. Springer Berlin Heidelberg. <https://doi.org/10.1198/tech.2004.s214>.
- Wu, H.T.J., Lau, W.K.M., 2016. Detecting climate signals in precipitation extremes from TRMM (1998–2013) - increasing contrast between wet and dry extremes during the “global warming hiatus. *Geophys. Res. Lett.* 43, 1340–1348. <https://doi.org/10.1002/2015GL067371>.
- Xavier, A.C.F., Blain, G.C., de Moraes, M.V.B., Sobierajski, G. da R., 2019. Selecting “the best” nonstationary generalized extreme value (Gev) distribution: on the influence of different numbers of GEV-models. *Bragantia* 78, 606–621. <https://doi.org/10.1590/1678-4499.20180408>.
- Xiao, C., Wu, P., Zhang, L., Song, L., 2016. Robust increase in extreme summer rainfall intensity during the past four decades observed in China. *Sci. Rep.* 6, 1–9. <https://doi.org/10.1038/srep38506>.