Revista de Matemática: Teoría y Aplicaciones **31**(1): 27-56
CIMPA - UCR ISSN: 1409-2433(Print), 2215-3373(Online)
DOI: https://doi.org/10.15517/rmta.v31i1.54869

WEATHER TYPES FOR THE SEASONAL TRANSITIONS IN CENTRAL AMERICA

TIPOS DE TIEMPO ATMOSFÉRICO DURANTE LAS TRANSICIONES ESTACIONALES EN AMÉRICA CENTRAL

Fernán Sáenz S.[1] Eric J. Alfaro[2] Hugo G. Hidalgo[3]

Received: 19/Jun/2023; Accepted: 30/Nov/2023

---

[1] Universidad de Costa Rica, Centro de Investigaciones Geofísicas (CIGEFI), San José, Costa Rica. E-Mail: fernan.saenzsoto@ucr.ac.cr
[2] Universidad de Costa Rica, Centro de Investigaciones Geofísicas (CIGEFI), Escuela de Física y Centro de Investigación en Ciencias del Mar y Limnología (CIMAR), San José, Costa Rica. E-Mail: erick.alfaro@ucr.ac.cr
[3] Universidad de Costa Rica, Centro de Investigaciones Geofísicas (CIGEFI), Centro de Investigación en Matemática Pura y Aplicada (CIMPA) y Escuela de Física, San José, Costa Rica. E-Mail: hugo.hidalgo@ucr.ac.cr

Abstract

Unsupervised learning techniques are employed to study the relationship between atmospheric circulation and precipitation over Central America and its surrounding areas. Specifically, the clustering algorithm k-means++ is applied to three coarse-grained datasets from ERA-interim reanalysis that are the candidates for representing the atmospheric state vector, each candidate contains its full temporal variability. Datasets are composed of: a) wind fields at 925, 800 and 200 hPa, b) same as "a)" plus convective available potential energy and c) same as "a)" plus total column water vapor. Clustering metrics, namely the variance ratio criterion, the silhouette criterion and the mean squared error, are computed to quantify clustering quality. Clusters are interpreted as weather types, recurrent configurations of the atmospheric state vector associated with observable weather states. The correct number of clusters for each dataset is determined with a Monte Carlo test of normality, to assure cluster existence. The main objective is to obtain a set of weather types containing elements that characterize the transition from and to the rainy season over the Pacific side of Central America as well as other elements of the seasonal cycle of regional precipitation, such as the Mid-Summer Drought. Besides the statistical metrics, in order to select between candidate datasets and plausible number of clusters, focus is given to the temporal characteristics of the clusters. Existing literature does not provide a set of weather types suitable to analyze seasonal transitions and the differences in the mechanisms associated with rainfall maxima.

Keywords: Central America; precipitation; weather types; cluster analysis; seasonal climate variability.

Resumen

Técnicas de aprendizaje no supervisado se emplean para estudiar la relación entre la circulación atmosférica y la precipitación sobre América Central y sus áreas circundantes. Específicamente, el algoritmo de agrupamiento k-means++ se aplica a tres conjuntos de datos de baja resolución del reanálisis ERA-interim, estos son candidatos a representar el vector de estado atmosférico y cada uno contiene su variabilidad temporal completa. Los conjuntos de datos probados son: a) campos de viento a 925, 800 y 200 hPa, b) lo mismo que "a)" más la energía potencial convectiva disponible y c) lo mismo que "a)" más el vapor de agua en la columna total. Se calculan métricas de agrupamiento, a saber, el criterio de relación de varianza, el criterio de silueta y el error cuadrático medio, para cuantificar la calidad del agrupamiento. Los grupos se interpretan como weather types, configuraciones recurrentes del vector de estado atmosférico asociadas con estados observables del tiempo atmosférico. El número correcto de grupos para cada conjunto de datos se determina con una prueba de normalidad de Monte Carlo para asegurar la existencia de grupos reales. El objetivo principal es obtener un conjunto de weather types que contengan elementos que caractericen la transición de y hacia la temporada de lluvias en la vertiente del Pacífico de América Central, así como otros elementos del ciclo estacional de precipitación regional, como las canículas. Además de las métricas estadísticas, para seleccionar entre conjuntos de datos y un número plausible de grupos, se presta atención a las características temporales de los grupos. La literatura existente no proporciona un conjunto de weather types adecuado para analizar transiciones estacionales y las diferencias en los mecanismos asociados con los máximos estacionales de lluvia.

Palabras clave: América Central; precipitación; tipos de tiempo atmosférico; análisis de conglomerados; variabilidad climática estacional.

Mathematics Subject Classification: Primary: 62H30; secondary: 86A10, 86-08.

1 Introduction

Precipitation is a crucial climatic factor in Central America (CA) because it directly affects the availability of water resources. Its variability significantly influences agricultural productivity, hydroelectric power generation, and freshwater availability for human consumption [40] [34]. Therefore, understanding the spatial and temporal variability of precipitation, which entails explaining its modulation by surrounding physical mechanisms, is essential for assessing and managing the impacts of climate variability and change on the region's water resources and ecosystems.

In general, the temporal variability of precipitation is characterized by variations across a continuous spectrum of time scales: from short-term variations (hours to days) to long-term trends (decades to centuries). According to [47], interannual climate variability in CA is mostly driven by El Niño-Southern Oscillation (ENSO) [77], the Atlantic Multidecadal Oscillation (AMO) [20] [24] and the Pacific Decadal Oscillation (PDO) [82]. On shorter time scales, the seasonal latitudinal migration of the solar radiation geographical maximum produces seasonal variations in the Sea Surface Temperature (SST) and these are related to variations in the regional atmospheric circulation and precipitation. These variations in the amount of rainfall that occur throughout the year are known as the seasonal cycle of precipitation. In general, it is one of the most relevant modes of the variability of precipitation within the tropics (ranging from 23.5°S to 23.5°N). In CA, the amount of rainfall can vary greatly depending on the season, with some seasons being characterized by heavy precipitation and others by dry conditions. These variations regulate agricultural practices and other socio-economic activities [37] [49] [40] [3]. For detailed reviews on the climate variability in CA see [4] [7] [47] [22].

The seasonal cycles of precipitation within the tropics are tightly connected with the seasonal variations of the atmospheric circulation [80] [10]. The atmospheric circulation transports water vapor and over oceanic surfaces may also enhance evaporation. These processes determine moisture availability and its vertical transport, both necessary conditions for precipitation to occur [11] [59]. Nevertheless, moisture arrangements that allow for organized convection to occur, in turn affect upper and lower tropospheric circulation patterns [72] [52] [51] [14].

The connection between the seasonal cycles of precipitation and lower tropospheric circulation over the tropical areas is known as the Global Monsoon [80] [77] [28] and is modulated by the heating gradients due to the unevenness of the solar radiation and its seasonal march. According to [28], the Global Monsoon is composed of the regional migrations of the Intertropical Convergence Zone (ITCZ), a belt of moist air and precipitation that converges within the larger circulation pattern of the tropics, known as the tropical atmospheric overturning circulation. However, [76] has pointed out that there are conceptual divergences pointing to underlying physical processes between these terms. For comprehensive analyses and technical details on the seasonal cycle of precipitation in the tropics see [10] and [50], while its relationships with the atmospheric circulation are addressed by [80].

The spatial variability of precipitation over this region is partly modulated by the mountain chain that crosses the isthmus from Northwest to Southeast defining two drainage basins, namely: the Pacific and Caribbean slopes [74] [57] [47] [57] [48]. A data-driven analysis through the application of Principal Component Analysis (PCA) [81] [19] to monthly precipitation series by [1] has shown that most of the spatial variability (80%) is accounted for by two regimes: one characteristic of the Pacific basin (72%) and the other of the Caribbean basin (8%).

The Pacific slopes are characterized by a marked contrast between wet and dry seasons. The wet season lasts from May to November-December with maximums during June and September and a relative minimum during July-August known as the Mid-Summer Drought (MSD) [46] [26]. Furthermore, the physical mechanisms responsible for these maxima differ [50] due to differences in large scale atmospheric stability and also to the energy exchanges at the ocean-atmosphere interface that are related with spatial variations of the sea-surface temperature (SST) [68]. This variability regime is tightly connected to the Global Monsoon or seasonal migrations of the ITCZ [80].

Over the Caribbean slopes, rainy conditions prevail throughout the year with mild seasonality featuring maxima during July and November, a minimum in March, as in the Pacific slope, and a Mid Autumn Decrease (MAD) in September-October [2]. This region, especially its areas in Costa Rica and Nicaragua, is subject to the direct impact of an intermittent east to west (easterly) flow located over the Caribbean Sea known as the Caribbean Low-Level Jet (CLLJ) [5] [6].

The CLLJ is vertically located between 925 and 700 hPa with a mean intensity that ranges from 12 to 14 m/s [27] and acts as moisture conveyor belt to CA from the Caribbean Sea, which is its main moisture source [21] [54]. According to [6], its seasonal cycle has its absolute (relative) maximum during July (January) and its absolute (relative) minimum during October (March). Periods of intense CLLJ favor mechanically forced convection over the Caribbean basin while favoring stable conditions (if present) over the Pacific slopes. On monthly time scales, the intensity of the CLLJ is negatively related with latitudinal center of mass of the ITCZ [35] affecting precipitation over the Pacific slopes. Furthermore, [26] found that the CLLJ boreal summer maximum is the main factor modulating the MSD. On interannual time scales, variations in CLLJ intensity, ENSO phase and precipitation anomalies are intertwined [6] [35].

Summing up, spatiotemporal patterns of SST, atmospheric circulation and precipitation interact in complex manners. These interactions are also modified by the topographic features of this region. Hence, it is key to identify and understand the regional scale covariability between these fields.

One way to describe the regional atmospheric circulation, its temporal variability and its relationships with other climate variables is by defining a set, or sets, of weather types that occur daily and are connected with changes in the spatial distributions of related variables [83] [9]. These weather types are recurrent configurations of the atmospheric state vector **(X)** associated with observable weather states. Statistically speaking, they may be viewed as cluttered regions of the atmospheric phase space or deviations from normality (multimodality) in the **X** multivariate probability density function (PDF)[15].

This work is concerned with finding a representation of $\mathbf{X_d}$ with its full temporal variability, that when partitioned, contains real clusters (weather types) characterized by seasonal cycles that help explain relevant features of the seasonal variability of precipitation in CA. Clusters representing features of key importance such as the seasonal transitions from dry to wet seasons and vice versa, the distinct physical mechanisms behind the annual maxima of precipitation over the Pacific slopes [50] and the MSD, are expected.

The clustering procedure is based on the k-means++ algorithm [45] [44] [8], which is recursively applied to a set of 3 approximations to $\mathbf{X_d}$. This procedure uses a Monte Carlo test for the normality of each $\mathbf{X_d}$ by also clustering surrogate datasets with their same temporal characteristics but drawn from Normal distributions [15]. Here the existence of clusters (weather types) will be assessed in conjunction with their relationships with precipitation variability. Specifically, their adequacy to describe the above-mentioned characteristics of the seasonal cycle of precipitation. Previous work [71] did not find any weather type characteristic for the dry to wet seasonal transitions neither a different weather type for the two rainfall maxima on the Pacific Slope, hence in this work additional variables and different normalization procedures are tested to see if those features can be identified in the new clustering output.

The implementation of circulation/weather types-based classifications has proven to be effective in analyzing atmospheric circulation in tropical Americas, revealing circulation patterns that exhibit significant seasonality. This has been demonstrated in studies by [13], [55], [70], [30], [61] and [71]. Table 2, located in Appendix A, summarizes the previous works applying clustering algorithms to dynamic variables in the Intra-Americas Seas.

This paper follows with a description of the datasets used for this research on section 2.1. Section 2.2 presents the methods used for defining the best approximation to the daily atmospheric state vector and its correct partitioning, computing the seasonal characteristics of the occurrence of the obtained clusters, and assessing the statistical significance of the cluster-conditioned atmospheric fields. In section 3 the results are presented, and these are discussed in section 4. The concluding remarks are presented in section 5.

2 Data and methods

2.1 Data

The representations of $\mathbf{X_d}$ studied in this work were composed of data extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis [18]. This dataset provides a high-quality, comprehensive representation of the Earth's atmosphere from 1979 to 2019, with a native horizontal resolution of 0.75° (approximately 80 km at the equator) and 60 vertical levels. This reanalysis incorporates data from various sources, including satellite and ground-based observations, which are physically balanced by the output from numerical weather prediction models. For this work we used wind fields at 200, 800 and 925 hPa, Convective Available Potential Energy (CAPE) and Total Column Water Vapour (TCWV).

The variable CAPE is the vertically integrated positive buoyancy of an atmospheric parcel ascending adiabatically [1] from the atmospheric level where buoyancy becomes positive to the level where buoyancy becomes neutral [36]. Hence, CAPE is related to the atmospheric updraft speed, and it is computed with the following equation [78]:

$$CAPE = \int_{p_{LFC}}^{nb} R_d \left(T_{vp} - T_{ve}\right) dlnp, \tag{2.1}$$

where $\mathbf{R_d}$ is the specific gas constant for dry air, $\mathbf{T_{vp}}$ is the virtual temperature [2] ($\mathbf{T_v = T(1 + 0.61r_v)}$), where $\mathbf{r_v}$ is the ratio of the mass of water vapor to the mass of dry air of the air parcel and $\mathbf{T_{ve}}$ is the virtual temperature of the environment.

The variable TCWV is the amount of water present in a vertical atmospheric column extending from the Earth surface to the top of the atmosphere. In ERA-interim it is computed from vertically integrating the specific humidity ($\mathbf{q}$), which is the mass of vapor in a unit of moist air, over the atmospheric column:

$$TCWV = \int_{p_s}^{p_{top}} q \, dp, \tag{2.2}$$

[1]In atmospheric physics, parcel theory is an abstraction that enables posing analytically solvable models of atmospheric convection. Adiabatic motions refer to motions without heat exchange between the parcel and its surrounding environment.

[2]Virtual temperature is a measure of temperature adequate to use dry-air equations with moist air.

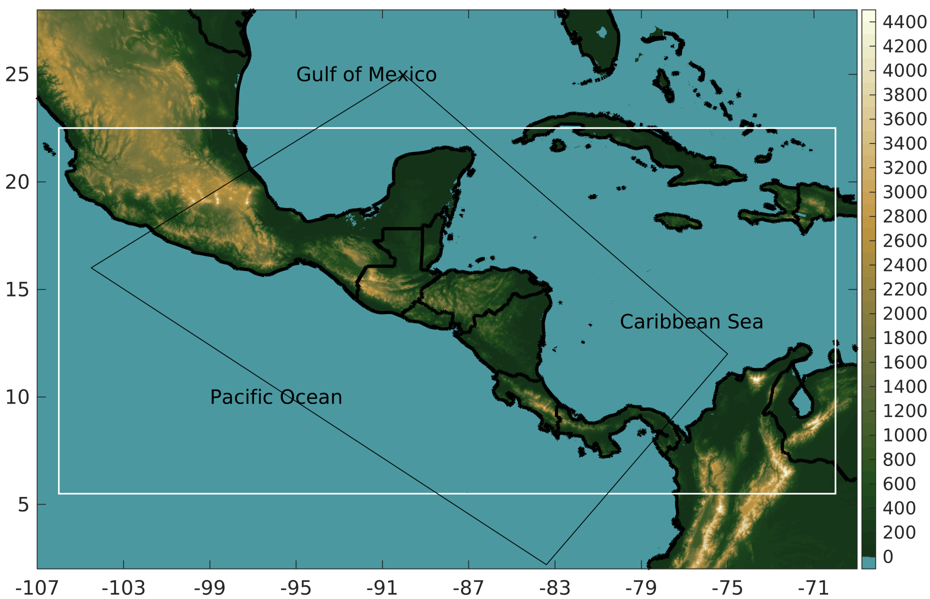

Figure 1: Study region. The white rectangle is the geographical domain of the clustering. The black polygon is the Southern Central America (SCA) region from the IPCC sub-regions [38].

where $\mathbf{p}$ is the atmospheric pressure ($\mathbf{p_s}$ at the surface and $\mathbf{p_{top}}$ at the top of the atmosphere).

The precipitation estimates were obtained from the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) dataset [25]. This is a global dataset that provides daily estimates of precipitation from 1981 to the present day, with a spatial resolution of 0.05° (approximately 5 km at the equator) and it is based on a combination of satellite and ground-based observations, specifically infrared and microwave sensors, rain gauges, and weather radar. For this study we employed a version of the CHIRPS dataset with a resolution of 0.25°, which has been already evaluated in our study region [73]. The grid points located inside the Southern Central America (SCA), Intergovernmental Panel for Climate Change (IPCC) subregion were selected [38]. For some computations, the dataset was split between Pacific and Caribbean grid points, derived from the Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0. [60].

The analysis period ranges from 1981 to 2015, it was chosen because it overlaps the ERA-interim and the CHIRPS periods with the best quality rain-gauge series over Central America [67] [71].

The geographical domain for the definition of the weather patterns is shown in Figure 1. This region allows for the incorporation of the effects of the CLLJ, the westerly winds from the eastern tropical Pacific ocean and the circulations forced during summer by the heating over the Sierra Madre in Mexico or Monsoon [72] [14].

2.2 Methods

2.2.1 Defining the approximations to $\mathbf{X_d}$

The procedure to determine weather types consists in finding an integer-valued surrogate variable that represents the affiliation of the daily atmospheric vector space ($\mathbf{X_d}$) to a partition of $\mathbf{X_d}$ space (coarse-graining) [29]. Hence, it poses the problem of finding a finite set of variables suitable to represent the practically infinite $\mathbf{X_d}$. Usually, $\mathbf{X_d}$ is represented by atmospheric variables of dynamic nature such as daily wind fields at specific vertical levels [13] [70], mean sea-level pressure [61] or geopotential height fields. Nonetheless, variables reflecting the thermodynamic state of the study region may also be included [55] [79] [71].

In this work, the best approximation to $\mathbf{X_d}$ is selected between a set of 3 combinations of variables that can be purely dynamic: wind fields at lower (925 and 800 hPa) and upper tropospheric levels (200 hPa) or TCWV and CAPE with daily resolution. In these datasets, we search for the smallest number of clusters that produce a seasonal distribution adequate to be related with the seasonal precipitation characteristics described above. The 3 possible representations of $\mathbf{X_d}$ tested in this work were:

1. Wind fields at 200, 800 and 925 hPa.
2. Wind fields at 200, 800 and 925 hPa + CAPE.
3. Wind fields at 200, 800 and 925 hPa + TCWV.

Prior to the coarse-graining phase, preprocessing routines such as time and/or space normalization [32], or standardization [19] may be applied to emphasize the spatiotemporal scales relevant for the study in question. Furthermore, some techniques for dimensionality reduction may also be applied to minimize the sampling issues associated with the large dimensionality of the representations of $\mathbf{X_d}$ [31]. However, some works neglect either or both steps. Here, each representation of $\mathbf{X_d}$ was subjected to a preprocessing phase that consist in:

1. Normalizing each variable by the square-root of the area-mean temporal standard deviation to reduce the influence that the data exerts in areas where the temporal variability is larger, has on the subsequent data analysis methods.
2. Each time series was multiplied by the cosine of its respective latitude [16]. This unifies the spatial representativity grid boxes, which are affected by the cartographic projection.

3. After the previous steps, each time series was centered by the removal of its temporal mean and each new representation of $\mathbf{X_d}$ was mapped onto a lower-dimensionality space via PCA by applying a Singular Value Decomposition implemented in the Python package EOFs [17]. The principal components that contained approximately 75% of the variability were retained.

2.2.2 Coarse-graining (clustering)

There exist a great variety of classification methods to coarse-grain a spatio-temporal dataset. For example, [65] listed and grouped 35 methods into 8 groups: subjective methods, threshold-based methods, based on principal component analysis, leader algorithms, hierarchical cluster analysis, optimization algorithms, mixture models and random-process. However, optimization algorithms such as k-means, self-organizing maps [53] and simulated annealing [66] are commonly used.

In this work, each preprocessed $\mathbf{X_d}$ was coarse-grained using a version of the unsupervised machine learning algorithm k-means that is initialized by a particular procedure that carefully selects the seeds for the original algorithm [8]. The method is known as k-means++ and is available in the Python package scikit-learn [64]. This overcomes one of the major drawbacks of the original algorithm: its sensitivity to seeding. The implementation used in this study works as follows:

1. From the $\mathbf{N}$ vectors to be clustered, let $\mu_{\mathbf{1}} = \mathbf{x_j}$, where $\mathbf{j}$ is a randomly chosen index between 1 and N, be the first initial centroid (seed).

2. For i = 2 to N (number of vectors):

   (a) Compute $\mathbf{d(x_i)^2}$, the minimal distance from vector $\mathbf{x_i}$ to each $\mu_{\mathbf{k}}$ centroid already chosen:
   $$d(x_i)^2 = min\left(||x_i - \mu_k||^2\right). \tag{2.3}$$

   (b) Compute the probability of selecting each vector $\mathbf{x_i}$ as the next centroid:
   $$p(x_i) = \frac{d(x_i)^2}{\sum_{j=1}^{N} d(x_i)^2}. \tag{2.4}$$

   Select the next centroid $\mu_{\mathbf{i}}$ with probability proportional to $\mathbf{p(x_i)}$.

3. Repeat until there is a seed for each $\mathbf{k}$ cluster.

4. Proceed with standard k-means, which works as follows:

   (a) With the initial (seeds) centroids $\mu_{\mathbf{k}}$, assign each vector to the nearest centroid using the Euclidean distance formula:
   For $\mathbf{i = 1}$ to $\mathbf{N}$, assign $\mathbf{x_i}$ to the cluster $\mathbf{j}$ that minimizes $||\mathbf{x_i} - \mu_{\mathbf{j}}||^{\mathbf{2}}$.

(b) Recalculate the cluster centers as the mean of the vectors assigned to them:

For $\mathbf{j = 1}$ to $\mathbf{k}$:

$$\mu_j = \frac{\sum_{i=1}^{s_j} x_i}{s_j}. \tag{2.5}$$

Where $\mu_{\mathbf{j}}$ the new cluster center.

(c) Repeat steps (a) and (b) until the cluster assignments no longer change.

The selection of the quantity of clusters is not a trivial question. This issue has been addressed at large by [15] in the context of analyzing circulation regimes away from the tropics. In the extratropics, circulation regimes are viewed as quasi-stationary or persistent anomalies with residence time scales larger than transition scales [63]. The existence of real clusters in a dataset, that are not artifacts of applying a clustering algorithm to a finite sized sample, requires the rejection of the null hypothesis of normality in the $\mathbf{X_d}$ PDF [15] and weather regimes correspond to bumps where the PDF deviates from normality. If clusters are found on $\mathbf{X_d}$, their representation of weather types or regimes is subjectively verified analyzing their characteristics. However, [15] remarked that the rejection of normality is not sufficient, and the skewness of the dataset should be investigated because highly skewed data can produce spurious clusters.

In the tropics, the theory of circulation regimes has not been developed, however, the application of clustering algorithms to find weather types suitable to analyze the atmospheric circulation, its variability and relationships to other variables is common (e.g. [56] [42] [62]). Furthermore, [58] employed a set of 30 observed weather types and their analogues in model predictions to produce seasonal precipitation forecast over India.

In this work, the theory of [15] was applied to assure statistical significance of clusters. The definitive representation of $\mathbf{X_d}$ and its optimal partitioning were selected by testing the null hypothesis that the clusters were mere artifacts of the application of the algorithm to finite sized data in PC-space that is in fact normally distributed. The alternative hypothesis is that the clusters exist and are the regions where the PDF deviates from normality. The procedure is described in the next section.

2.2.3 Definition of the optimal $\mathbf{X_d}$ and its partitioning

The procedure followed for the definition of the optimal representation of $\mathbf{X_d}$ and its optimal partitioning is outlined next. First, for each $\mathbf{X_d}$ candidate a set of 200 surrogate normally distributed datasets [75] with the same spectral features as the original data were generated using the Fourier phase randomization technique. This can be achieved by randomly shuffling the phase values of the Fourier transform. Mathematically, given a vector $\mathbf{x}$ of length $\mathbf{N}$, its Fourier transform $\mathbf{FT(k)}$

can be expressed as:

$$FT(k) = \sum_{n=0}^{N-1} x(n)\, e^{-2\pi \frac{ink}{N}}, \tag{2.6}$$

where **i** is the imaginary unit and **k** is the frequency index. The Fourier phase randomization method involves randomizing the phase of $\mathbf{FT(k)}$ while preserving its magnitude [23]. This can be achieved by multiplying each complex Fourier coefficient $\mathbf{FT(k)}$ by a complex number of unit magnitude, $\mathbf{e^{i\theta_k}}$, where $\theta_\mathbf{k}$ is a random phase angle for each frequency index **k**.

The randomized Fourier coefficients, $\mathbf{Y(k)}$, can be expressed as:

$$Y(k) = |FT(k)| e^{i\theta_k}, \tag{2.7}$$

where $|\mathbf{FT(k)}|$ is the magnitude of the original Fourier coefficient. To obtain the randomized signal $\mathbf{y(n)}$, the inverse Fourier transform is applied to $\mathbf{Y(k)}$:

$$y(n) = \frac{1}{N} \sum_{k=0}^{N-1} Y(k) e^{2\pi \frac{nk}{N}}, \tag{2.8}$$

which results in a vector that has the same magnitude as the original $\mathbf{x(n)}$, but with a randomized phase.

The k-means++ algorithm was then executed 100 times for each value between 3 and 9 on both the original data and each of the 200 surrogate datasets. For each of these executions the variance ratio criterion (VRC)[12], silhouette criterion (SC)[69] and mean-square-error (MSE) were calculated. The smallest k for which the candidate $\mathbf{X_d}$ outperforms the 99% of the surrogates in all metrics is taken as its optimal partitioning factor. Then the collection of parameter vectors describing the clusters ($\boldsymbol{\Theta} = \{\theta_1, ..., \theta_c\}$) is analyzed to find if its seasonality exhibits the required characteristics, as described in the following section.

The VRC criterion or Calinski-Harabasz score, is a metric used to evaluate the quality of clustering results in k-means++ clustering. It measures the ratio of the between-cluster dispersion and the within-cluster dispersion, which indicates how well separated the clusters are. The formula for VRC is:

$$VRC(k) = \frac{B(k)\frac{k}{k-1}}{W(k)\frac{k}{N-K}}, \tag{2.9}$$

where **k** is the number of clusters, **N** is the total number of data points, $\mathbf{B(k)}$ is the between-cluster sum of squares, which measures the variability between the means of the clusters and $\mathbf{W(k)}$ is the within-cluster sum of squares, which measures the variability within the clusters.

The silhouette score measures the similarity of a vector to its own cluster compared to other clusters, and ranges from -1 to +1. A score of +1 indicates that a vector is well-matched to its own cluster and poorly matched to neighboring clusters, while a score of -1 indicates the opposite. A score of 0 indicates that the object is like its own cluster and neighboring clusters. The formula for the silhouette score for a single object is:

$$s_i = \frac{b_i - a_i}{max\{a_i, b_i\}}, \tag{2.10}$$

where $\mathbf{s_i}$ is the silhouette score for vector $\mathbf{x_i}$, $\mathbf{a_i}$ is the average dissimilarity between $\mathbf{x_i}$ and all other vectors in its own cluster, and $\mathbf{b_i}$ is the minimum average dissimilarity between $\mathbf{x_i}$ and all vectors in any other cluster. The $\mathbf{max\{a_i, b_i\}}$ is used to normalize the score between -1 and +1. Here the average silhouette score across all objects is used:

$$S = \frac{1}{N}\sum_{i=1}^{N} s_i, \tag{2.11}$$

where **N** is the total number of objects in the dataset. The MSE metric refers to the error done when replacing each vector in PC space by the centroid of the cluster it belongs to. This can be expressed as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N-1} (x_i - c_k)^2\,, \tag{2.12}$$

where $\mathbf{c_k}$ is the centroid of the cluster where $\mathbf{x_i}$ was assigned.

2.2.4 Annual cycles of occurrence for the optimal X partition clusters (weather types)

The condition that our optimal $\mathbf{X_d}$ partition should have at least one cluster characteristic of the beginning of the rainy season and at least another for its ending implies that the partitionings done for this work would result in clusters characterized by seasonality. This assumption stems from the fact that previous work (e.g. [71]) have shown that if the annual cycles are not removed from the **X** representation, the clusters obtained from the partitioning do have marked seasonalities. The annual cycles or climatologies of the probabilities of occurrence were computed by computing the frequency of occurrence of each cluster, for each Julian day (1-365 or 1-366 for leap-years), during the 1981-2015 period. These series were smoothed with a 10-day moving average. Weather types were numbered by the order of occurrence of their climatological annual maximum.

2.2.5 Assessing field significances

The conditional composites of the standardized anomalies of precipitation were computed to assess if a cluster from the partitioning of $\mathbf{X_d}$ is characterized by wet or dry conditions over the Pacific, Caribbean or both slopes of the isthmus. Over each sub-region, wet (dry) weather types are defined as those with more (less) than 75 (25)% of grid points characterized by positive (negative) precipitation mean standardized anomalies, transition weather types are those between 25% to 75% of positive mean standardized anomalies. The statistical significance of these fields is tested with a Monte Carlo approach with 1000 trials, analogous to the applied by [41]. For each cluster a sample with its same quantity of elements was randomly drawn from the precipitation data and the sample mean was computed and stored. The procedure was repeated 1000 times, if the original conditional composite of standardized anomalies were larger than the 97.5 sample percentile or smaller than the 2.5 percentile, the conditional composite is significant at the 0.05 level. The same procedure was applied to all the fields shown.

3 Results

3.1 Best dataset ($\mathbf{X_d}$) and its optimal partitioning

3.1.1 Clustering: significance and quality

Figure 2 shows, for each dataset, all clustering metrics plotted as functions of the number of clusters. Also, the value of the 99^{th} percentile of the corresponding set of surrogates for VRC and SI and the 1^{th} percentile in the MSE case, are plotted. These plots show that all datasets can be partitioned in 6 clusters, rejecting the null hypothesis of spurious clustering. Hence, we take 6 clusters as an adequate solution for all $\mathbf{X_d}$ representations.

An analysis of the 3 $\mathbf{X_d}$ representations partitioned into 6 clusters, based on the selected metrics shows that the partition of WND+CAPE outperforms the remaining 6-cluster partitions. It is characterized by better quality clustering than the other partitions (larger VRC), its clusters are more self-similar that those from the other partitions (larger SI) and its centroids are more representative of their respective clusters than those of the other partitions (smaller MSE).

3.1.2 Weather types seasonality

Once the number of clusters was defined for each dataset and a candidate dataset (WND+CAPE) had been selected, the seasonal cycles of the probabilities of occurrence of the weather types from this dataset were computed, and are displayed in Figure 3. This figure shows that the seasonal cycles are characterized by high seasonality, i.e. periods of high frequency occurring in a restricted set of months and periods of almost no activity restricted to another set of months. Four weather types feature a unimodal seasonal cycle, hence their occurrence tends to peak at a certain time of the year surrounded by periods of increasing and decreasing frequency (weather types numbered 1, 2, 3 and 4). The 2 remaining weather types (5 and 6) feature bimodal seasonal cycles, characterized by relative maximum that tends to occur before the absolute maximum. This maximum surrounds a relative minimum of frequency of occurrence.

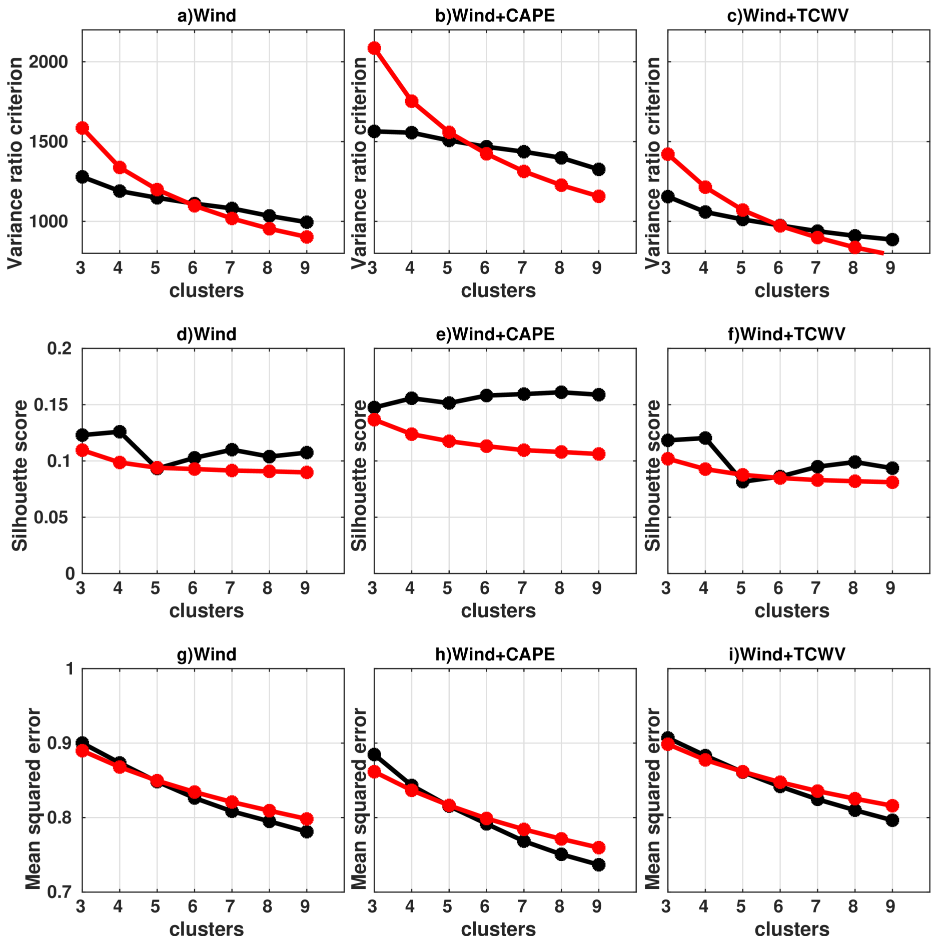

Figure 2: Clustering metrics. Variance ratio criterion (VRC; panels a, b and c), silhouette scores (SI; panel d, e and f) and mean squared errors (MSE; panels g, h and f). Each metric was computed for the clustering, with k values from 3 to 9, of each dataset (black dots and lines) and its respective 200-member surrogate dataset. For VRC and SI (MSE), red dots and lines represent the 99th (1th) percentile of the values computed from the clustering of the surrogate datasets.

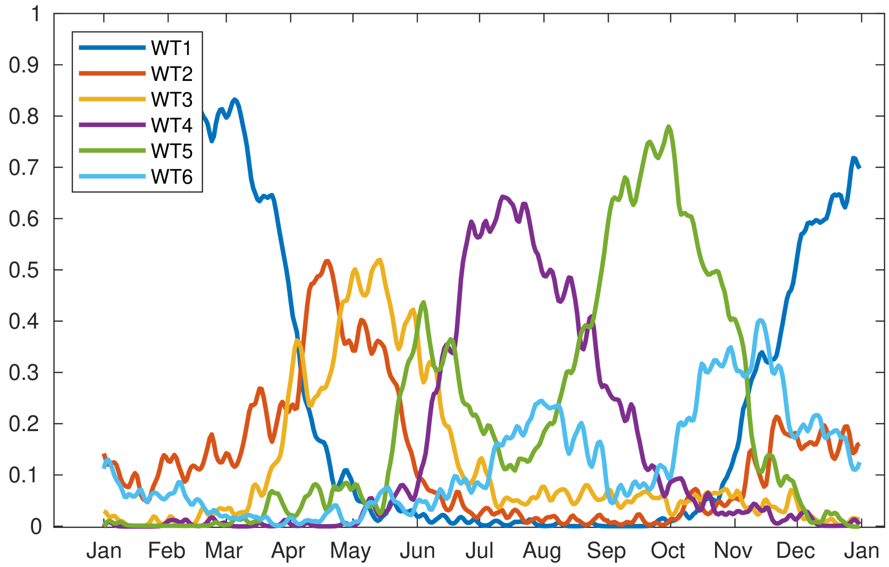

Figure 3: Climatology of observed occurrence probabilities for each cluster of the optimal X_d partition. Each line represents the smoothed (10-day moving average) probabilities of occurrence of each weather type for each Julian day.

The relative wetness of the weather types from all datasets is summarized in Table 1. Here the areal percentage of positive standardized anomalies of conditional precipitation over both the Pacific and the Caribbean is tabulated.

	X_1	X_2	X_3
WT1	0.0|0.0	0.0|0.0	0.0|0.0
WT2	0.0|0.0	0.0|1.2	0.0|0.0
WT3	0.0|0.0	65.3|14.4	0.0|0.0
WT4	100.0|99.7	100.0|99.2	100.0|99.7
WT5	100.0|97.4	100.0|99.1	100.0|97.3
WT6	3.0|15.5	30.5|80.9	6.8|19.1

Table 1: Spatial percentage of positive mean conditional standardized anomalies of precipitation for each weather type. Only anomalies significant at a 0.05 level are accounted. The gridded precipitation dataset was divided between basins using the GLOBE Digital Elevation Model [60] as a reference.

Table 1 shows that only WND+CAPE has weather types that characterize the seasonal transitions over the Pacific in a way that is consistent with precipitation anomalies. Recalling that in subsection 2.2.6 we defined a spatial criterion for transition weather types: those with between 25% and 75% of the total basin area with positive anomalies of conditional precipitation. From this dataset, WT3

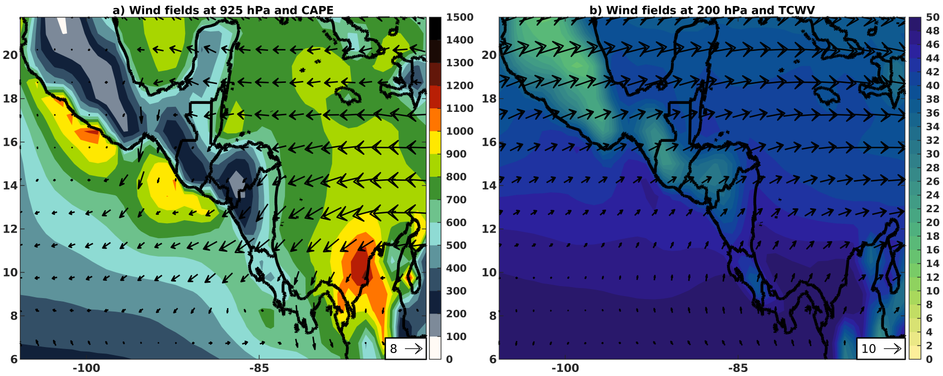

Figure 4: Climatological mean daily fields (1981-2015). a) Wind fields at 925 hPa (vectors) [$\mathbf{ms^{-1}}$] and CAPE (shaded) [$\mathbf{JKg^{-1}}$]. b) Wind fields at 200 hPa (vectors) [$\mathbf{ms^{-1}}$] and TCWV (shaded) [**mm**]

and WT6 present spatial distributions of precipitation consistent with the above definition. Then its partitioning into 6 clusters is selected as the definitive $\mathbf{X_d}$ partition. From a perspective of precipitation, WT3 and WT6 have fractions of positive anomalies that imply mixed signals. While WT1 and WT2 (WT4 and WT5) are purely dry (wet) weather types. Furthermore, the transition weather types, as opposed to the wet and dry weather types, feature opposed fractions of wetness between the Pacific and Caribbean basins. The transition from dry-to-wet season (WT3) is characterized by a Pacific (Caribbean) slope dominated by wet (dry) conditions. The transition from wet-to-dry in the Pacific features a wetter Caribbean.

3.2 Spatial characteristics of the weather types

3.2.1 Mean climatological conditions

The climatological daily values of the wind fields at 925 and 200 hPa levels, CAPE and TCWV are shown in Figure 4. These fields are shown as reference, because the weather types are discussed in terms of their conditional standardized anomalies.

Figure 4.a shows that the CA region is climatologically subject to an easterly wind regime, the trade winds, that are enhanced by the presence of the CLLJ. This is coherent with the picture presented in the same panel for CAPE. Enhanced winds at lower tropospheric levels induce evaporation over the oceanic areas: larger water vapor mixing ratios and then larger CAPE values than over land areas.

Figure 4.b shows that, climatologically, upper tropospheric wind fields over CA region are westerly with a south-north gradient of its magnitude, connected to the northern hemispheric westerly regime. The climatological distribution of TCWV features north-south gradient reflecting the effects of the trade winds and

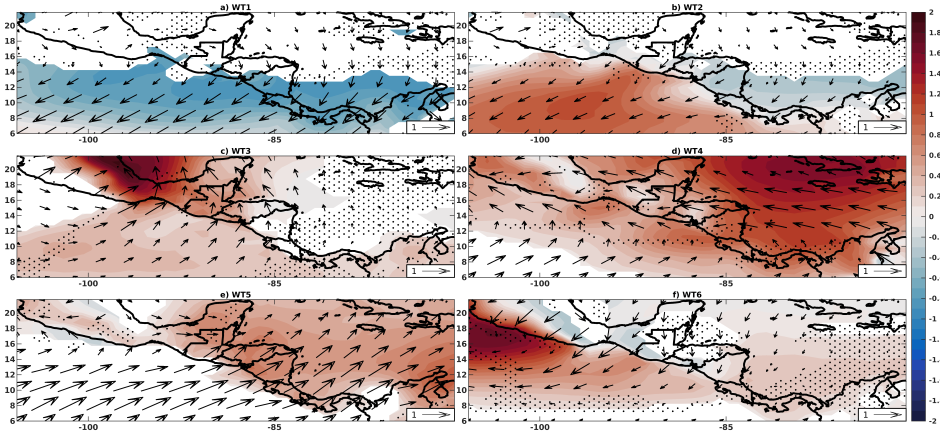

Figure 5: Conditional mean standardized anomalies of wind fields at 925 hPa level (vectors) and CAPE (shaded). Stippling means NOT significant wind anomalies. Only significant CAPE anomalies are drawn.

the Inter-Tropical Convergence Zone (ITCZ). Furthermore, there is a land-sea contrast reflecting oceanic evaporation and atmospheric column thickness.

3.2.2 Conditional atmospheric circulation and thermodynamic features

The mean conditional standardized anomalies of the wind fields at 925 hPa and CAPE are shown in Figure 5 while the same computations for wind fields at 200 hPa and TCWV are shown in Figure 6.

Dry season weather types, WT1 and WT2, are characterized by an anticyclonic anomalous wind field at 925 hPa that signal an enhancement of the easterly wind regime and the influence of episodic northerly winds (Figure 5.a and Figure 5.b). While at 200 hPa level the anomalous wind fields are southwesterly, which imply an enhancement of the upper-tropospheric westerly regime connected to the midlatitudes circulation (Figure 6.a and Figure 6.b). For both circulation types, CAPE anomalies are only significant South of nearly 14°N but differ between the two weather types. For WT1, CAPE significant anomalies are mostly negative while for WT2 they are mostly positive, especially over the eastern tropical Pacific. In terms of atmospheric column moisture, both weather types feature negative TCWV anomalies, with WT1 being the driest weather type.

The WT3 is characteristic of the transition from dry-to-wet, its wind field at 925 hPa shows southwesterly anomalies over the eastern tropical Pacific and the northmost portion of the isthmus, that converge over the Gulf of México, where there are large positive CAPE anomalies (Figure 5.c). Column moisture presents mostly neutral anomalies over the domain, with patches of positive anomalies south of 8°N, where the ITCZ is located and over the northern part of the CA isthmus

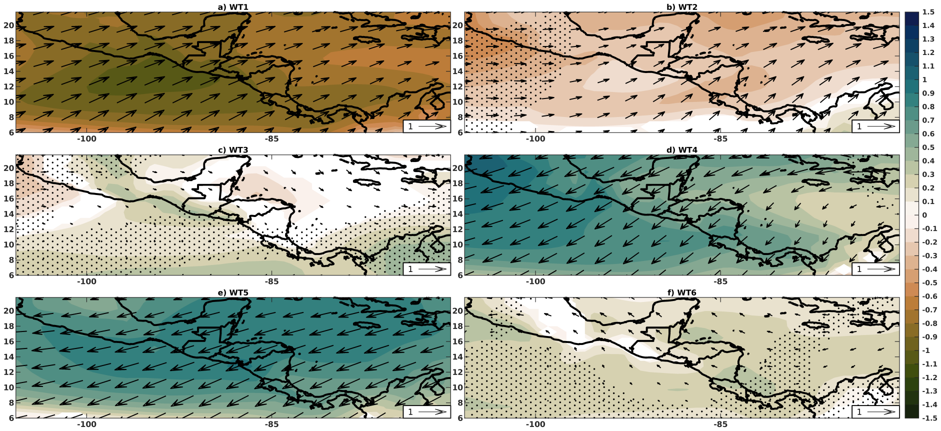

Figure 6: Conditional mean standardized anomalies of wind fields at 200 hPa level (vectors) and TCWV (shaded). Stippling means NOT significant wind anomalies. Only significant TCWV anomalies are drawn.

(Figure 6.c). Over the Caribbean there are no significant anomalies neither of low-level winds nor CAPE. The 200 hPa level wind field is close to its climatological values (Figure 6.c).

The weather types characteristic of the established rainy season, WT4 and WT5, feature conditional anomalies with similar structures for CAPE, TCWV and upper tropospheric circulations. For these weather types, CAPE anomalies are mostly positive with larger anomalies over the Caribbean (Figure 5.d and Figure 5.e), TCWV anomalies are also positive (Figure 6.d and Figure 6.e) and the upper tropospheric circulation features easterly anomalies. The main difference between these weather types is in the 925 hPa level circulations: WT4 is characterized by large southeasterly anomalies over the Caribbean and the northern areas of CA while WT5 is characterized by southwesterly anomalies over the Pacific, the southern areas of CA and the Caribbean.

The WT6 is characteristic of the wet-to-dry seasons transition, its significant wind anomalies on the 925 hPa level are northeasterly (Figure 5.f). CAPE anomalies are mostly positive over the ocean while negative anomalies are found over the northern region of CA (Figure 5.f). As with WT3, TCWV anomalies are small and positive and the 200 hPa level wind field is close to its climatological values (Figure 6.f).

3.2.3 Conditional precipitation fields

The conditional mean standardized anomalies for the weather types are presented in Figure 7 while Figure 8 shows the difference between the mean precipitation fields for climatologically consecutive weather types.

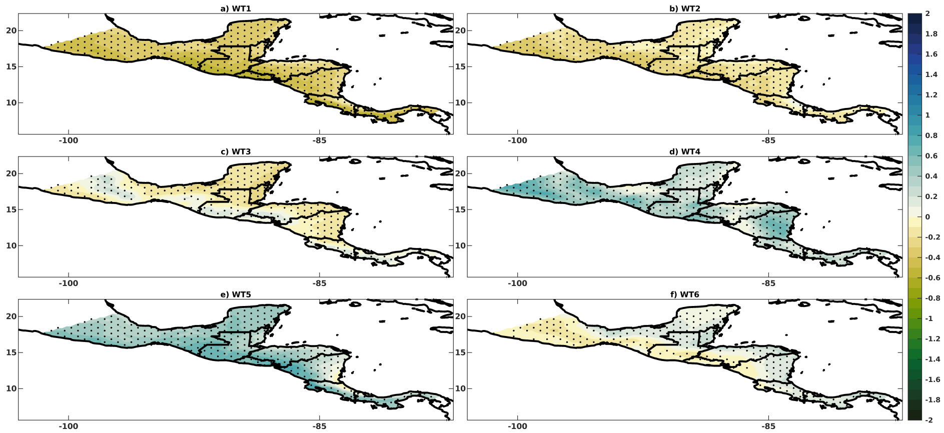

Figure 7: Conditional mean standardized anomalies of precipitation for each weather type. CHIRPS over SCA. Black dots denote significant anomalies at 0.05 from a Montecarlo test.

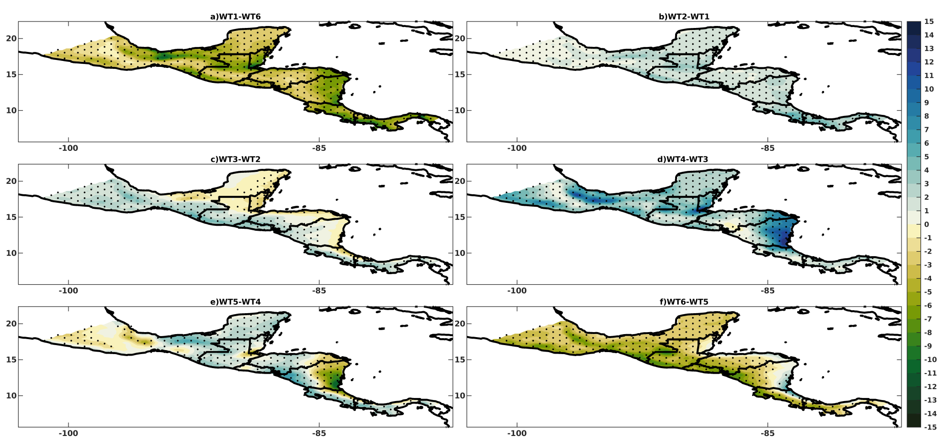

Figure 8: Differences between mean conditional precipitation of climatologically subsequent weather types [**mm**].

Dry season weather types, WT1 and WT2, feature negative precipitation anomalies along the entire domain (Figure 7.a and Figure 7.b). However, WT2 is characterized by wetter conditions (Figure 8.b), with greater difference over southeastern Costa Rica and Panama.

Weather type WT3 is characterized by positive precipitation anomalies over the coastal regions of the Pacific slopes of CA and negative anomalies over the Caribbean (Figure 7.c). The difference with WT2 (Figure 8.c) highlights this contrast. The wet season WT4 is characterized by positive precipitation anomalies along the entire domain, except for small patches over the Nicaragua, Honduras and southernmost Gulf of México coast (Figure 7.d). However, it features drier conditions than WT3 in some regions along the Pacific slope of CA (Figure 8.d). The other wet season weather type, WT5 is also characterized by generalized positive precipitation anomalies, except for a small patch along the Caribbean coast of Nicaragua and Costa Rica (Figure 7.e). The difference between these weather types shows that WT5 is characterized by wetter conditions than those of WT4 for most CA, except in southern México and the Caribbean coasts of Nicaragua and Costa Rica, where WT4 is significantly wetter.

Weather type WT6 features positive precipitation anomalies over the Caribbean slopes of CA and across the whole isthmus south of 10°N. Negative anomalies are found along the Pacific slopes. The differences in mean precipitation with respect to WT5 show that this is a weather type characteristic of the transition from wet-to-dry season over the Pacific but not over the Caribbean where WT6 intensifies the wet conditions.

4 Discussion

The first phase of this work consisted of the selection of a dataset, and its partition into clusters, that provided the best statistical qualities and some climatological characteristics. The Monte Carlo tests rejected the hypothesis of the unimodality of the selected $\mathbf{X_d}$ probability distribution (Figure 2, panels b, e and h) for all cluster numbers larger than 6. However, the principle of parsimony, that urges to explain natural phenomena using the smallest possible number of elements, was evoked to select the definitive number of clusters. Furthermore, none of the 9 PCs that contained 75% of the variability has a skewness with absolute value larger than 0.33. Hence the skewness of the distributions is discarded as the cause underlying the existence of clusters. The dataset composed of wind fields at 925, 800 and 200 hPa and CAPE was suitable to be partitioned into 6 clusters with better quality, more self-similarity and more representativeness than the corresponding to the remaining datasets.

The climatological constraints that were imposed are: that some weather types should represent the transition to and from the rainy season over the Pacific slopes of CA and that different weather types should represent the rainy season maxima over this region. Weather Types from the above-mentioned dataset feature these climatological characteristics. WT3 represents the transition from dry to wet season (Figure 3), the first precipitation maximum that tends to occur during June is produced by the alternation of WT3, WT4 and WT5. While WT5 dominates the circulation during the second precipitation maximum on the Pacific slope in concordance with the MAD on the Caribbean slope in September-October. Previous work [71] did not feature such weather types. Assuming that ERA-interim and ERA5 reanalysis datasets are consistent, differences in data resolution and spatial standardization may explain this contrast. Compared to previous works presenting weather types over CA and its surroundings, the partition presented in this work and that of [71] used the fewest weather types (6) to describe the regional atmospheric circulation variability. Seven circulation types were presented by [13], [55] found a solution with 8 weather types, [70] presented 11 circulation types and [61] presented 20 circulation patterns but for a synoptic scale domain much larger than the previously presented. See Table A.1 for details about these works. The weather types presented in this work feature varying degrees of consistency with the previously reported classifications over the CA region.

WT1 features the lower-level circulation with easterly anomalies that resembles the winter maximum of the CLLJ. This weather type is similar in its spatial and temporal characteristics to CT3 and CT4 from [13], weather regime #2 from [55], Winter North Eastern Winds (WNEW) regime from [70], WT20 from [61] and WT1 from [71]. The conditional lower-tropospheric circulation of WT2 features anomalies with signs similar to those of WT1 but smaller in magnitude. CAPE anomalies are larger in WT2, consistently with wetter conditions. This weather type is similar in its spatial and temporal characteristics to CT1 and CT2 from [13], weather regime #1 from [55], Spring NASH West (SPNW) regime from [70], WT13 from [61] and no weather type from [71].

The transition weather type WT3 is characterized by anomalies that imply a de-intensification of the trade winds regime, which reduces ventilation and allows for convection to occur. This weather type is similar in its spatial and temporal characteristics to weather regime #6 from [55], East North Atlantic High (ENAH) regime from [70] and WT5 from [61].

WT4 is characterized by enhanced 925 hPa easterlies over the Caribbean with a seasonal cycle of occurrence that peaks in July. Hence, it represents the summer enhancement of the CLLJ and the MSD, recalling that the first maximum of precipitation over the Pacific slope is represented by the alternation of WT3 and WT5; WT4 is drier than WT5 (Figure 8.e). This pattern is consistent across the whole range of classifications: CT7 from [13], weather regime #4 from [55], Summer Low Level Jet (SLLJ) regime from [70], WT5 from [61] and WT5 from [71].

Weather Type WT5 represents the most intense conditions of the rainy season with cyclonic anomalies over CA and a seasonal cycle that follows the bimodal distribution characteristic of the rainfall over the Pacific slope of CA and is associated also with the MAD in the Caribbean slope [2]. This weather type is also consistent across the whole range of classifications: CT6 from [13], weather regime #5 from [55], Summer Monsoonal Winds Regime (SMWR) regime from [70] and WT4 from [71]. In the synoptic classification of [61] it is divided into 4 weather types: WT3, WT4, WT6, and WT7, showing that westerly anomalies over CA can arise under various synoptic regimes.

Weather Type 6 is characterized by northeasterly 925 hPa anomalies over the northern areas of CA and neutral anomalies elsewhere. This pattern could be related to periods of lessening of CLLJ after the MSD and intensification after the second peak of WT5, hence signaling the transition out of the rainy season. This weather type has no analog in either of the referred studies.

From the above comparison, it is clear that dry season weather types, as well as transition ones, have a degree of similarity between the different classifications. However, only wet season patterns, here WT4 representing the summer CLLJ and WT5 representing cyclonic anomalies over CA, show a consistent signal across classifications. Hence, this technique may be useful to assess the capacity of climate models to represent the circulation patterns representative of the rainy season over CA.

5 Conclusions

In this study, we applied k-means++ to three datasets that were candidates for representing $\mathbf{X_d}$, the daily atmospheric state vector time series; hence, each dataset is coarse-grained into an integer-valued time series that represents the belonging to a certain cluster generated from $\mathbf{X_d}$. This is interpreted as the daily occurrence of a specific weather type. A Monte Carlo test was applied to ensure that deviations from normality existed in each $\mathbf{X_d}$ representation and, for each, the smallest number of clusters that assure non-normality was selected. The dataset composed of wind fields at 925, 800, and 200 hPa and CAPE was suitable to be partitioned into 6 clusters with better quality, more self-similarity, and more representativeness than the remaining datasets.

From this procedure, a dataset composed of wind fields at 925, 800, and 200 hPa and CAPE from the ERA-interim reanalysis, partitioned into 6 clusters was selected as the coarse-grained representation of $\mathbf{X_d}$. This dataset provided weather types for the dry and wet seasons as well as the seasonal transitions. The wet season weather types are consistent with some of the other classifications reported for the region [13] [55] [70] [61] [71]. Furthermore, the weather types characteristic of the seasonal transitions (WT3 and WT6) have no clear analog in the literature and provide the opportunity to study the typical atmospheric circulation patterns for these transitions.

Acknowledgements

The authors wish to acknowledge the funding of this research through the following Vicerrectoría de Investigación, Universidad de Costa Rica grants: C3991 (UCREA), C3195, C2806, C2103, C0130, B9454 (VI-Grupos), A5719 (Mini Congreso -CIGEFI), A4906 (Programa de Estudios Sociales de la Ciencia, la Técnica y el Medio Ambiente), A1715 y B0810. Eric Alfaro wishes to acknowledge the Escuela de Física, Vicerrectoría de Docencia, Universidad de Costa Rica, for his support during his sabbatical semester.

A Previous applications of cluster analysis to extract weather/circulation types over Central America

Reference	Variables	Latitude Longitude	Spatial domain	Period	Sources	Number of clusters
[13]	Wind fields at 850 hPa	2.5∘ x 2.5∘	30∘N 0∘N 110∘W 40∘W	1979-2010	NCEP/DOE reanalysis [39]	7
[55]	Wind fields at 925 hPa and Outgoing long-wave radiation	2.5∘ x 2.5∘	31.25∘N 8.75∘N 98.75∘W 56.25∘W	1979-2013	NCEP/DOE reanalysis [39], NOAA-interpolated [43]	8
[70]	Wind fields at 925 and 850 hPa	0.7∘ x 0.7∘	25∘N 8∘N 100∘W 65∘W	1979-2012	ERA-Interim reanalysis [18]	11
[61]	Mean sea-level pressure	0.7∘ x 0.7∘	50∘N 0∘N 150∘W 10∘W	1982-2016	ERA-Interim reanalysis [18]	20
[71]	Wind fields at 925, 850 and 200 hPa and CAPE	0.25∘ x 0.25∘	25∘N 8∘N 100∘W 65∘W	1979-2019	ERA5 reanalysis[33]	6

Table 2: Weather/circulation types over Central America and its surroundings using dynamical variables (wind, pressure, geopotential).

References

[1] E. Alfaro, *Some Characteristics of the Annual Precipitation Cycle in Central America and their Relationships with its Surrounding Tropical Oceans*. Tópicos Meteorológicos y Oceanográficos **9**(2002), no. 2, 88–103. URL: https://www.kerwa.ucr.ac.cr/handle/10669/15428.

[2] E. J. Alfaro et al., *Caracterización climática y análisis de mecanismos moduladores del descenso de las lluvias en la vertiente Caribe de América Central durante septiembre-octubre*. Revista de Ciencias Ambientales **58**(2024), no. 1, 1–24. DOI: 10.15359/rca.58-1.4

[3] E. Alfaro, X. Chourio, Á. Muñoz, S. Mason, *Improved Seasonal Prediction Skill of Rainfall for the Primera Season in Central America*. Int. J. Climatol **38**(2018), e255–e268. DOI: 10.1002/joc.5366

[4] J. A. Amador et al., *The easternmost tropical Pacific. Part I: A climate review*. Revista de Biología Tropical **64**(2016), no. 1, 1–22. DOI: 10.15517/rbt.v64i1.23407

[5] J. Amador, *A Climatic Feature of Tropical Americas: The Trade Wind Easterly Jet*. Tópicos Meteorológicos y Oceanográficos **5**(1998), no. 2, 91–102. URL: https://www.kerwa.ucr.ac.cr/handle/10669/76623.

[6] J. Amador, *The Intra-Americas Sea Low-Level Jet*. Annals of the New York Academy of Sciences **1146**(2008), no. 1, 153–188. DOI: 10.1196/annals.1446.012

[7] J. Amador et al., *The Easternmost Tropical Pacific, Part II: Seasonal and Intraseasonal Modes of Atmospheric Variability*. Revista de Biología Tropical **64**(2016), 23–57. DOI: 10.15517/rbt.v64i1.23409

[8] D. Arthur, S. Vassilvitskii, *k-means++: the advantages of careful seeding*. ACM-SIAM Symposium on Discrete Algorithms. 2007. URL: https://dl.acm.org/doi/10.5555/1283383.1283494.

[9] C. Beck, A. Philipp, F. Streicher, *The Effect of Domain Size on the Relationship Between Circulation Type Classifications and Surface Climate*. Int. J. Climatol. **36**(2013), 2692–2709. DOI: 10.1002/joc.3688

[10] R. Bombardi, V. Moron, J. Goodnight, *Detection, Variability, and Predictability of Monsoon Onset and Withdrawal Dates: A Review*. Int J Climatol **40**(2020), 641–667. DOI: 10.1002/joc.6264

[11] C. Bretherton, M. Peters, L. Back, *Relationships Between Water Vapor Path and Precipitation Over the Tropical Oceans*. J. Climate **17**(2004), 1517–1528. DOI: 10.1175/1520-0442(2004)017<1517:RBWVPA>2.0.CO;2

[12] T. Caliński, J. Harabasz, *A Dendrite Method for Cluster Analysis*. Communications in Statistics **3**(1974), no. 1, 1–27. DOI: 10.1080/03610927408827101

[13] X. Chadee, R. Clarke, *Daily Near-Surface Large-Scale Atmospheric Circulation Patterns Over the Wider Caribbean*. Climate Dynamics **44**(2015), no. 11, 2927–2946. DOI: 10.1007/s00382-015-2621-2

[14] K. Chang, K. Bowman, L. Siu, A. Rapp, *Convective Forcing of the North American Monsoon Anticyclone at Intraseasonal and Interannual Time Scales*. Journal of the Atmospheric Sciences **78**(2021), no. 9, 2941–2956. DOI: 10.1175/JAS-D-21-0009.1

[15] B. Christiansen, *Atmospheric Circulation Regimes: Can Cluster Analysis Provide the Number?* Journal of Climate **20**(2007), no. 10, 2229–2250. DOI: 10.1175/JCLI4107.1

[16] C. Chung, S. Nigam, *Weighting of Geophysical Data in Principal Component Analysis*. Journal of Geophysical Research: Atmospheres **104**(1999), no. D14, 16925–16928. DOI: 10.1029/1999JD900234

[17] A. Dawson, *EOFS: A Library for EOF Analysis of Meteorological, Oceanographic, and Climate Data*. Journal of Open Research Software **4**(2016), no. 1, e14. DOI: 10.5334/jors.122

[18] D. Dee et al., *The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System*. Q.J.R. Meteorol. Soc. **137**(2011), 553–597. DOI: 10.1002/qj.828

[19] T. Delsole, M. Tippett, *Statistical Methods for Climate Scientists*. 1st ed. Cambridge University Press, 2022. DOI: 10.1017/9781108659055

[20] T. L. Delworth, M. E. Mann, *Observed and simulated multidecadal variability in the Northern Hemisphere*. Climate Dynamics **16**(2000), 661–676. DOI: 10.1007/s003820000075

[21] A. Duran-Quesada, L. Gimeno, J. Amador, *Role of Moisture Transport for Central American Precipitation*. Earth System Dynamics **8**(2017), no. 1, 147–161. DOI: 10.5194/esd-8-147-2017

[22] A. Duran-Quesada, R. Sorí, P. Ordoñez, L. Gimeno, *Climate Perspectives in the Intra-Americas Seas*. Atmos **11**(2020), no. 9, 959. DOI: 10.3390/atmos11090959

[23] W. Ebisuzaki, *A Method to Estimate the Statistical Significance of a Correlation When the Data Are Serially Correlated*. J. Climate **10**(1997), 2147–2153. DOI: 10.1175/1520-0442(1997)010<2147:AMTETS>2.0.CO;2

[24] D. B. Enfield, A. M. Mestas-Nuñez, P. J. Trimble, *The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S.* Geophysical Research Letters **28**(2001), no. 10, 2077–2080. DOI: 10.1029/2000GL012745

[25] C. Funk et al., *The Climate Hazards Infrared Precipitation with Stations - A New Environmental Record for Monitoring Extremes*. Sci Data **2**(2015), 150066. DOI: 10.1038/sdata.2015.66

[26] J. García-Franco, R. Chadwick, L. Gray, *Revisiting Mechanisms of the Mesoamerican Midsummer Drought*. Clim Dyn **60**(2023), 549–569. DOI: 10.1007/s00382-022-06338-6

[27] I. García-Martínez, M. Bollasina, *Sub-Monthly Evolution of the Caribbean Low-Level Jet and Its Relationship with Regional Precipitation and Atmospheric Circulation*. Climate Dynamics **54**(2020), 4423–4440. DOI: 10.1007/s00382-020-05237-y

[28] R. Geen, S. Bordoni, D. Battisti, K. Hui, *Monsoons, ITCZs, and the Concept of the Global Monsoon*. Reviews of Geophysics **58**(2020), e2020RG000700. DOI: 10.1029/2020RG000700

[29] D. Glannakis, A. Majda, *Data-driven methods for dynamical systems*. R. Melnik (Ed.). Mathematical and Computational Modeling. 2015. DOI: 10.1002/9781118853887.ch7

[30] I. Gouirand, V. Moron, B. Sing, *Seasonal Atmospheric Transitions in the Caribbean Basin and Central America*. Climate Dynamics **55**(2020), 1809–1828. DOI: 10.1007/s00382-020-05356-6

[31] A. Hannachi, *Patterns Identification and Data Mining in Weather and Climate*. Springer Cham, 2021. DOI: 10.1007/978-3-030-67073-3

[32] F. Hansen, D. Belušić, *Tailoring Circulation Type Classification Outcomes*. International Journal of Climatology **41**(2021), no. 14, 6145–6161. DOI: 10.1002/joc.7171

[33] H. Hersbach et al., *The ERA5 Global Reanalysis*. Quarterly Journal of the Royal Meteorological Society **146**(2020), 1999–2049. DOI: 10.1002/qj.3803

[34] H. Hidalgo, E. Alfaro, *Some Physical and Socio-Economic Aspects of Climate Change in Central America*. Progress in Physical Geography: Earth and Environment **36**(2012), no. 3, 379–399. DOI: 10.1177/0309133312438906

[35] H. Hidalgo, A. Durán-Quesada, J. Amador, E. Alfaro, *The Caribbean Low-Level Jet, the Inter-Tropical Convergence Zone and Precipitation Patterns in the Intra-Americas Sea: A Proposed Dynamical Mechanism*. Geografiska Annaler: Series A, Physical Geography **97**(2015), no. 1, 41–59. DOI: 10.1111/geoa.12085

[36] J. R. Holton, G. J. Hakim, *An Introduction to Dynamic Meteorology*. 5th. Academic Press, Waltham, MA, 2013. DOI: 10.1016/C2009-0-63394-8

[37] K. T. Ingram, M. C. Roncoli, P. H. Kirshen, *Opportunities and Constraints for Farmers of West Africa to Use Seasonal Precipitation Forecasts with Burkina Faso as a Case Study*. Agricultural Systems **74**(2002), no. 3, 331–349. DOI: 10.1016/S0308-521X(02)00044-6

[38] M. Iturbide et al., *An Update of IPCC Climate Reference Regions for Subcontinental Analysis of Climate Model Data: Definition and Aggregated Datasets*. Earth System Science Data **12**(2020), no. 4, 2959–2970. DOI: 10.5194/essd-12-2959-2020

[39] M. Kanamitsu et al., *NCEP–DOE AMIP-II Reanalysis (R-2)*. Bulletin of the American Meteorological Society **83**(2002), no. 11, 1631–1644. DOI: 10.1175/BAMS-83-11-1631

[40] A. Karmalkar, R. Bradley, H. Diaz, *Climate Change in Central America and Mexico: Regional Climate Model Validation and Climate Change Projections*. Clim Dyn **37**(2011), 605–629. DOI: 10.1007/s00382-011-1099-9

[41] A. Laurie, M. Barlow, S. B. Feldstein, W. J. Gutowski, *Identification of Large-Scale Meteorological Patterns Associated with Extreme Precipitation in the US Northeast*. Climate Dynamics **50**(2018), no. 5, 1819–1839. DOI: 10.1007/s00382-017-3724-8

[42] P. Laux, B. Böker, E. Martins, et al., *A Semi-Objective Circulation Pattern Classification Scheme for the Semi-Arid Northeast Brazil*. Int J Climatol **41**(2021), 51–72. DOI: 10.1002/joc.6608

[43] B. Liebmann, C. A. Smith, *Description of a complete (interpolated) outgoing longwave radiation dataset*. Bull. Amer. Meteor. Soc. **77**(1996), 1275–1277.

[44] S. Lloyd, *Least Squares Quantization in PCM*. IEEE Transactions on Information Theory **28**(1982), no. 2, 129–137. DOI: 10.1109/TIT.1982.1056489

[45] J. MacQueen, *Some Methods for Classification and Analysis of Multivariate Observations*. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Statistics, Vol. 1. 1967.

[46] V. Magaña, A. Amador, S. Medina, *The Midsummer Drought over Mexico and Central America*. Journal of Climate **12**(1999), no. 6, 1577–1588. DOI: 10.1175/1520-0442(1999)012%3C1577:TMDOMA%3E2.0.CO;2

[47] T. Maldonado, E. Alfaro, H. Hidalgo, *Revision of the Main Drivers and Variability of Central America Climate and Seasonal Forecast Systems*. Revista de Biología Tropical **66**(2018), 153. DOI: 10.15517/rbt.v66i1.33294

[48] T. Maldonado, E. J. Alfaro, H. Hidalgo, *Analysis of Precipitation Clusters and Their Seasonal Changes over Central America for the 1976-2015 Period*. Revista de Matemática Teoría y Aplicaciones **28**(2021), no. 2, 337–362. DOI: 10.15517/rmta.v28i2.42322

[49] J. Marengo et al., *Onset and End of the Rainy Season in the Brazilian Amazon Basin*. Journal of Climate **14**(2001), no. 5, 833–852. DOI: 10.1175/1520-0442(2001)014<0833:OAEOTR>2.0.CO;2

[50] C. Martinez, Y. Kushnir, L. Goddard, et al., *Interannual Variability of the Early and Late-Rainy Seasons in the Caribbean*. Clim Dyn **55**(2020), 1563–1583. DOI: 10.1007/s00382-020-05341-z

[51] J. Mejía, J. Yepes, J. Henao, et al., *Towards a Mechanistic Understanding of Precipitation over the Far Eastern Tropical Pacific and Western Colombia, One of the Rainiest Spots on Earth*. Journal of Geophysical Research: Atmospheres **126**(2021), no. 5, e2020JD033415. DOI: 10.1029/2020JD033415

[52] D. Mesa-Sánchez, O. Rojo-Hernández, *On the General Circulation of the Atmosphere around Colombia*. Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales **44**(2020), no. 172, 857–875. DOI: 10.18257/raccefyn.899

[53] S. Michaelides, F. Liassidou, C. Schizas, *Synoptic Classification and Establishment of Analogues with Artificial Neural Networks*. Pure and Applied Geophysics **164**(2007), 1347–1364. DOI: 10.1007/s00024-007-0222-7

[54] J. Morales, P. Arias, J. Martínez, A. Durán-Quesada, *The Role of Low-Level Circulation on Water Vapour Transport to Central and Northern South America: Insights from a 2D Lagrangian Approach*. International Journal of Climatology **41**(2021), E2662–E2682. DOI: 10.1002/joc.6873

[55] V. Moron, I. Gouirand, M. Taylor, *Weather Types across the Caribbean Basin and Their Relationship with Rainfall and Sea Surface Temperature*. Climate Dynamics **47**(2015), no. 1, 601–621. DOI: 10.1007/s00382-015-2858-9

[56] V. Moron et al., *Weather Types and Hourly to Multiday Rainfall Characteristics in Tropical Australia*. J. Climate **32**(2019), 3983–4011. DOI: 10.1175/JCLI-D-18-0384.1

[57] R. Muñoz-Jiménez, J. Giraldo-Osorio, A. Brenes-Torres, et al., *Spatial and Temporal Patterns, Trends and Teleconnection of Cumulative Rainfall Deficits across Central America*. International Journal of Climatology **39**(2019), 1940–1953. DOI: 10.1002/joc.5925

[58] R. Neal et al., *The application of predefined weather patterns over India within probabilistic medium-range forecasting tools for high-impact weather*. Meteorological Applications **29**(2022), no. 3, e2083. DOI: 10.1002/met.2083

[59] J. D. Neelin, O. Peters, K. Hales, *The transition to strong convection*. J. Atmos. Sci. **66**(2009), 2367–2384. DOI: 10.1175/2009JAS2962.1

[60] NOAA National Geophysical Data Center, *Global Land One-kilometer Base Elevation (GLOBE) v.1*. Tech. rep. NOAA National Centers for Environmental Information, 1999. DOI: 10.7289/V52R3PMS

[61] C. A. Ochoa-Moya et al., *Climatological large-scale circulation patterns over the Middle Americas region*. Atmosphere **11**(2020), no. 7, 745. DOI: 10.3390/atmos11070745

[62] M. E. Olmo et al., *Circulation patterns and associated rainfall over south tropical South America: GCMs evaluation during the dry-to-wet transition season*. Journal of Geophysical Research: Atmospheres **127**(2022). DOI: 10.1029/2022JD036468

[63] T. N. Palmer, *A Nonlinear Dynamical Perspective on Climate Prediction*. Journal of Climate **12**(1999), no. 2, 575–591. DOI: 10.1175/1520-0442(1999)012<0575:ANDPOC>2.0.CO;2

[64] F. Pedregosa et al., *Scikit-learn: machine learning in python*. Journal of Machine Learning Research **12**(2011), 2825–2830.

[65] A. Philipp, C. Beck, R. Huth, J. Jacobeit, *Development and comparison of circulation type classifications using the COST 733 dataset and software*. International Journal of Climatology **36**(2016), 2673–2691. DOI: 10.1002/joc.3920

[66] A. Philipp et al., *Long-Term Variability of Daily North Atlantic–European Pressure Patterns since 1850 Classified by Simulated Annealing Clustering*. J. Climate **20**(2007), no. 16, 4065–4095. DOI: 10.1175/JCLI4175.1

[67] B. Quesada-Montano et al., *Characterising droughts in Central America with uncertain hydro-meteorological data*. Theoretical and Applied Climatology **137**(2019), 2125–2138. DOI: 10.1007/s00704-018-2730-z

[68] N. Ramesh, Q. Nicolas, W. R. Boos, *The Globally Coherent Pattern of Autumn Monsoon Precipitation*. J. Climate **34**(2021), no. 14, 5687–5705. DOI: 10.1175/JCLI-D-20-0740.1

[69] P. Rousseeuw, *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics **20**(1987), 53–65. DOI: 10.1016/0377-0427(87)90125-7

[70] F. Sáenz, A. M. Durán-Quesada, *A climatology of low level wind regimes over Central America using a weather type classification approach*. Frontiers in Earth Sciences **3**(2015), 1–18. DOI: 10.3389/feart.2015.00015

[71] F. Sáenz et al., *Atmospheric circulation types controlling rainfall in the Central American Isthmus*. International Journal of Climatology **43**(2023), no. 1, 197–218. DOI: 10.1002/joc.7745

[72] L. Siu, K. Bowman, *Forcing of the upper-tropospheric monsoon anticyclones*. Journal of the Atmospheric Sciences **76**(2019), no. 7, 1937–1954. DOI: 10.1175/JAS-D-18-0340.1

[73] I. T. Stewart, E. P. Maurer, K. Stahl, K. Joseph, *Recent evidence for warmer and drier growing seasons in climate sensitive regions of Central America from multiple global datasets*. International Journal of Climatology **42**(2021), no. 3, 1399–1417. DOI: 10.1002/joc.7310

[74] M. Taylor, E. Alfaro, *Central America and the Caribbean, climate of*. J. Oliver (Ed.). Encyclopedia of World Climatology. Encyclopedia of Earth Sciences Series. Springer, Dordrecht, 2005. DOI: 10.1007/1-4020-3266-8_37

[75] J. Theiler et al., *Testing for nonlinearity in time series: the method of surrogate data*. Physica D: Nonlinear Phenomena **58**(1992), no. 1, 77–94. DOI: 10.1016/0167-2789(92)90102-S

[76] L. Tomassini, *The interaction between moist convection and the atmospheric circulation in the Tropics*. Bulletin of the American Meteorological Society **101**(2020), no. 8, E1378–E1396. DOI: 10.1175/BAMS-D-19-0180.1

[77] K. Trenberth, D. Stepaniak, J. Caron, *The global monsoon as seen through the divergent atmospheric circulation*. Journal of Climate **13**(2000), no. 22, 3969–3993. DOI: 10.1175/1520-0442(2000)013<3969:tgmast>2.0.co;2

[78] A. Tsonis, *Vertical stability in the atmosphere*. An Introduction to Atmospheric Thermodynamics. 2nd ed. Cambridge University Press, 2007, 143–158. DOI: 10.1017/CBO9780511619175.009

[79] N. Vigaud, A. Robertson, *Convection regimes and tropical-midlatitude interactions over the intra-American seas from May to November*. International Journal of Climatology **37**(2017), no. S1, 987–1000. DOI: 10.1002/joc.5051

[80] P. Wang et al., *The global monsoon across timescales: coherent variability of regional monsoons*. Climate of the Past **10**(2014), no. 6, 2007–2052. DOI: 10.5194/cp-10-2007-2014

[81] D. Wilks, *Statistical Methods in the Atmospheric Sciences*. Elsevier, Amsterdam, Netherlands, 2020. DOI: 10.1016/C2017-0-03921-6

[82] Y. Zhang, J. Wallace, D. Battisti, *ENSO-like Interdecadal Variability: 1900–93*. Journal of Climate **10**(1997), no. 5, 1004–1020. DOI: 10.1175/1520-0442(1997)010<1004:ELIV>2.0.CO;2

[83] E. Zorita, J. P. Hughes, D. P. Lettemaier, H. von Storch, *Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation*. Journal of Climate **8**(1995), no. 5, 1023–1042. DOI: 10.1175/1520-0442(1995)008<1023:SCORCP>2.0.CO;2