© Revista de Matemática: Teoría y Aplicaciones 2021 **28**(2) : 337–362
cimpa – ucr issn: 1409-2433 (Print), 2215-3373 (Online)
doi: 10.15517/rmta.v28i2.42322

ANÁLISIS DE LOS CONGLOMERADOS DE PRECIPITACIÓN Y SUS CAMBIOS ESTACIONALES SOBRE AMÉRICA CENTRAL PARA EL PERIODO 1976-2015

ANALYSIS OF PRECIPITATION CLUSTERS AND THEIR SEASONAL CHANGES OVER CENTRAL AMERICA FOR THE 1976-2015 PERIOD

Tito Maldonado* Eric J. Alfaro† Hugo G. Hidalgo‡

Received: 3/Jul/2020; Revised: 25/Nov/2020;
Accepted: 25/Nov/2020

*Universidad de Costa Rica, Centro de Investigaciones Geofísicas (CIGEFI), San José, Costa Rica. E-Mail: tito.maldonado@ucr.ac.cr

†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

‡Universidad de Costa Rica, Centro de Investigaciones Geofísicas (CIGEFI) y Escuela de Física, San José, Costa Rica. E-Mail: hugo.hidalgo@ucr.ac.cr

Resumen

La ubicación geográfica de América Central juega un papel importante en la descripción de la variabilidad climática de la región. Está rodeada por dos grandes masas de agua, el Pacífico Tropical del Este en el lado occidental y el Mar Caribe en el lado oriental. La región es sensible al efecto de los sistemas dinámicos tanto a gran escala como a escala regional que actúan en sus proximidades, siendo la topografía el principal modulador local de la variabilidad en la región. Para tener en cuenta esta variabilidad climática espacial se utilizaron 57 estaciones meteorológicas con datos diarios de precipitación en América Central para el periodo 1976-2015. Se definieron índices mensuales para describir cuánto y cómo llueve: acumulado total mensual (ACU), cantidad de días con lluvia (DCP), porcentaje de días que no sobrepasan el percentil 20 (extremos secos), y el porcentaje de días que exceden el percentil 80 (extremos húmedos). Por medio de técnicas de aprendizaje automatizado (análisis de conglomerados) se estimaron un número óptimo de grupos para cada variable. La optimización se realizó utilizando el estadístico de brecha. Se encontró un patrón de grupos localizados principalmente en la vertiente Pacífico y Caribe, mientras que en todas las variables se identificó un grupo localizado en la región del Caribe de Costa Rica. Al analizar los cambios en los 40 años de análisis, no se encontraron cambios ni tendencias significativos en las escalas de tiempo mensual estacional y anual ni a nivel de estación, grupal ni regional.

Palabras clave: América Central; precipitación; análisis de conglomerados; aprendizaje automatizado; estadístico de brecha; análisis de tendencias; variabilidad climática.

Abstract

The geographical location of Central America plays a significant role in describing the climate variability of the region. It is surrounded by two large water masses, the Eastern Tropical Pacific ocean on the western side and the Caribbean Sea on the eastern side. The region is sensitive to the effect of both large-scale and regional-scale dynamical systems acting in its vicinity, being the topography the main local modulator of the variability in the region. To account this spatial climate variability we used 57 meteorological stations with daily precipitation data in Central America for the period 1976-2015. Monthly indices were defined to describe how much and how it rains: monthly total accumulated (ACU), number of days with rain (DCP), percentage of days that do not exceed the 20th percentile (dry extremes), and the percentage of days that exceeds the percentile 80th (wet extreme). Using automated learning techniques (cluster analysis), an optimal number of groups was estimated for each variable. Optimization was performed using the gap statistic. A pattern of groups located primarily on the Pacific and Caribbean slopes was found, while in all variables a group located in the Caribbean region of Costa Rica was identified. When analyzing the changes in the 40 years of analysis, no significant changes or trends were found in the monthly seasonal or annual time scales or at the station, group or regional level.

Keywords: Central America; precipitation; cluster analysis; trend analysis; unsupervised machine learning; gap statistic; climate variability.

Mathematics Subject Classification: 62H30, 91C20, 86A10, 86-08.

1 Introducción

En América Central la precipitación presenta una alta variabilidad temporal y espacial, lo que dificulta generalizar su comportamiento para toda la región. Es por eso que tanto para determinar su ciclo anual, estacionalidad y estudiar posibles cambios observados o estimar variaciones a futuro, es de relevancia identificar regiones en las cuales la lluvia muestre similitud y represente de forma más adecuada el clima local. La variabilidad de la precipitación ha sido previamente estudiada, y se ha determinado que tanto los factores orográficos, como procesos de gran escala tienen incidencia en las anomalías de lluvia en las diferentes escalas de tiempo [7], [8], [9], [10], [21], [23], [29].

El régimen de lluvias en América Central está principalmente determinado por la interacción de los vientos alisios con la cadena montañosa que atraviesa el istmo, por lo que se pueden identificar a groso modo, dos ciclos de lluvia anuales (ver el detalle que explica este mecanismo en [23]). Según [2], [29] y [38], el primero, y por lo general más conocido, es un ciclo anual bimodal en la cuenca del Pacífico centroamericano, que cuenta con una estación seca bien definida durante el invierno boreal (enero a marzo), con un primer máximo de lluvia durante mayo-junio, y un segundo máximo de precipitación (por lo general más intenso) durante los meses de setiembre a octubre. Estos dos máximos son separados por una reducción en las lluvias que ocurre durante julio y agosto conocida como el veranillo o canícula (MSD por sus siglas en inglés, [7], [28], [32]). Simultáneo con este proceso se observa un incremento en los alisios en bajo nivel (aprox. 925 hPa) asociado con la corriente en chorro de bajo nivel sobre el Caribe o CLLJ por sus siglas en inglés, ([7], [6], [23], [31]). Por su parte, el otro ciclo anual se localiza en la vertiente Caribe, y principalmente presenta lluvias en todos los meses de año, sin embargo tiene un máximo en julio, y dos reducciones significativas de la lluvia en marzo-abril y luego en setiembre-octubre.

Entender como afecta el cambio climático en la región es de relevancia debido a que es la zona tropical de más alta vulnerabilidad en cuanto a respuesta climática se refiere y es definida por [14] como un "punto caliente". A nivel socio-económico es vital conocer los efectos el cambio climático, ya que en muchos países de la región su economía depende de la agricultura ([25]) además de sus efectos en la biodiversidad ([24]) debido a la gran variedad de especies y ecosistemas que posee el istmo centroamericano. Estudios recientes han encontrando en escenarios de cambio climático, que los modelos predicen una reducción de lluvia al norte del istmo ([13], [20], [22]), contrario al sur donde se encuentran condiciones más húmedas según los modelos de circulación general y modelos regionales utilizados.

Estudios previos han analizado cambios observados de lluvia en la región centroamericana y el Caribe. Por ejemplo, [1] en su estudio analizaron las variables: precipitación y temperatura de estaciones meteorológicas, para determinar las tendencias históricas en estas para la región. Esos autores encontraron patrones de temperatura con tendencias positivas significativas al 5%, mientras que para precipitación no hay un patrón claro de cambio en los 40 años (1961 – 2003) de análisis en las series anuales regionales, debido a la alta variabilidad año a año en la precipitación. Sin embargo, no detectaron una reducción en la precipitación total en la región. Cabe destacar, que estos autores al analizar la serie agregada encontraron algunos patrones de cambio en el ciclo hidrológico principalmente relacionados a eventos extremos húmedos de precipitación. Otro estudio reciente de tendencias hidrometeorológicas en la región se encuentra en [5], quienes encontraron tendencias significativas en la temperatura en América Central, mientras que la aridez no muestra un cambio tan consistente en la región. Al igual que el estudio de [21], las tendencias de precipitación no son significativas, excepto en unas pocas localidades aisladas.

Similarmente, [37] encontraron que los índices de precipitación utilizados para estudiar los cambios en la tendencia de precipitación para la región del Caribe en el periodo de 1961-2010, fueron poco consistentes, con tendencias débiles y no significativas. Sin embargo en un estudio reciente, [11] al analizar los cambios en el MSD en un periodo de 37 años (1981-2018), enfatizan que en escala regional se observan cambios positivos en la duración, mas no así, en la magnitud, la cual es dominada por tendencias espaciales heterogéneas y variabilidad interanual enlazada a modos de gran-escala de la circulación océano-atmósfera. Guillén-Oviedo et al. en [16], por otra parte, encontraron al analizar los eventos extremos de precipitación que, si bien aparecen las dos regiones climáticas descritas anteriormente, se observa además una tercera región de transición intermontana.

Por tanto, el objetivo de este estudio es determinar por medio de técnicas de aprendizaje automatizado (*unsupervised machine learning* en inglés, [27]) subregiones climáticas de acuerdo a las características estadísticas de precipitación. Para este trabajo además, se analizarán variables que describen el comportamiento de las lluvias, como la frecuencia de días con lluvia, el porcentaje mensual de días que no exceden el percentil 20 de lluvia acumulada por mes, como indicador de extremos secos, y el porcentaje mensual de días que sobrepasan el percentil 80 de lluvia acumulada como índice de extremos húmedos. Además, se estudiarán los cambios de cada una de estas variables en las subregiones o grupos encontrados para el istmo durante el periodo de 1976 al 2015. En la Sección 2 se describen los datos y métodos utilizados para el análisis, la Sección 3 muestra los resultados obtenidos, y en la Sección 4 algunos comentarios finales y recomendaciones.

2 Métodos

2.1 Adquisición y procesamiento de datos

Se utilizaron un total de 186 estaciones con datos diarios proporcionados por los Servicios Meteorológicos e Hidrológicos Nacionales (SMHN) de América Central, con registros desde enero de 1976 a diciembre de 2015 (4 décadas).

Se aplicó un control de calidad para seleccionar solamente aquellas estaciones que tuvieran al menos el 60% de los datos durante el periodo mencionado anteriormente. Después del control de calidad, se seleccionaron 57 estaciones de 186 posibles localizadas en el istmo centroamericano (Figura 1).

2.2 Variables de precipitación

De los datos diarios, cuatro series mensuales son estimadas para cada una de las estaciones: precipitación total acumulada (ACU), frecuencia de días con lluvia (DCP), el porcentaje de días que sobrepasan el percentil 80 mensual climático (p80, extremo húmedo), y el porcentaje de días por debajo del percentil 20 mensual climático (p20, extremo seco). Sobre los detalles en el cálculo de estas variables, se puede consultar [30], [32], [31]. Se define como día con lluvia, aquel en que la precipitación observada en un día tiene valores mayores a 0.1 mm día^{-1}.

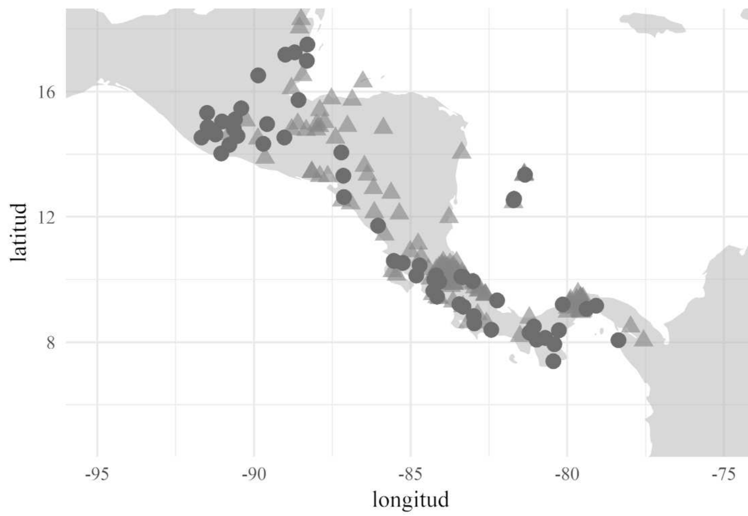


Figura 1: Distribución espacial de las 186 estaciones meteorológicas en América Central. En círculos azules se distinguen las 57 estaciones que contienen al menos el 60% de los datos diarios de lluvia para el periodo 1976-2015, que se utilizaron para este estudio, mientras que los triángulos grises representan las estaciones que se descartaron en el control de calidad.

Las medianas para cada variable son calculadas para las cuatro décadas dentro del periodo de estudio: 1976-1985, 1986-1995, 1996-2005, y 2006-2015, usando el valor anual en cada estación para los doce meses del año. La partición por décadas responde a evidencia de procesos de baja frecuencia que afectan al clima en América Central en escalas decenales o mayores como la Oscilación Decenal del Pacífico, y la Oscilación Multidecenal del Atlántico [21], [29], [30], [31], [32], [33].

Se escogen estaciones con al menos seis valores en una década en particular (estaciones con al menos 60% de los datos), para analizar los cambios interdecenales y tendencias. La década 1976-1985 se utiliza como período base. Los cambios para el resto de las décadas son estimados sustrayendo a cada década lo observado en el período base. La significancia estadística para las diferencias de las medianas es establecida de acuerdo al *valor p* de una prueba de suma de categoría de Wilcoxon de dos colas [40].

Las tendencias de cada mes son calculadas para el valor de la serie de tiempo anual de las cuatro variables desde enero de 1976 hasta diciembre de 2015, y su significancia estadística es establecida usando el *valor p* de una prueba *t-student*.

2.3 Conglomerados y aprendizaje automatizado

Para identificar regiones en las cuales la precipitación se comporta estadísticamente similar, se utiliza el análisis de agrupamientos o conglomerados (*cluster analysis*, en inglés) para las cuatro variables de precipitación. Esta técnica es una importante herramienta dentro del área de "aprendizaje automatizado" (*unsupervised machine learning* en inglés) [27], para determinar grupos en datos sin la ayuda de la respuesta de una variable [3], [4]. Cabe resaltar que el análisis de conglomerados no es un único método, sino una familia de varias decenas de métodos [19].

Para este estudio se usa el método *k-means*, el cual es el algoritmo muy utilizado en el ámbito del aprendizaje automatizado [17] para particionar un conjunto de datos en un conjunto de grupos k, donde k representa el número de grupos pre-especificados por el analista. Cada objeto (estación) dentro de cada uno de los grupos son lo más parecido posible, mientras que objetos de diferentes grupos tienen las mayores disparidades posibles entre ellos. En esta técnica, cada grupo es representado por su centroide, el cual corresponde al promedio de los puntos asignados a cada grupo. La idea básica detrás de *k-means* consiste en definir grupos de tal forma que las variaciones intra grupales son minimizadas. La técnica de *k-means* podría encontrar óptimos locales del conjunto de datos.

Existen muchos algoritmos para determinar el número de clases para *k-means*, como el índice de [12] o los estudiados en [34], [35] para el caso jerárquico. Para el objetivo de este estudio se utilizará el algoritmo de [18] el cual define el total de la variación dentro del grupo como la suma del cuadrado de las distancias euclidianas entre los objetos y su correspondiente centroide. En este algoritmo primero se selecciona un número aleatorio de objetos de los datos, como los centros iniciales para los grupos. Estos objetos seleccionados son conocidos como los centroides o promedio de grupo. Luego, cada uno de los objetos restantes son asignados a los centroides más cercanos de acuerdo a su distancia euclidiana entre el objeto y el promedio de grupo. Después de la primera etapa de asignación, el algoritmo calcula un nuevo promedio para cada grupo. Una vez se realiza este paso, los objetos son nuevamente examinados para detectar si es posible una cercanía a otros grupos. Este proceso es iterativo, y se realiza hasta que se alcance una convergencia, es decir, que los grupos obtenidos son iguales a los anteriores. Además, se empleó un método de múltiples configuraciones iniciales aleatorias, y que para este caso se utilizaron 25 inicializaciones por recomendaciones computacionales del software R [27].

En este estudio, los grupos se estiman utilizando toda la serie de tiempo de los datos mensuales, i.e. enero 1976 a diciembre 2015. Para determinar el número de grupos o regiones óptimo en cada una de las variables, se empleó un método de cálculo desatendido conocido como el cálculo del estadístico de brecha (*gap statistic* en inglés) [39], el cual se describe a continuación.

El estadístico de brecha estima el total de variaciones intra grupales para diferentes valores k con sus valores esperados bajo la distribución referencia nula de los datos (i.e. una distribución sin agrupamiento obvio). El conjunto de datos de referencia es generado usando simulaciones de Monte Carlo del proceso de muestreo. Es decir, para cada variable (x_i) en el conjunto de datos, se calcula su ámbito $[\min(x_i), \max(x_i)]$ y se generan valores para cada n puntos uniformemente entre los intervalos $\min$ y $\max$.

Para los datos observados y de referencia, la variación intra grupal total es calculada usando diferentes valores de k. El estadístico de brecha para un k dado es definido como:

$$Gap_n(k) = E_n^* \log(W_k) - \log(W_k), \tag{1}$$

donde E_n^* denota la expectación bajo una muestra de tamaño n de la distribución de referencia. E_n^* es definida bajo bootstrapping (B), generando B copias de los datos referencia, y calculando el promedio $\log(W_k^*)$. W_k es la variación intra-clases para una partición en k clases. El estadístico de brecha mide la desviación del valor observado W_k desde su valor esperado bajo la hipótesis nula. El número óptimo de grupos $(\hat{k})$ será el valor que maximice $Gap_n(k)$. Estos significa que la estructura de agrupamiento está lejos de una distribución uniforme de puntos.

El algoritmo se compone de los siguientes pasos:

1. Se agrupan los datos, variando el número de grupos desde $k = 1, ..., k_{\max}$, y se calcula el correspondiente W_k.

2. Se generan B conjuntos de datos referencia y se agrupan cada uno de ellos con diferentes números de grupos $W_{kb}^*, b = 1, 2, ..., B, k = 1, ..., k_{\max}$. Se calcula el estadístico de brecha estimada con la ecuación (1).

3. Sea $\bar{w} = (1/B) \sum_b \log(W_{kb}^*)$, se calcula la desviación estándar

$$sd(k) = \sqrt{(1/b) \sum_b (\log(W_{kb}^*) - \bar{w})^2}, \tag{2}$$

y se define $s_k = sd_k \times \sqrt{(1 + 1/B)}$.

4. Se escoge el número de grupos para el k más pequeño tal que:

$$Gap(k) \geq Gap(k+1) - s_{k+1}. \tag{3}$$

Cabe destacar que para cada uno de los grupos encontrados, se aplica un análisis similar al que se hace para cada estación, con el fin de evidenciar cambios en la precipitación en los 40 años del período de estudio.

3 Resultados

3.1 Análisis de conglomerados

Los resultados de la estadística de brecha se muestran en la Figura 2 para la variable ACU. Se desprende de esta imagen que el valor óptimo de grupos para esta ACU (y DCP no se muestra) es de 6. Esto difiere ligeramente de [3] y [4] quienes encontraron 5 conglomerados para la precipitación acumulada, pero utilizando datos mensuales, mismo número encontrado para el p20 en este estudio (no se muestra). Por otra parte, [16] encontraron 3 grupos al analizar los eventos extremos, mismo número de grupos encontrados para el p80 en este estudio (no se muestra). Dos de estos grupos coinciden en identificar a las regiones Pacífico y Caribe de Centroamérica, sin embargo el tercer grupo no coincide, ya que en este estudio, el tercer grupo se localiza en Caribe de Costa Rica (como se muestra a continuación), mientras que [16] encontraron un grupo que lo relacionan con una zona de transición intermontana de América Central. Nótese que no necesariamente el número de grupos coincide entre variables.

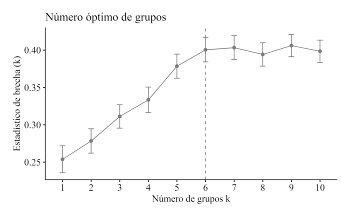


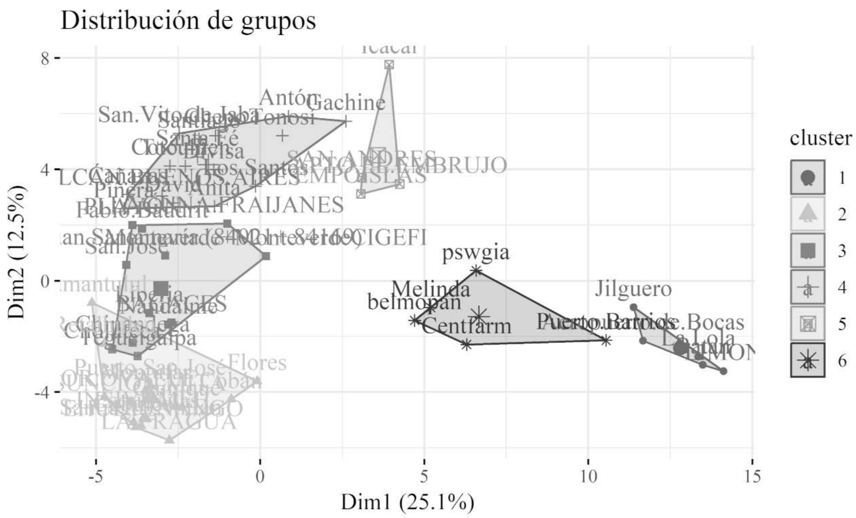
Figura 2: Número de grupos calculados para el acumulado total mensual de precipitación en América Central. En la gráfica el número de grupos óptimo ocurre en el primer punto de inflexión.

Para examinar si las agrupaciones muestran una distribución razonable, se analiza la desviación con respecto a las dos primeras componentes principales en cada conjunto de datos. Se nota una superposición entre los grupos 2 y 3 de la variable ACU (Figura 3a), mientras que el resto de variables (Figura 3b-d) no se encontraron superposiciones de grupos.

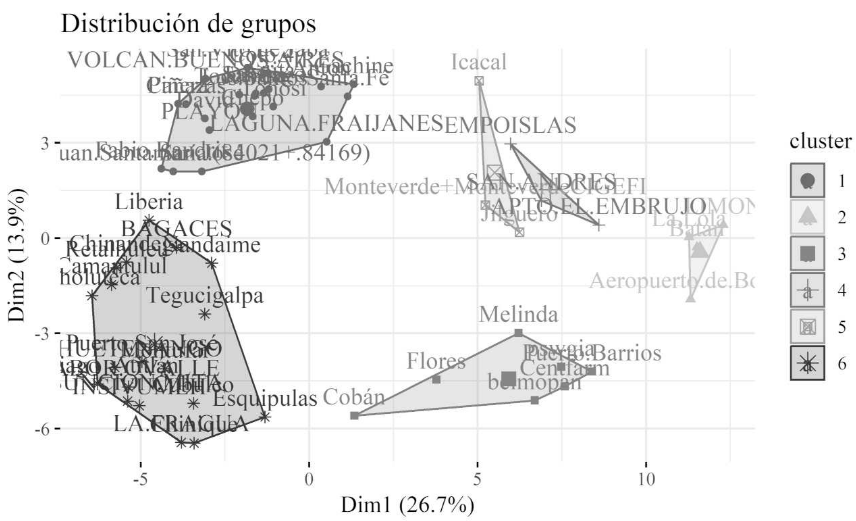
Al inspeccionar la distribución espacial de los grupos en cada una de las variables (Figura 4), se nota un patrón consistente, el cual divide a los conglomerados entre estaciones localizadas en la vertiente Pacífico y Caribe. Para ACU, DCP y p20 se observan además otras subregiones formados por estaciones localizadas al norte y al sur de ambas vertientes (Figura 4a,c,d), es decir, las estaciones se pueden agrupar de norte a sur y de este a oeste en la región de América Central. Es de notar que el p80 (Figura 4c) identifica un conglomerado de estaciones localizado principalmente en la región Caribe de Costa Rica.

El ciclo anual de cada uno de los grupos derivados del análisis aplicado a cada variable presenta algunas distinciones que merecen una revisión. En la variable ACU (Figura 5a) se puede observar que los grupos pertenecientes a la vertiente Pacífico (1, 2, 4 y 7) presentan un ciclo bimodal, con la presencia del MSD. La diferencia principal es la ocurrencia de los máximos y mínimos de precipitación, y la cantidad de precipitación que se observa en cada uno de los puntos críticos de los ciclos anuales, por ejemplo el grupo 7 (localizado en el Pacífico Sur de Costa Rica), presenta las cantidades más altas de precipitación, con máximos en mayo y octubre (por sobre los 400 mm mes^{-1}), y un mínimo relativo en julio (cerca de 350 mm mes^{-1}), mientras que los grupos 1 y 2 (localizados en el Pacífico Norte) presentan máximos de lluvia en junio y setiembre con valores cercanos a los 350 mm mes^{-1}, y su mínimo relativo en julio (1) y agosto (2) con valores cercanos a los 200 mm mes^{-1}.

El grupo 3 (localizado en Belice) presenta un ciclo anual de lluvias con un primer máximo en junio y un segundo máximo en octubre (alrededor de 350 mm mes^{-1}), sin embargo la presencia del MSD no es tan notoria como en las estaciones localizadas en el Pacífico centroamericano. El grupo 5 (localizados en la región Caribe de Costa Rica) no presenta un ciclo bimodal bien definido como su contra parte de la vertiente Pacífico, sin embargo, hay dos reducciones relativas de las lluvias en los meses de marzo y abril con valores cercanos a los 175 mm mes^{-1}, y octubre (unos 200 mm mes^{-1}). En el grupo 6 (localizado entre la zona norte de Costa Rica y en San Andrés, Colombia) el ciclo anual de precipitación presenta una temporada seca bien definida de enero a abril, y los máximos de lluvia se localizan en junio-agosto ($\approx$ 200 mm mes^{-1}), y octubre ($\approx$ 250 mm mes^{-1}) con un mínimo relativo en setiembre (menos de 200 mm mes^{-1}).

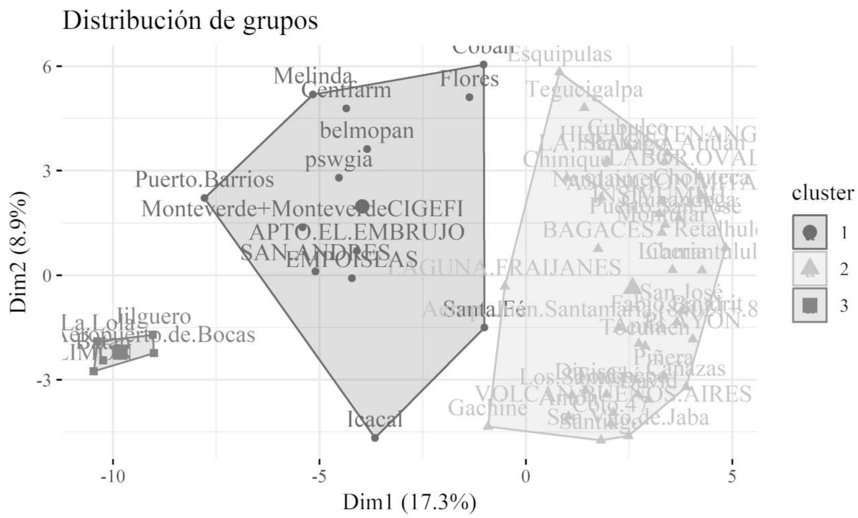


(a) ACU

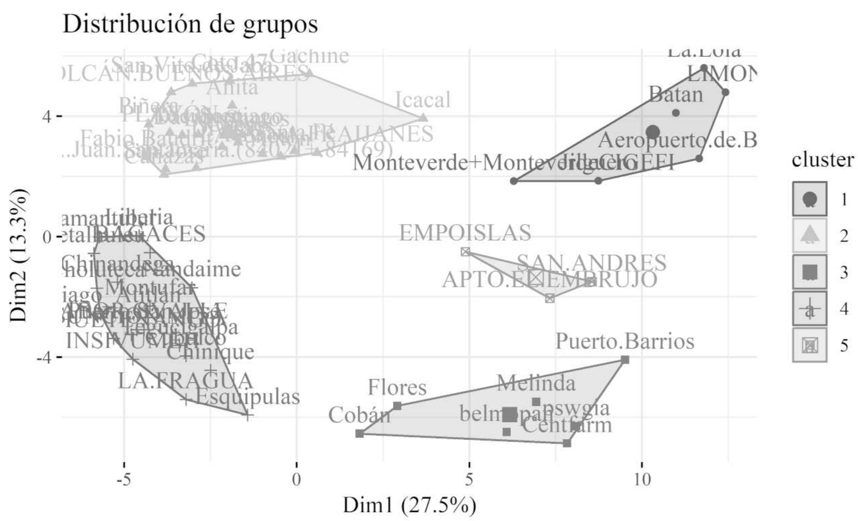


(b) DCP

Figura 3: Distribución de los grupos en cada una de las variables **a)** ACU, **b)** DCP, **c)** p80 y **d)** p20 en el espacio definido por las primeras dos componentes principales de cada conjunto de datos.

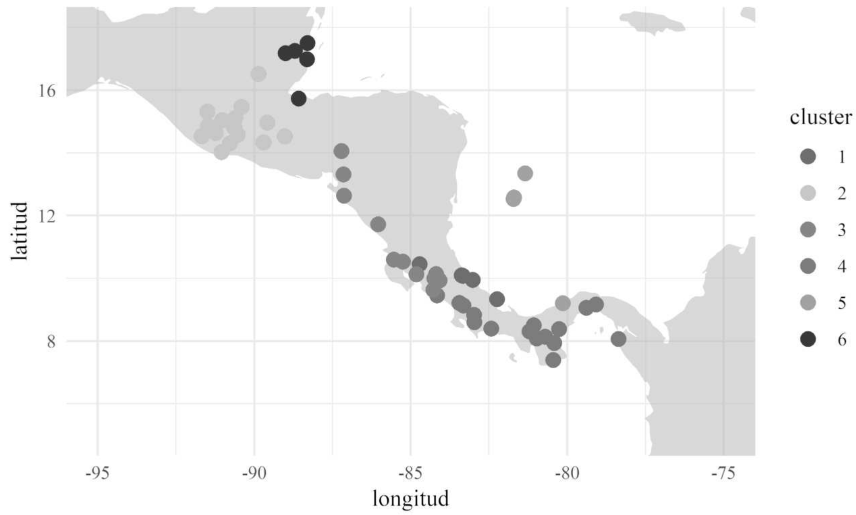


(c) p80

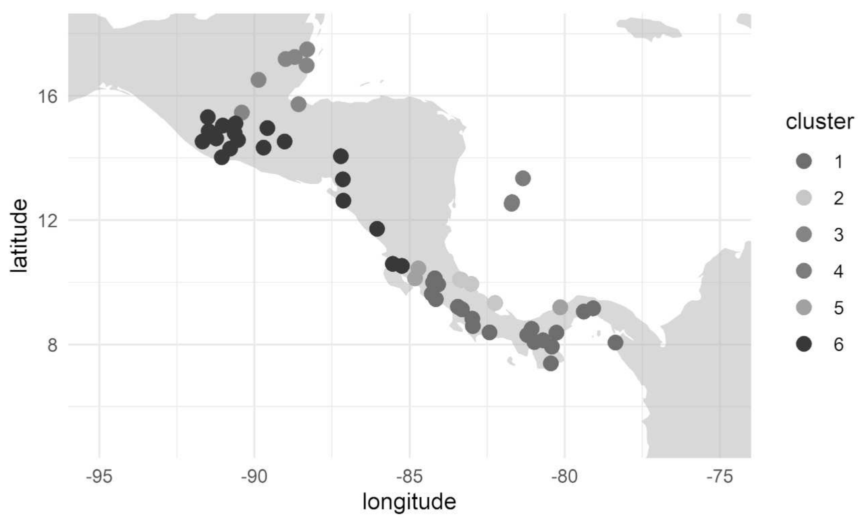


(d) p20

Figura 3: Distribución de los grupos en cada una de las variables **a)** ACU, **b)** DCP, **c)** p80 y **d)** p20 en el espacio definido por las primeras dos componentes principales de cada conjunto de datos. (cont.)

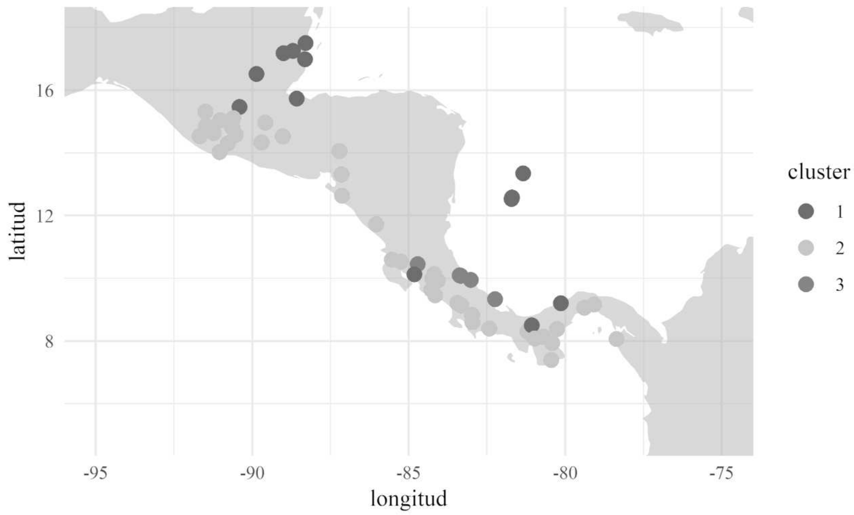


(a) ACU

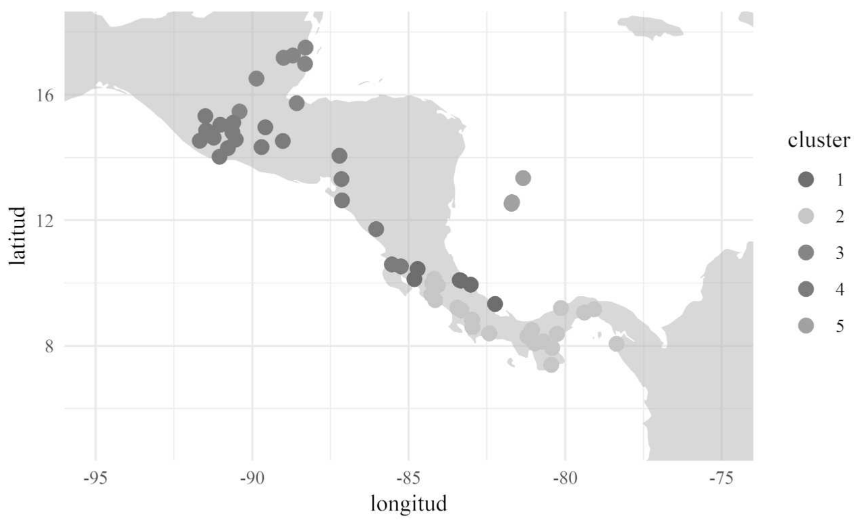


(b) DCP

Figura 4: Distribución espacial de los grupos en cada una de las variables **a)** ACU, **b)** DCP, **c)** p80 y **d)** p20.



(c) p80



(d) p20

Figura 4: Distribución espacial de los grupos en cada una de las variables **a)** ACU, **b)** DCP, **c)** p80 y **d)** p20. (cont.)

Los DCP (Fig. 5b) muestran tanto para las estaciones localizadas en la vertiente Pacífico como Caribe, una regularidad entre mayo y octubre donde se observa constantemente que hay al menos 20 días con lluvia. Sin embargo, nótese que las estaciones relacionadas a la vertiente Pacífico (clusters 1 y 6), presentan una disminución de días con precipitación entre los meses de junio-julio y agosto, coincidente con el MSD. Sin embargo, la disminución es más marcada en el cluster 6, localizado al norte de Centroamérica, y en la zona del Corredor Seco Centroamericano ([15],[21],[36]). Los grupos 2, 4 y 5 muestran una similitud en el ciclo anual ya que están localizados en la zona Caribe de Costa Rica (cluster 2), y en las islas de San Andrés y Providencia, los cuales evidencian una reducción de los DCP aproximadamente de febrero a abril, y una segunda reducción de agosto a octubre. El cluster 5, aparece como una zona de transición entre las vertientes Caribe y Pacífico de Costa Rica y Panamá. Cabe señalar que este último es el que presenta más cantidad de días con lluvia, siendo el mínimo durante los meses de febrero a abril, con 20 días con lluvia al menos, y llegando a tener cerca de los 30 días durante julio a diciembre. Por último, el cluster 3 se encuentra localizado en el oeste de Guatemala y Belice, y revela una reducción importante de febrero a mayo, con mínimo en abril, muy similar al ciclo de las estaciones ubicadas en el Pacífico; sin embargo, no exhibe una reducción de los días con precipitación que se pueda relacionar con el MSD.

Para los extremos húmedos (p80, Figura 5c), la división es menos compleja. Se observa que hay una división clara entre las estaciones de la vertiente Pacífico (cluster 2) y las del Caribe (cluster 1), excepto en el Caribe costarricense (cluster 3).

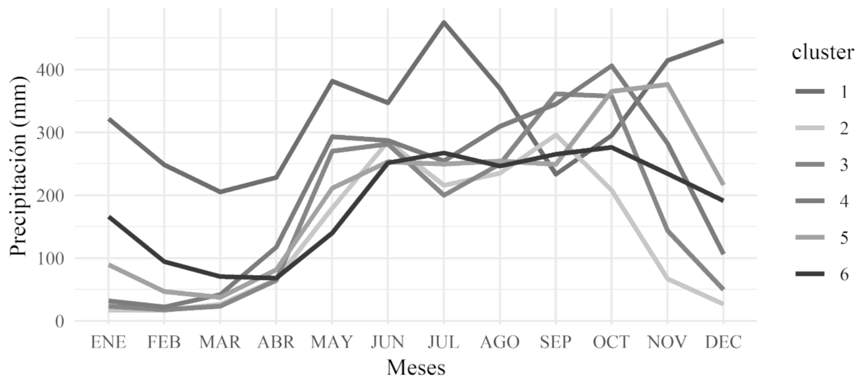
Nótese que estos ciclos anuales son similares a los ciclos de ACU y DCP, localizando los máximos y mínimos valores de porcentajes en los mismos meses correspondientes para cada cluster asociado entre cada una de las variables. Por ejemplo, se encuentran mínimos porcentajes en el cluster 2 de enero a marzo, correspondiente a la estación seca del Pacífico, y una reducción en julio correspondiente al MSD. Los máximos son alcanzados en junio y setiembre. Mientras que, para los clusters caribeños, hay una reducción del p80 entre los meses de marzo a abril, pero el rasgo más sobresaliente es el máximo en julio y mínimo relativo en setiembre del cluster 3, que no están presentes en el 1. Cabe destacar que el porcentaje de días en todos los clusters, que alcanzan extremos húmedos, apenas sobrepasa el 15% de los días como máximo, principalmente en el cluster 3. Los extremos secos (Figura 5d), muestran que los clusters localizados en el Pacífico(2y4), tienen un mayor porcentaje de días secos (por arriba del 80%) durante los meses de diciembre a abril, y un menor porcentaje de mayo a octubre, con un ligero incremento durante julio a agosto. Los clusters3y5 (que se pueden identificar con los clusters3y4 de DCP), presentan un ciclo similar con máximos en marzo, abril y valores mínimos de mayo a enero. El cluster 1, captura nuevamente la región Caribe de Costa Rica, la cual se diferencia con el resto de grupos en el Caribe por el mínimo relativo en julio y la aparición de un máximo relativo en setiembre. Nótese que los grupos hallados en los extremos secos (p20) contrastan correctamente los grupos encontrados para las otras variables (ACU, DCP y p80). La Tabla 1 contiene un resumen de los grupos identificados para cada variable y su similitud en localización, y ciclo anual.

Tabla 1: Resumen de los cluster encontrados por variables divididos por región. Se utilizaron las dos regiones climáticas conocidas tradicionalmente Pacífico y Caribe, pero se incluye la región Caribe de Costa Rica, ya que el método la diferenció del resto, particularmente por lo que se observa en los meses de julio y setiembre.

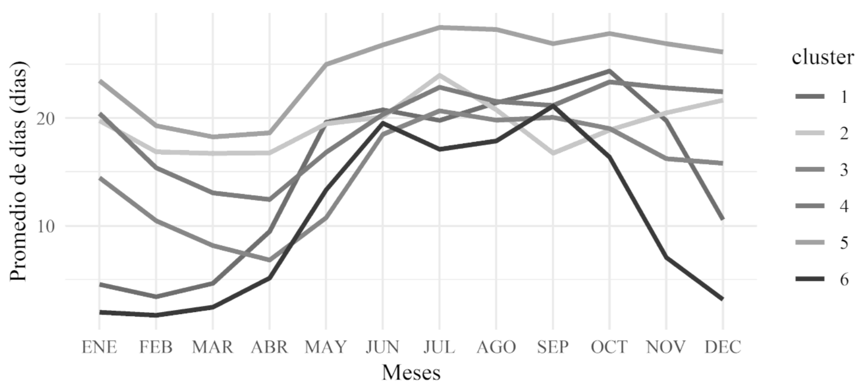
Vertiente	ACU	DCP	p80	p20
Pacífico	2,3,4	1,6	2	2,4
Caribe	5,6	3,4,5	1	3,5
Caribe Sur	1	2	3	1

Tabla 2: Promedio de anomalías estacionales para las décadas 1990 (1986–1995), 2000 (1996–2005) y 2010 (2006–2015) calculadas para la variable ACU, utilizando como periodo base la década de 1980 (1976–1985). En negrita las anomalías con un valor p menor a 0.05.

	cluster 1			cluster 4		
	1990	2000	2010	1990	2000	2010
DEF	-6	**128**	**105**	1	20	14
MA	31	42	42	-4	-4	6
MJ	19	**138**	64	-11	0	18
JA	-2	-20	-17	31	21	57
SON	-5	16	40	14	14	**46**
	cluster 2			cluster 5		
DEF	1	-4	2	-15	20	15
MA	7	2	**13**	-6	-11	0
MJ	-4	12	43	-1	-3	-26
JA	10	3	41	-25	-22	-18
SON	-3	**44**	**40**	11	24	**64**
	cluster 3			cluster 6		
DEF	0	**12**	7	-14	**-50**	-20
MA	-2	1	4	10	-2	-4
MJ	-15	-8	-17	-41	-36	-4
JA	14	-8	8	-20	-17	-43
SON	-13	51	20	13	14	3

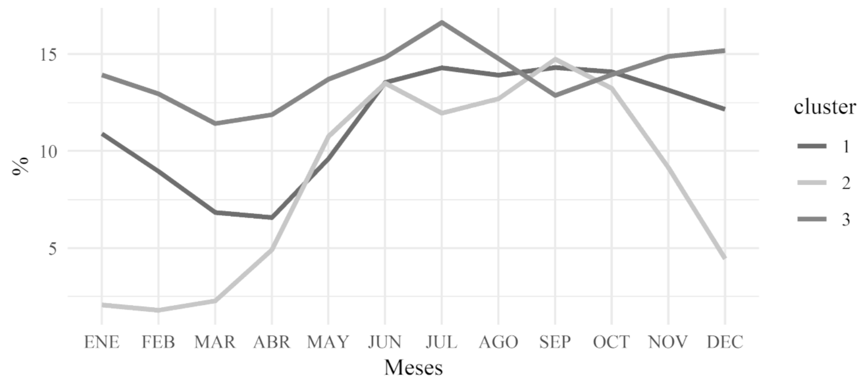


(a) ACU

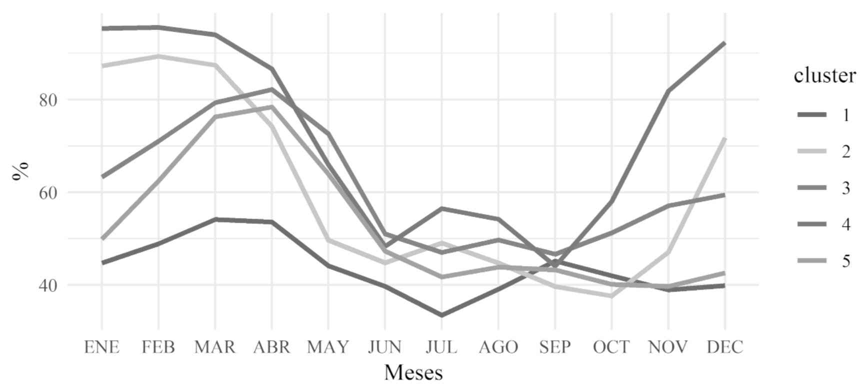


(b) DCP

Figura 5: Climatología del periodo 1975 - 2015 de las variables **a)** acumulado total de precipitación (ACU), **b)** frecuencia de días con precipitación (DCP), **c)** porcentaje de días que exceden el percentil 80 (p80) y **d)** porcentaje de días que no exceden el percentil 20 (p20), mostradas por cada grupo calculado en cada una de las variables. Las climatologías se calcularon como el valor mensual promedio de las series de tiempo.



(c) p80



(d) p20

Figura 5: Climatología del periodo 1975 - 2015 de las variables **a)** acumulado total de precipitación (ACU), **b)** frecuencia de días con precipitación (DCP), **c)** porcentaje de días que exceden el percentil 80 (p80) y **d)** porcentaje de días que no exceden el percentil 20 (p20), mostradas por cada grupo calculado en cada una de las variables. Las climatologías se calcularon como el valor mensual promedio de las series de tiempo. (cont.)

Tabla 3: Promedio de anomalías estacionales para las décadas 1990 (1986–1995), 2000 (1996–2005) y 2010 (2006-2015) calculadas para la variable DCP, utilizando como periodo base la década de 1980 (1976–1985). En negrita las anomalías con un valor p menor a 0.05.

	cluster 1			cluster 4		
	1990	2000	2010	1990	2000	2010
DEF	0	**2**	1	-1	3	4
MA	0	0	1	-1	2	4
MJ	-1	0	0	0	2	3
JA	1	0	2	-1	2	3
SON	0	1	1	-1	2	3
	cluster 2			cluster 5		
DEF	0	0	1	**-3**	-1	-2
MA	0	0	1	-2	-3	0
MJ	1	1	0	-1	0	-2
JA	0	0	-1	-1	0	**-1**
SON	1	-1	0	0	**-1**	-1
	cluster 3			cluster 6		
DEF	1	0	**2**	0	0	0
MA	1	0	**2**	0	0	1
MJ	-1	0	1	0	0	1
JA	1	0	-1	0	0	1
SON	**2**	1	1	0	2	1

3.2 Cambios observados desde 1976 hasta 2015

Para estudiar los cambios experimentados por la precipitación durante el periodo de 1975 al 2015, se dividió la serie de tiempo en cuatro décadas, utilizando como línea base el periodo de 1976-1985. Las siguientes décadas seleccionadas corresponden a los periodos 1986-1995 (1990), 1996-2005 (2000) y 2006-2015 (2010). Las anomalías se estimaron para cada estación (no se muestran los resultados) y para cada cluster, en escalas de tiempo mensual, estacional y anual.

Las tendencias son estimadas de los valores anuales de cada uno de las escalas de tiempo mensual, estacional y anual. La significancia estadística de la tendencia es determinada por el *valor p*.

Las Tablas 2 a 5 muestran un resumen de las anomalías estacionales calculadas para cada variable, y por grupos. De estas tablas se observan cambios

Tabla 4: Promedio de anomalías estacionales para las décadas 1990 (1986–1995), 2000 (1996–2005) y 2010 (2006-2015) calculadas para la variable p80, utilizando como periodo base la década de 1980 (1976–1985). En negrita las anomalías con un valor p menor a 0.05.

	cluster 1			cluster 2		
	1990	2000	2010	1990	2000	2010
DEF	-1	0	0	0	1	0
MA	0	0	1	0	0	1
MJ	-2	-1	-1	0	0	1
JA	0	0	0	1	0	2
SON	0	2	2	0	2	2
	cluster 3					
DEF	-1	**6**	4			
MA	2	2	3			
MJ	1	**6**	3			
JA	0	0	0			
SON	1	**1**	3			

heterogéneos en cada uno de los casos y la mayoría sin alguna significancia estadística. Esto último también se encontró en los resultados mensuales y por estaciones (no se muestra), lo cual es derivado a la alta variabilidad espacial y temporal que presenta una variable como la lluvia en esta región. Nótese que a pesar de que algunas variables, grupos y estaciones exhiben algunos resultados significativos, no es algo que necesariamente se replica en las demás variables. Por ejemplo, los grupos 1 ACU (Tabla 2), 2 DCP (Tabla 3), 3 p80 (Tabla 4) y 1 p20 (Tabla 5), representan la región del Caribe costarricense. Para la estación DJF (un periodo de bastante lluvia en esta región), se obtienen anomalías positivas y significativas en la variable ACU en las últimas dos décadas del análisis (2000 y 2010). Sin embargo, ni en el cluster 2 de DCP, 3 de p80 o 1 de p20 se observan cambios significativos, a pesar de haber un aumento en el p80 y una disminución en el p20 en estos clusters durante el mismo periodo. Cabe destacar que las tendencias fueron también analizadas (no se muestran) en las diferentes escalas de tiempo, sin embargo los resultados revelaron la misma heterogeneidad comentada anteriormente.

Tabla 5: Promedio de anomalías estacionales para las décadas 1990 (1986–1995), 2000 (1996–2005) y 2010 (2006-2015) calculadas para la variable p20, utilizando como periodo base la década de 1980 (1976–1985). En negrita las anomalías con un valor p menor a 0.05.

	cluster 1			cluster 4		
	1990	2000	2010	1990	2000	2010
DEF	2	-2	-4	0	1	0
MA	0	-1	-4	0	1	-2
MJ	-3	-6	-1	1	0	-2
JA	0	1	4	-1	0	-3
SON	-2	0	-2	1	-4	-2
	cluster 2			cluster 5		
DEF	1	**-6**	-3	0	-3	-2
MA	1	0	-2	4	1	-2
MJ	3	2	2	-1	-3	1
JA	-	-2	-5	1	3	2
SON	0	-2	-2	2	1	-1
	cluster 3					
DEF	-2	4	-1			
MA	-1	1	-3			
MJ	5	3	1			
JA	2	4	6			
SON	-3	1	1			

4 Conclusión

Utilizando el estadístico de brecha, se encontraron diferentes subregiones climáticas para América Central en cada una de las variables que se definieron para describir la precipitación. Sobresale que la técnica separa adecuadamente estaciones del Pacífico y del Caribe centroamericano, pero además se identificó una subregión localizada en el Caribe costarricense. Los ciclos anuales mostraron principales diferencias en las cantidades de lluvia, y en los tiempos de ocurrencia de procesos climáticos que afectan la región, principalmente el MSD. Esto último, es evidenciado principalmente en la región Caribe de Costa Rica, en la cual se encontró diferencias importantes con el resto del Caribe, en un aumento de la lluvia durante julio y una disminución importante durante setiembre. Al estudiar los cambios de precipitación en el periodo 1976-2015, utilizando tanto las estaciones como las subregiones, y a escalas de tiempo mensual, estacional o anual no se encontró evidencia de cambios significativos, ni de una tendencia clara, a excepción de casos puntuales, por lo que no se puede llegar a un resultado generalizado para la región, en concordancia con estudios previos ([1], [20], [26], [37]).

Financiamiento y agradecimientos

Los autores agradecen a los siguientes proyectos inscritos en la Universidad de Costa Rica: 805-C0-610 (Fondo de Estímulo), C0-074, B8-766 (Redes), B9-454 (Grupos), B0-810 y EC-497 (VarClim).

Referencias

[1] E. Aguilar, T.C. Peterson, P.R. Obando, R. Frutos, J.A. Retana, M. Solera,..., R. Mayorga, *Changes in precipitation and temperature extremes in Central America and northern South America*, Journal of Geophysical Research: Atmospheres (2005), 1961–2003. Doi: 10.1029/2005JD006119

[2] E.J. 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. Doi: 10669/15428

[3] E. Alfaro, L. Cid, *Ajuste de un modelo VARMA para los campos de anomalías de precipitación en Centroamérica y los índices de los océanos Pacífico y Atlántico Tropical*, Atmósfera **12**(1999), no. 4, 205–222. Doi: 10669/15236

[4] E. Alfaro, F. Soley, *Ajuste de un modelo VAR como predictor de los campos de anomalías de precipitación en Centroamérica*, Rev. Mate. Teor. Aplic. **8**(2001), no. 1, 99–116. Doi: 10.15517/RMTA.V8I1.199

[5] M. Alfaro-Córdoba, H.G. Hidalgo, E.J. Alfaro, *Aridity trends in Central America: A spatial correlation analysis*, Atmosphere **11**(2020), no. 4, 427. Doi: 10.3390/atmos11040427

[6] J.A. Amador, *A climatic feature of the tropical Americas: The trade wind easterly jet*, Tópicos Meterológicos y Oceanográficos **5**(1998), no. 2, 91–102. Doi: 10669/76623

[7] J.A. Amador, *The intra–Americas sea low–level jet: Overview and future research*. Annals of the New York Academy of Sciences **1146**(2008), no. 1, 153–188. Doi: 10.1196/annals.1446.012

[8] J.A. Amador, A.M. Durán-Quesada, E.R. Rivera, G. Mora, F. Sáenz, B. Calderón, N. Mora, *The easternmost tropical Pacific. Part II: Seasonal and intraseasonal modes of atmospheric variability*, Revista de Biología Tropical **64**(2016), no. 1, S23–S57. Doi: 10.15517/RBT.V64I1.23409

[9] J.A. Amador, E.R. Rivera, A.M. Durán-Quesada, G. Mora, F. Sáenz, B. Calderón, N. Mora, *The easternmost tropical Pacific. Part I: A climate review*, Revista de Biología Tropical **64**(2016), no. 1, S1–S22. Doi: 10.15517/RBT.V64I1.23407

[10] J.A. Amador, E.J. Alfaro, O.G. Lizano, V.O. Magaña, *Atmospheric forcing of the eastern tropical Pacific: A review*, Progress in Oceanography **69**(2006), no. 2-4, 101–142. Doi: 10.1016/j.pocean.2006.03.007

[11] T.G. Anderson, K.J. Anchukaitis, D. Pons, M. Taylor, *Multiscale trends and precipitation extremes in the Central American Midsummer Drought*, Environmental Research Letters **14**(2019), no. 12, 124016. Doi: 10.1088/1748-9326/ab5023

[12] T. Calinski, J. Harabasz, *A dendrite method for cluster analysis*, Communications in Statistics **3**(1974), no. 1, 1–27. Doi: 10.1080/03610927408827101

[13] R. Fuentes-Franco, E. Coppola, F. Giorgi, E.G. Pavia, G.T. Diro, F. Graef, *Inter-annual variability of precipitation over Southern Mexico and Central America and its relationship to sea surface temperature from a set of future projections from CMIP5 GCMs and RegCM4 CORDEX simulations*, Climate Dynamics **45**(2015), 425–440. Doi: 10.1007/s00382-014-2258-6

[14] F. Giorgi, *Climate change hot-spots*, Geophysical Research Letters **33**(2006), no. 8, L08707. Doi: 10.1029/2006GL025734

[15] Y. Gotlieb, P.M. Pérez-Briceño, H.G. Hidalgo, E.J. Alfaro, *The Central American Dry Corridor: A consensus statement and its background*, "Yu'am" Revista Mesoamericana de Biodiversidad y Cambio Climático **3**(2019), no. 5, 42–51.

[16] H. S. Guillén-Oviedo, L. Cid-Serrano, E.J. Alfaro-Martínez, *Comparison of parameters of the generalized extreme value distribution associated with*

extreme rainfall events in Central America, Uniciencia **34**(2020), no. 1, 111–128. Doi: 10.15359/ru.34-1.7

[17] J.A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, New York, 1975.

[18] J.A. Hartigan, M.A. Wong, *Algorithm AS 136: A k-means clustering algorithm*. Journal of the Royal Statistical Society, Series C (Applied Statistics) **28**(1979), no. 1, 100-108. Doi: 10.2307/2346830

[19] C. Hennig, M. Meila, F. Murtagh, R. Rocci, *Handbook of Cluster Analysis*, 1era Ed. Chapman & Hall, London, 2015. Doi: 10.1201/b19706

[20] H. Hidalgo, E. Alfaro, B. Quesada-Montano, *Observed (1970-1999) climate variability in Central America using a high-resolution meteorological dataset with implication to climate change studies*, Climatic Change **141**(2017), 13–28. Doi: 10.1007/s10584-016-1786-y

[21] H.G. Hidalgo, E.J. Alfaro, J.A. Amador, A. Bastidas, *Precursors of quasi-decadal dry-spells in the Central America Dry Corridor*, Clim. Dyn. **53**(2019), 1307–1322. Doi: 10.1007/s00382-019-04638-y

[22] H.G. Hidalgo, J.A. Amador, E.J. Alfaro, B. Quesada, *Hydrological climate change projections for Central America*, Journal of Hydrology **495**(2013), 94–112. Doi: 10.1016/j.jhydrol.2013.05.004

[23] H.G. Hidalgo, A.M. Durán-Quesada, J.A. Amador, E.J. Alfaro, *The Caribbean low level jet, the inter-tropical convergence zone and precipitation patterns in the intra Americas sea: a proposed dynamical mechanism*, Geogr. Ann. Series A Phys. Geogr. **97**(2015), no. 1, 41–59. Doi: 10.1111/geoa.12085

[24] P.A. Imbach, B. Locatelli, L.G. Molina, P. Ciais, P.W. Leadley, *Climate change and plant dispersal along corridors in fragmented landscapes of Mesoamerica*, Ecology and Evolution **3**(2013), no. 9, 2917–2932. Doi: 10.1002/ece3.672

[25] P. Imbach, M. Beardsley, C. Bouroncle, C. Medellin, P. Läderach, H. Hidalgo,...,C.I. Donatti, *Climate change, ecosystems and smallholder agriculture in Central America: an introduction to the special issue*, Climatic Change **141**(2017), 1–12. Doi: 10.1007/s10584-017-1920-5

[26] P.D. Jones, C. Harpham, I. Harris, C.M. Goodess, A. Burton, A. Centella-Artola,..., T. Baur, *Long-term trends in precipitation and temperature*

across the Caribbean, International Journal of Climatology **36**(2016), no. 9, 3314-3333. Doi: 10.1002/joc.4557

[27] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning (Vol. 1)*, STHDA, 2017. https://www.amazon.com/-/es/Mr-Alboukadel-Kassambara/dp/1542462703

[28] V. Magaña, J.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<1577:TMDOMA>2.0.CO;2

[29] T.J. Maldonado, E.J. Alfaro, H.G. Hidalgo, *A review of the main drivers and variability of Central America's Climate and seasonal forecast systems*, Revista de Biología Tropical **66**(2018), no. 1, S153–S175. Doi: 10.15517/RBT.V66I1.33294

[30] T. Maldonado, E. Alfaro, B. Fallas, L. Alvarado, *Seasonal prediction of extreme precipitation events and frequency of rainy days over Costa Rica, Central America, using canonical correlation analysis*, Adv. Geosci. **33**(2013), 41–52. Doi: 10.5194/adgeo-33-41-2013

[31] T. Maldonado, E. Alfaro, A. Rutgersson, J.A. Amador, *The early rainy season in Central America: The role of the tropical North Atlantic SSTs*, Int. J. Climatol. **37**(2017), no. 31, 3731–3742. Doi: 10.1002/joc.4958

[32] T. Maldonado, A. Rutgersson, E. Alfaro, J. Amador, B. Claremar, *Interannual variability of the midsummer drought in Central America and the connection with sea surface temperatures*, Adv. Geosci. **42**(2016), 35–50. Doi: 10.5194/adgeo-42-35-2016

[33] T. Maldonado, A. Rutgersson, R. Caballero, F.S.R. Pausata, E. Alfaro, J. Amador, *The role of the meridional sea surface temperature gradient in controlling the Caribbean low-level jet*, J. Geophys. Res. Atmos. **122**(2017), 5903–5916. Doi: 10.1002/2016JD026025

[34] G.W. Milligan, M.C. Cooper, *An examination of procedures for determining the number of clusters in a data set*, Psychometrika **50**(1985), no. 2, 159–179. Doi: 10.1007/BF02294245

[35] R. Mojena, *Hierarchical grouping methods and stopping rules: an evaluation*, The Computer Journal **20**(1977), no. 4, 359–363. Doi: 10.1093/comjnl/20.4.359

[36] L.E. Quesada-Hernández, O.D. Calvo-Solano, H.G. Hidalgo, P.M. Pérez-Briceño, E. J. Alfaro, *Dynamical delimitation of the Central America Dry Corridor (CADC) using drought indices and aridity values*, Progress in Physical Geography **43**(2019), no. 5, 627–642. Doi: 10.1177/0309133319860224

[37] T.S. Stephenson, L.A. Vincent, T. Allen, C.J. Van Meerbeeck, N. McLean, T.C. Peterson,..., A.R. Trotman, *Changes in extreme temperature and precipitation in the Caribbean region, 1961-2010*, International Journal of Climatology **34**(2014), no. 9, 2957–2971. Doi: 10.1002/joc.3889

[38] M.A. Taylor, E.J. Alfaro, *The climate of Central America and the Caribbean*, in: J.E. Oliver (Ed.), Encyclopedia of World Climatology, Springer Netherlands, Dordrecht, 2005, pp. 183–189. Doi:

[39] R. Tibshirani, G. Walther, T. Hastie, *Estimating the number of clusters in a data set via the gap statistic*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**(2001), no. 2, 411–423. Doi: 10.1111/1467-9868.00293

[40] D.S. Wilks, *Statistical Methods in the Atmospheric Sciences*, 4ta Ed. Academic Press, Oxford, 2019. https://www.elsevier.com/books/statistical-methods-in-the-atmospheric-sciences/wilks/978-0-12-8158