

# Determinación de áreas hidroclimáticamente homogéneas. Una propuesta técnica

## *Determination of hydroclimatically homogeneous areas. A technical proposal*

De León Pérez, D. <sup>a,b\*</sup>, Domínguez, E. <sup>c</sup>

<sup>a</sup>Grupo de Investigación ciencia e ingeniería del agua y el ambiente, Facultad de Ingeniería, Pontificia Universidad Javeriana, Bogotá, Colombia.

<sup>b</sup>Instituto de Investigación de Ingeniería del Agua y Medio Ambiente, Universitat Politècnica de València, Valencia, España.

<sup>c</sup>Facultad de Estudios Ambientales y Rurales, Pontificia Universidad Javeriana, Bogotá, Colombia.

E-mail: <sup>a</sup>daviddeleon@javeriana.edu.co, <sup>b</sup>ddeleo1@doctor.upv.es, <sup>c</sup>e.dominguez@javeriana.edu.co

\*Autor para correspondencia

Recibido: 22/11/2020

Aceptado: 12/03/2021

Publicado: 30/04/2021

**Citar como:** De León Pérez, D., Domínguez, E. 2021. Determination of hydroclimatically homogeneous areas. A technical proposal. *Ingeniería del agua*, 25(2), 97-114. <https://doi.org/10.4995/la.2021.14659>

## RESUMEN

Diferentes investigadores en hidroclimatología aplican técnicas basadas en autovectores para comprimir grandes volúmenes de información mientras conservan la estructura invariante de los datos originales. La presente investigación desarrolló una metodología que aplica una de estas técnicas, Análisis de Componentes Principales, a los elementos de variabilidad en series de tiempo hidroclimáticas y luego se identifican grupos o clústers mediante el método “*k-means*”. El resultado es un mapa regionalizado por variable. Finalmente se hace la intersección de estos mapas y obteniéndose áreas que presentan una estructura hidroclimática homogénea debido que las variables comparten su estructura de varianza. En el caso de estudio se evaluaron 8 variables para Colombia (9268 series de tiempo), obteniendo como resultado 26 regiones hidroclimáticas. Obtener regiones hidroclimáticamente homogéneas brinda la posibilidad de generar, entre otros, proyectos de adaptación al cambio climático de forma localizada con el fin de dar soluciones cuasi particulares que maximicen los resultados.

**Palabras clave** | regionalización hidroclimática, áreas homogéneas, hidroclimatología, climatología, variabilidad hidroclimática, series de tiempo, análisis de componentes principales.

## ABSTRACT

*Different hydroclimatology researchers apply eigenvectors-based techniques to compress large volumes of information while preserving the invariant structure of the original data. This research developed a methodology applying one of these techniques, Principal Component Analysis, on the elements of variability in hydroclimatic time series, and then to identify clusters with the “k-means” method. The result is a regionalized map by variable. Finally, these maps are intersected, obtaining areas with a homogeneous hydroclimatic structure, because the variables have similarity in their variance structure. In the case study, 8 variables were evaluated for Colombia (9268 time series), obtaining as a result 26 hydroclimatic regions. Obtaining hydroclimatically homogeneous regions offers the possibility of generating, among others, projects for climate change adaptation in a localized way to provide quasi-specific solutions that maximize results.*

**Key words** | hydroclimatic regionalization, homogenous areas, hydroclimatology, climatology, hydroclimatic variability, time series, principal components analysis.

## INTRODUCCIÓN

El clima es un factor importante en el desarrollo de las actividades humanas, ya que de acuerdo con las condiciones prevalecientes el hombre determina uno u otro lugar para establecerse o para el desarrollo de ciertas actividades productivas. Por definición el clima es una caracterización promedio de las condiciones variantes en el tiempo de un área o región o lugar determinado y se caracteriza con el valor esperado para las variables hidroclimáticas (Lorenz, 1970; Hannachi, 2004). Es conocido que en el planeta tierra a lo largo del tiempo se han producido diferentes variaciones u oscilaciones en el clima, estas se producen a diferentes escalas de tiempo con lo cual aparecen los conceptos de variabilidad y cambio climático (IDEAM, 2001).

La variabilidad climática incluye los eventos extremos, las fluctuaciones mensuales, estacionales y anuales con respecto a los valores esperados para las variables climáticas, enmarcada en ventanas temporales relativamente cortas, de otra parte la magnitud de la variabilidad se define al comparar los momentos estadísticos de la serie de tiempo en la ventana temporal analizada con los mismos momentos estadísticos calculados para una ventana temporal lo suficientemente extensa para caracterizar el clima (Poveda, 2004).

Por su parte el cambio climático incluye variaciones evaluadas durante periodos mucho más largos que van desde décadas hasta siglos, donde al evaluar una ventana de tiempo determinada *versus* el total de la serie de tiempo (ST), se notan cambios muy marcados en los primeros dos momentos estadísticos, de acuerdo con esto se puede inferir que cambios climáticos drásticos pueden dar origen a nuevos patrones “normales” del clima (IDEAM, 2001). Algunas fuentes que generan cambios climáticos vienen dadas por variaciones en parámetros orbitales, en la radiación solar o por movimientos en la corteza terrestre, además por ajustes del sistema atmosférico global en su búsqueda natural de un punto de equilibrio dinámico (Mesa *et al.*, 1997; León *et al.*, 2000; IDEAM, 2001).

La dinámica atmosférica a nivel global recibe la energía dentro de un esquema termodinámico en el que se desarrolla la propagación y difusión de propiedades de la atmósfera, tanto a nivel espacial como temporal, lo cual trae como consecuencia la aparición de sistemas de circulación confluyente que en general son concentradores de masas de aire, las que a su vez son reguladoras del estado del tiempo y el clima (León *et al.*, 2000; Montoya Gaviria, 2008). Por ejemplo, en el caso de la parte tropical de América del sur (donde se ubica Colombia), los ciclos climáticos son directamente influenciados por la oscilación anual de la Zona de Confluencia Intertropical-ZCIT, presentando adicional a esta variabilidad temporal, una variabilidad espacial debido a la presencia de la gran cuenca del río Amazonas, la cordillera de Los Andes y por influencia de la circulación atmosférica asociada a los océanos Atlántico tropical y Pacífico (Poveda y Mesa, 1997; Poveda *et al.*, 2001).

Debido a que la dinámica atmosférica tiende a seguir patrones o ciclos, que en la mayoría de los casos no se pueden detectar a simple vista (y tampoco son necesariamente previsibles), los hidrólogos, climatólogos y meteorólogos han trabajado de manera ardua con el fin de encontrar métodos o criterios que permitan clasificar patrones del clima que conecten la dinámica atmosférica global con la hidroclimatología local; gracias a esto aparecen técnicas como las basadas en aproximaciones a través de los autovectores, con las cuales comprimen la información original, pero guardan la estructura invariante de los datos y permiten ver núcleos en los cuales se agrupan las variables analizadas de acuerdo a sus patrones estructurales (Hotellin, 1933).

Las técnicas de Regionalización son metodologías que permiten dividir un área determinada en unidades territoriales más pequeñas y que sean homogéneas respecto a algún criterio establecido (hidroclimatología en este caso), regionalizar brinda una herramienta importante para el análisis y evaluación de sistemas ambientales, ya que permite establecer patrones climáticos espaciales para desarrollar una planificación más acorde a las condiciones particulares de cada región (White *et al.*, 1991; Gómez Lende, 2011). Por las brechas encontradas en la literatura, el objetivo principal de este trabajo fue determinar una metodología para la regionalización de áreas que hidroclimáticamente tengan una estructura homogénea para todas las variables hidroclimáticas evaluadas, respetando la dinámica atmosférica en cuanto a que la difusión de las propiedades varía tanto espacial como temporalmente. Un concepto clave para el desarrollo, es que en la metodología los límites de las regiones no dependen necesariamente de las líneas divisorias de cuenca, sino de la estructura invariante de los elementos de variabilidad que constituyen las ST de las variables hidroclimáticas.



## Recolección de la información base

Debido a que en el desarrollo del estudio involucra análisis y evaluación de las ST existentes para las variables hidroclimáticas, es necesario que la recopilación de datos sea de fuentes confiables de medición y/o cálculo, debido a esto los datos de las ST analizadas corresponden a las de la base de datos del El Instituto de Hidrología, Meteorología y Estudios Ambientales-IDEAM. Para tener datos homogéneos se tomaron entre 1970 y 2014, Colombia se vinculó al protocolo de la Organización Meteorológica Mundial-OMM hacia 1966. Las variables hidroclimáticas fueron analizadas a una resolución mensual con el fin de capturar los invariantes climáticos que, según Fedorov (2015) pueden estar asociados a los movimientos de la tierra (Traslación, Rotación, Nutación y Precesión). La descripción de las variables hidroclimáticas evaluadas se presentan en la Tabla 1.

**Tabla 1** | Variables hidroclimáticas analizadas. El numeral 1 indica que es media mensual y el 4 indica que es total mensual.

Variable Hidroclimática	Descripción
Brillo Solar (BS4)	Cantidad de tiempo que un determinado lugar está expuesto a la radiación solar directa durante un periodo de tiempo determinado (p.e. h/mes)
Evaporación (EV4)	Disminución en lámina de agua en un tanque de medición con un área de sección cercana a 1 m <sup>2</sup> , es medido en unidades de longitud que equivalen a un volumen por unidad de área, p.e. milímetros (mm) que equivalen a 1 litro por cada m <sup>2</sup>
Humedad Relativa (HR1)	Proporción de agua en estado gaseoso real en el aire relacionada con la cantidad de agua necesaria para que el aire llegue al punto de saturación para determinada temperatura; su medida se da en porcentaje (%)
Punto de Rocío (PR1)	Temperatura a la cual el agua contenida en una porción de aire inicia el proceso de condensación formando la neblina; la unidad de medida es la misma de la temperatura (p.e. °C)
Precipitación (PT4)	Es el resultado solido o liquido de la condensación del vapor de agua que se libera por el aire o las nubes cuando han alcanzado el punto de saturación; se mide en unidades de longitud (p.e. mm)
Caudal Líquido (QL1)	Es el volumen de agua que pasa por un lugar de referencia en un determinado periodo de tiempo, se mide en unidades de volumen por unidad de tiempo (p.e. m <sup>3</sup> /s)
Temperatura del aire (TS1)	Es una medida de la energía cinética de un objeto con respecto a su capacidad de intercambiar calor con su entorno, se destacan 3 diferentes unidades de medición grados Celsius (°C), Kelvin (K) y Fahrenheit (°F)
Tensión de Vapor (TV1)	Es la presión de la fase gaseosa de un sólido o liquido respecto a la fase líquida a determinada temperatura; su medición se realiza en unidades estándar de presión (p.e. N/m <sup>2</sup> )

## Análisis y evaluación de la información base

La información de ST que se recopila de las bases de datos mencionadas, requiere ser evaluada y depurada para trabajar con las series que sean de mejor calidad en búsqueda que los resultados obtenidos sean fiables. Debido a lo anterior se aplicaron las siguientes pruebas para realizar su depuración:

- Prueba de datos anómalos con los tests de Grubbs (Grubbs, 1950) y de dobles medianas (Leys *et al.*, 2013)
  - Test de Grubbs:

$$\Gamma = \frac{\max |\chi_i - \bar{\chi}|}{\sigma} \quad (1)$$

De donde  $\Gamma$ : estadístico de Grubbs;  $\bar{\chi}$ : promedio de las observaciones;  $\chi_i$ : observación  $i$ ;  $\sigma$ : desviación estándar de la muestra

Cuando se realiza un análisis de distribución de dos colas el Estadístico de Grubbs se evalúa para valores máximos (cola derecha) como:

$$\Gamma = \frac{\chi_{\max} - \bar{\chi}}{\sigma} \quad (2)$$

Y para los valores mínimos (cola izquierda) como:

$$\Gamma = \frac{\bar{\chi} - \chi_{\min}}{\sigma} \quad (3)$$

De donde  $\chi_{\max}$ : valor máximo de las observaciones;  $\chi_{\min}$ : valor mínimo de las observaciones

El estadístico de Grubbs es calculado y para evaluar la hipótesis nula (el valor no es un anómalo) se debe cumplir que:

$$\Gamma < \frac{(\eta - 1)}{\sqrt{\eta}} \sqrt{\frac{(t_{(\alpha/2\eta), \eta - 2})^2}{\eta - 2 + (t_{(\alpha/2\eta), \eta - 2})^2}} \quad (4)$$

De donde  $\eta$ : número de datos;  $t_{(\alpha/2\eta), \eta - 2}$ : valor crítico de la distribución  $t_{student}$  con  $\eta - 2$  grados de libertad y nivel de significancia  $\alpha/2\eta$ .

- Test de dobles Desviaciones Absolutas de la Mediana (DAM): Consiste en estimar la mediana de las observaciones y a partir de este valor dividir la muestra en dos grupos de valores. Con estos dos grupos generados se procede a realizar de manera independiente un test de DAM para cada uno de los grupos y de acuerdo al valor de corte determinar que valores pueden ser anómalos. Para el valor de corte determinado se puede acceder a la tesis de maestría del ingeniero Andrés Velasco donde se presenta un análisis detallado (Velasco, 2016).

$$DAM = \text{median} (|\chi_i - \bar{\chi}|) \quad (5)$$

De donde  $\bar{\chi}$ : mediana de todas las observaciones (percentil 50)

- Evaluación de ventana tiempo más larga con datos disponibles, en esta evaluación se verifica que no existan más de dos años seguidos sin datos registrados (De León Pérez, 2017)
- Análisis de consistencia por longitud, en esta evaluación se seleccionan las ST con más de 30 años de registro (WMO, 1994)
- Análisis de consistencia por completitud, en esta evaluación se seleccionan las ST con menos de 30% de datos faltantes en el registro (WMO, 1994)
- Complementación de los datos faltantes mediante una red neuronal artificial (Govindaraju, 2000; Coulibaly y Evora, 2005). La complementación de las ST se desarrolló mediante una red neuronal artificial (RNA) teniendo como predictores iniciales las ST de estaciones vecinas que presentan mayor correlación con la estación a complementar. Sobre estas series de tiempo, y con el fin de optimizar los cálculos matemáticos, se aplicó una reducción dimensional de la matriz de predictores mediante un ACP (al 85% de la varianza total). De las diferentes configuraciones se utilizó una RNA del tipo perceptrón multicapa. Este es un tipo de algoritmo de aprendizaje supervisado, ya que se le ingresan valores de entrenamiento con los cuales el algoritmo genera la estructura matemática que calcula a partir de unos datos de entrada (predictores, neuronas de entrada) los valores de salida (objetivos, neuronas de salida). En un sistema de RNA del tipo perceptrón multicapa las entradas a las unidades de proceso para una capa, son las salidas de las unidades de proceso de la capa que la precede. Matemáticamente para una RNA de perceptrón multicapa con una sola capa oculta se expresaría como (Samarasinghe, 2016):

$$y_i = \gamma_1 \left( \sum_{j=1}^N w_{ij} s_j \right) = \gamma_1 \left( \sum_{j=1}^N w_{ij} \left( \gamma_2 \left( \sum_{r=1}^M w_{jr} x_r \right) \right) \right) \quad (6)$$

De donde  $y_i$ : valor pronosticado;  $w_{ij}$ : peso sináptico de la conexión entre la unidad de salida  $i$  y la unidad de proceso  $j$  de la capa oculta;  $N$ : unidades de proceso de la capa oculta;  $\gamma_1$ : función de transferencia de las unidades de salida;  $M$ : unidades de proceso de la capa de entrada;  $w_{jr}$ : peso sináptico de la conexión entre la unidad de proceso  $j$  de la capa oculta con el sensor de entrada  $r$ ;  $\gamma_2$ : función de transferencia de las unidades de capa oculta.

Para evaluar el error en el proceso de entrenamiento se utiliza un método de retropropagación, es decir, ajustar de acuerdo con el error obtenido en un proceso inverso formulado así:

$$E = \sum_{\kappa=1}^{\rho} \sum_{i=1}^M (Y_i(\kappa) - y_i(\kappa)) \quad (7)$$

De donde  $E$ : error entre lo medido y lo pronosticado;  $\rho$ : número de iteraciones;  $Y_i$ : función de transferencia de las unidades de salida;  $M$ : unidades de proceso de la capa de entrada.

Teniendo la actualización de pesos la ecuación de cálculo queda así:

$$w_{ij}(\kappa+1) = w_{ij}(\kappa) + \Delta w_{ij}(\kappa) \quad (8)$$

De donde:

$$\Delta w_{ij}(\kappa) = -\eta \frac{\partial E}{\partial w_{ij}(\kappa)} = \eta (Y_i(\kappa) - y_i(\kappa)) \cdot \gamma_i(h_{ij}) \cdot s_j(\kappa) \quad (9)$$

Siendo:

$$h_i = \sum_{j=1}^N w_{ij}(\kappa) \cdot s_j(\kappa) \quad (10)$$

- Prueba de aleatoriedad mediante la prueba de rachas (Bradley, 1968). El primer paso consiste en sustraer la media (o la mediana) de la muestra a cada una de las observaciones, a los resultados se les aplica una transformación binaria. Así, si el valor es mayor que cero se coloca un 1; si no, se coloca un 0. Sobre esta nueva serie de tiempo codificada de forma binaria se realiza el conteo de rachas resultante de esta codificación y se obtiene el número de rachas empíricas. Luego para calcular el número de rachas teóricas y su correspondiente desviación estándar se aplican las siguientes ecuaciones:

$$R_t = \frac{\eta + 1}{2} \quad (11)$$

$$\sigma_t = \frac{\sqrt{\eta - 1}}{2} \quad (12)$$

De donde  $R_t$ : conteo de rachas teóricas;  $\sigma_t$ : desviación estándar de rachas teóricas;  $\eta$ : número de datos.

Con estos valores se procede a calcular el intervalo de confianza así:

$$R_t - t_{(\alpha/2\eta), \eta-2} \cdot \sigma_t < R_e \leq R_t + t_{(\alpha/2\eta), \eta-2} \cdot \sigma_t \quad (13)$$

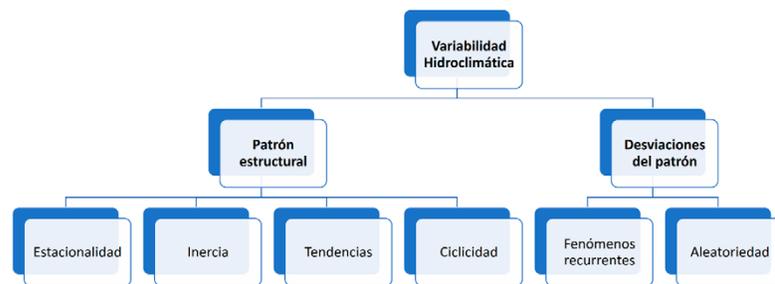
De donde  $t_{(\alpha/2\eta), \eta-2}$ : valor crítico de la distribución  $t_{student}$  con  $\eta-2$  grados de libertad y nivel de significancia  $\alpha/2\eta$ ;  $R_e$ : conteo de rachas empíricas.

Así si el número de rachas empíricas de la ST se encuentra dentro del intervalo de confianza, se considera aleatoria la ST. Este test se aplicó a cada uno de los conjuntos estadísticos analizados (12, 1 por cada mes del año) tomando como criterio que, si más del 70% de los meses resultaban aleatorios, se podría considerar la ST como aleatoria.

## Análisis y evaluación de los elementos de la variabilidad hidroclimática

Con el fin de capturar la información que puede ser invariante en la estructura de las ST, se utilizaron los elementos del concepto de variabilidad hidroclimática como los datos a ser utilizados para la generación de las regiones homogéneas. La variabilidad hidroclimática es formada por dos componentes, la variabilidad climática y la hidrológica. La variabilidad climática

evalúa oscilaciones del clima respecto a su estado medio en diferentes escalas tanto espaciales como temporales (IPCC *et al.*, 2013). El otro componente, variabilidad hidrológica, representa variaciones en tiempo y espacio del ciclo hidrológico, por lo tanto, es indispensable conocer el fenómeno a evaluar para seleccionar la escala temporal y espacial que capture el proceso en estudio. Variabilidad climática no implica que exista consigo variabilidad hidrológica, debido a que en esta última entra en juego la cuenca, que es una reguladora dentro del ciclo hidrológico gracias a la interacción de los seres vivos al interior de ella, por lo cual, la morfometría de la cuenca debe ser tenida en cuenta para evaluar la variabilidad hidrológica. Basados en lo anterior, la respuesta del sistema hidroclimático puede ser representada o expresada como el resultado de combinar los 6 fenómenos que se observan en la Figura 2.



**Figura 2** | Esquema de componentes de la variabilidad hidroclimática (Velasco, 2016; De León, 2017).

La forma como se combinan estos 6 componentes tiene un patrón determinístico, sin embargo, si se acepta que el principio de superposición es válido para los procesos del ciclo hidrológico, la construcción de una serie de tiempo hidroclimática podría esquematizarse de la siguiente forma:

$$\phi(t) = E(t) + \rho + \tau(t) + \zeta(t) + \lambda(t) + \xi \quad (14)$$

De donde:  $\phi(t)$ : serie de tiempo;  $E(t)$ : estacionalidad;  $\rho$ : inercia;  $\tau(t)$ : tendencia;  $\zeta(t)$ : ciclicidad (ciclos de largo plazo);  $\lambda(t)$ : fenómenos recurrentes;  $\xi$ : aleatoriedad

Para desarrollar la regionalización se trabajó con los elementos del patrón estructural de la ST, ya que las desviaciones del patrón principal dependen de eventos macroclimáticos más complejos y de la aleatoriedad, no obstante, al evaluar de manera probabilística los componentes del patrón estructural, en cierta forma se tiene en cuenta la aleatoriedad del proceso.

## Caracterización estocástica

Con el fin de definir los elementos del patrón estructural se desarrolló una caracterización estocástica de las ST previamente depuradas, para trabajar con los datos de mejor calidad. Para la caracterización estocástica se realizaron los siguientes pasos:

- Ajuste de Funciones de Densidad de Probabilidad (FDP) para cada conjunto estadístico válido (Haan, 1977), para cada mes en este caso: Existen muchas distribuciones teóricas de amplio uso con soporte y fundamentación matemática, que además están bien implementados en diferentes librerías en códigos de programación, p.e. la librería ScipyStats (Jones *et al.*, 2001) del lenguaje de programación Python.
- Evaluación de la inercia,  $\rho$ : La mayoría de los procesos físicos presentan una inercia que describe la memoria del proceso, es decir, cuan dependientes son los valores de un instante de tiempo “ $i$ ” respecto a otro instante “ $j$ ” presentando entre estos dos instantes de tiempo un rezago ( $\tau$ ). Debido a que es un análisis de correlación de la serie consigo misma el método matemático de análisis que se utiliza es la Función de Autocorrelación, que evalúa la correlación de la serie de tiempo con ella misma en un determinado rezago  $\tau$  (Montgomery *et al.*, 2015; De Gooijer, 2017). Esto permite determinar la

presencia de estacionalidad (o ciclicidad) en las ST, lo que sería la memoria del proceso, pero también permite determinar la memoria de corto plazo o inercial cuando se trabaja con la ST desestacionalizada.

- Estimación de la estacionalidad,  $E(t)$ : Para el cálculo de la estacionalidad en primera instancia se asume el proceso como estacionario en el sentido débil. En el caso de caracterización de clima, la estacionalidad se mide con el primer momento al origen y su franja de variabilidad como la raíz cuadrada del segundo momento central (Malcolm and McDonnell, 2005; Domínguez *et al.*, 2010; Dingman, 2015).
- Análisis de ciclos seculares,  $\zeta(t)$ : Los ciclos seculares representan ciclos de baja frecuencia en las variables hidroclimáticas y su análisis se realiza para evaluar posibles efectos del cambio climático o encontrar una estructura temporal de las variables ante el equilibrio dinámico e inestable de la atmósfera (Malcolm and McDonnell, 2005). Estos se calculan mediante la curva de diferencias integradas (CDI), la cual es un artificio matemático que transforma los datos en las anomalías de los mismos respecto a la media, los escala respecto a la desviación estándar y luego se acumulan. Cuando se realizan todos los cálculos se construye la gráfica de diferencias integradas (Fecha vs. CDI) o si se quiere mirar rachas de humedad y sequía se puede construir la curva de diferencias (Fecha vs. CDi). Su cálculo se desarrolla mediante una tabla de la siguiente forma:

*Columna 1: Fecha; Columna 2:  $x_i$ ; Columna 3:  $k_i = x_i / \bar{x}$ ; Columna 4:  $CD_i = (k_i - 1) / C_v$ ; Columna 5:  $CDI = \sum_{i=1}^{\eta} CD_i$*

De donde  $x_i$ : valor registrado;  $\bar{x}$ : media de las observaciones;  $CD_i$ : diferencias relativas;  $C_v$ : coeficiente de variación;  $\eta$ : número de observaciones;  $CDI$ : diferencias integradas

## Regionalización

Para la regionalización de las variables hidroclimáticas se utilizó el método del Análisis de Componentes Principales (ACP) con el fin de agrupar las ST de acuerdo con la estructura de los datos. La información de entrada son las ST y algunas transformaciones de estas (Percentil 50 y Curva de diferencias) con el fin de verificar que los grupos generados fueran consecuentes y darle robustez al método planteado. El método desarrollado toma como datos de entrada las ST (o su transformación) y aplica una reducción dimensional mediante un ACP (Varianza explicada >70%), luego se agrupan las estaciones mediante el método de los K-means.

El principio básico de la formulación del K-means es dividir una muestra de  $\eta$  observaciones en  $C$  clústers con la mayor homogeneidad posible, para esto el algoritmo de K-means tiene como objetivo elegir los centroides que minimizan la inercia al cuadrado (Ecuación (15)). Primero se determinan los centroides iniciales, luego se genera un bucle que asigna cada observación al centroide más próximo, ahí calcula el valor  $r$  y genera una segunda iteración produciendo nuevos centroides y calculando el valor  $r$ , se valúa la diferencia entre el primer  $r$  y el segundo de tal forma que el bucle se detiene cuando la diferencia sea menor que el umbral. La inercia, o el criterio de la suma de cuadrados dentro del grupo, puede ser reconocido como una medida de la coherencia interna del clúster (Arthur y Vassilvitskii, 2007).

$$\phi = \sum_{i=0}^{\eta} \underset{u_j \rightarrow C}{\text{MIN}} \left| |x_i - \mu_j| \right|^2 \quad (15)$$

De donde  $\phi$ : criterio de la suma mínima de inercia al cuadrado;  $x_i$ : observación  $i$ ;  $\mu_j$ : centroide  $j$ ;  $C$ : número de clústers;  $\eta$ : número de observaciones

Un detalle importante es que el método de los K-means requiere como dato inicial el número de grupos (clústers) para los cuales se va a evaluar la suma de la mínima inercia al cuadrado, sin embargo, el número de grupos será el número de regiones para cada variable hidroclimática, dato que la metodología planteada precisamente requiere determinar. Debido a esto se evaluó entre 2 y 12 clústers mediante el método del Coeficiente de Siluetas con el fin de hallar era el número óptimo de grupos. El proceso es iterativo, se propone el número de grupos (iniciando con 2 y hasta 12), con el método del “K-means” se calculan los centroides, luego se calcula el Coeficiente de Siluetas y finalmente se comparan los coeficientes obtenidos, seleccionando el número de clústers que tenga el menor valor para el Coeficiente de Siluetas. El Coeficiente de Siluetas se calcula utilizando la distancia media al

interior de clúster ( $\alpha$ ) versus la distancia media del clúster más próximo ( $\beta$ ) para cada observación (ver Ecuación (16)). El algoritmo de cálculo de la metodología se presenta en la Figura 3.

$$sc = \frac{1}{\eta} \sum_{i=0}^{\eta} \frac{\beta_i - \alpha_i}{MAX |\beta, \alpha|} \tag{16}$$

De donde  $sc$ : coeficiente de siluetas;  $\eta$ : número de observaciones;  $\beta_i$ : distancia media de la observación  $i$  al clúster más próximo;  $\alpha_i$ : distancia al interior del clúster al que pertenece la observación  $i$ .

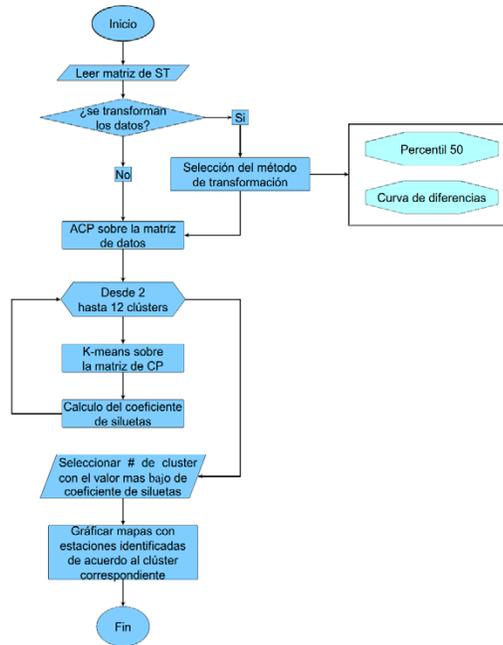


Figura 3 | Algoritmo resumen de la metodología planteada para la construcción de los mapas de regiones.

## RESULTADOS

### Evaluación de la información básica

El análisis de la información base (IDEAM) se muestra en la Tabla 2, donde se puede ver que solo cerca del 34% de ST son aptas para evaluación hidroclimática de acuerdo con criterios de la Organización Meteorológica Mundial:

Tabla 2 | Resumen cantidad de ST filtradas. Columna 1: Variable, columna 2: Cantidad de ST iniciales, columna 3: ST con datos antes de 1970, columna 4: ST que pasan análisis de consistencia, columna 5: ST que pasan el análisis por ventana óptima.

Variable hidroclimática	Inicial	< 1970	Consistencia	Ventana óptima
BS4	589	0	203	174
EV4	604	0	185	145
HR1	770	2	287	244
PR1	770	1	267	234
PT4	3802	1	1583	1497
QL1	1211	0	431	373
TS1	775	2	310	280
TV1	747	1	262	231
Total	9268	7	3528	3178

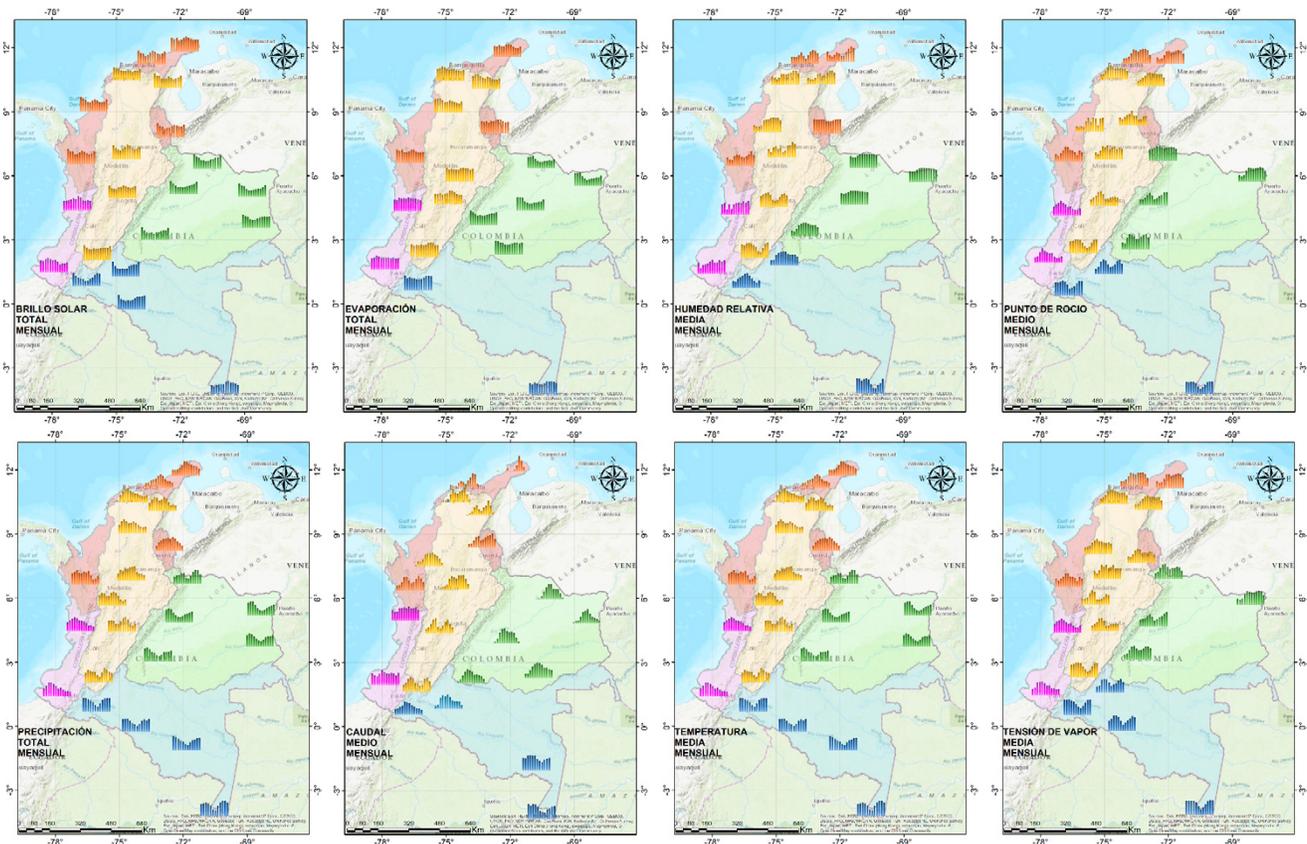
## Caracterización estocástica

Prueba de aleatoriedad: La prueba de aleatoriedad permite conocer si una serie de tiempo se comporta o no como aleatoria, ya que si no tienen un patrón aleatorio para sus realizaciones no es posible asignarle una FDP que la describa. Se puede observar en la Tabla que de las 3178 ST que habían pasado el análisis de consistencia, 2662 pasaron la prueba de aleatoriedad (considerando aleatorias las series de tiempo que al menos 8 meses del año resultaron aleatorios), es decir cerca del 84% de las ST efectivamente según la metodología del test de rachas se pueden considerar aleatorias.

**Tabla 3** | ST que pasan la prueba de aleatoriedad, discriminadas por variable.

Variable	Serie aleatorias
BS4	170
EV4	138
HR1	157
PR1	102
PT4	1451
QL1	345
TS1	201
TV1	98
Total	2662

Estacionalidad: En la Figura 4 se presenta el resultado de la evaluación de la estacionalidad (mediante el valor esperado para cada mes) en cada una de las 8 variables analizadas en el desarrollo de la presente investigación.

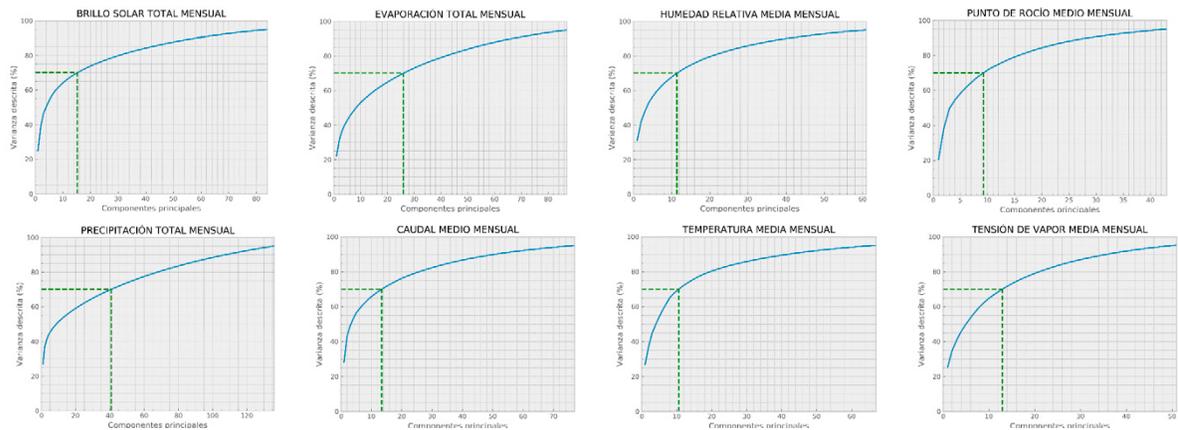


**Figura 4** | Estacionalidad en Colombia para las 8 variables analizadas. Las series anuales están escaladas de forma que sea esquemática la representación solo para ver los ciclos.

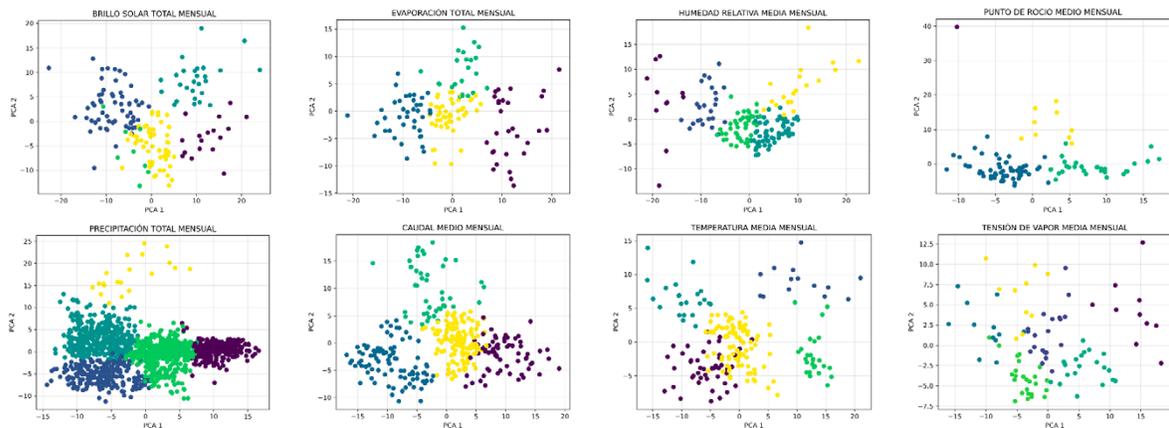
Inercia: La memoria inercial de la precipitación y el brillo solar a nivel general para todo el país es nula, es decir que los valores de un mes no están asociados con el mes que le precede o con el que el sucede. En el caso de los caudales existe una memoria inercial en prácticamente todas las regiones del país, presentándose en las regiones donde existen reservorios (naturales o artificiales, superficiales o subterráneos) valores de memoria inercial entre 4 y 7 meses con correlación significativa. El caso de la temperatura es similar al del caudal donde se presenta a nivel general de todo el país una memoria inercial en las ST analizadas, presentando en la mayoría de la zona andina y el pacífico del país una memoria inercial de entre 6 y 9 meses, hacia la Orinoquía a nivel general entre 1 y 3 meses, en el caribe en las zonas próximas a las Sierra Nevada de Santa Marta memoria inercial entre 2 y 5 meses.

Ciclos seculares: A nivel general para todo el país el análisis no se observa presencia de estos, esto puede ser debido a que la longitud de las series no es tan extensa y por ello no se observan de manera clara estructuras cíclicas de baja frecuencia a nivel regional. Sin embargo, algunas ST de forma aislada presentan estructuras que podrían asociarse con ciclos seculares.

## Análisis de Componentes Principales y generación de clústers



**Figura 5 |** Resultado del ACP aplicado en la base de datos, En azul número de componentes principales versus varianza descrita y en verde componentes principales para describir el 70% de la varianza.



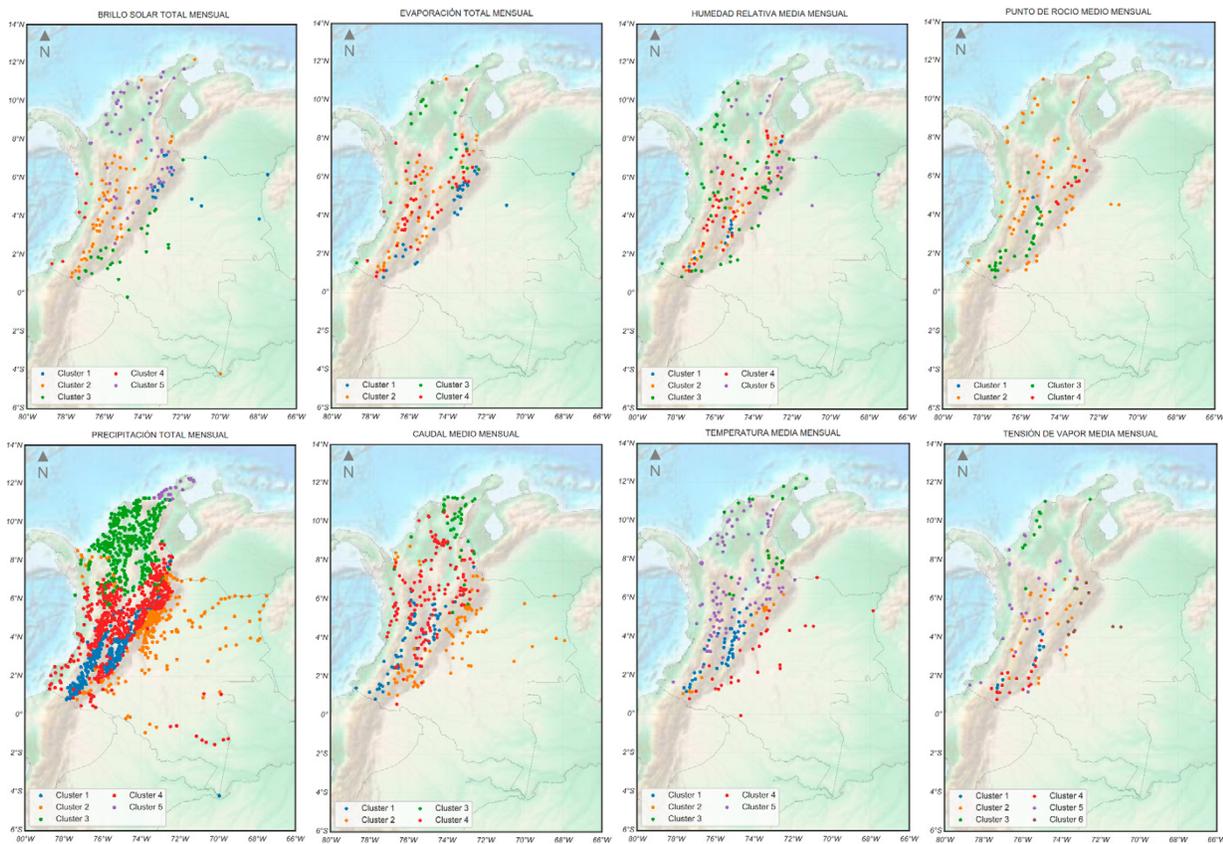
**Figura 6 |** Grupos generados mediante la metodología K-means para los dos primeros Componentes Principales.

La reducción dimensional aplicada mediante el ACP comprime en gran medida la cantidad de información que debe ser utilizada, pero guardando la estructura invariante de las series de tiempo y su variabilidad. De un total de 480 meses de información en cada serie de tiempo, la precipitación total mensual (PT 4) es la que más componentes principales requiere para comprimir la

información, como se observa en la Figura 5, con cerca de 40 componentes principales de la base de datos se obtiene el 70% de la varianza explicada (cerca de 8,33% del total de datos de la base de datos), mientras que las demás variables requieren muchos menos componentes principales para describir el mismo porcentaje de varianza, lo cual refuerza la hipótesis de usar el ACP como elemento para comprimir la información y mejorar los tiempos de procesamiento de información, con lo que se reduce el costo computacional.

Sobre los resultados del ACP, filtrado para el 70% de la varianza explicada, se aplicó para cada una de las variables el método de los K-means con el fin de obtener los grupos de estaciones que presentan una estructura “similar”. Como se explicó en la metodología se siguió un proceso iterativo donde se evaluó entre 2 y 12 clústers por el método del Coeficiente de Siluetas de forma que se obtuviera el número óptimo de grupos. Como se observa en la Figura 6, para la base de datos el número óptimo de grupos estuvo entre 4 y 6 (ver también Figura 7).

## Regionalización

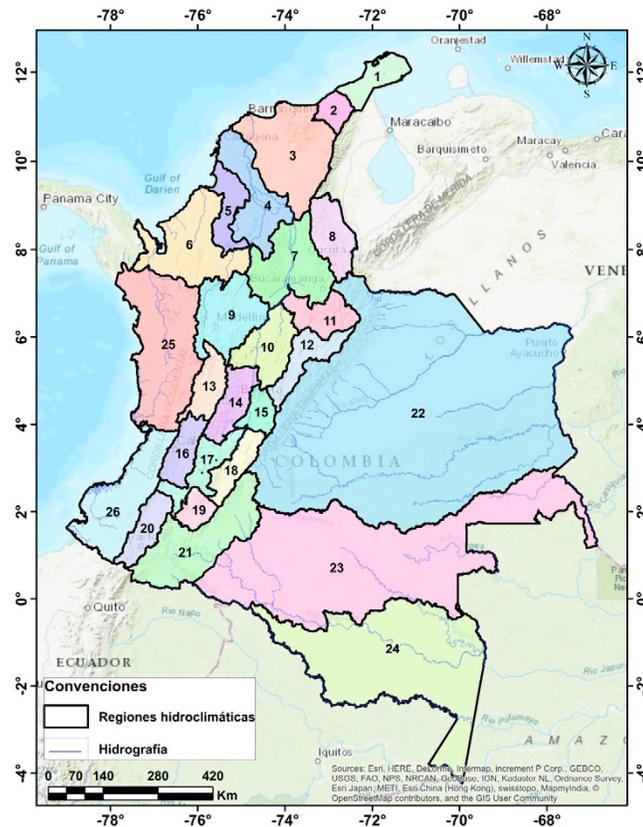


**Figura 7** | Grupos determinados por la metodología planteada para cada variable hidroclimática de forma aislada.

Los resultados obtenidos, si bien clarifican la existencia de núcleos (Figura 6), es necesario pasar este agrupamiento a una vista espacial donde se verifique que las series que se agrupan responden a estructuras similares, pero no solo desde la matemática (expresada en el ACP) sino incluyendo el comportamiento físico de las variables y su limitación o dependencia de los diferentes accidentes geográficos que son modificadores del clima.

Con los resultados obtenidos del método de K-means se procede a construir el mapa de regionalización para cada una de las variables (Figura 7), esto además permite visualizar que los grupos formados sean consecuentes con la orografía del país, cada uno de forma independiente. Luego de tener estos resultados individuales, se procede a realizar la intercepción de los 8 mapas

individuales (uno por cada variable hidroclimática analizada), con el fin de construir el mapa de áreas o regiones hidroclimáticamente homogéneas para todo el país que se presenta Figura 8.



**Figura 8** | Áreas hidroclimáticamente homogéneas para Colombia.

El norte del país se divide en 8 regiones bien definidas. La zona desértica de La Guajira que es influenciada directamente por la temporada de Huracanes (Región 1, Figura 8). El pie de la cara norte de La Sierra Nevada de Santa Marta que también hace parte del departamento de La Guajira forma otra región bien definida (Región 2, Figura 8) que físicamente se distingue de La Guajira desértica debido a que son áreas sin aridez evidente como si se observa en la zona desértica. Los departamentos del Magdalena, Atlántico y el centro y norte del Cesar hidroclimáticamente responden de forma muy similar para las 8 variables evaluadas por lo cual la metodología los captura como una región homogénea (Región 3, Figura 8).

La ecorregión de La Mojana como uno de los Hidrosistemas más complejos del planeta, es un símil de una cadena de embalses que regula en gran medida el clima y la escorrentía hacia aguas abajo de ella. Debido a esto La Mojana y el centro y norte del departamento de Bolívar (ubicado aguas abajo de La Mojana) forman una región hidroclimática que básicamente responde a la regulación del complejo de ciénagas, lagos y lagunas de esta ecorregión (Región 4, Figura 8).

La presencia del fenómeno de la baja anclada de Panamá que se manifiesta con gran intensidad cerca de mitad de año genera un aumento en las precipitaciones en el departamento de Córdoba, en el Urabá Antioqueño, en el nudo del Paramillo y sobre el litoral Caribe desde el final del Golfo de Morrosquillo hasta el Golfo del Darién (incluyendo la cara norte de la Serranía del Darién), formando otra región hidroclimática de condiciones homogéneas (Región 6, Figura 8). En medio de esta región afectada por la baja anclada de Panamá (Región 6, Figura 8) y la región formada por el centro norte del departamento de Bolívar y la ecorregión de La Mojana (Región 4, Figura 8) existe una pequeña región (Región 5, Figura 8) de un ancho aproximado entre 50 y 55 kilómetros que exhibe los mismos patrones hidroclimáticos que la región 3.

El sur de Cesar y Bolívar, la zona oriental de Antioquia y el norte del departamento de Santander forman otra región que hidroclimáticamente responde de forma homogénea (Región 7, Figura 8). Colindante con la región 7 y separada por la extensión de la Cordillera Oriental (antes de formar La Serranía del Perijá), aparece como una región homogénea la zona del Catatumbo Colombiano (Región 8, Figura 8).

La zona que abarca toda el área enmarcada entre las tres Cordilleras de los Andes (Zona Andina), evidencia la gran influencia que tienen estas sobre la hidroclimatología del país. Es así como, aparecen 12 regiones homogéneas, pero bien separadas por la estructura orográfica de las tres cordilleras. Sobre la cordillera oriental se diferencian 5 regiones desde el departamento de Boyacá hasta el departamento del Huila (Regiones 11, 12, 15, 18 y 19, Figura 8); en el caso de la cordillera central se presentan 3 regiones que van desde el suroccidente del departamento de Santander hasta el oriente del departamento del Cauca (Regiones 10, 14 y 17, Figura 8) y finalmente para la cordillera occidental la metodología planteada captura 4 regiones homogéneas (Regiones 9, 13, 16 y 20, Figura 8).

La zona oriental y sur del país (Llanos orientales y Amazonía) se divide en 4 regiones homogéneas (Regiones 21, 22, 23 y 24, Figura 8). La primera (Región 21, Figura 8) ubicada en el pie de monte de la cordillera oriental en la zona aledaña al departamento del Putumayo, es formada por el sur de los departamentos de Nariño y Cauca y el norte de los departamentos de Putumayo y Caquetá. Los llanos orientales se dividen en dos regiones, una conformada por la zona oriental del departamento de Boyacá, el norte de los departamentos de Guaviare y Guainía, más los departamentos del Meta, Vichada y Arauca (Región 22, Figura 8); y otra conformada por el sur de los departamentos del Putumayo, Guaviare, Guainía, el centro y sur del departamento de Caquetá y el departamento de Vaupés (Región 23, Figura 8). Y El departamento del Amazonas que conforma a nivel general una región (Región 24, Figura 8).

Finalmente, el Pacífico Colombiano se divide en dos regiones, una conformada por el departamento del Chocó y el occidente de Antioquia (Región 26, Figura 8), y otra región conformada por la parte centro-occidental de los departamentos de Valle del Cauca, Cauca y Nariño (Región 25, Figura 8).

---

## DISCUSIÓN

El mapa de regiones hidroclimáticamente homogéneas responde a la estructura orográfica y a algunos condicionantes climáticos, presentando una clara diferencia con las áreas hidrográficas (AH) oficialmente establecidas para el país, no porque estas estén mal planteadas, sino porque su delimitación no fue desarrollada bajo características hidroclimáticas sino como áreas estratégicas para el desarrollo del país enmarcadas en las macrocuencas superficiales. Los resultados evidencian la no uniformidad de las respuestas hidroclimáticas en Colombia, siendo una de las posibles causas que dificultan las tareas de pronóstico que desarrollan los meteorólogos.

Una investigación en Mongolia Central como método central aplica el ACP, pero en su modalidad de Componentes Principales Rotados (Leland *et al.*, 2013), sin embargo, en este estudio no se tiene en cuenta las variables hidroclimáticas para generar las regiones, sino que realizan una correlación del desarrollo de los anillos del tronco en algunos tipos de árboles (que es asociada definitivamente con la hidroclimatología) la que sería válida y muy relevante en casos que no se disponga de estaciones de medición. En otro proyecto que tomó como estudio de caso el continente africano un grupo de investigadores desarrolló un software de código abierto en el lenguaje “R” (R Core Team, 2017) denominado: *HiClimR: Hierarchical Climate Regionalization* (Badr *et al.*, 2015), el cual usa un método de agrupamiento jerárquico aplicado sobre un análisis de correlaciones, si bien un análisis de correlaciones puede brindar una evaluación inicial se debe definir el motivante de esta correlación, ya que se pueden presentar correlaciones espurias con alta significancia, pero que no presenten sentido físico del proceso, es así como al analizar cada una de las variables evaluadas en la presente investigación se hizo especial énfasis en garantizar que los resultados fueran consecuentes con la orografía en cada región, ya que esto es un factor determinante en el clima, hecho que se refuerza con la investigación de Aliaga *et al.* (2017) donde toman para la región de Las Pampas Argentinas 33 estaciones climáticas (precipitación, temperatura, velocidad del viento y humedad relativa) más la topografía y generan mediante el método de Agrupamiento Jerárquico de Ward (Ward y Marion, 1963) diferentes regiones homogéneas.

Si bien en la presente investigación no se aplican métodos jerárquicos para la clasificación por clústers, los resultados son bastante acertados con ciertas estructuras climáticas o topográficas que existen en Colombia, las técnicas matemáticas aplicadas son reconocidas por su fiabilidad y buenos resultados, es así como Abadi *et al.* (2019) en su investigación aplican también técnicas de Análisis de Componentes Principales para reducción dimensional y el método de k-means para determinar los grupos que determinaran la regionalización hidroclimática (con diferencia en el método de selección del óptimo número de clústers, determinación de regiones hidroclimáticas y técnicas de depuración de la información base).

## CONCLUSIONES

La metodología planteada para la regionalización arrojó resultados satisfactorios, ya que en los mapas preparados se observan áreas de comportamiento homogéneo que responden a la estructura orográfica de Colombia, presentando una clara diferencia con las áreas hidrográficas (AH) oficialmente establecidas para el país, no porque estas estén mal planteadas, sino porque su delimitación no fue desarrollada por consideraciones o análisis hidroclimáticos, fueron concebidas realmente como áreas estratégicas para el desarrollo del país enmarcadas en las macrocuencas superficiales.

Para evaluar la regionalización hubo dos elementos de la estructura de variabilidad hidroclimática que no arrojaron resultados satisfactorios, la inercia y los ciclos seculares. La inercia quizás porque no haya ninguna condicionante regional para la memoria de los procesos más allá que algunos parámetros geomorfológicos de las cuencas, que dependen de cuestiones diferentes al clima. Los ciclos seculares pueden haber sido un elemento que no exhiba regiones porque la longitud de las ST aún no exhibe ciclos completos, debido a esto en medio del contexto general no aparecen regiones definidas.

La estacionalidad para las épocas húmedas en Colombia está dividida a nivel general en dos tipos de estructura anual. La primera estructura se da en las regiones que presentan solo un máximo anual cerca de mitad de año entre los meses junio y julio, ubicadas en la parte oriental del país en toda la región de la Orinoquía y la región Amazónica Colombiana aproximadamente hasta el eje ecuatorial (0° de Latitud), además del área entre el golfo del Darién y el Parque Nacional Natural del Paramillo. El resto del país (excepto el Pacífico) presenta una estructura de dos máximos al año el primero entre los meses de abril y mayo, y el segundo ubicado entre octubre y noviembre. En el caso de esta estructura de dos máximos al año se puede hablar de otra subdivisión, ya que las regiones ubicadas más hacia el sur presentan valores más altos para el máximo del primer semestre del año, caso contrario ocurre en las estaciones ubicadas más hacia el norte del país, donde el máximo del segundo semestre es más alto que el del primero, lo cual se asocia con la llegada de la temporada de huracanes que dejan sistemas de bajas presiones en el caribe generando que las precipitaciones del segundo semestre se incrementen.

La precipitación como elemento regulador del clima da un panorama general de cómo es la estructura hidroclimática del país, ya que cuando aumenta la precipitación trae consigo aumento en los caudales de escorrentía (son variables directamente proporcionales), mientras que la Temperatura (junto a las variables que dependen de ella: tensión de vapor, evaporación, punto de rocío...) se manifiesta de manera inversamente proporcional, al igual que el Brillo Solar para el que es claro que a mayor precipitación hay mayor nubosidad y disminuyen las horas de incidencia directa de los rayos solares.

La propuesta técnica para definir regiones hidroclimáticamente homogéneas puede ser útil para discretizar de manera más específica con el fin de ajustar los pronósticos teniendo en cuenta que la respuesta hidroclimática no es homogénea para todo el país y que si se refina la escala aparecerán subregiones en las cuales sea más acertado desarrollar modelos de simulación para pronósticos de corto, mediano y largo plazo.

Los resultados se confirman que la Amazonía Colombiana requiere mayor monitoreo, ya que para algunas variables no existen estaciones de medición suficientes para realizar una regionalización más detallada, esto es un punto de análisis clave, ya que siendo la Amazonía una región tan importante a nivel mundial, se requieren más esfuerzos de inversión en monitoreo debido a que es un gran regulador de riqueza hídrica, de humedad y de la producción de oxígeno a través de la copiosa vegetación que posee la selva Amazónica.

Definir una metodología que permita construir regiones hidroclimáticamente homogéneas es aún un camino para seguir investigando, pero de importante énfasis toda vez que no se pueden generalizar los efectos a escala más local de las variables

macroclimáticas, ni de la variabilidad o el cambio climático, adicionalmente para el desarrollo de proyectos de inversión es totalmente necesario el conocimiento de la respuesta hidroclimática de cada zona en particular, de forma que se pueda prever a futuro que condiciones pueden prevalecer y como se puede mejorar o mitigar el impacto antrópico en la sostenibilidad del planeta. Este estudio también confirma que si el propósito es regionalizar hidroclimáticamente necesariamente hay que incluir las variables directas de la hidroclimatología a fin de que los resultados sean robustos.

---

## AGRADECIMIENTOS

A la Gobernación del Magdalena y a COLCIENCIAS por seleccionar al ing. David De León Pérez como beneficiario de una beca-crédito condonable mediante la convocatoria 672 de COLCIENCIAS “Formación de capital humano de alto nivel para el departamento del Magdalena 2014” (Maestría Nacional). Al equipo del Taller S-Multistor que brindó su apoyo en medio de la Cooperación Programática entre la Dirección General de Cooperación Internacional (DGIS) del Ministerio de Asuntos Exteriores de los Países Bajos e IHE Delft, a través de la participación de la Facultad de Estudios Ambientales y Rurales de la Pontificia Universidad Javeriana (Bogotá, Colombia).

---

## REFERENCIAS

- Abadi, A.M., Rowe, C.M., Andrade, M. 2019. Climate regionalization in Bolivia: A combination of non-hierarchical and consensus clustering analyses based on precipitation and temperature. *International Journal of Climatology*, 40(10), 4408-4421. <https://doi.org/10.1002/joc.6464>
- Aliaga, V.S., Ferrelli, F., Piccolo, M.C. 2017. Regionalization of climate over the Argentine Pampas. *International Journal of Climatology*, 37(S1), 1237-1247. <https://doi.org/10.1002/joc.5079>
- Arthur, D., Vassilvitskii, S. 2007. k-means++: The advantages of careful seeding. In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, pp. 1027–1035. <https://dl.acm.org/doi/10.5555/1283383.1283494>
- Badr, H., Zaitchik, B.F., Dezfuli, A.K. 2015. A tool for hierarchical climate regionalization. *Earth Science Informatics*, 8, 945-958. <https://doi.org/10.1007/s12145-015-0221-7>
- Bradley, J. 1968. *Distribution-Free Statistical Tests*. Prentice-Hall, Englewood Cliffs, NJ, 283-310.
- Coulibaly, P., Evora, N.D. 2005. Comparison of neural network methods for infilling missing daily weather records. *Journal of Hydrology*, 341 (1-2), 27-41. <https://doi.org/10.1016/j.jhydrol.2007.04.020>
- De Gooijer, J. 2017. *Elements of Nonlinear Time Series Analysis and Forecasting*. Springer Series in Statistics. Springer International Publishing. <https://doi.org/10.1007/978-3-319-43252-6>
- De León Pérez, D. 2017. *Análisis de sincronización espacio-temporal de señales océano-atmosféricas y variables hidroclimatológicas de Colombia*. Tesis de maestría, Facultad de Ingeniería, Pontificia Universidad Javeriana. <http://hdl.handle.net/10554/35731>.
- Dingman, S.L. 2015. *Physical Hydrology*: Third Edition. Waveland Press. <https://www.waveland.com/browse.php?t=382>
- Domínguez, E., Angarita, H., Rivera, H. 2010. Viabilidad para pronósticos hidrológicos de niveles diarios, semanales y decadales en Colombia. *Ingeniería e investigación*, 30(2), 178-187. Disponible en <http://www.scielo.org.co/pdf/iei/v30n2/v30n2a18.pdf>. Último acceso: marzo de 2021.
- Dubois, P., Hinsen, K., Hugunin, J. 1996. Numerical python. *Computers in Physics*, 10(3), 262. <https://doi.org/10.1063/1.4822400>
- Environmental Systems Research Institute (ESRI). 2015. *Arcgis for desktop*. <http://www.esri.com>

- Fedorov, V. 2015. Spatial and temporal variations in solar climate of earth in the present epoch. *Izvestiya, Atmospheric and Oceanic Physics*, 51(8), 779-791. <https://doi.org/10.1134/S0001433815080034>
- Gómez Lende, S. 2011. Región y regionalización: Su teoría y su método. El nuevo orden espacial del territorio argentino. *Universidad del Bio-Bio*, 26, 83-122. <http://www.ubiobio.cl/miweb/webfile/media/222/Tiempo/2011/%2327.05.pdf>
- Govindaraju, R.S., 2000. Artificial neural networks in hydrology. ii: Hydrologic applications. *Journal of Hydrologic Engineering*, 5(2), 124-137. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2000\)5:2\(124\)](https://doi.org/10.1061/(ASCE)1084-0699(2000)5:2(124))
- Grubbs, F.E. 1950. Sample criteria for testing outlying observations. *Annals of Mathematical Statistics*, 21(1), 27-58. <https://doi.org/10.1214/aoms/1177729885>
- Haan, C. 1977. *Statistical Methods in Hydrology*. Iowa State University Press. <http://hdl.handle.net/1969.3/24532>
- Hannachi, A. 2004. *A primer for eof analysis of climate data*. Department of Meteorology, University of Reading, 1-33.
- HIMAT. 1978. Resolución 00337, por el cual se adopta un sistema de codificación para las estaciones hidrometeorológicas. *Instituto Colombiano de Hidrología, Meteorología y Adecuación de Tierras*, Bogotá, Colombia.
- Hotellin, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441. <https://doi.org/10.1037/h0071325>
- Hunter, J.D. 2007. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3), 90-95. <https://doi.org/10.1109/MCSE.2007.55>
- IDEAM. 2001. *El medio ambiente en Colombia, 2nd Edición*. Instituto de Hidrología, Meteorología y Estudios Ambientales, IDEAM, Bogotá D.C., Colombia. <http://www.ideam.gov.co>
- IPCC, Stocker, T., Qin, D., Plattner, G.K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. 2013. *Annex III: Glossary*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, book section AIII, p. 1447-1466. <https://www.climatechange2013.org>
- Jones, E., Oliphant, T., Peterson, P., et al. 2001. SciPy: *Open source scientific tools for Python*. <http://www.scipy.org/>
- Leland, C., Pederson, N., Hessel, A., Nachin, D., Davi, N., D'Arrigo, R., Jacoby, G., 2013. Climate regionalization in Bolivia: A combination of non-hierarchical and consensus clustering analyses based on precipitation and temperature. *Dendrochronologia*, 31(3), 205-215. <https://doi.org/10.1016/j.dendro.2012.11.003>
- León, G., Zea, J., Eslava, J. 2000. Circulación general del trópico y la zona de confluencia intertropical en Colombia. *Meteorología Colombiana*, 1, 31-38. [http://168.176.14.11/fileadmin/content/geociencias/revista\\_meteorologia\\_colombiana/numero01/01\\_05.pdf](http://168.176.14.11/fileadmin/content/geociencias/revista_meteorologia_colombiana/numero01/01_05.pdf)
- Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764-766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- Lorenz, E.N. 1970. Climatic change as a mathematical problem. *Journal of Applied Meteorology*, 9(3), 325-329. [https://doi.org/10.1175/1520-0450\(1970\)009%3C0325:CCAAMP%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(1970)009%3C0325:CCAAMP%3E2.0.CO;2)
- MADS. 2012. Decreto 1640, por medio del cual se reglamenta los instrumentos para la planificación ordenación y manejo de las cuencas hidrográficas y acuíferos, y se dictan otras disposiciones. *Ministerio de Ambiente y Desarrollo Sostenible*, Bogotá, Colombia. Disponible en <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49987>
- Malcolm, G.A., McDonnell, J.J. 2005. *Encyclopedia of hydrological science*. John Wiley & Sons Ltd. <https://doi.org/10.1002/0470848944>
- McKinney, W. 2010. Pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9), 1-9.
- Mesa, O.J., Poveda, G., Carvajal, L. 1997. *Introducción al clima de Colombia*. Universidad Nacional de Colombia.

- Montgomery, D., Jennings, C., Kulahci, M. 2015. *Introduction to Time Series Analysis and Forecasting*. Wiley Series in Probability and Statistics. Wiley. <https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12203>
- Montoya Gaviria, G. d. J. 2008. *Lecciones de meteorología dinámica y modelamiento atmosférico*. Universidad Nacional de Colombia. Facultad de Ciencias. [http://ciencias.bogota.unal.edu.co/fileadmin/Facultad\\_de\\_Ciencias/Publicaciones/Archivos\\_Libros/Libros\\_Geociencias/Lecciones\\_de\\_Meteorologia\\_dinamica/Metereologia.pdf](http://ciencias.bogota.unal.edu.co/fileadmin/Facultad_de_Ciencias/Publicaciones/Archivos_Libros/Libros_Geociencias/Lecciones_de_Meteorologia_dinamica/Metereologia.pdf)
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, Microtome Publishing*. hal-00650905v1. Disponible en <https://hal.inria.fr/hal-00650905v1>. Último acceso: marzo de 2021.
- Poveda, G. 2004. La Hidroclimatología de Colombia: Una síntesis desde la escala inter-decadal hasta la escala diurna. *Revista de la Academia Colombiana de Ciencias*, 28(107), 201-222. Disponible en [https://www.acefyn.com/revista/Vol\\_28/107/201-222.pdf](https://www.acefyn.com/revista/Vol_28/107/201-222.pdf). Último acceso: marzo de 2021.
- Poveda, G., Jaramillo, A., Mantilla, R. 2001. Asociación entre el fenómeno el niño y las anomalías de humedad del suelo y del índice “ndvi” en Colombia. In: *IX Congreso Latinoamericano e Ibérico de Meteorología y VIII Congreso Argentino de Meteorología*. p. 8. [http://www.acefyn.com/revista/Vol\\_28/107/201-222.pdf](http://www.acefyn.com/revista/Vol_28/107/201-222.pdf)
- Poveda, G., Mesa, O.J. 1997. Feedbacks between hydrological processes in tropical south america and large-scale ocean-atmospheric phenomena. *Journal of Climate*, 10(10), 2690-2702. [https://doi.org/10.1175/1520-0442\(1997\)010%3C2690:FBHPIT%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010%3C2690:FBHPIT%3E2.0.CO;2)
- Python Core Team. 2017. Python: A dynamic, open source programming language. <https://www.python.org/>.
- R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rossum, G. 1995. *Python reference manual*. Tech. rep., CWI (Centre for Mathematics and Computer Science), Amsterdam, The Netherlands, The Netherlands.
- Samarasinghe, S. 2016. *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*. CRC Press. 25, 33
- Velasco, A. 2016. *Integración del concepto de variabilidad hidroclimática en pronósticos hidrológicos de largo plazo de resolución mensual en Colombia*. Tesis de maestría, Facultad de Ingeniería, Pontificia Universidad Javeriana. <http://hdl.handle.net/10554/19628>
- Ward, J.H., Marion, JR. 1963. Application of an Hierarchical Grouping Procedure to a Problem of Grouping Profiles. *Educational and Psychological Measurement*, 23(1), 69-81. <https://doi.org/10.1177/001316446302300107>
- White, D., Richman, M., Yarnal, B. 1991. Climate regionalization and rotation of principal components. *International Journal of Climatology*, 11(1), 1-25. <http://doi.org/10.1002/joc.3370110102>
- WMO. 1994. *Guía de prácticas hidrológicas: adquisición y proceso de datos, análisis, predicción y otras aplicaciones*. Organización Meteorológica Mundial. [http://www.whycos.org/hwrp/guide/index\\_es.php](http://www.whycos.org/hwrp/guide/index_es.php)