

# Análisis de datos: Tarea 1

*Diana Marcela Álvarez Murillo*

*Ingeniera Ambiental y Civil*

*Especialización en Aprovechamiento de Recursos Hidráulicos*

*Universidad Nacional de Colombia, sede Medellín*

El uso de la estadística permite afianzar el entendimiento sobre diferentes fenómenos que ocurren a nuestro alrededor, a partir de una serie de observaciones o datos que admiten evaluar y cuantificar la incertidumbre asociada a ellos con el propósito de pronosticar posibles eventos de ocurrencia de dichos fenómenos en el futuro. Es por esta razón que la estadística juega muchos roles en las ciencias atmosféricas dado que la incertidumbre en el comportamiento atmosférico es lo que hace la atmósfera interesante (Wilks, 2006).

Una aplicación importante de la estadística en la meteorología y climatología es darle sentido a nuevo conjunto de datos (Wilks, 2006). Por lo cual, en este trabajo se pretende tomar una serie de datos, de precipitación en este caso, y evaluar diferentes técnicas estadísticas básicas como los índices de localización, dispersión y simetría; histogramas, percentiles y probabilidades; estacionariedad de los índices; y establecer si los datos tienen tendencia o no, con el fin de mejorar el entendimiento sobre ellos.

## Objetivos

- Analizar una serie de datos de cualquier variable atmosférica con más mil datos.
- Estimar índices de localización dispersión y simetría.
- Estimar histogramas, percentiles y probabilidades de la serie.
- Analizar la estacionariedad de la serie.
- Analizar si la serie tiene tendencia.

## Análisis

### *Serie de datos de precipitación diaria*

Para el cálculo de los diferentes estadísticos se procedió con la descarga de una serie de precipitación diaria desde KNMI Climate explorer que es una aplicación web para analizar datos climáticos estadísticamente (<https://climexp.knmi.nl/start.cgi?id=someone@somewhere>). Los datos de la precipitación diaria pertenecen a un lugar llamado Las Gaviotas en el departamento de Vichada, Colombia (ver Figura 1), lo que da un indicio de lo que se esperaría encontrar en la magnitud de esta variable en el sector. En la Tabla 1 se presentan las principales variables climatológicas y en la Figura 2 se muestra la serie de datos de precipitación diaria del departamento.

Tabla 1. Principales variables climatológicas en el departamento de Vichada, Colombia.  
Modificado de: (IDEAM, 2005)

Precipitación promedio anual (mm)	2166
Promedio de evapotranspiración potencial anual	2136
No. De días con lluvia anual	166
Promedio de humedad relativa anual	70%
Temperatura media mínima anual (°C)	23,7
Temperatura media máxima anual (°C)	26,0



Figura 1. Localización de la serie diaria de precipitación.

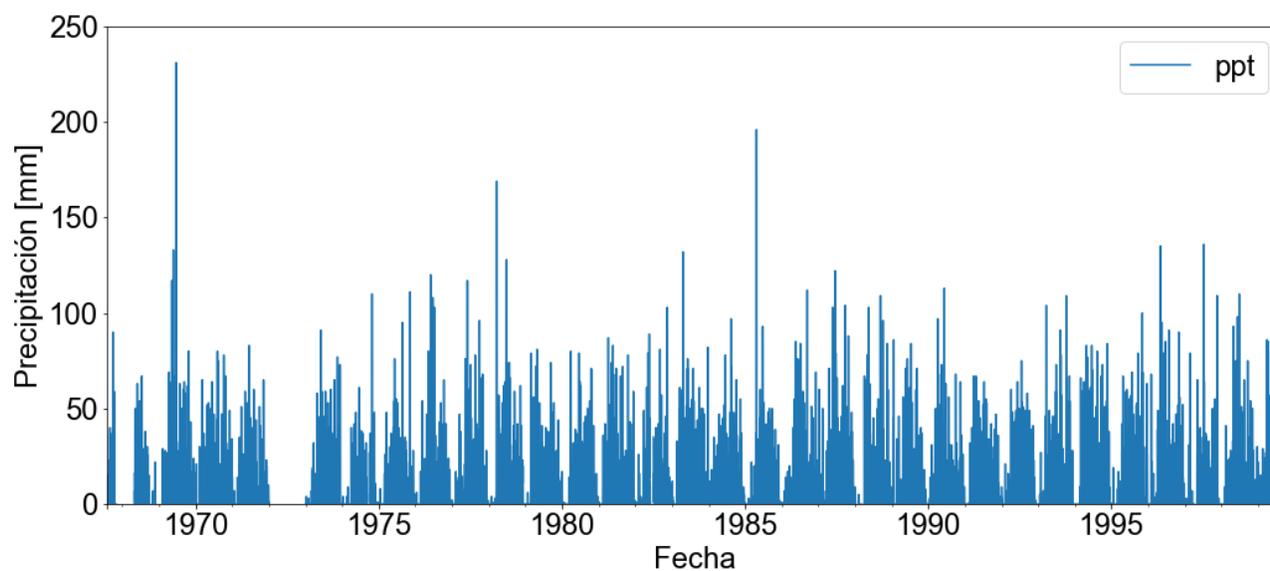


Figura 2. Serie de precipitación para Las Gaviotas en Vichada, Colombia.

### Estimación de índices de localización dispersión y simetría.

El índice de localización se refiere a la tendencia central o la **magnitud general de los datos**, el índice de dispersión denota el grado de variación respecto a un valor central y la simetría describe el balance en el cual los datos se distribuyen alrededor de su centro (Wilks, 2006). En la Tabla 2 se presentan los índices de localización, dispersión y simetría para la serie total de datos de precipitación

Tabla 2. Índices de localización, dispersión y simetría

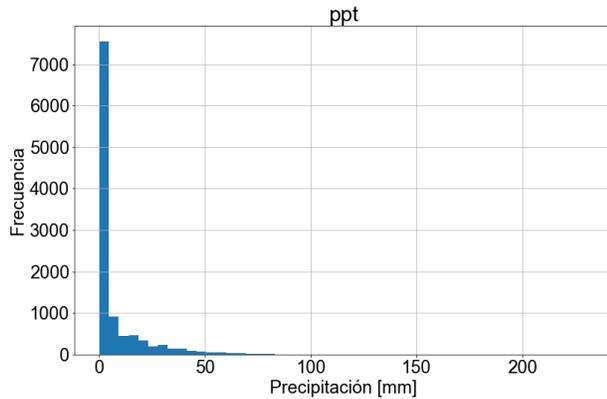
Índice de Localización	
Media [mm]	7,46
Mediana [mm]	0,00
Índice de dispersión	
Desviación estándar [mm]	15,18
Rango Inter-cuartil (IQR)	8,00
Índice de simetría	
Coficiente de asimetría	3,51
Yule Kendall	1,00

Como se observa de la tabla anterior la media y la mediana son diferentes, esto da a entender que la serie tiene valores extremos que hace que la media de la serie se vea influenciada por estos. De la misma manera puede corroborarse al comparar el IQR y la desviación estándar y el coeficiente de asimetría y el índice de Yule-Kendall; en que al ser de diferente magnitud se entiende que la serie tiene valores extremos o outliers.

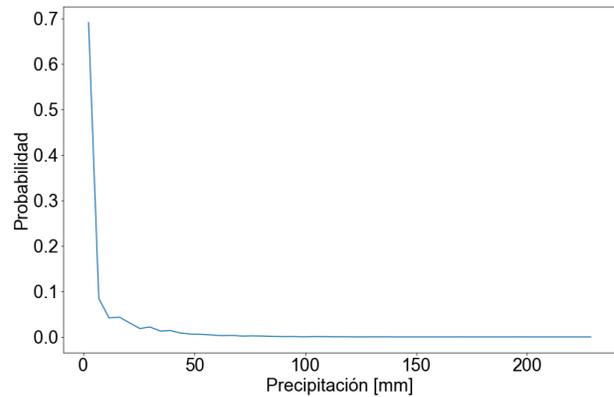
En cuanto a la simetría, tanto desde el coeficiente de asimetría como desde el índice de Yule-Kendall y a partir de que la media de los datos es mayor a la mediana, se puede decir que los datos tienen una simetría positiva; es decir, **tiene una asimetría con cola a la derecha (sesgados a la derecha)**. La distancia entre los cuantiles más bajos y la mediana es menor que para los cuantiles altos; en otras palabras, hay una probabilidad más alta de tener valores bajos de precipitación.

### Estimación de histogramas, percentiles y probabilidades de la serie.

En la Figura 3 se observa el histograma de frecuencias y la distribución de probabilidades. De las figuras se observa que, efectivamente, los datos están **sesgados a la derecha** por lo que hay una probabilidad mayor de que se presente un valor diario de precipitación que sea bajo. Una lluvia de 80 mm/día o mayor es estadísticamente más improbable que si se presenta una lluvia entre 0 y 50 mm/día; es decir, como se mencionó anteriormente, hay una probabilidad mayor de obtener valores bajos de precipitación diaria.



a) Histograma de frecuencias



b) Distribución de probabilidad

*Figura 3. Histograma de frecuencias y distribución de probabilidad.*

### *Evaluación de la estacionariedad de la serie.*

Para la evaluación de la estacionariedad de la serie de datos se tomaron diferentes ventanas de tiempo que al moverse día a día permiten observar cuál es la variación de los índices a medida que se avanza en el tiempo. Para analizar cómo varían los índices en el tiempo se tomaron diferentes ventanas con el fin de observar no solo si los índices varían en el tiempo sino también en cómo afecta la longitud de la ventana en la variación de estos. Para este análisis se tomaron ventanas de tiempo de 1, 5, 10 y 15 años.

En las Figura 4, Figura 5, Figura 6, Figura 7, Figura 8 y Figura 9 se observa cómo varían, según las diferentes ventanas de tiempo, la función de probabilidad en un gráfico de colores y en una superposición de los histogramas para cada ventana; la media y la desviación estándar; la media y la mediana; y la mediana y el rango Inter-cuartil, respectivamente.

Es de resaltar que la precisión del sistema de medida de la precipitación en Las Gaviotas no tiene cifras decimales, por lo que los valores que arroja dicho sistema consisten en números enteros; es decir, no hay precipitaciones de 0.5 mm/día sino de 0 mm o 1 mm según sea el caso. Por esta razón, en muchos de los gráficos presentados se observan saltos bruscos en la serie.

Cuando se observa en cada una de las figuras presentadas para la ventana de un año, se diría que los índices de localización, dispersión y simetría, y las funciones de probabilidad no son estacionarias puesto que en todas se observan cambios en la magnitud de las variables analizadas y es precisamente con esta ventana de tiempo en que se observa mayor ruido en los resultados. Esta variación puede deberse a la oscilación meridional de la Zona de Convergencia Intertropical (ZCIT) que se da como resultado del ciclo anual de temperaturas superficiales por efecto de la insolación, constituye el mecanismo físico de mayor importancia para explicar el ciclo anual de la hidro-climatología de Colombia (Poveda, 2015). Dado que este efecto explica el ciclo anual, estas variaciones son captadas por esta ventana de tiempo y se observa una variación mayor.

De la misma manera se observan cambios, aunque menores, en las variables si se toma una ventana de 5 años. Estos cambios en los índices analizados pueden deberse, ya no a la ZCIT sino a la variabilidad climática global a escala interanual que en Colombia está controlada por la ocurrencia del evento El Niño/Oscilación del sur (ENSO, por sus siglas en inglés). El ENSO tiene una variabilidad estacional (asociada con su enfasamiento con el ciclo anual) además de que también tiene un componente cuasi-bienal con un tiempo de recurrencia entre 2 a 2,5 años, así como una baja frecuencia, con una recurrencia entre 4-5 años (Poveda, 2015).

En ese sentido, a medida que la ventana es más grande, 10 y 15 años por ejemplo, el ruido en las variables se hace menor y parece que se volviera un poco más estacionarias. Lo anterior puede asociarse con la variabilidad climática global a escala de varias décadas que está dominada por el cambio climático y el cambio ambiental global de largo plazo, así como por los fenómenos macro-climáticos que operan en esta escala temporal, como la Oscilación Decadal del Pacífico (PDO) y la Oscilación del Atlántico Norte (NAO), que tienen una influencia significativa sobre el clima en Colombia (Poveda, 2015). Mientras mayor sea la ventana de tiempo, menor es la variabilidad de los índices y probabilidades puesto que abarca los períodos de ocurrencia de cada uno de los eventos.

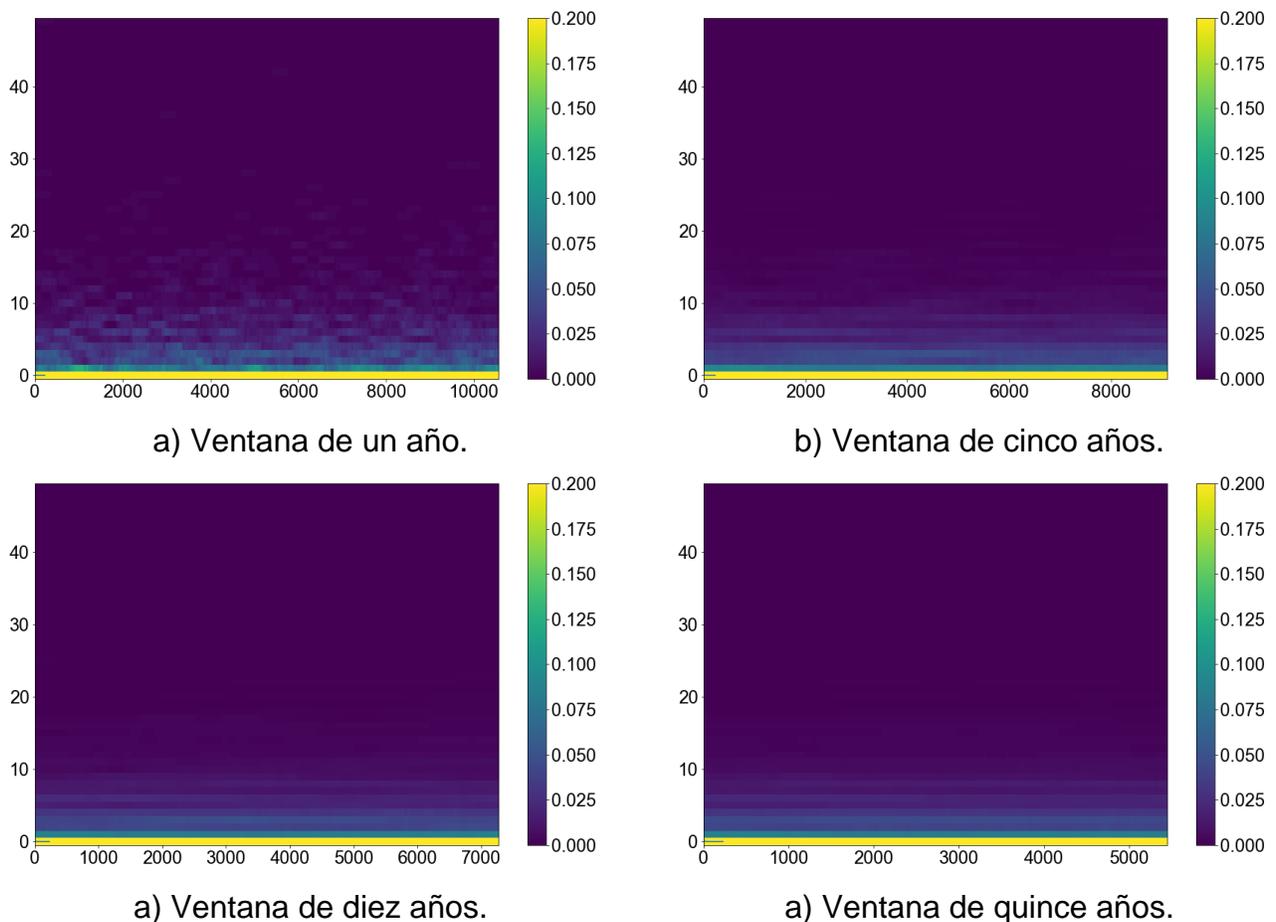
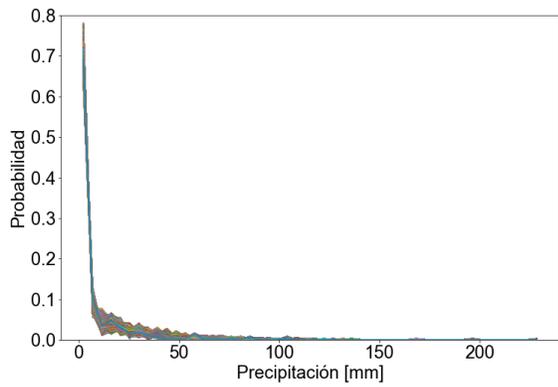
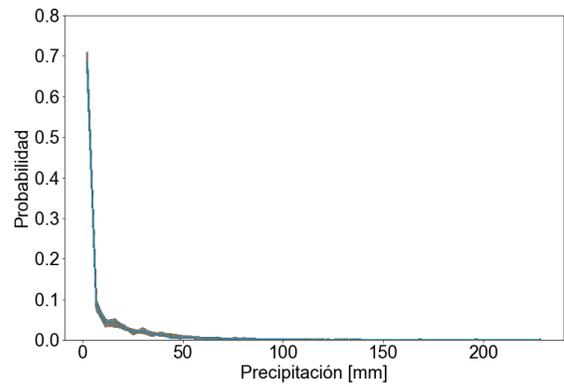


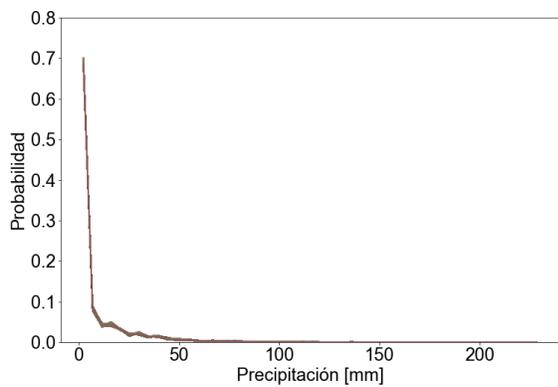
Figura 4. Estacionariedad de la distribución de probabilidad en gráfico de colores.



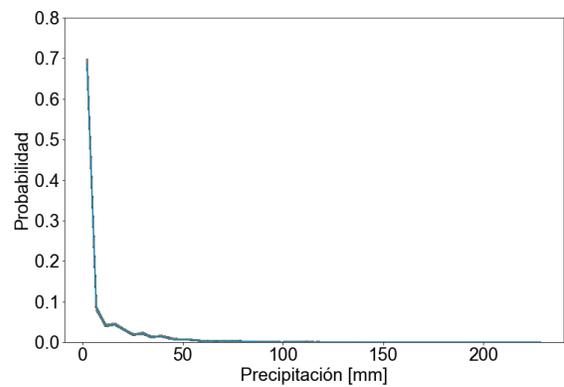
a) Ventana de un año.



b) Ventana de cinco años.

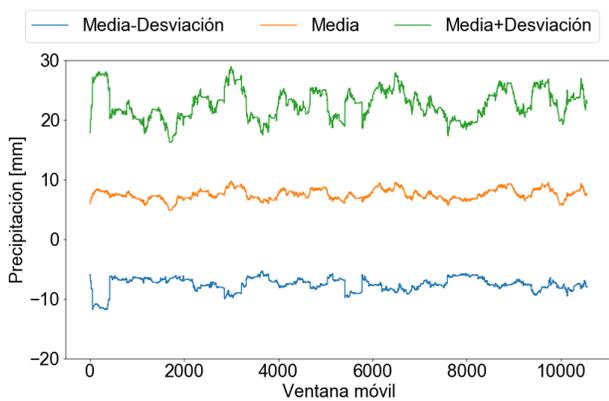


a) Ventana de diez años.

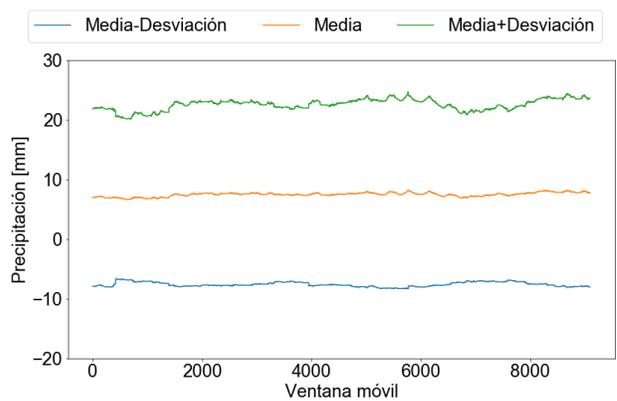


a) Ventana de quince años.

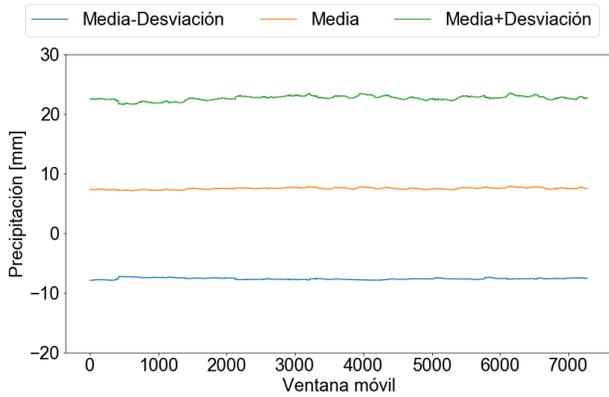
**Figura 5.** Estacionariedad de la función de probabilidad con las probabilidades de todas las ventanas de tiempo superpuestas.



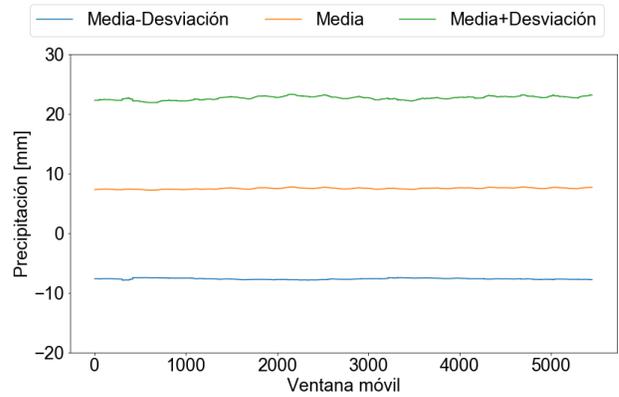
a) Ventana de un año.



b) Ventana de cinco años.

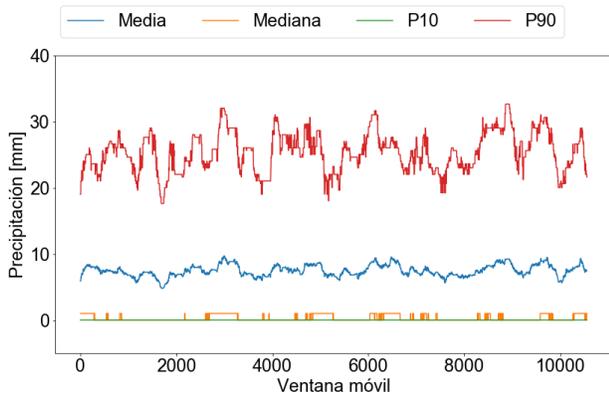


a) Ventana de diez años.

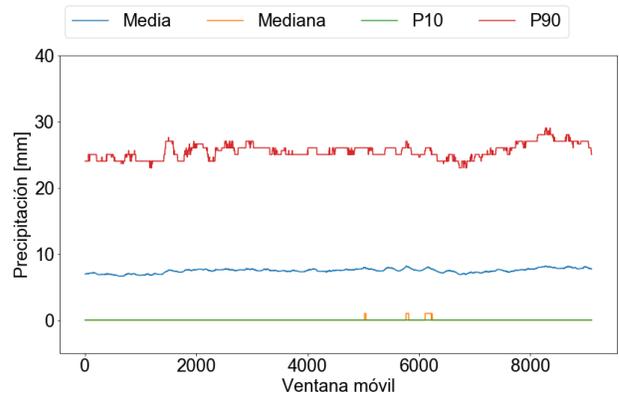


a) Ventana de quince años.

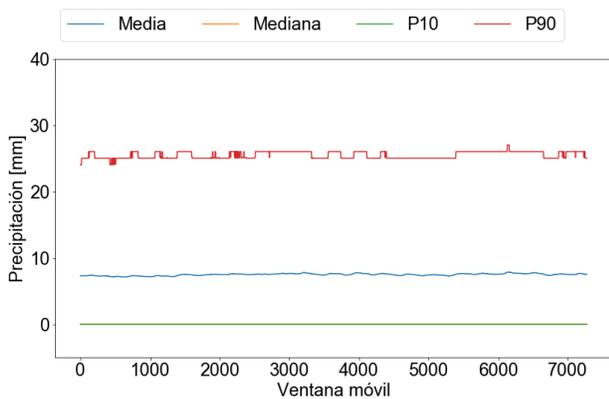
**Figura 6. Estacionariedad de la media y la desviación estándar.**



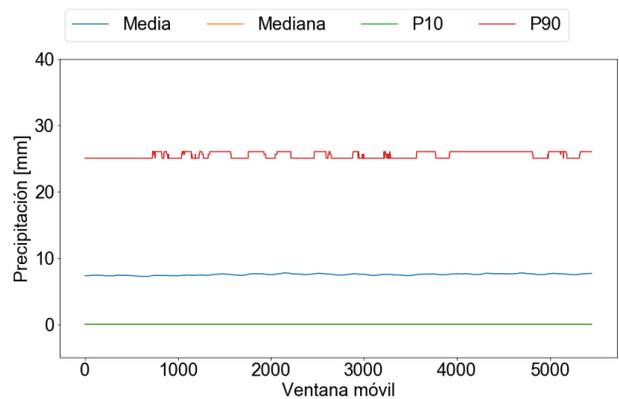
a) Ventana de un año.



b) Ventana de cinco años.

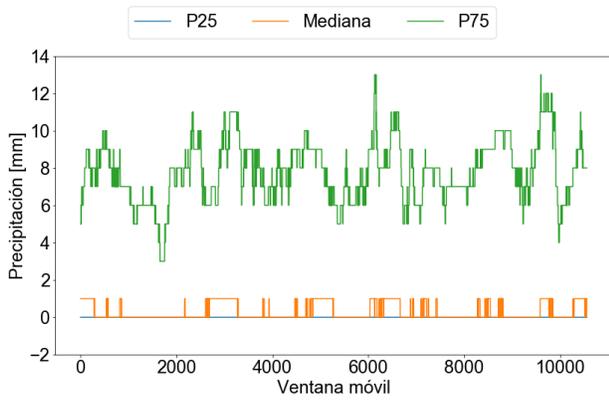


a) Ventana de diez años.

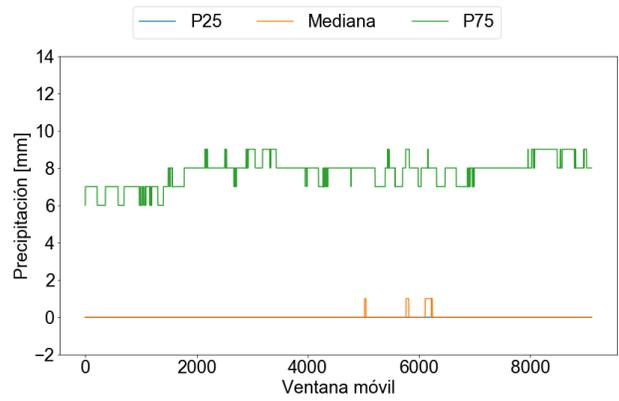


a) Ventana de quince años.

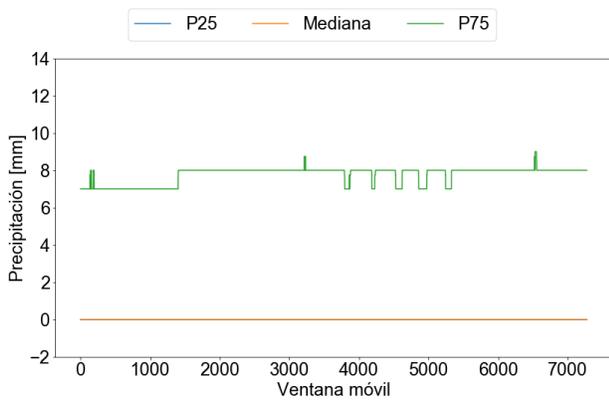
**Figura 7. Estacionariedad de la media, mediana y percentiles 10 y 90.**



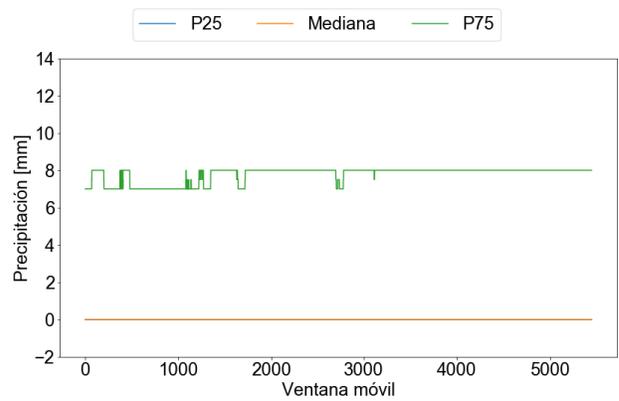
a) Ventana de un año.



b) Ventana de cinco años.

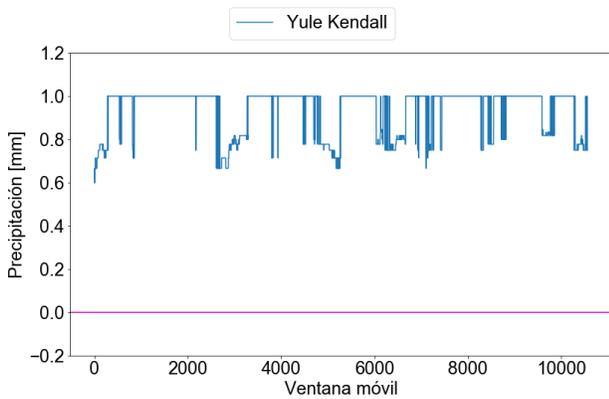


a) Ventana de diez años.

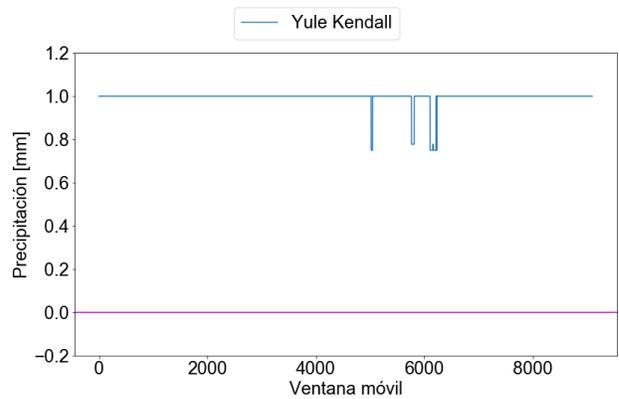


a) Ventana de quince años.

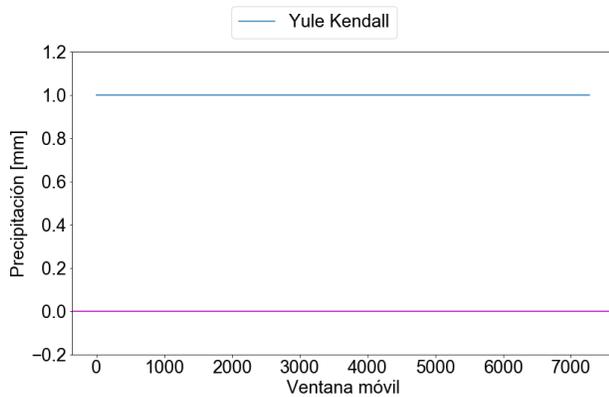
Figura 8. Estacionariedad de la mediana y los percentiles 25 y 75.



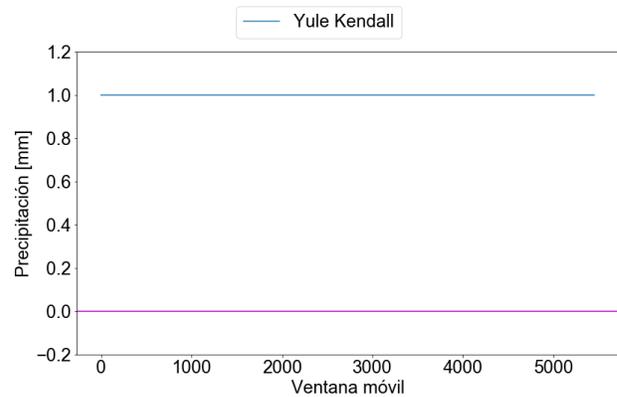
a) Ventana de un año.



b) Ventana de cinco años.



a) Ventana de diez años.



a) Ventana de quince años.

Figura 9. Estacionariedad del índice de Yule-Kendall

*Evaluación de la tendencia de la serie.*

Para evaluar la tendencia de la serie se hizo uso de la prueba de Mann-Kendall. La prueba de Mann-Kendall sirve para ver la tendencia no paramétrica de una serie de datos. Es muy utilizado para una serie de datos que no muestra variaciones estacionales (Mozejko, 2012).

En la Tabla 3 se presentan los resultados de la prueba. En la tabla se observa que Z de la prueba de Mann-Kendall es cero y teniendo en cuenta que para obtener una tendencia estadísticamente significativa se necesita un Z asociado a una probabilidad de  $\alpha/2$  donde  $\alpha$  es el 5% se concluye que esta serie no tiene tendencia. Puesto que el estadístico de prueba se debe comparar con una distribución de probabilidad gaussiana donde en ambos lados de la distribución hay dos colas de igual probabilidad entonces es por esto por lo que  $\alpha$  se divide entre dos, para que se mantenga el error que se quiere asumir (5%).

Tabla 3. Resultados de la prueba de Mann-Kendall

S	1,0
Varianza	165059424823,67
z	0

*Conclusiones*

A partir del análisis de datos y del cálculo de los índices de localización, dispersión y simetría y de las distribuciones de probabilidad es posible conocer el comportamiento de una serie de datos. Para este informe se toma una serie de precipitación diaria para un lugar llamado Las Gaviotas en el departamento de Vichada, Colombia. Por medio del análisis de esta serie de datos, se puede decir que tiene un comportamiento con una simetría hacia la derecha, comportamiento típico de una serie de precipitación, en que hay una mayor probabilidad de obtener valores menores de precipitación. Valores extremos de esta variable son improbables.

Para verificar la estacionariedad de los índices analizados, se realizan diferentes ventanas de tiempo; de 1, 5, 10 y 15 años con el fin no solo de ver si los índices analizados varían con el tiempo sino también en cómo influye la longitud de la ventana la variación -o estacionariedad- de los mismos. Se observa que a medida que aumenta la longitud de la ventana disminuye la variación de los índices y que este comportamiento puede estar asociado a los diferentes fenómenos hidro-climáticos que afectan la climatología en Colombia. Es decir, si se observa los resultados obtenidos para una ventana de un año, se observa mucha variación en los índices puesto que, en este rango de tiempo, la variación puede estar asociada al paso de la ZCIT. De la misma manera se observa con las ventanas de 5, 10 y 15 años en que, aunque la variación va disminuyendo según la ventana, esta misma variación está asociada a fenómenos que actúan en mayores escalas de tiempo. En ese sentido, debido a la presencia de los fenómenos micro y macro-climáticos que tienen efectos sobre la climatología en Colombia los índices analizados para la precipitación no son estacionarios.

Cuando se realiza la evaluación de la prueba de Mann-Kendall se observa que la serie no tiene tendencia para un error asumido del 5%.

### *Bibliografía*

IDEAM. (2005). *ATLAS CLIMATOLÓGICO DE COLOMBIA*.

Mozejko, J. (2012). Detecting and Estimating Trends of Water Quality Parameters. *Water Quality Monitoring and Assessment*. <https://doi.org/10.5772/33052>

Poveda, G. (2015). La Hidroclimatología De Colombia, (December).

Wilks, D. S. (Department of E. and A. S. C. U. (2006). *Statistical methods in the atmospheric sciences. International Geophysics Series* (Vol. 59). <https://doi.org/10.1002/met.16>