

Mucho cuidado con la estructura del informe, me parece un poco redundante. Lo ideal es que se siga una línea "lógica".
Cuidado con el lenguaje técnico.
Buen trabajo en general :)

3.9

* Sobre 4.5

Tarea 1 Análisis de datos

Jhonny Alexander Puerta Mazo¹

Universidad Nacional de Colombia, Medellín, Antioquia

TASK INFO

Task history:

Keywords:

Media
Percentil
Localización
Dispersión
Simetría
Histograma
Estacionariedad

R E S U M E N

Una serie de herramientas se presentarán a lo largo de este informe, para evaluar y determinar el grado de estacionariedad, variaciones y comportamiento de la evaporación en Baja California, México. En una primera instancia se determinarán índices que darán idea del proceder de los datos a lo largo de los años como lo son los índices de localización, dispersión y simetría. Se presentará un análisis tanto cualitativo (gráfico) como cuantitativo sobre los índices, lo que nos ofrecerá varias conclusiones de cómo ha variado este parámetro hidrológico a lo largo de tantos años.

INTRODUCCIÓN

La evaporación es uno de los factores más importantes en el ciclo hidrológico, un pilar a la hora de realizar un buen balance hídrico y realizar estimaciones precisas sobre las cantidades de agua en cualquier tipo de ambiente, siendo la conexión entre las fuentes hídricas y la atmósfera.

Entender el comportamiento de una serie de datos, y más de una temática tan importante como es la evaporación en el ciclo hidrológico, puede darnos idea del proceder hídrico en algún sitio, es por tal razón que si se puede encontrar patrones en la información que nos permitan realizar estudios, análisis y conclusiones de variables climatológicas; se tendrán oportunidades de encontrar las causas, razones y probabilidad de sucesos determinados en la temática ambiental.

1. Series

Se tomó una serie de datos de evaporación en el estado de Baja California en México, con una resolución temporal diaria desde julio de 1970 hasta diciembre de 2012.

2. Lectura

3. Gráfica

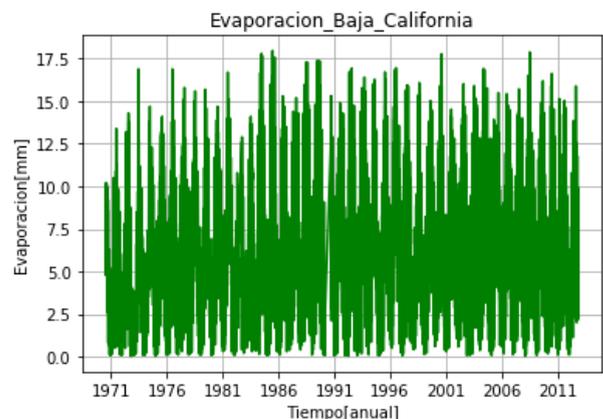


Figura 1. Evaporación Baja California

En la figura 1, observando los datos de la serie, esta se encuentra con unos valores máximos de evaporación en los mismos rangos usualmente, no presentando picos abruptos (outliers) que puedan presentar alguna inquietud en los datos. Alrededor del 2005 se encuentra una acumulación de evaporación que puede demandar un estudio más detallado (no objetivo de este trabajo). Otras de las posibilidades que se presenta es que el umbral que se alcance se deba sólo a la cantidad disponible de agua a evaporar, por consiguiente no se logra encontrar valores más altos al máximo que visualmente se observa en la figura 1.

4. Percentiles, Probabilidades e Histograma

En la figura 2 se tienen los percentiles de la evaporación tomando el total de datos de la serie, marcando lo que son gráficamente los percentiles de 25, 50 y 75, percentiles importantes a la hora de interpretar los datos:

¹ Estudiante de Posgrado de Recursos Hidráulicos E-mail address: japuertam@unal.edu.co

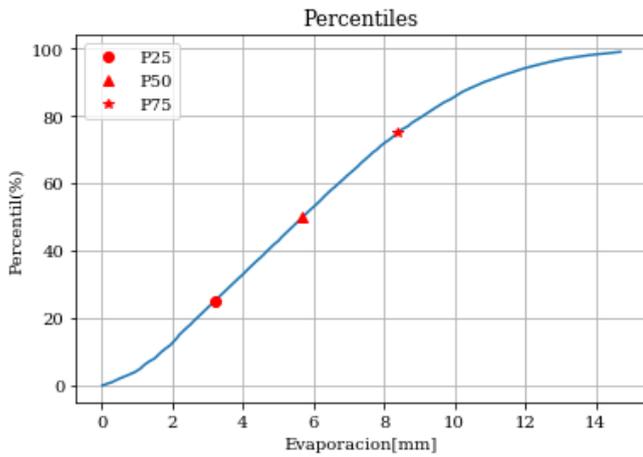


Figura 2. Percentiles

Los percentiles destacados son

$$P_{25} = 3,2 \text{ mm}$$

$$P_{50} = 5,7 \text{ mm}$$

$$P_{75} = 8,39 \text{ mm}$$

Los percentiles muestran que los datos tienden a ser muy uniformes, y que los mayores valores de evaporación tienden a ser correspondientes a los menores en cuanto a sus respectivos percentiles.

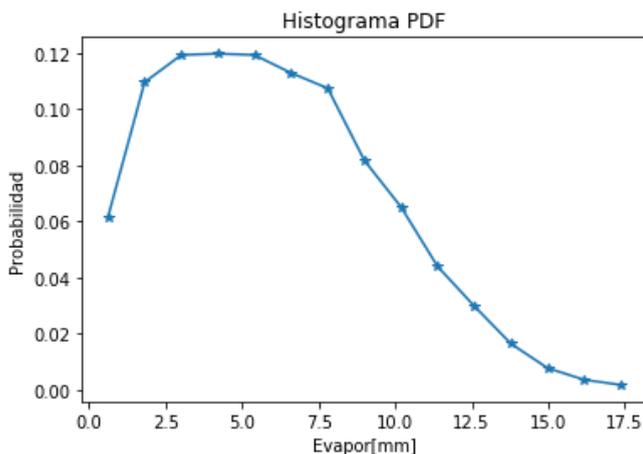


Figura 3. PDF

En la figura 3 se muestra el histograma de la evaporación desde 1970 a 2012, con una división de 15 intervalos, observando que la mayor probabilidad a lo largo del tiempo evaluado está entre 2,5 y 6 mm, con un valor del 12%, es decir; la evaporación que más se ha presentado ha sido la de 5 mm/día aproximadamente.

A continuación se mostrará un mapa de densidades donde se observará como varían las probabilidades a lo largo de los años, tomando un tamaño de ventana aproximado de 10 años, ya que al tener más de 10 años de datos diarios, se pueden observar mucho mejor tendencias que se puedan presentar, o eventos importantes (la elección del tamaño de ventana se escogió a ojo, sin embargo, cabe destacar que existe un tamaño de ventana adecuado, más no se enfatizó mucho en su determinación, se eligió siguiendo un concepto hidrológico).

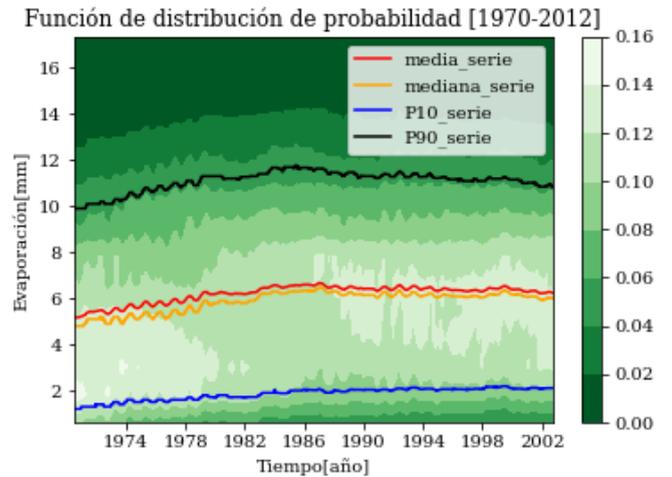


Figura 4. Histograma en función de la ventana

Aunque los valores presentes de la función de distribución tienden a ser muy horizontales, se puede observar que el valor de la media no se sostiene en el tiempo al igual que el ancho de las franjas que tiene unas variaciones no muy importantes pero notorias. Esto da una primera idea a que la serie no tenga tendencia.

5. Índices

5.1 Localización

Para la serie de datos de evaporación desde 1970 a 2012, los índices de localización que se determinaron fueron la media, mediana, percentiles 90, y 10, y la trimedia, en la figura 5, puede observarse, gracias a la comparación con la serie original, que los índices de localización dan entre los rango de en qué se encuentran los datos de la serie original, aclarando que como estos valores son para la serie completa de evaporación, los valores determinados para cada índice son un valor único para todos los años.

Esta quizás no es la mejor forma de mostrar estos índices... ya lo hiciste en los histogramas y lo harás cuando muestres los índices móviles

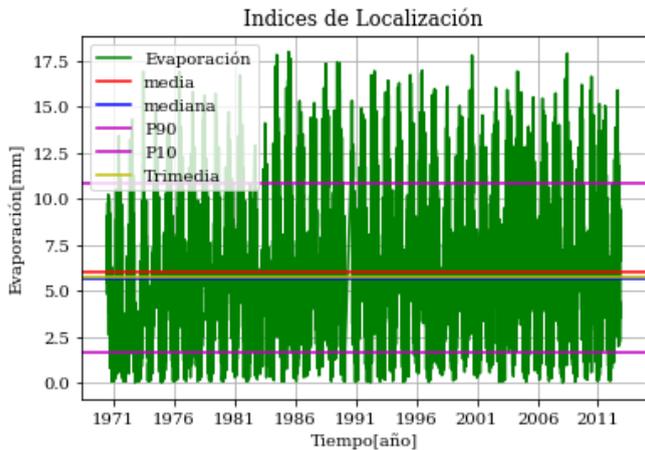


Figura 5. Índice de Localización serie Evaporación [1970-2012]

Cabe destacar que debido a la ausencia de outliers, la media no se ve afectada y en este caso, índice de localización de la media puede tomarse robusto, comparándolo con índices que indiferente de sus datos serán robustos como la trimedia y la mediana (Wilks, 2006).

Media = 6,03 mm

Mediana = 5,7 mm

Trimedia = 5,74 mm

Claramente se observa que el dato más afectado por la variabilidad de los datos es la media, comparándolo con la mediana y la trimedia; sin embargo, en este caso estos índices de localización dan una idea de los valores referencia de la serie ya que todos presentan entre si diferencias muy mínimas.

5.2 Dispersión

Para la serie de datos de evaporación desde 1970 a 2012, en la figura 6 se muestra la media, con su desviación estándar, remarcando que la desviación estándar más la media, se encuentra dentro del rango de valores de la serie original con valores de 9,5 y 2,55 mm/día.

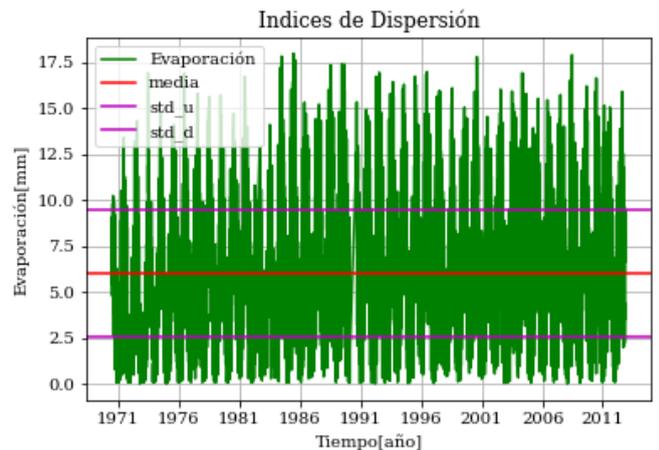


Figura 6. Dispersión de la desviación estándar comparada con la media y la serie

En la figura 7, encontramos sólo los valores de la desviación estándar, la desviación absoluta de la mediana (MAD) y el IQR.

El IQR aunque es un índice que nos dice en que valores se encuentra el 50% de los datos con una robustez importante, es muy complejo de visualizar, ya entraríamos a comparar mejor entre el MAD y el Std.

IQR = 5,19 mm

Std = 3,47 mm

MAD = 2,6 mm

Comparando con la desviación estándar que de por si no es un índice robusto por su volatilidad, se toma mejor el valor de comparación de la MAD como el valor de dispersión a tomar para análisis. El valor de MAD al ofrecer una confiabilidad mayor, arroja franjas alrededor de la mediana con una robustez suficiente para dar o reflejar un comportamiento de toda la serie en base a este índice.



Figura 7. Índice de dispersión serie Evaporación [1970-2012]

5.3 Simetría

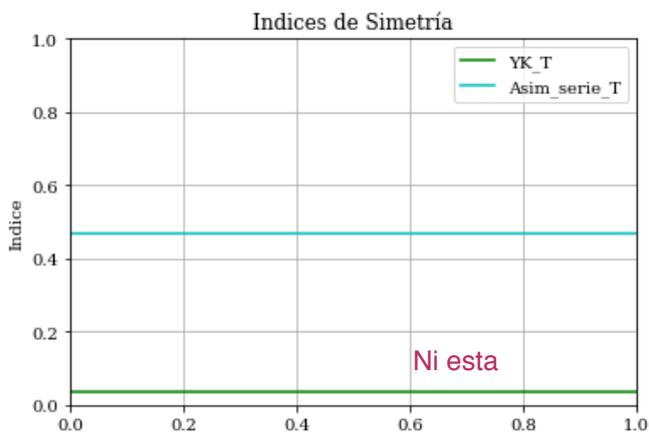


Figura 8. Índice de simetría serie Evaporación [1970-2012]

Se toma el valor de cero para ver hacia donde está la masa de la serie, si la simetría es positiva o negativa, para nuestro caso de estudio, tanto el índice de asimetría como el Yule- Kendall son positivos, por lo que en la función de distribución u observando la gráfica 3, se hace evidente que la cola de distribución se alarga con valores superiores a la media.

Para la serie de datos en función de la ventana seleccionada:

5.4 Localización Ventana

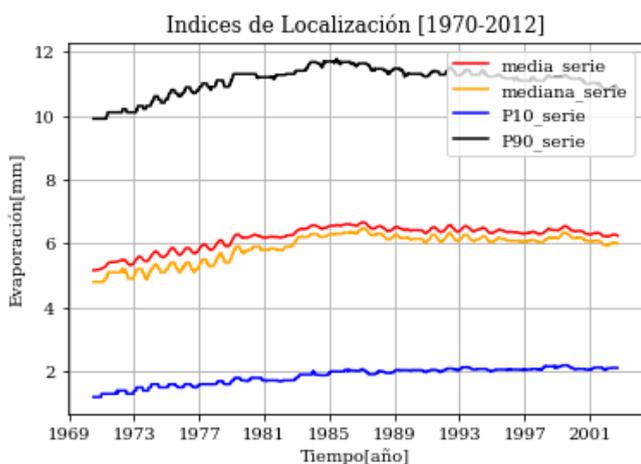


Figura 9. Índice de Localización en función de ventana [1970-2012]

La figura 9 nos muestra los índices de localización móvil (en función de la ventana), observando que tienen un periodo de incremento entre 1970 y 1985, para después tomar un valor cuasi-estacionario alrededor de 6 mm de evaporación (lo cual podría darnos idea de estacionariedad de los índices).

5.5 Dispersión Ventana

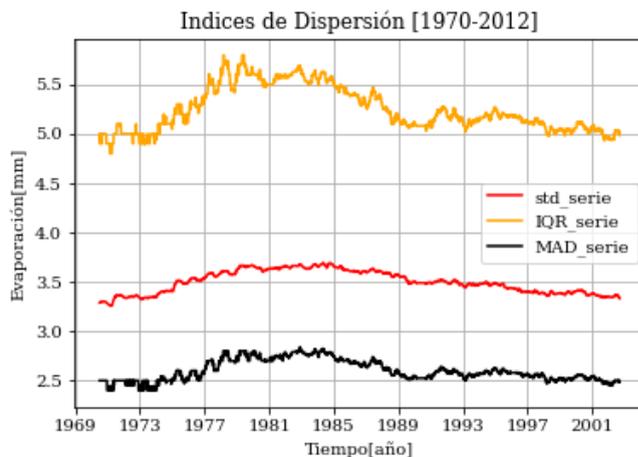


Figura 10. Índice de dispersión en función de ventana [1970-2012]

La figura 10 permite visualizar un periodo de tiempo donde los índices de dispersión tiene un máximo en la época de 1973 a 1989, lo cual nos llevaría a analizar porque las fluctuaciones que afectan los valores usualmente normales, fueron mucho mayores en esa época (una incremento de temperatura, la precipitación fue mayor que en las otras épocas por lo que había más agua disponible por evaporar).

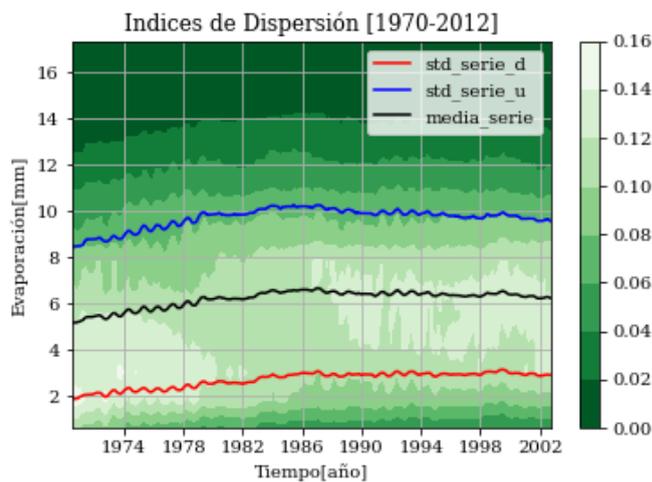


Figura 11. Dispersión de la desviación estándar comparada con la media y la serie en función de la ventana [1970-2012]

La figura 11 se graficó para mostrar como son los valores presentes a la hora de aplicarle el valor de desviación estándar a la media, y se observa que siguen teniendo un periodo de crecimiento, pero que las fluctuaciones grandes de la desviación no afectan el comportamiento general de la media, dejando que está se encuentre muy estable.

5.6 Simetría Ventana

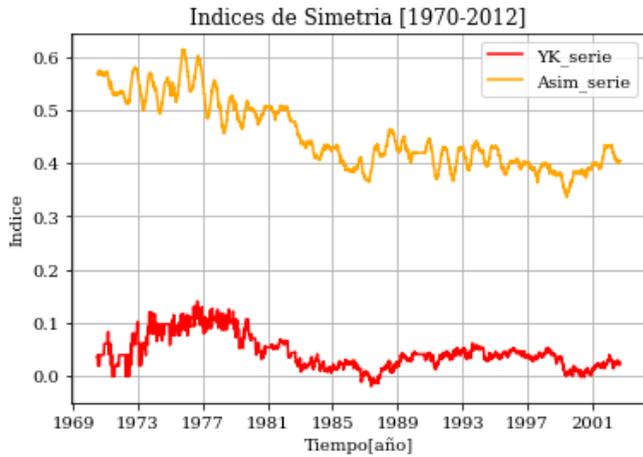


Figura 12. Índice de simetría en función de la ventana [1970-2012]

La figura 12 toma el valor de cero para ver hacia donde está la masa de la serie, si la simetría es positiva o negativa, para nuestro caso de estudio, tanto el índice de asimetría como el Yule- Kendall son positivos, por lo que en la función de distribución u observando la gráfica 11, se hace evidente que la cola de distribución se alarga con valores superiores a la media. Todo esto quiere decir que hay altas probabilidades a la hora de encontrar valores menores a la media (≈ 6 mm/día) de evaporación a lo largo de los años.

6. ¿Son estacionarios los histogramas?

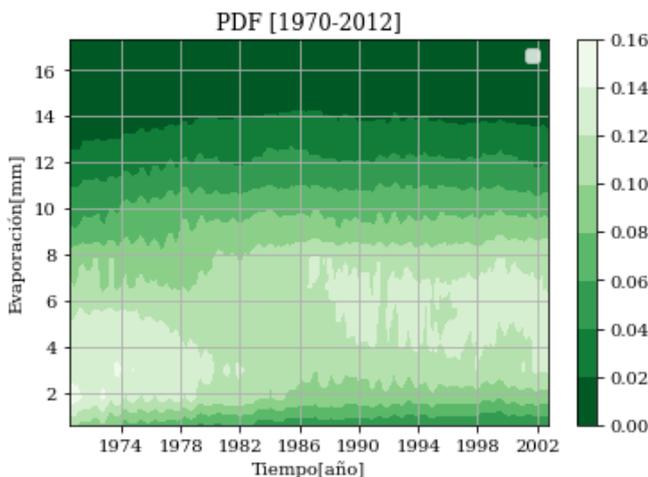


Figura 13. Densidad de histograma [1970-2012]

Tomando de base 6 mm de evaporación, se puede observar que la gráfica es no-estacionaria hasta 1986, después de ese año en mención, encontramos unas franjas con un ancho más uniforme, lo que nos puede decir que hay gran probabilidad de que sea estacionario el histograma desde 1986 a 2002.

7. ¿Son estacionarios los índices?

7.1 Localización

Se entiende por estacionario a una serie que se mantiene constante en su media y su desviación estándar; se tomará entonces la serie que se obtuvo correspondiente a la media de la serie original de evaporación y se evaluará lo mencionado. La figura 14 nos permite observar como es la variación en el tiempo de los índices de localización, y observamos que los 4 índices tiene un comportamiento muy similar, así que evaluaremos la estacionariedad del índice más robusto, concluyendo que los demás índices tienen el mismo comportamiento.

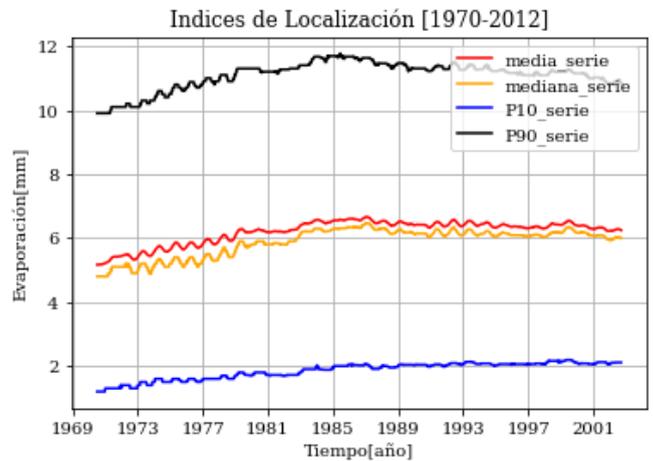


Figura 14. Índice de Localización en función de ventana [1970-2012]

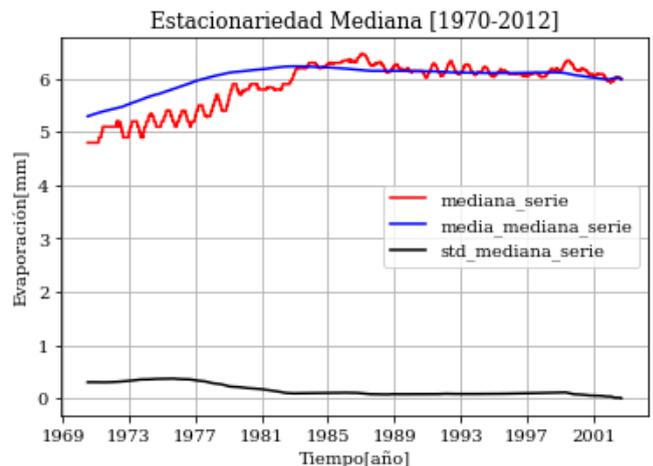


Figura 15. Evaluación estacionariedad índice Localización Mediana [1970-2012]

Se concluye entonces que la serie no es estacionaria en toda su longitud, ya que la media de la mediana tiene un tramo creciente hasta que se estabiliza en 1985, lo mismo pasa en el comportamiento de la desviación estándar de la mediana, es decreciente, hasta que deja de fluctuar en 1985; sin

embargo tiene un periodo quasi estacionario entre 1985 y 1997, siendo estrictos; ya que la variación no es significativa (de una décima) y podría decirse que es estacionaria en su totalidad. Lo mismo sucede con los índices de la media, y los percentiles P10 y P90.

7.2 Dispersión

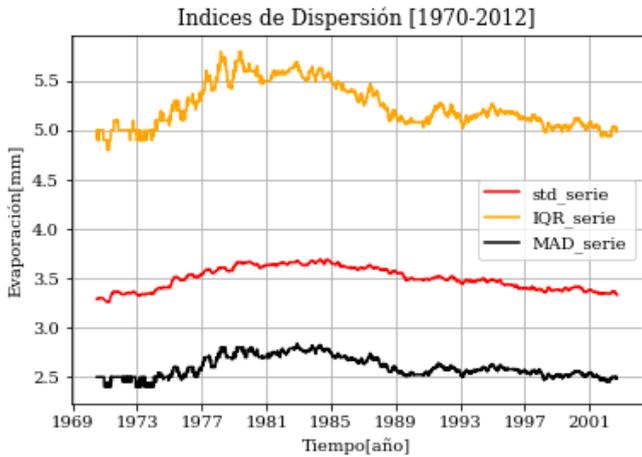


Figura 16. Índice de dispersión en función de ventana [1970-2012]

Se aplicará la misma metodología que se hizo con los índices de localización, tomando entonces la MAD por ser el índice más robusto y evaluar la estacionariedad de este.

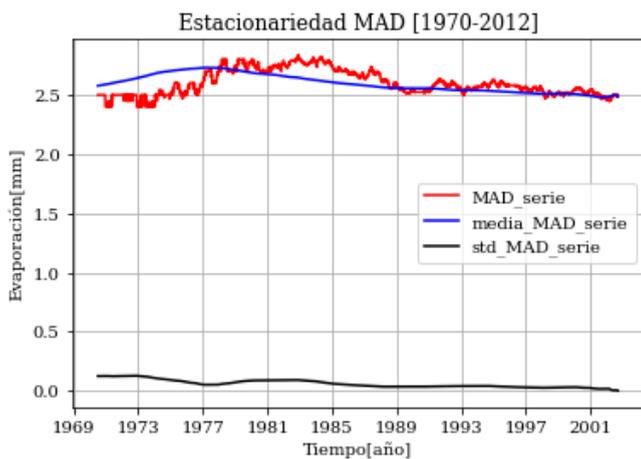


Figura 17. Evaluación estacionariedad índice dispersión MAD [1970-2012]

Se encuentra que el índice de dispersión más robusto tiene a tener un comportamiento similar a los índices de localización, manteniendo estacionariedad sólo entre el lapso de tiempo 1985 y 2001, siendo estrictos; ya que la variación no es significativa (de una décima) y podría decirse que es estacionaria en su totalidad. Lo mismo sucede con el índice IQR, más no es el caso de la desviación estándar que si

presenta desviaciones muy marcadas y cambios en su media, por lo que este parámetro no es estacionario.

7.3 Simetría

En la siguiente figura podremos observar los índices de YK y asimetría.

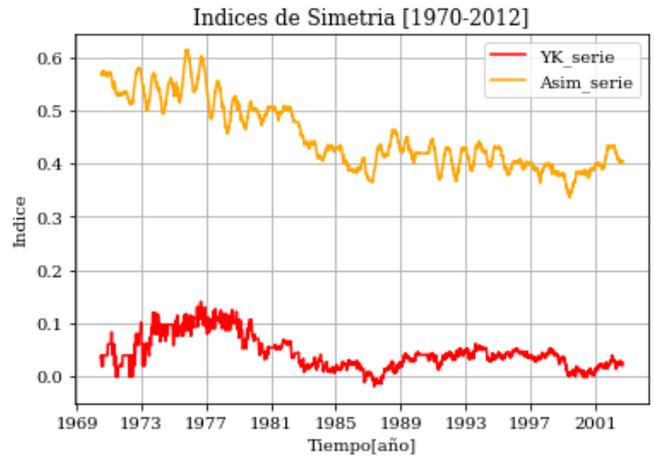


Figura 18. Índice de simetría en función de la ventana [1970-2012]

Aplicando lo mencionado anteriormente, se encuentra en la siguiente figura que el índice YK no es estacionario.

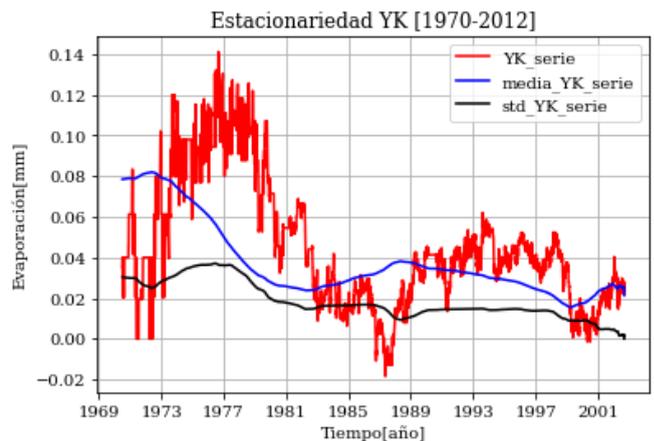


Figura 19. Evaluación estacionariedad índice asimetría YK [1970-2012]

No entiendo bien el objetivo de calcular la media a los índices

8. ¿Existe tendencia en la serie?, ¿En los percentiles?

Mediante el método de Mann-Kendall se evaluará la tendencia de la serie de Evaporación, al igual que la de percentiles; destacando que la tendencia de los percentiles se evaluará en la que pueda presentarse en el IQR, por ser el índice más robusto que presenta la serie.

Ya que queremos evaluar la existencia o no de tendencia, nuestra forma de llegar a la respuesta es evaluando con un incertidumbre de $\alpha/2$, si lo que quisiéramos es evaluar si la tendencia es creciente o decreciente, se evaluaría comparando el estadística con una incertidumbre a 1 cola, es decir, se toma el α completo.

Para determinar si hay o no tendencia se calcula la probabilidad del estadístico que determinamos con la prueba de Mann-Kendall y si es menor que el 97,5%, no hay tendencia (Khambhammettu & Lukas, 2005)

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad (1)$$

$$\text{Var}(S) = \frac{[n * (n-1) * (2n+5) - \sum_{p=1}^g t_p * (t_p-1) * (2 * t_p - 5)]}{18} \quad (2)$$

$$Z = \frac{S - 1}{\sqrt{\text{Var}(S)}}, \text{ si } s > 0 \quad (2)$$

$$Z = 0, \text{ si } s = 0 \quad (3)$$

$$Z = \frac{S + 1}{\sqrt{\text{Var}(S)}}, \text{ si } s < 0 \quad (4)$$

Se evalúa entonces este valor de Z que se tiene tanto para la serie como para los percentiles (IQR), encontrando y mostrando los resultados presentes en la siguiente tabla:

Tabla 1. Evaluación de Tendencia

	Mann Kendall S	Estadístico Z	Probabilidad	Tendencia al 97,5%	Ho: Aleatoriedad, Ha = Tendencia
Serie	55265	0,751941725	0,751941725	No hay tendencia	Se rechaza Ha
Percentiles	-14420475	-36,6046553	1,21E-293	No hay tendencia	Se rechaza Ha

**Dónde está la comparación
con la distribución normal?
no es claro**

9. Discusión

Al graficar los datos de evaporación desde 1970 a 2012, se tiene que la información base no presenta outliers, que generen fluctuaciones erróneas en la serie, sin embargo es objetivo de futuros estudios, unos valores más altos de la media alrededor del 2005, debiéndose posiblemente a mediciones mal hechas u otras de las posibilidades que se presenta es que el umbral que se alcance se deba sólo a la cantidad disponible de agua a evaporar, por consiguiente si

existen estos valores de la media puede deberse a que la precipitación fue mayor (mayor cantidad disponible para evaporar). Los histogramas muestran una uniformidad a la hora de la distribución temporal de la evaporación, con algunos lapsos que puede tener base en el aumento de temperatura o eventos meteorológicos no considerados. Los índices de localización, dispersión y simetría, dan unos rangos que a la hora de comparar no se salen de la muestra de datos por lo que se considera que se ha realizado unos cálculos idóneos. Los histogramas de la evaporación con un valor del 12% de probabilidad, que la evaporación es de 5 mm/día aproximadamente. Los histogramas en función de la ventana podrían presentar otra distribución según la elección de la ventana, por lo que encontrar un tamaño de ventana óptimo sería el ideal de este trabajo o de trabajos similares. Los índices aunque presentan una estacionariedad, en algunos segmentos del tiempo tienden a ser no-estacionarios, por lo que siendo rigurosos en el análisis, la serie debería no arrojar que no es estacionaria; sin embargo, evaluando la tendencia mediante el método de Mann-Kendall, se encontró que la serie no tiene tendencia, reafirmando la conclusión visual que se determinó mediante los índices y la serie misma.

10. Conclusiones

- La serie aunque no estacionaria en algunos segmentos del tiempo, se determinó la no tendencia de la misma.
- Se observa unos rangos de evaporación dentro de los reales, lo que da la conclusión que se trabajó con unos datos de una fuente información, y con una resolución óptima para sacar buenas conclusiones.
- El número de datos con los que se contó pudo dar la oportunidad de realizar un estudio que nos arroje un comportamiento lo más acertado posible, dando la posibilidad a estudios más complejos para determinar posibles tendencias futuras.

• Referencias

- Wilks, D.S. 2006. Statistical methods in the atmospheric sciences. Department of Earth and Atmospheric sciences, Cornell University. Second Edition
- Khambhammettu, P., Lukas, R., 2005. Mann-Kendall analysis for the FORT ORD site. Disponible en: <http://www.statisticshowto.com/wp-content/uploads/2016/08/Mann-Kendall-Analysis-1.pdf>