

Chicos, recuerden que esta es una clase de análisis de datos, deben darle sentido físico a sus resultados, no solo explicar los números que aparecen en las gráficas... sino tratar de explicar por qué suceden las cosas... en eso faltó un poquito más de detalle.

Tarea 1- Análisis de datos ambientales

Antonio Miguel González Dumar CC.: 1069501101
Verónica Rivera López CC.: 1039468637

4.0

1. Teniendo en cuenta que el requisito principal para la selección de la base de datos a trabajar para la tarea era que esta tuviera una cantidad de datos mayor a 1000, se optó por una donde se encontraron mediciones de velocidad promedio del viento, velocidad máxima, dirección promedio y dirección máxima de la estación 201 Torre Siata entre el 01 y el 31 de julio del 2018 con resolución minutal, disponibles en la página Web del Siata. Se seleccionaron solamente 2944 datos en formato .CSV, que corresponden a la velocidad promedio del viento entre la medianoche del 01 de julio y la 1:33 am del 03 de julio.

2. En la figura 1 se hace una primera inspección del comportamiento de los datos con el fin de identificar posibles *outliers* y datos faltantes. A partir de esta se puede identificar que para el conjunto de datos seleccionados no se encontraron datos de este tipo. Se puede percibir que los valores tienen una tendencia marcada que hace referencia al ciclo diurno, mostrando claramente que en las horas de iluminación los vientos alcanzan el valor máximo, en contraste con las **horas de penumbras**, donde los vientos tienen un valor mínimo.

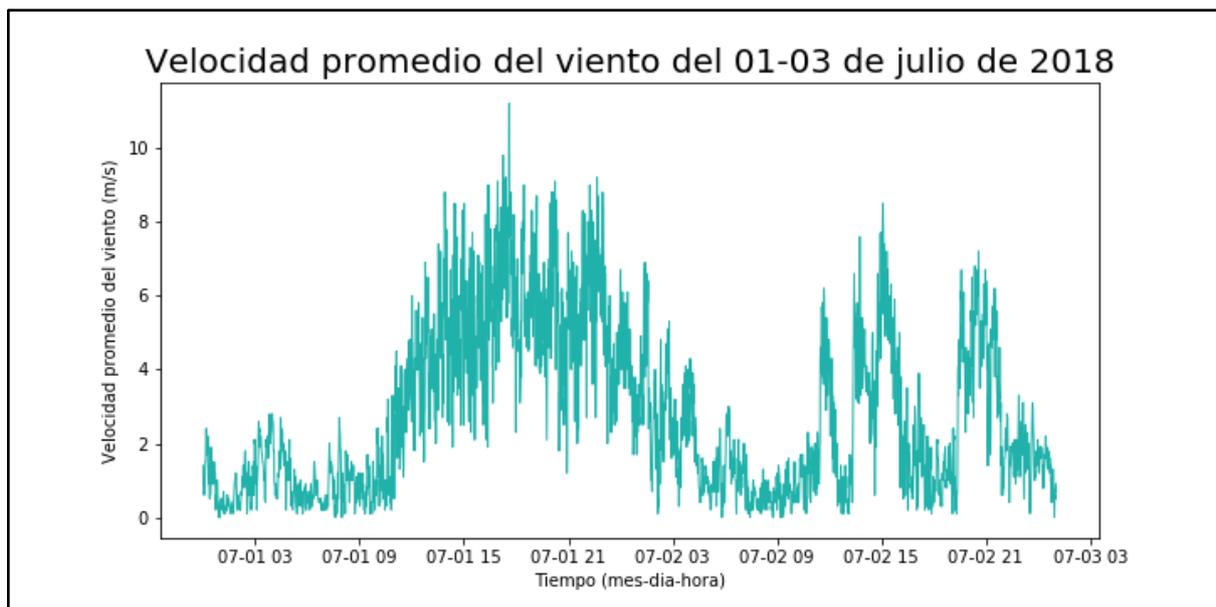


Figura 1. Comportamiento de la variable *Velocidad del viento promedio*. Fuente: Elaboración propia.

3. Los índices estadísticos para la base de datos seleccionada se muestran en la tabla 1. Se puede observar que la media y la mediana de la serie son aproximadamente diferentes. Esto conlleva a que la elección única de una u otra variable como índice absoluto de localización puede generar errores, pues, no es correcto suprimir a primera vista la información de uno de estos índices, dado que, para ello, se requiere un análisis más profundo. Esta diferencia se puede generar debido a la naturaleza de los datos.

Es importante recordar que en general, la media es más susceptible a la presencia de datos extremos en la serie y que la mediana ofrece un análisis más resistente para la misma.

Medidas de localización	
Media	2.7827
Mediana	2.1
Medidas de desviación	
Desviación estándar	2.2198
Rango Intercuartil	3.6
Medidas de asimetría	
Coef. Yule-Kendall	0.33
Coef. Asimetría	0.7595

Tabla 1. Coeficientes estadísticos de la serie. Fuente: Elaboración propia.

En el caso de la desviación estándar y el rango intercuartil, se observa que tampoco existe un grado de similitud entre ambos estimadores usados. Tomando como referencia el valor arrojado de la media y la figura 1 se observa que, en la figura 2, la serie no se encuentra concentrada cerca de la media, indicando que esta se encuentra dispersa.

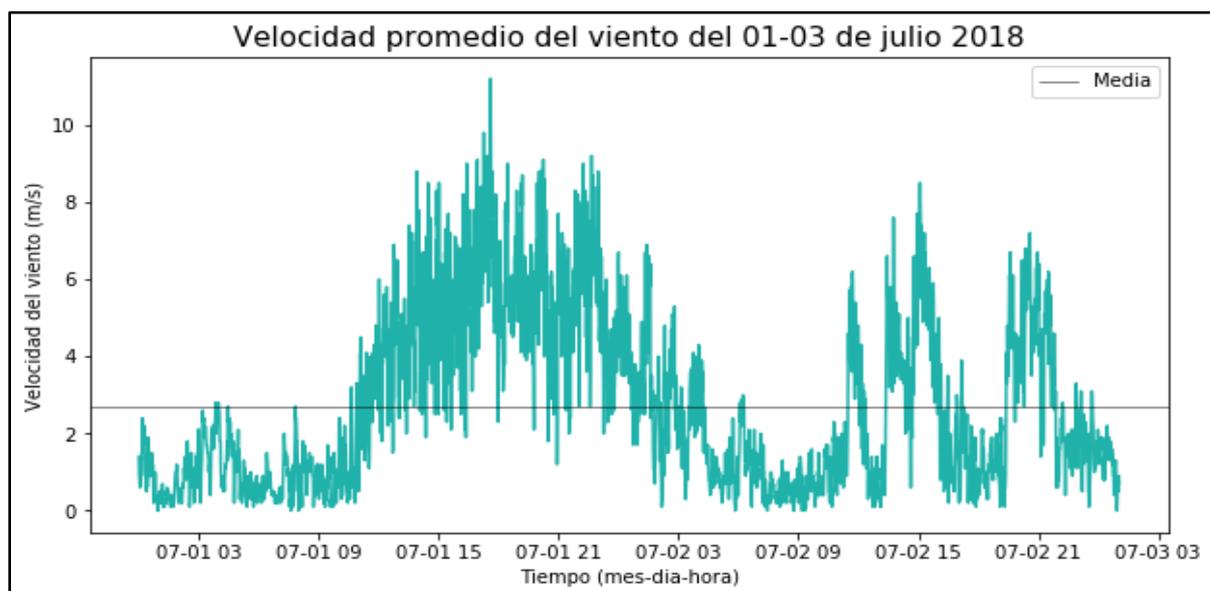


Figura 2. Comportamiento de la variable *Velocidad del viento promedio* más media. Fuente: Elaboración propia.

Adicionalmente de la figura 2 se puede observar que en general la serie es asimétrica con respecto a la media ya que predominan valores por debajo de esta. Por tanto, se espera que el histograma esté concentrado en la parte izquierda del mismo.

4. En la figura 3 se presenta la función de densidad de probabilidad acumulada de la serie y 5 de los percentiles más representativos.

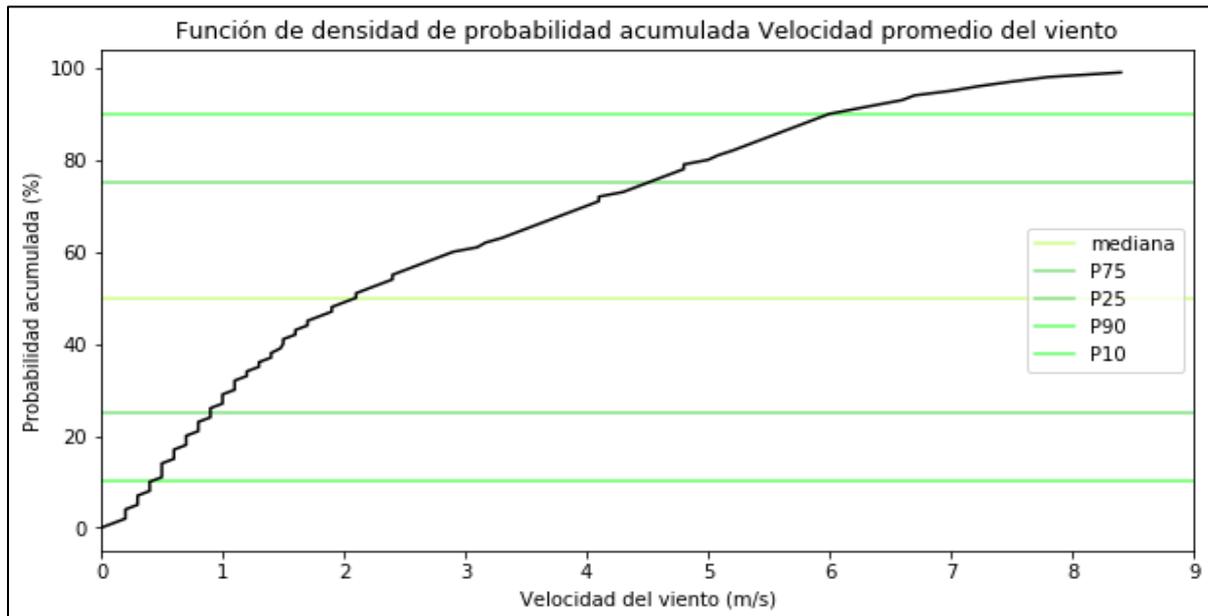


Figura 3. Función de densidad de probabilidad de la variable *Velocidad del viento promedio*. Fuente: Elaboración propia.

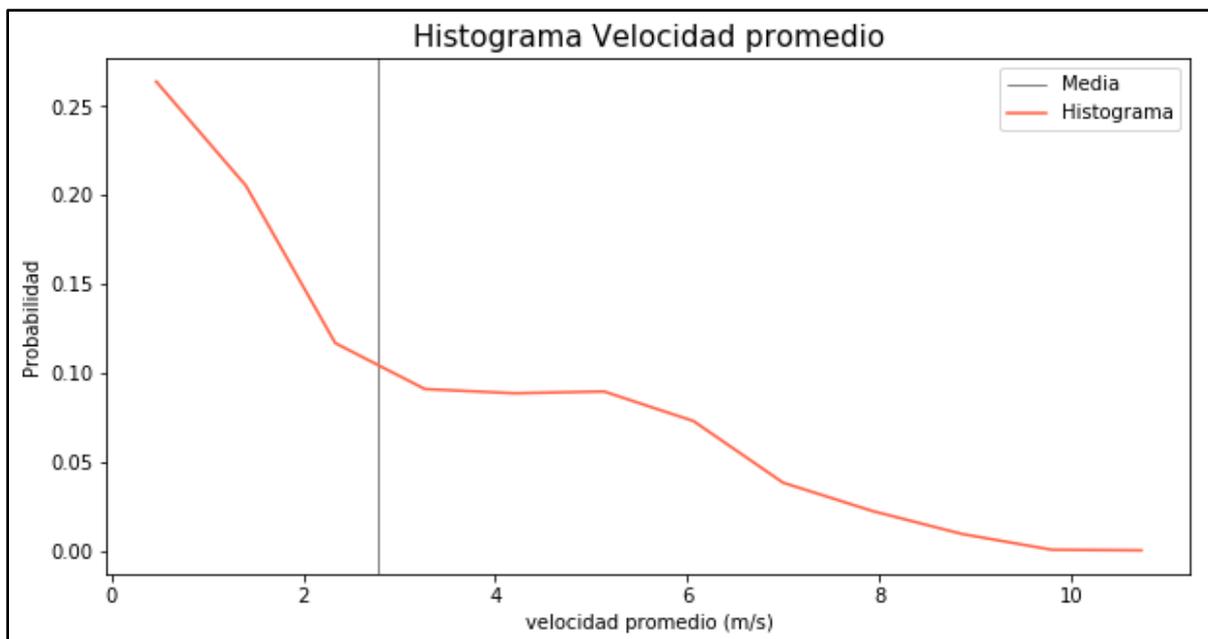


Figura 4. Función de densidad de probabilidad de la variable *Velocidad del viento promedio*. Fuente: Elaboración propia.

Así como se comentó anteriormente, teniendo presente los valores arrojados por el coeficiente de asimetría, podemos comprobar, en la figura 4, que efectivamente la serie se concentra en valores por debajo del de la media, allí las probabilidades de obtención de datos dentro de ese rango son mayores.

5. Para la estacionariedad se evaluó el comportamiento de los percentiles, media, mediana y demás momentos estadísticos en una ventana aproximada de 6 horas. Este tamaño se decidió teniendo en cuenta la frecuencia de los datos evaluados (minutal). Así mismo, se encontró que, a partir de ventanas de este tamaño, los histogramas comenzaron a mostrar diferencias con respecto al histograma de la serie original, presentado en la figura 4.

El análisis por ventanas es útil a la hora de evaluar estacionariedad porque permite visualizar si una fracción de la serie se comporta de la misma manera en que se comporta la serie entera. En este sentido, se define que una serie es estacionaria cuando ninguno de sus momentos estadísticos muestra una variación respecto al tiempo. La figura 5 muestra el histograma conjunto de la serie, es decir una simplificación de todos los histogramas obtenidos a partir de la definición de la ventana de evaluación. Teniendo presente la escala de color es claro que en la figura 5 las concentraciones de los datos varían de rango en función del tiempo, esta mayor densidad en diferentes puntos modifica directamente la media de la velocidad promedio del viento para los tiempos definidos. Como ya se mencionó, la variación de este parámetro es suficiente para afirmar que no hay estacionariedad para la media, y por lo tanto tampoco hay para la serie.

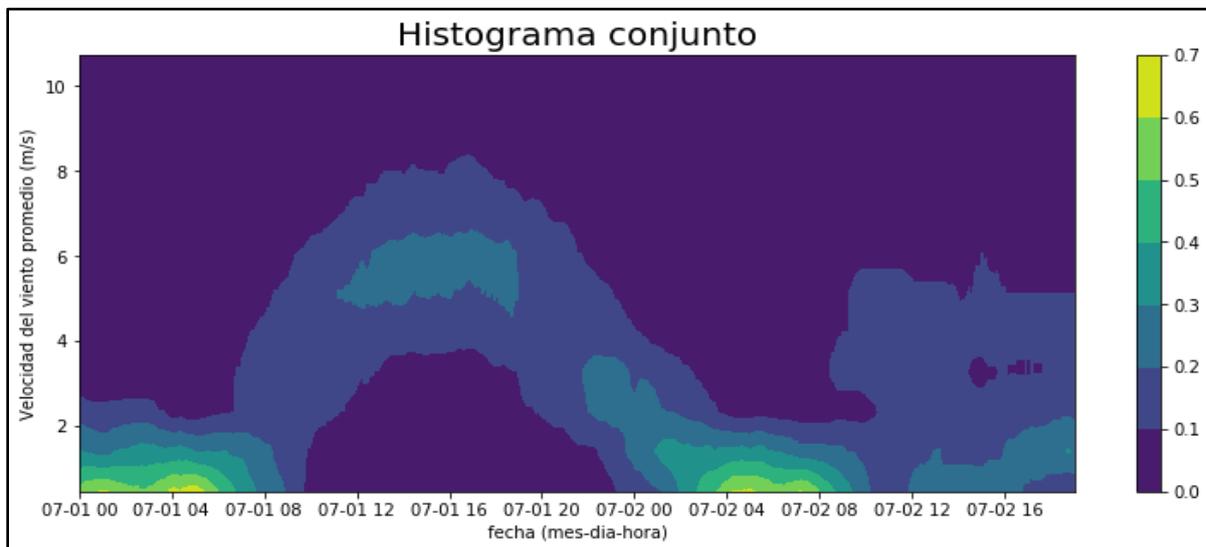


Figura 5. Histograma conjunto *Velocidad del viento promedio*. Fuente: Elaboración propia.

6. Teniendo presente a variación de los momentos estadísticos evaluados a lo largo del tiempo, y presentados en las figuras 6 a 9, se puede concluir que ninguno de estos es estacionario. Así mismo, en la figura 10 se puede observar como es el comportamiento de la desviación estándar con respecto a la media (bandas de confianza)

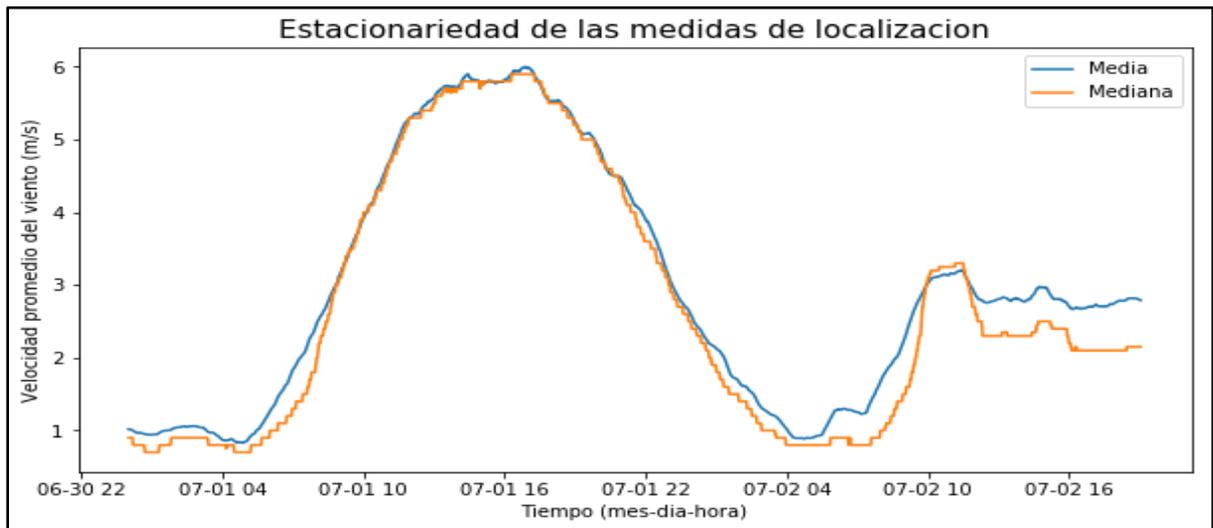


Figura 6. Media y mediana estacionaria *Velocidad del viento promedio*. Fuente: Elaboración propia

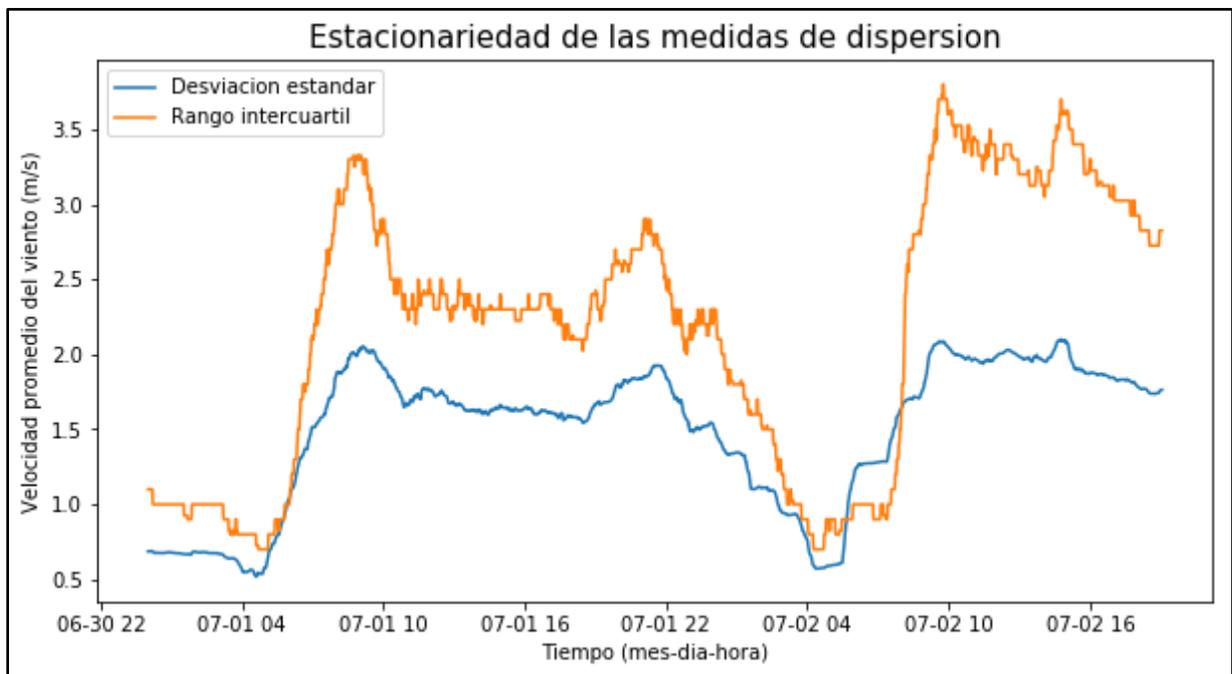


Figura 7. Desviación estándar y rango intercuartil estacionario *Velocidad del viento promedio*. Fuente: Elaboración propia

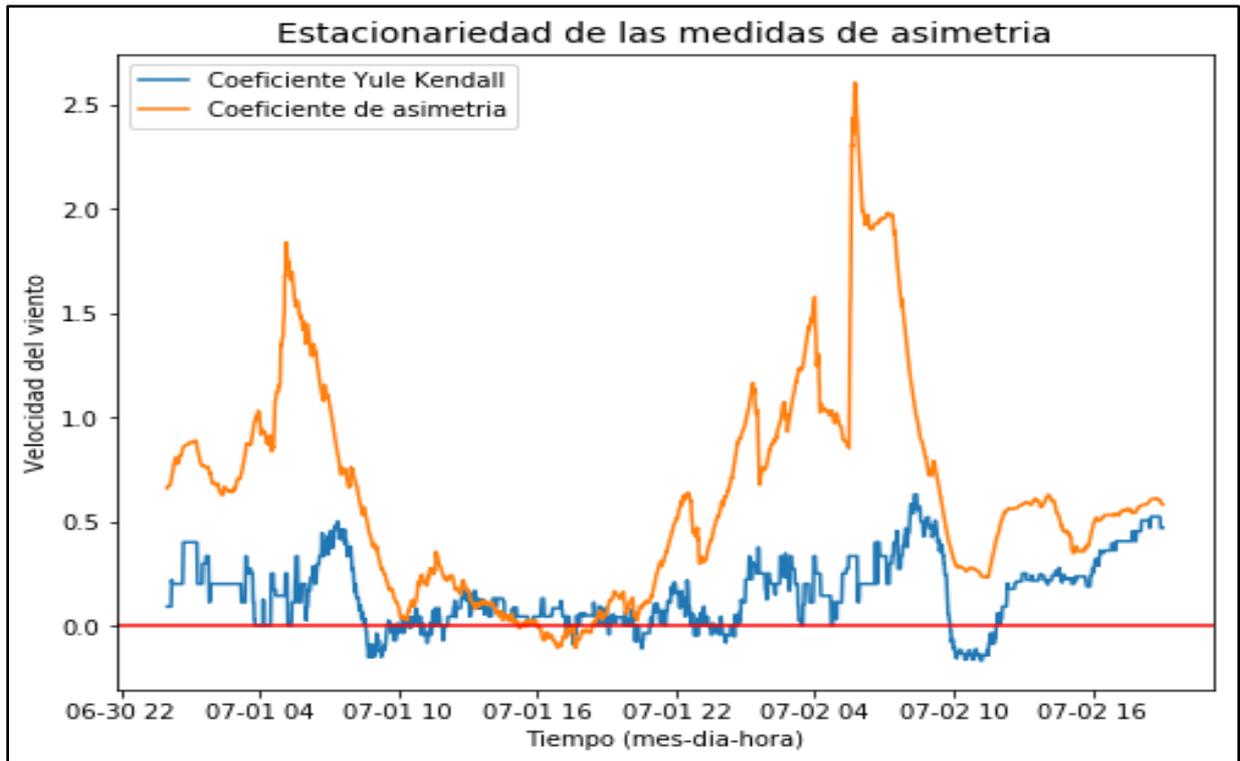


Figura 8 .Coeficiente Yule Kendall y asimetría estacionario *Velocidad del viento promedio*. Fuente: Elaboración propia

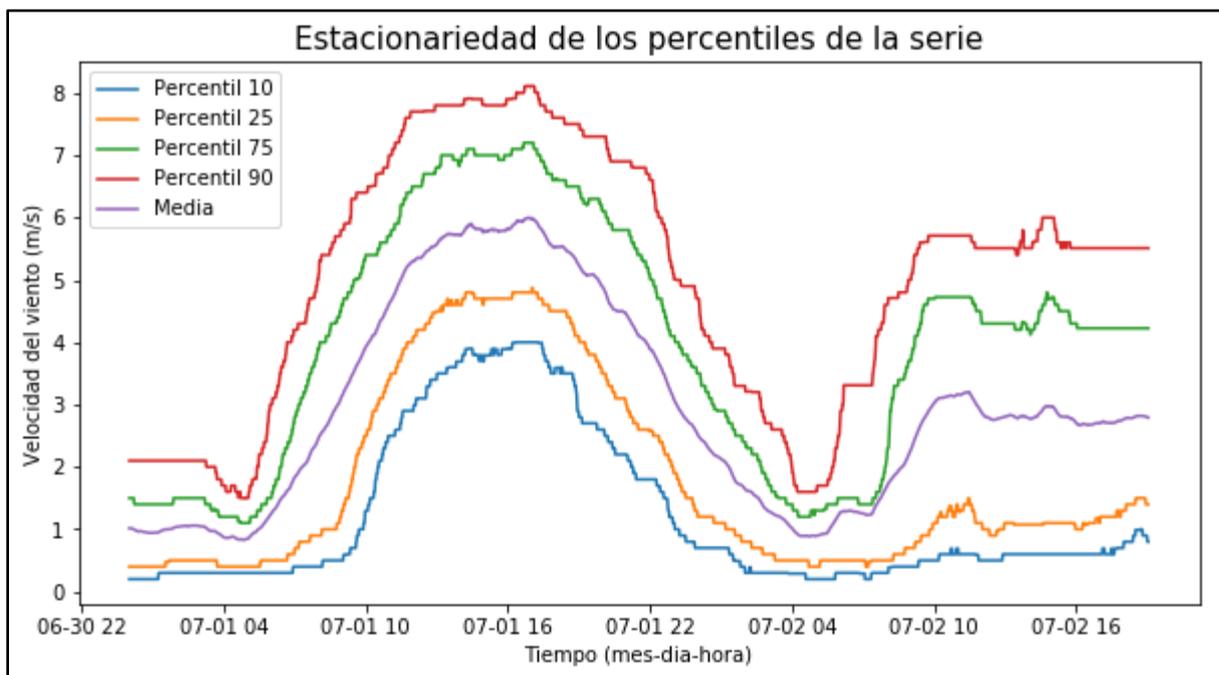


Figura 9. Percentiles estacionarios *Velocidad del viento promedio*. Fuente: Elaboración propia.

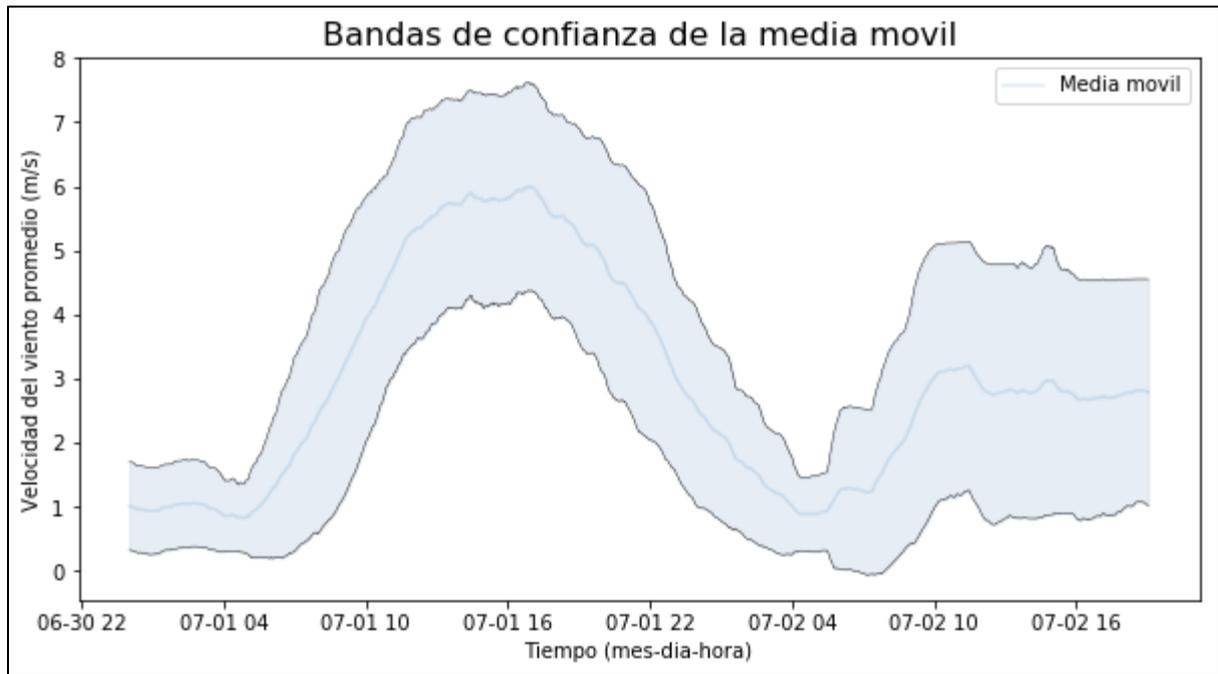


Figura 10. Banda de confianza media estacionaria *Velocidad del viento promedio*. Fuente: Elaboración propia

7. Se definió la función Mann Kendal a partir de dos argumentos: uno que corresponde a la serie a la cual se le desea evaluar la tendencia, y otro al valor *alfa* que, es igual a la incertidumbre con que se quiere rechazar la hipótesis nula, en este caso: la distribución aleatoria de los valores en la serie. Se calculó la tendencia en la serie conformada por la media móvil, percentil 10 móvil, percentil 25 móvil, percentil 75 móvil, y el percentil 90 móvil para la ventana móvil ya definida. Es importante aclarar que para la evaluación en el test se tuvieron presentes los empates, estos se definen como las repeticiones de valores dentro de los datos. La presencia de estos, indica que varios valores de un x comparten un mismo valor en y . Los resultados de esta evaluación en las series se presentan a continuación, para la comparación con la función normal, se usó una incertidumbre de 5% lo cual arroja un valor de comparación $Z_{1-\alpha}$ de 0.828943. Teniendo en cuenta que en las series evaluadas el estadístico de prueba es positivo, la hipótesis nula (distribución aleatoria de la serie) puede rechazarse cuando este es mayor a $Z_{1-\alpha}$ (VSP Main, 2018).

Test de tendencia para la media móvil		
Valor S	Varianza S	Zmk
121313	1918168719	2.7699
alpha	Conclusión	
0.05	Con una tendencia de 95% se rechaza la hipótesis nula (VSP Main, 2018)	

Tabla 2. Resultados test de Mann Kendall para la media móvil. Fuente: Elaboración propia

Test de tendencia para el percentil 10		
Valor S	Varianza S	Zmk
117646	1892383263	2.7044
alpha	Conclusión	
0.05	Con una tendencia de 95% se rechaza la hipótesis nula (VSP Main, 2018)	

Tabla 3. Resultados test de Mann Kendall para el percentil 10. Fuente: Elaboración propia.

Test de tendencia para el percentil 25		
Valor S	Varianza S	Zmk
132863	1905281456	3.0438
alpha	Conclusión	
0.05	Con una tendencia de 95% se rechaza la hipótesis nula (VSP Main, 2018)	

Tabla 4. Resultados test de Mann Kendall para el percentil 25. Fuente: Elaboración propia.

Test de tendencia para el percentil 75		
Valor S	Varianza S	Zmk
102371	1914850803	2.3394
alpha	Conclusión	
0.05	Con una tendencia de 95% se rechaza la hipótesis nula (VSP Main, 2018)	

Tabla 5. Resultados test de Mann Kendall para el percentil 75. Fuente: Elaboración propia.

Test de tendencia para el percentil 90		
Valor S	Varianza S	Zmk
176956	1912355729	4.0464
alpha	Conclusión	
0.05	Con una tendencia de 95% se rechaza la hipótesis nula (VSP Main, 2018)	

Tabla 6. Resultados test de Mann Kendall para el percentil 90. Fuente: Elaboración propia.

8. Conclusiones

En el ejercicio se pudo identificar que tanto la media como los demás percentiles evaluados en la función Mann Kendall tuvieron un comportamiento creciente. A primera impresión, el valor obtenido como estadístico de prueba, podría no ser coherente con el resultado de los estadísticos calculados, pues, aparentemente estos tienen un comportamiento periódico que podría sugerir la no existencia de una tendencia. Sin embargo, es importante tener presente que las variables evaluadas en el test tienen empates, los cuales alteran la varianza del parámetro S de la función, siendo necesario su remoción.

En la figura 5 se encontró que la serie que describe la velocidad del viento promedio no es estacionaria, ya que esta depende de la hora del día en la que se esté evaluando. Así mismo, se observa que esta aumenta aproximadamente en horas de la tarde y/o principios de la noche y disminuye en horas de la madrugada, como consecuencia del ciclo diurno. Se observa que este comportamiento es coherente con lo presentado en la figura 1.

Referencias

VSP Main. (07 de Julio de 2018). *Mann-Kendall Test For Monotonic Trend*. Obtenido de https://vsp.pnnl.gov/help/vsample/design_trend_mann_kendall.htm