

Informe final de Métodos estadísticos

Estudiante: Noelia Santos

Fecha de entrega: 15/07/2021

En el presente trabajo se analiza la serie temporal mensual de la temperatura en el nivel de presión correspondiente a 100 hPa en la base Marambio (64°14'S, 56°38'O) para el período 1980-2019 obtenida a partir del reanálisis ERA 5 [1].

La motivación para la elección de esta serie se enmarca dentro del trabajo de doctorado de la estudiante, cuyo objetivo principal es el análisis del flujo de rayos cósmicos galácticos observado por un detector instalado en la base Marambio en el año 2019.

Los rayos cósmicos galácticos son partículas que adquieren velocidades relativistas en diferentes eventos energéticos ocurridos en la galaxia. Aquellas partículas que logran alcanzar el entorno terrestre y precipitarse hacia la atmósfera se denominan rayos cósmicos primarios. Debido a la interacción entre los rayos cósmicos primarios y las partículas de la atmósfera se produce una reacción en cadena donde se generan diferentes tipos de partículas o rayos cósmicos secundarios conocida como cascada o lluvia de partículas.

El detector mencionado anteriormente permite contar el número de rayos cósmicos secundarios que depositaron su energía en él por segundo. Naturalmente el número de partículas secundarias generadas en las lluvias depende de las condiciones atmosféricas del sitio de observación.

Analizar el flujo de partículas secundarias observado e identificar eventos transitorios de origen interplanetario es de interés para Meteorología del Espacio, que es el área en el cual se enmarca el proyecto antártico del detector de partículas, por lo cual para ello primero es necesario remover los efectos atmosféricos presentes en la señal.

Dos de los efectos más importantes nombrados en diferentes trabajos del área son el debido a la variación de la presión barométrica y a la variación del perfil vertical de la temperatura atmosférica [2].

El primer efecto se puede remover experimentalmente a partir de la medición de la presión barométrica del sitio. El segundo efecto, relacionado con la probabilidad de interacción y decaimiento de las partículas secundarias, requiere conocer la temperatura en las diferentes capas de la atmósfera. En los datos que se tienen del detector de la base Marambio se observa una modulación anual del flujo de partículas como así también variabilidades en escalas temporales más pequeñas que, a priori, podrían estar relacionadas con, por ejemplo, el calentamiento repentino de la estratosfera. Para modelar el efecto de la temperatura es útil considerar los datos de reanálisis como ERA 5 o GDAS.

En este contexto se presenta el trabajo “Atmospheric Temperature Effect in Secondary Cosmic Rays Observed With a 2 m² Ground-Based tRPC Detector” publicado en la revista Earth and Space Science en el año 2020 [3]. Los autores utilizan el perfil vertical de temperatura con resolución de 6 horas para los 37 niveles de presión del reanálisis ERA-Interim y la metodología de componentes principales (CP) para modelar el efecto de la temperatura en el flujo observado por un detector de rayos cósmicos secundarios que se encuentra en Santiago de Compostela, España. El “método integral” para estimar el efecto de la temperatura tiene en cuenta la variación de la temperatura en cada capa de la atmósfera y cada una de ellas posee un determinado peso. La integral se aproxima por una sumatoria obteniéndose la siguiente expresión:

$$\left. \frac{\Delta R}{R_0} \right|_T = \sum_{i=1}^n W_T(h_i) \Delta T_i \Delta h_i \quad (1)$$

El término de la izquierda de la ecuación 1 representa la variación relativa del conteo de partículas ΔR respecto a la media del período de estudio R_0 debido al efecto de la temperatura. En el término de la derecha se considera que la atmósfera posee n capas y para cada capa i se necesita conocer: la

variación de la temperatura ΔT_i , el espesor Δh_i y el peso asociado $W_T(h_i)$. Los autores argumentan que no puede realizarse una regresión lineal múltiple directamente para estimar los pesos dado que las variables predictoras ΔT_i no son independientes. En la Figura 1 se muestra la correlación lineal entre los diferentes niveles de presión, evidenciando esto último (no aclaran nada sobre la significancia de estos coeficientes). Por este motivo, proponen la metodología CP para encontrar un nuevo conjunto de variables predictoras que sean independientes entre sí. Posteriormente realizan una regresión lineal múltiple con estas nuevas variables y estiman el valor de los pesos. En la Figura 2 (a) se presentan los pesos obtenidos a partir de utilizar CP y se compara con 3 estimaciones teóricas (la diferencia entre cada una es el rango de energía de las partículas consideradas). En la Figura 2 (b) se presentan los pesos obtenidos para una regresión lineal múltiple sin utilizar CP, dejando en evidencia las diferencias, sobre todo en los niveles de presión más baja. Respecto a la metodología, los autores afirman que eventos excepcionales ocurridos en la atmósfera como por ejemplo, calentamientos estratosféricos repentinos, y otros eventos transitorios afectan la determinación de las CP y es por eso que utilizan la serie de tiempo de los últimos 30 años del perfil de temperatura para determinarlas.

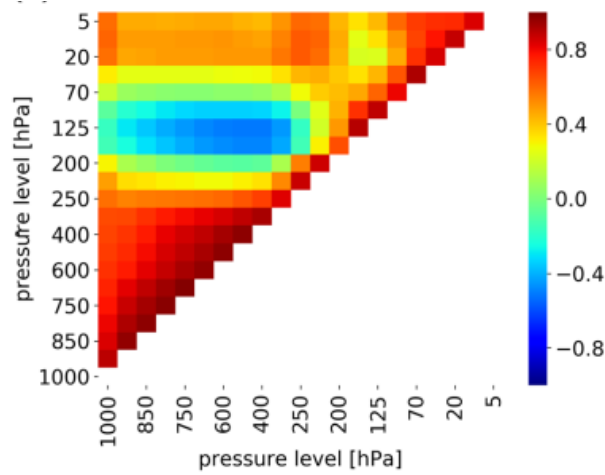


Figura 1: Correlación lineal entre la temperatura de los diferentes niveles de presión en Santiago de Compostela (desde enero 2015 a diciembre de 2016) [3].

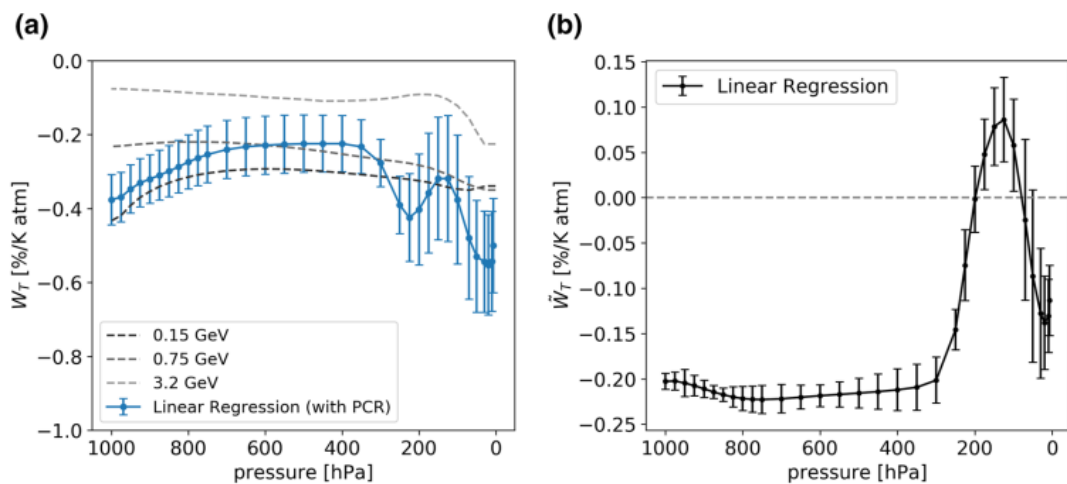


Figura 2: Pesos estimados a partir de realizar una regresión lineal múltiple (a) utilizando como predictores las dos primeras CP (b) utilizando como predictores la variación de temperatura en cada nivel de presión. Las líneas punteadas en (a) representan estimaciones teóricas de los pesos para diferentes energías de las partículas secundarias [3].

Si bien para realizar este tipo de análisis se necesita la temperatura en los diferentes niveles y una resolución temporal al menos diaria, en el presente trabajo final se eligió trabajar con la temperatura mensual en el nivel de 100 hPa, que es el nivel promedio en el cual se generan el tipo de partículas secundarias que presentan mayor sensibilidad a la variación de temperatura [2]. De esta forma el análisis de la serie va a permitir familiarizarse con la variable de interés y dar el primer paso para avanzar en esta línea de investigación.

Para comenzar con el análisis de la serie primero se debe graficar la misma y realizar el control de calidad de los datos. Dado que la serie proviene de un reanálisis no posee datos faltantes. En la Figura 3 se observa que la serie presenta tanto una componente estocástica como determinista. Respecto a esta última componente, a priori, gráficamente se pueden identificar ciclos siendo el más evidente el ciclo anual.

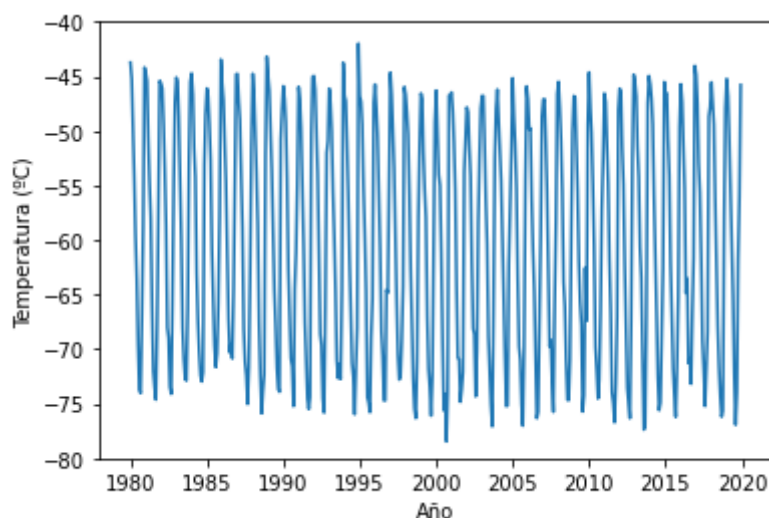


Figura 3: Serie temporal mensual de la temperatura para el nivel 100 hPa en la base Marambio para el período 1980-2019 del reanálisis ERA 5.

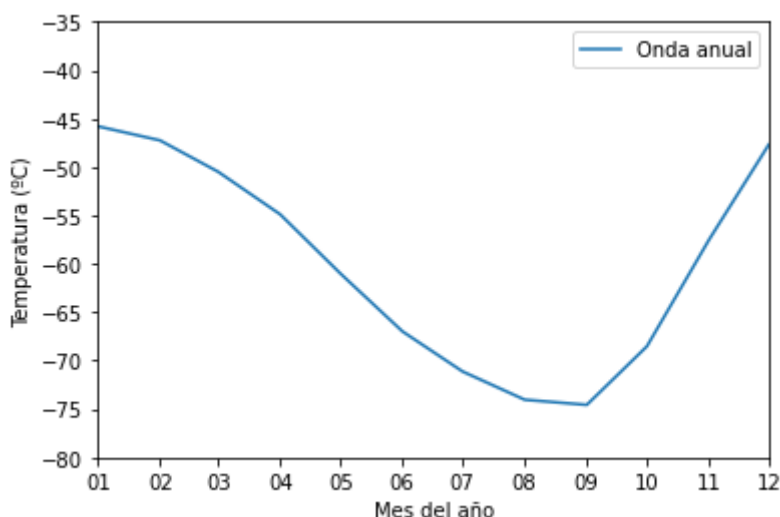


Figura 4: La serie original y la onda anual construida a partir de los promedios sobre cada uno de los meses del año para los 40 años de estudio.

En la Figura 4 se observa la onda anual construida a partir de considerar la media de cada uno de los meses del año para los 40 años de estudio. En la Figura 5 se presentan las anomalías de la temperatura para el año 2019 (período de interés para este trabajo dado que es el inicio de las mediciones del detector de base Marambio). El área sombreada representa la región definida por una desviación estándar respecto a la media asociada con cada uno de los meses del año. Utilizando esta medida como criterio puede decirse que en Julio - Agosto de 2019 se observan anomalías negativas y en Septiembre - Octubre de 2019 anomalías positivas, siendo ésta última de mayor amplitud.

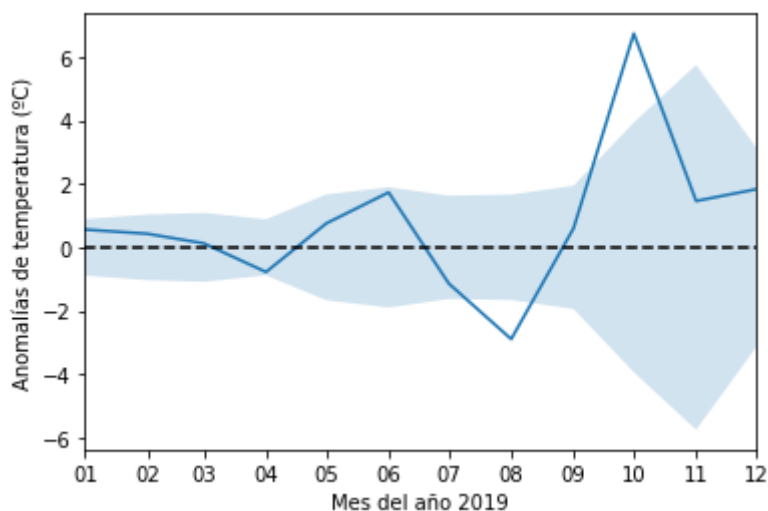


Figura 5: Anomalías de la temperatura mensual para el nivel 100 hPa en la base Marambio para el año 2019. El área sombreada representa la región definida por una desviación estándar respecto a la media asociada con cada mes.

Por otra parte, se dividió la serie original en dos períodos iguales, cada uno de 20 años, y se calculó la media y desviación para los meses de enero y julio de ambos períodos para evaluar si hubo cambios significativos. Los resultados se presentan en la Tabla 1.

Períodos	Media (°C)	Desviación (°C)
Enero (1980 - 1999)	− 45.54	0.86
Enero (2000 - 2019)	− 46.08	0.84
Julio (1980 - 1999)	− 70.82	1.45
Julio (2000 - 2019)	− 71.66	2.03

Tabla 1: Media y desviación de la serie de temperatura de los eneros y julios en los períodos 1980-1999 y 2000-2019.

Se realizó un test de medias tanto para enero como para julio. Se tiene evidencia suficiente para decir que hay un cambio de media en enero con un 5 % de significancia, pero no para julio.

También se analizó la correlación lineal entre la serie original y el índice de oscilación sur SOI [4], obteniéndose $r = 0.01$. Considerando sólo los meses de enero el coeficiente de correlación lineal es $r = -0.11$ y considerando los meses de julio $r = -0.22$. Se realizó un test normal en el primer caso y

un test t-student para los últimos dos casos considerando un 5% de significancia. En ninguno de los tres casos se cuenta con evidencia suficiente para afirmar que el coeficiente de correlación de la población no es cero.

Se extendió el análisis de la relación con el índice SOI para toda la península antártica. A partir de los datos de la variable del modelo NCEP [5] se calculó la correlación espacial entre el índice SOI y la temperatura en $h = 100$ hPa en la península antártica para los meses de enero y julio en el período 1980-2019. El mapa de correlación para ambos meses se presenta en la Figura 6. Si bien se trata de un modelo diferente al utilizado en este trabajo, en la base Marambio (indicada con un punto negro en la Figura 6) se obtienen los mismos coeficientes de correlación reportados anteriormente para enero y julio. En ambos meses la correlación es negativa, siendo mayor en módulo en Julio.

En la península antártica predomina una correlación negativa con el índice SOI para los meses de Julio. Para los meses de enero la correlación en general es muy pequeña tendiendo a ser negativa para las latitudes más bajas y positiva para latitudes altas. De todas formas en todos los casos hay que evaluar la significancia. Considerando el test de student con 5% de significancia y 38 grados de libertad el coeficiente de correlación a partir del cual se tiene significancia estadística es $|r_c| = 0.27$, por lo cual no se cuenta con evidencia suficiente para decir que el coeficiente de correlación de la población en cada sitio de la península no es cero.

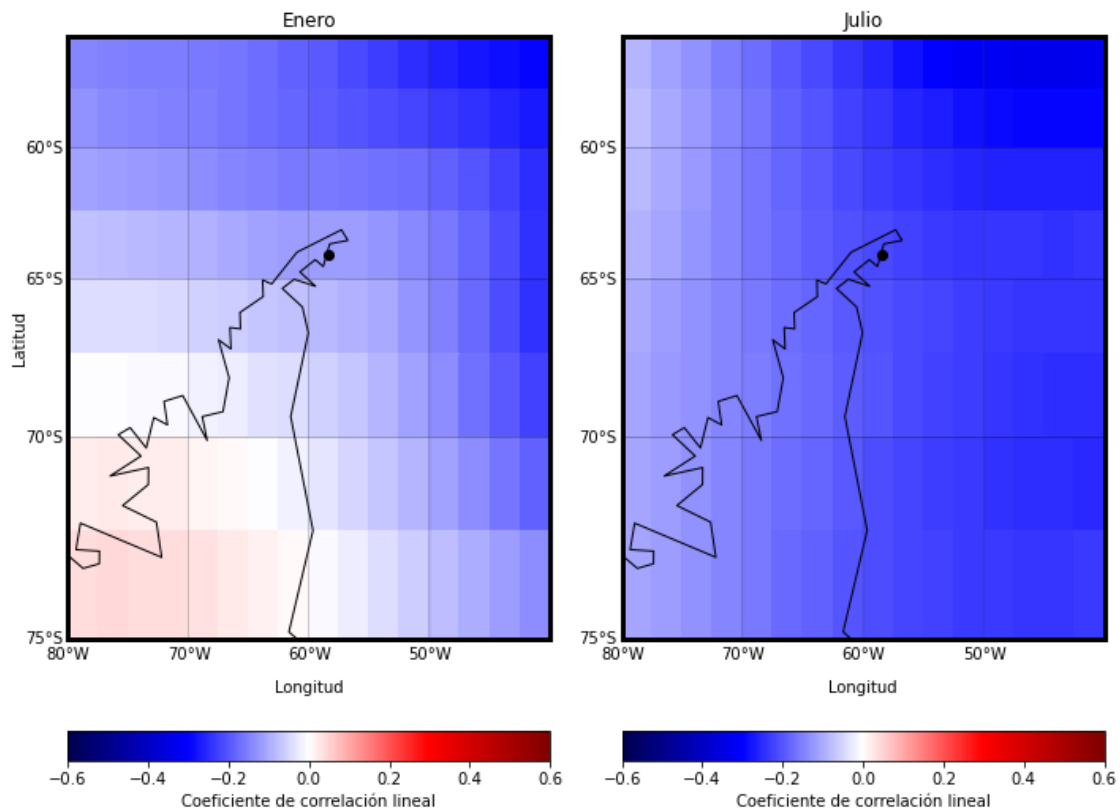


Figura 6: Mapa de correlación espacial entre el índice SOI y la temperatura en 100 hPa en la península antártica para enero y Julio en el período 1980-2019.

Para avanzar con el análisis de la parte determinística de la serie se consideró realizar un test de Mann-Kendall para evaluar si la serie es al azar o no. Para ello se consideró la serie de promedios anuales que se presenta en la Figura 7. Con un 5% de significancia puede afirmarse que la serie no es al azar, por lo cual a priori la serie presenta una tendencia pero al menos visualmente no queda claro que tipo de tendencia podría ser. Además podría considerarse la posibilidad de que también haya un ciclo de período mayor a 40 años.

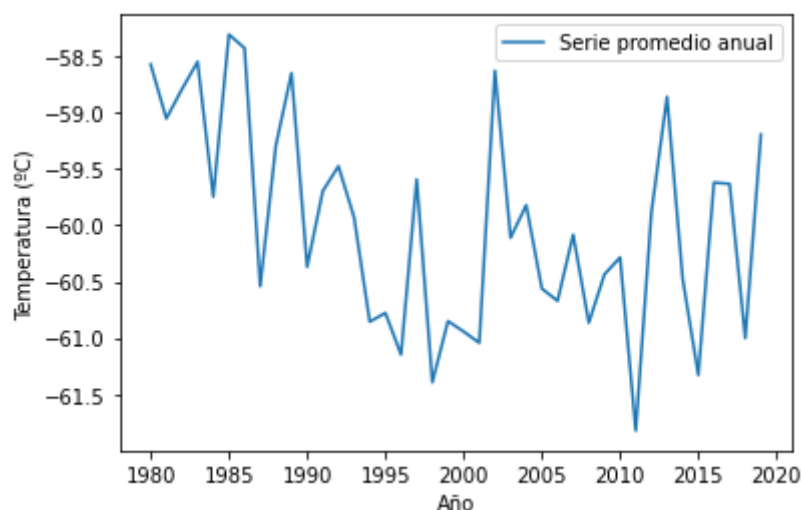


Figura 7: Serie temporal anual de la temperatura para el nivel 100 hPa en la base Marambio para el período 1980-2019 del reanálisis ERA 5.

Continuando con el análisis de la parte determinística de la serie se realizó un análisis armónico de forma tal de detectar posibles ciclos en la misma. Para realizarlo se debe tener en cuenta que la cantidad de datos sea par, en este caso $N = 480$, por lo tanto se tienen 240 armónicos. Además como la resolución de la serie original es mensual, el período fundamental es de 480 meses. Esto sugiere que, de encontrarse la onda anual, deberá estar representada por el armónico 40. En la Figura 8 se presenta el periodograma obtenido para la serie original. Puede observarse que, efectivamente, la onda anual está presente y representada por el armónico 40 que explica el 91% de la varianza de la serie.

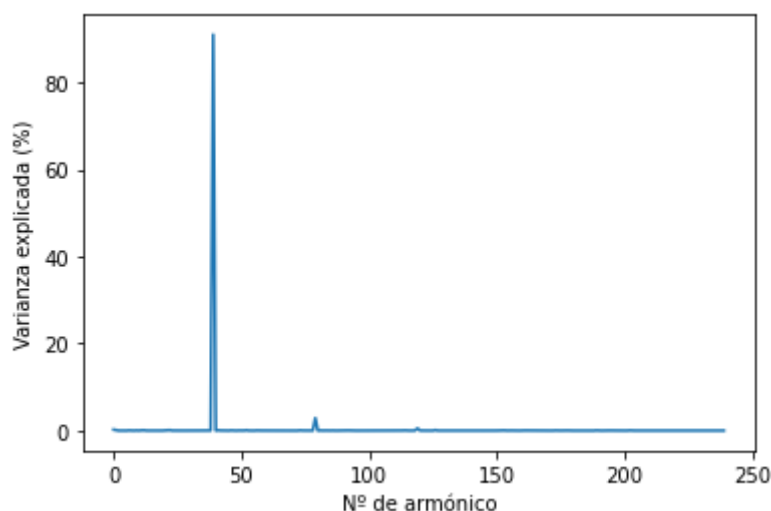


Figura 8: Periodograma de la serie original.

Para identificar que otras periodicidades están presentes en la serie, se procedió a remover la onda anual a partir de la construcción de la onda teórica (sin evaluar la significancia aún). A la serie con la onda anual filtrada se le realizó nuevamente el análisis armónico. En la Figura 9 se presenta el periodograma para la serie filtrada. Se observa una periodicidad que explica el 32% de varianza asociada al armónico 80 que se corresponde con un período de 6 meses. También se observa una periodicidad que explica una varianza mucho menor, del orden del 5%, asociada con el armónico 12 que se corresponde con un período de 4 meses.

La presencia de la onda semi-anual podría deberse a que la onda anual “real” difiere de la onda anual teórica que se reconstruyó con el análisis armónico, y por lo tanto podrían quedar rastros de este ciclo.

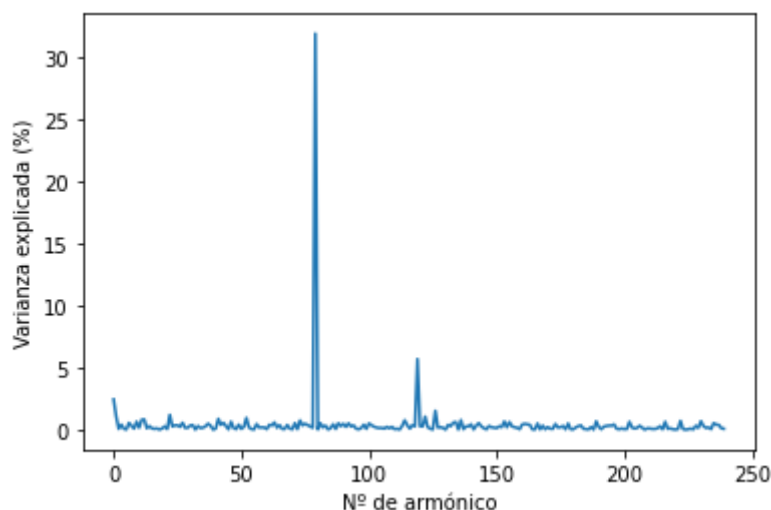


Figura 9: Periodograma de la serie sin la onda anual.

Para evaluar la significancia de las periodicidades observadas en el análisis armónico se procede a realizar un análisis espectral. Se consideró la ventana de Hann y un 95% de nivel de confianza. Además se determinó que el espectro teórico sigue un modelo de ruido rojo. En la Figura 10 se grafican el espectro empírico, el teórico y las bandas de significancia calculadas.

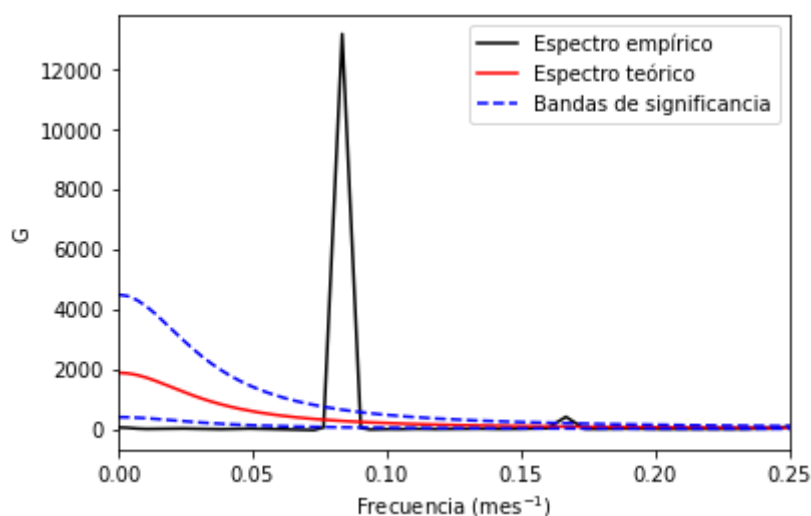


Figura 10: Análisis espectral de la serie original. Se observa el espectro empírico de la señal, el espectro teórico considerando un modelo de ruido rojo y las bandas de significancia.

Teniendo en cuenta la relación entre frecuencia y periodo, las bandas de periodicidades que resultaron significativas (las que se observan por encima de la banda de significancia superior) son 12.52 - 11.52 meses y 6.12 - 5.88 meses. Este resultado es consistente con los dos armónicos que se vio anteriormente que explican la mayor cantidad de varianza: la onda anual (armónico 40) y la semi-anual (armónico 80). Ahora que ya se verificó que ambas son significativas pueden removerse de la serie efectivamente.

Finalmente se utilizó la metodología de Wavelet, considerando la Wavelet de Morlet, para analizar cómo es la variación temporal de las periodicidades de la serie original. En la Figura 11 se presenta el espectro de poder de Wavelet. Solo se debe considerar lo graficado en el interior del cono de influencia, en particular las regiones con mayor potencia que resultaron significativas identificadas con líneas negras. Se puede observar que la banda de periodicidad entre 8 y 16 meses es significativa a lo largo de todo el período de estudio. Este rango de periodicidades está asociado a la onda anual, y es consistente con lo esperado. Por otra parte también hay otras periodicidades significativas pero intermitentes entre 2 y 6 meses, en principio habría que analizar qué proceso físico puede estar presente en esas situaciones o también repetir el análisis considerando otras ventanas para ver si siguen estando presentes y siendo significativas. Aclaración: no se esperaba obtener valores negativos de potencia, pero no se encontró cuál es el motivo en el código.

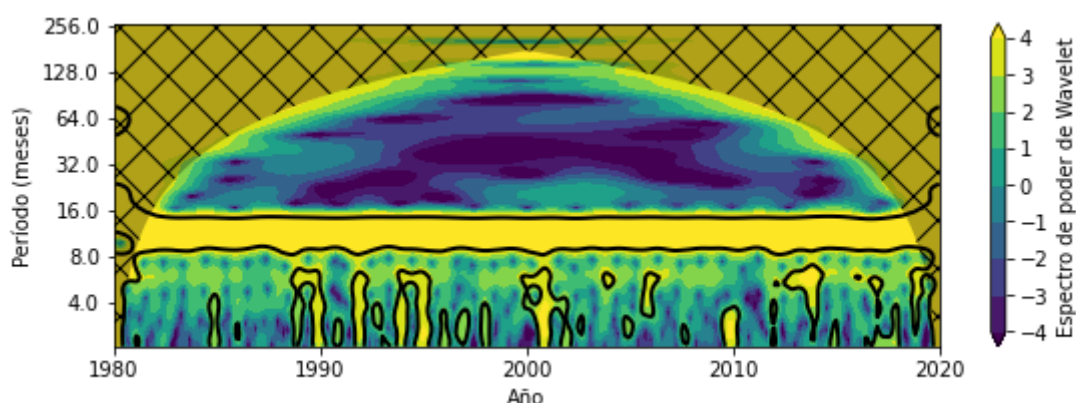


Figura 11: Espectro de Wavelet para la serie original.

Referencias

- [1] <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview>
- [2] De Mendonça, R. y otros (2016). Temperature Effect in Secondary Cosmic Rays (MUONS) Observed at the Ground: Analysis of the Global MUON Detector Network Data. *The Astrophysical Journal*, 830(2), 88.
- [3] Irma Riádigos, I. y otros (2020). Atmospheric Temperature Effect in Secondary Cosmic Rays Observed With a 2 m² Ground-Based tRPC Detector. *Earth and Space Science*, 7, e2020EA001131.
- [4] <https://psl.noaa.gov/data/correlation/soi.data>
- [5] <http://www.esrl.noaa.gov/psd/data/correlation/>