

estacionariedad

Elio

July 20, 2018

Demostración matemática de las propiedades de la estacionariedad

La idea es tener un fundamento matemático robusto para la medida de estacionariedad que estoy usando. La medida es la razón entre el promedio de la amplitud de las ondas 3 mensuales (AM) y la amplitud de la onda 3 del promedio las ondas mensuales (MA). Es decir, si la amplitud de una onda w_i es $A(w_i)$, se tiene que

$$\overline{A(w_i)} = \frac{1}{N} \sum_{i=1}^N A(w_i)$$
$$A(\overline{w_i}) = A\left(\frac{1}{N} \sum_{i=1}^N w_i\right)$$

La primera expresión no tiene demasiada complicación, pero la segunda hay que desarrollarla.

Empecemos con el caso $N = 2$. Se tiene entonces que

$$w_1 = A_1 \cos(k(\phi - \alpha_1)) \quad w_2 = A_2 \cos(k(\phi - \alpha_2))$$

La suma de las ondas será una tercer onda con igual período pero distinta amplitud y fase. [Se puede demostrar](#) que la amplitud A_3 de dicha suma es

$$A(w_1 + w_2) = A_3 = \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos(\alpha_1 - \alpha_2)}$$

En el caso de ondas perfectamente estacionarias se tiene que $\alpha_1 = \alpha_2 = \alpha_i = \alpha_0$ (o equivalentemente, $\alpha_i \neq \alpha_0 \rightarrow A_i = 0$) de manera que $A_3 = A_1 + A_2$ y generalizando a la suma en N , se llega a que

$$A\left(\frac{1}{N} \sum_{i=1}^N w_i\right) = \frac{1}{N} \sum_{i=1}^N A(w_i)$$
$$A(\overline{w_i}) = \overline{A(w_i)}$$

Es decir, para ondas perfectamente estacionarias, $AM/MA = 1$.

Para ondas con fase completamente aleatoria... ???

ok, acá se me acaba la matemática pura y viene la parte empírica. Para distintos valores de N , calculo la amplitud de la suma de N ondas con amplitud constante (2, en este caso) y fase aleatoria ($\alpha \sim U(-\pi, \pi)$). Repito eso 1000 veces y calculo el promedio. Esto me da una estimación de la esperanza matemática de $A(\overline{w_i})$ para distintos tamaños muestrales.

Se puede ver en la Figura 1 que $\lim_{N \rightarrow \infty} A(\overline{w_i}) = 0$ y que va como $\sim N^{-1/2}$ aunque con una constante multiplicativa que en este caso es $k = 1.77$. La parte más empírica viene ahora, porque *jugando* con este y otros casos, se puede ver que esa contante es $k = \frac{\sqrt{A}}{2} \sqrt{\frac{\pi}{N}}$. De manera que, *empíricamente* se puede ver que en el caso de ondas con fase totalmente aleatorias

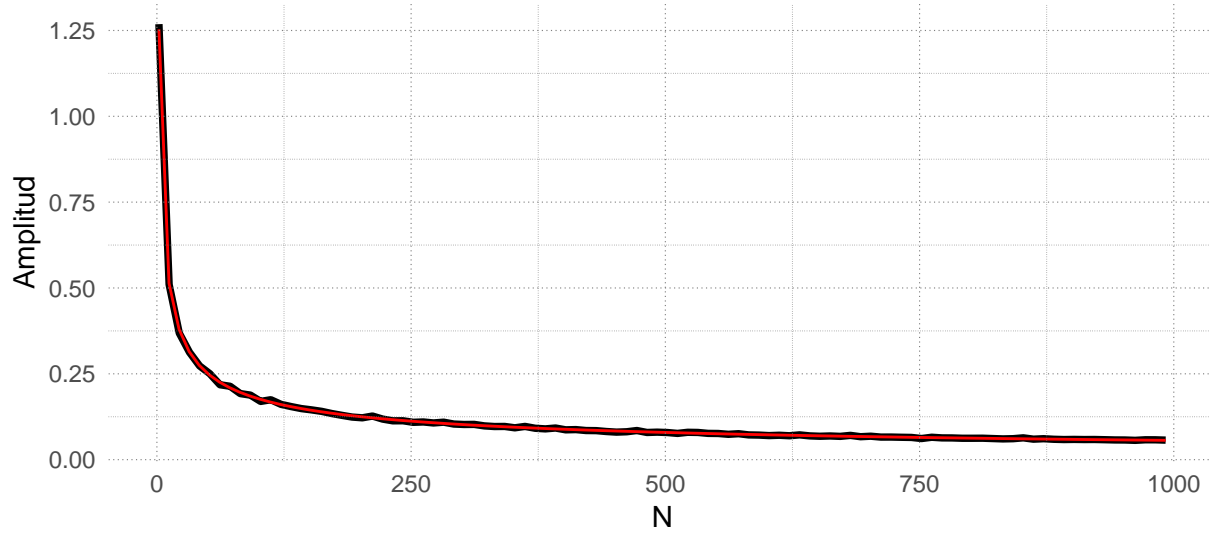


Figure 1: Amplitud del promedio de N ondas con amplitud = 2. La línea roja es la línea $y = \sqrt{\frac{\pi}{x}}$.

$$A(\overline{w_i}) = \overline{A(w_i)} \frac{1}{2} \sqrt{\frac{\pi}{N}}$$

Poniendo todo en limpio, se demostró (con una mezcla de teoría y práctica), que si se define $S = \frac{A(\overline{w_i})}{\overline{A(w_i)}}$ como medida de estacionariedad, se tiene que

$$\frac{1}{2} \sqrt{\frac{\pi}{N}} \leq S \leq 1$$

y que las igualdades izquierda y derecha valen para el caso de pura inestacionariedad y pura estacionariedad respectivamente.

Otras posibles medidas de estacionariedad

Desvío estándar de la fase

Una forma intuitiva de medir la estacionariedad es simplemente ver la variabilidad en la fase. Una onda perfectamente estacionaria debería tener nula variabilidad en la fase mientras que una onda totalmente “antiestacionaria” tendría variabilidad ¿infinita?

El problema de esta medida es que no tiene en cuenta lo que sucede con la amplitud. Puede darse el caso de fase con variabilidad infinita pero donde la amplitud es función de la fase. Es decir, que la onda tenga amplitud máxima para una fase determinada. En ese caso, es válido hablar de que la onda tiene un nivel de estacionariedad y se vería así en los promedios.

Correlación entre fase y amplitud

Siguiendo el razonamiento anterior, es posible pensar en utilizar la correlación entre la amplitud y la desviación absoluta entre la fase y la fase media. Una correlación negativa sería entonces indicio de estacionariedad ya que indicaría que si la onda está lejos de la fase media, la amplitud debería ser pequeña.

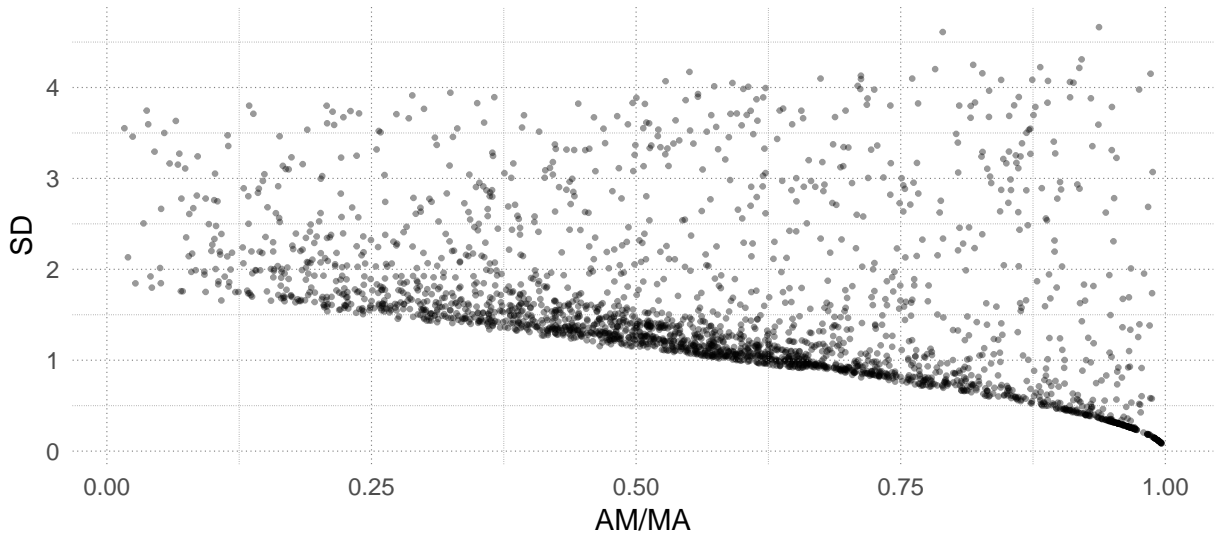


Figure 2: Relación entre SD y AM/MA

Es una medida que funciona relativamente bien (basado en probarlo en datos reales) pero es bastante más ruidosa. Además, al usar la correlación se elimina el efecto de la variabilidad en la fase; dos grupos de ondas pueden tener la misma correlación entre fase y amplitud pero si uno tiene mayor variabilidad en la fase, implicaría menor estacionariedad.

Por otro lado, es difícil interpretar correlaciones positivas.

Covarianza entre fase y amplitud

Una alternativa a la medida anterior. En principio tiene el mismo problema que la correlación, pero además se mezcla la magnitud de la variabilidad de la amplitud. Regiones con pequeña varianza en la amplitud de las ondas (como en los trópicos) van a tener valores pequeños de covarianza aún cuando las ondas sean perfectamente estacionarias.

Desvío estándar de la fase pesado por la amplitud

La primer alternativa tiene el problema que no tiene en cuenta lo que sucede con la amplitud en relación a la fase. Una posible forma de corregirlo sería la de pesar la fase por la amplitud. De esta manera, una onda con fase muy distinta a la media pero con poca amplitud, contribuye poco a la variabilidad.

Por lo que estuve viendo, esta métrica sirve bastante bien aunque tiene un problema en que no es una medida necesariamente lineal, como se muestra en la Figura 2. Además, si bien una onda perfectamente estacionaria tendría un sd nulo, una no estacionaria alcanzaría un máximo que no conozco.

Tendría que ver qué propiedades tiene.

Correlación media entre la onda 3 mensual y la onda 3 media

Otra medida posible es la correlación media entre la onda 3 de cada mes y la onda 3 media. Un pequeño detalle es que algunos autores recomiendan primero transformar los valores usando la transformada de Fischer, promediar esos valores y luego hacer la transformación inversa. Para evaluar ambos, llamo r a la correlación media y r_z a la inversa de la correlación media transformada.

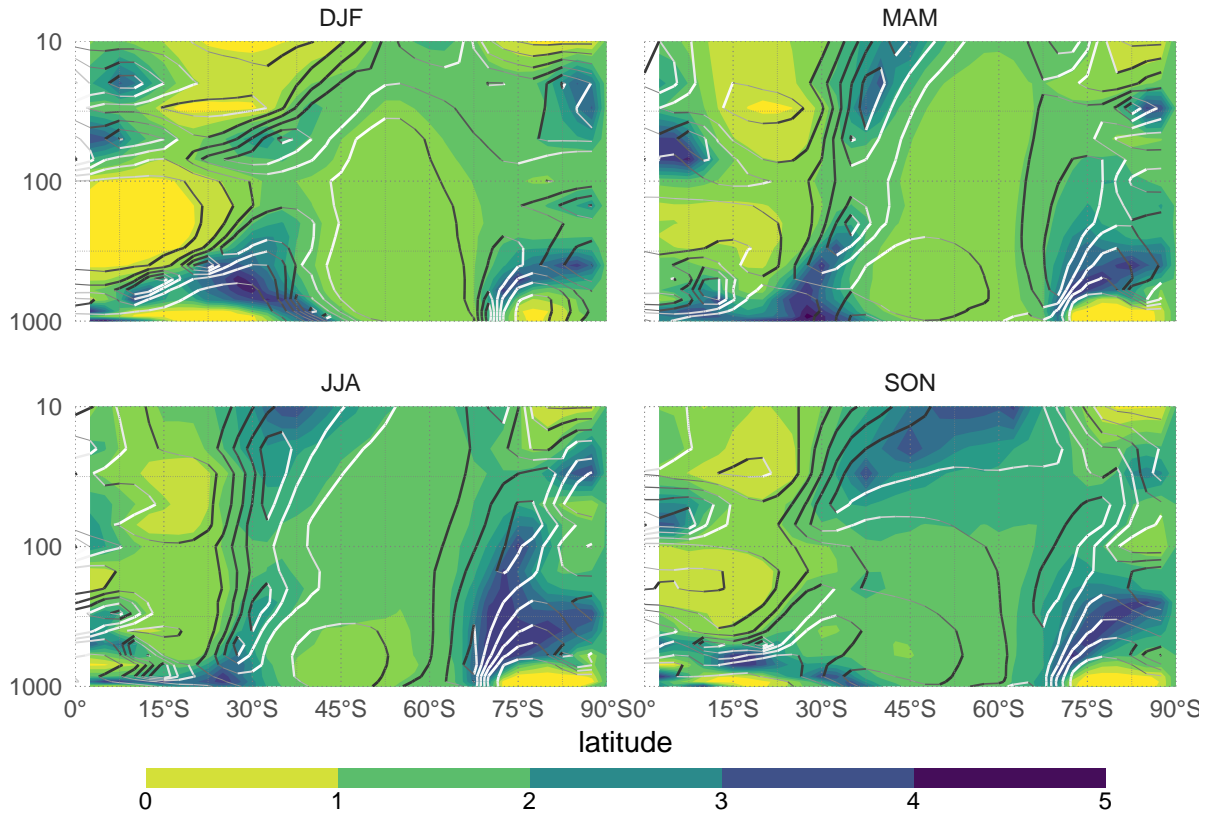


Figure 3: Desvío estándar pesado de la fase en color, amplitud de la media sobre media de la amplitud en contornos

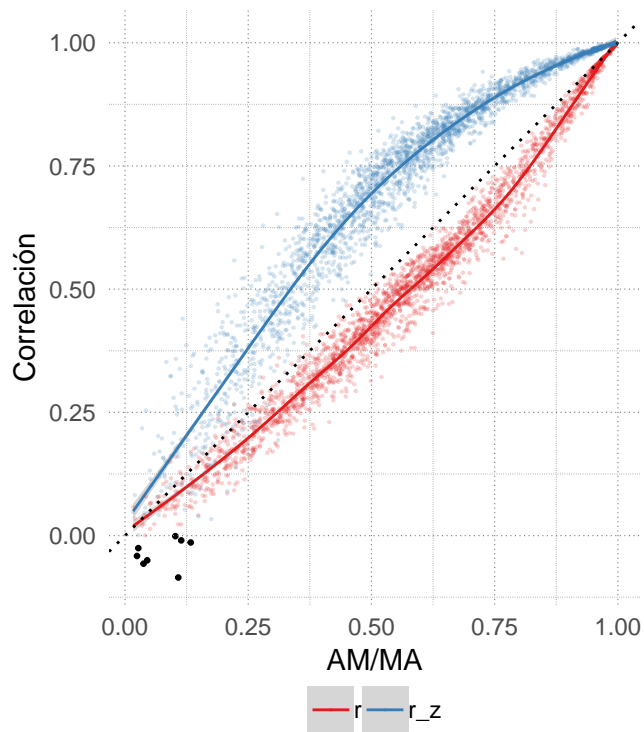


Figure 4: Relación entre la correlación media sin transformar y las otras dos medidas

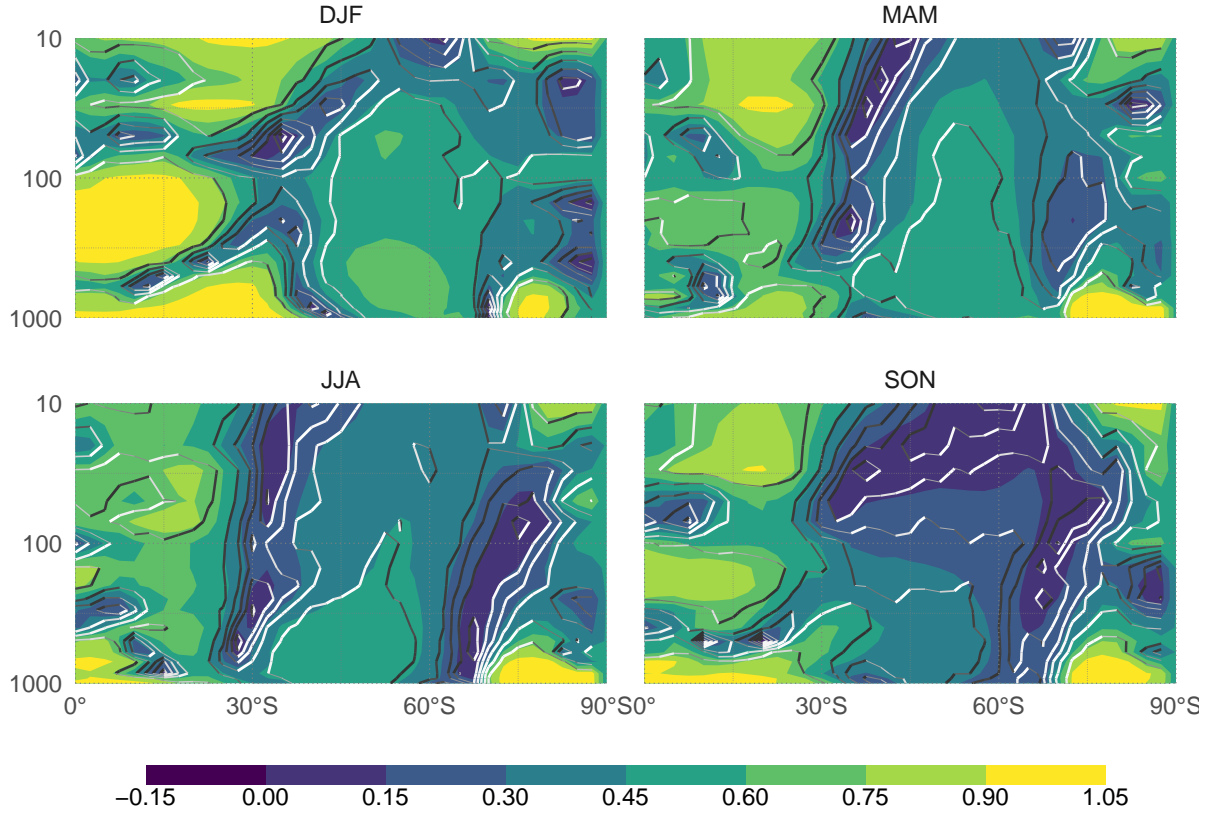


Figure 5: r en color, r_z en contornos

Como se ve en la Figura 4, donde se muestran las dos medidas de correlación en función de AM/MA, r tiene una relación más lineal con AM/MA y además es más similar (más cercana a la línea punteada) que r_z . En comparación a AM/MA, r_z tiende a sobreestimar la estacionariedad mientras que r tiende a subestimarla, aunque en menor medida.

En principio hablar de sub o sobreestimar no es del todo correcto, ya que, en principio, también podría decirse que tanto r como AM/MA sobreestiman la estacionariedad en comparación a r_z . Sin embargo, en la literatura dice que r tiende a subestimar la correlación poblacional mientras que r_z tiende a sobreestimarla, algo que es consistente con lo que se observa en la Figura 4. Sin embargo, la literatura dice que el sesgo de r_z es *menor* que el sesgo de r y esto no se cumple en este caso *asumiendo que AM/MA representa la correlación verdadera*.

Un problema es que, a diferencia de AM/MA que es no negativa por construcción, tanto r como r_z puede tomar valores negativos (marcados con puntos negros en la Figura 4) que no son fáciles de interpretar. La interpretación más consistente, en mi opinión, es que se trata de correlaciones poblacionales nulas o muy pequeñas que son negativas por variabilidad en el sampleo.

Observando los campos (Figura 5), es notable que r_z es más *ruidoso* que r , especialmente en regiones con baja estacionariedad. Supuestamente esto también está reflejado en la literatura, que dice que r_z tiene mayor varianza.

Lo bueno de esta medida es que puede aplicarse a meses individuales y analizar la variabilidad temporal. Por ejemplo, la tendencia lineal (Figura 6 y Figura 7).

Críticas ex post facto

Pensando sobre esto, ¿tiene sentido usar la correlación como medida de estacionariedad a nivel mensual?

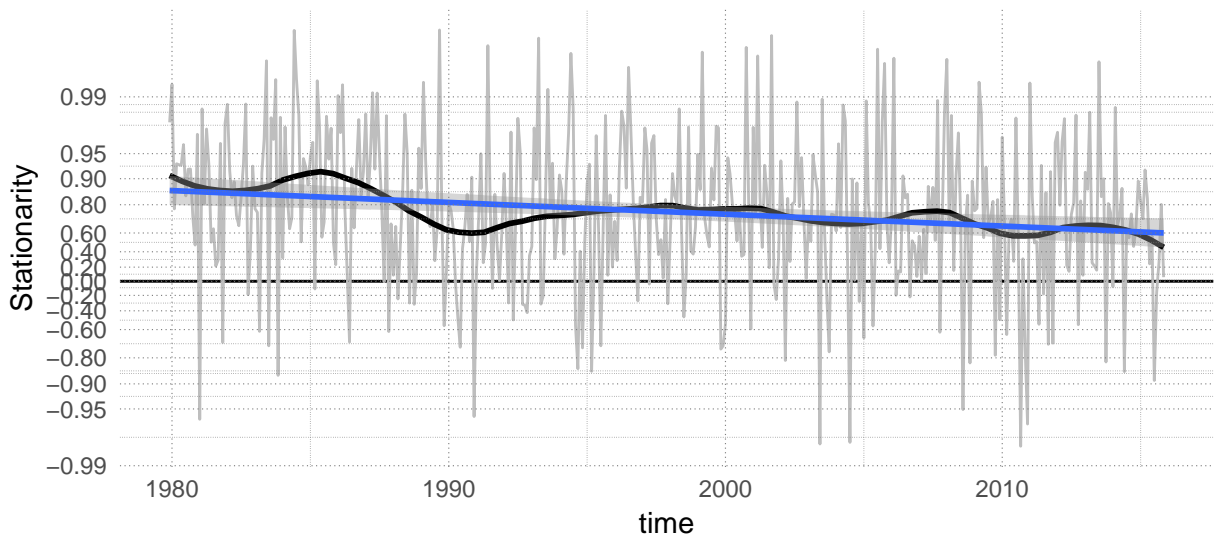


Figure 6: Mean correlation for latitude between 65°S and 40°S and level between 100hPa 700hPa.

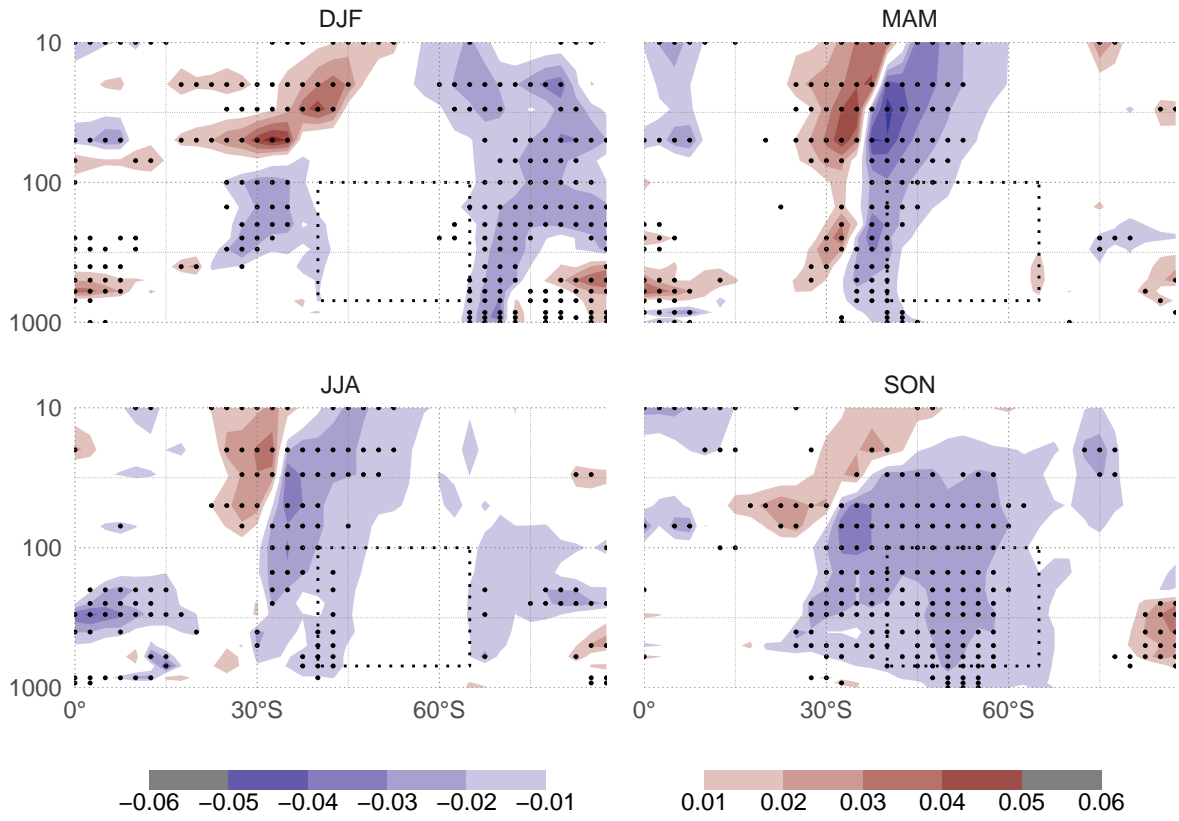


Figure 7: Linear regression estimate for stationarity (correlation per decade).

Lo que estoy haciendo es básicamente comparando qué tan parecida es la onda mensual con la onda media de todo el período, pero no me dice si durante ese mes o los meses cercanos la onda se movió mucho o no. Aún una onda totalmente estacionaria “localmente” (en el tiempo) aparece como poco estacionaria si la onda medio de ese período está ubicada en un lugar distinto al promedio de todo el período general.

Correlación media pesada por la amplitud

Luego de reflexionarlo un poco más, me parece que la correlación media entre las ondas y la onda media tiene problemas conceptuales como medida de estacionariedad. Se puede demostrar matemáticamente que la correlación entre dos ondas con igual frecuencia es idéntica al coseno de la diferencia entre las fases. De ahí que esta medida tampoco esté teniendo en cuenta la relación entre amplitud y fase. Una onda de muy baja amplitud pero corrida 180° en fase con respecto a la onda media tiene la misma magnitud de correlación que una de alta amplitud y completamente en fase con la onda media y, por tanto, pesa lo mismo cuando se promedia.

Una forma de corregir eso es hacer un promedio pesado por la amplitud de la onda. Y ahora viene el chiste: hacer eso es idéntico a calcular AM/MA (queda en deuda la demostración matemática, pero se ve “empíricamente”).

Con esto, creo que se justifica la decisión del AM/MA de forma más sólida y que es un muy buen candidato para la estacionariedad.

Yapa: S a lo largo del tiempo

Concluyo que la estacionariedad no puede estar definida para un mes en particular sino que sólo se define con respecto a un intervalo de tiempo sobre el cual se hace el promedio que se correlaciona con las ondas individuales. Lo más directo es usar todo el período, pero si interesa saber el nivel de estacionariedad en una década determinada, por ejemplo, se debe usar el promedio de esa década. Más aún, la correlación de cada campo mensual debe hacerse con el promedio del período *de ese mes* para que el ciclo anual no aparezca como una no estacionariedad.

El procedimiento para calcular la estacionariedad es calcular la onda media de cada mes durante el período, calcular la correlación¹ de cada onda con su promedio correspondiente y finalmente hacer un promedio pesado por la amplitud de cada onda. Para obtener una serie temporal se aplica este procedimiento con una ventana de tiempo móvil (por ejemplo, de 5 años).

Eso es lo que se muestra en la Figura 8 para 55°S y 200hPa. Es interesante que a pesar de usar definiciones ligeramente distintas y tener los cuidados mencionados en los párrafos anteriores, se consigue un patrón muy similar al de la línea suave de la Figura 6 (que se muestra en gris en la Figura 8). Una tendencia a la baja con mínimos locales en los 90, mediados de 2010 y máximo a principios de los 2000.

¹Es computacionalmente más eficiente y exacto calcular el coseno de la diferencia de fase en vez de la correlación de campos.

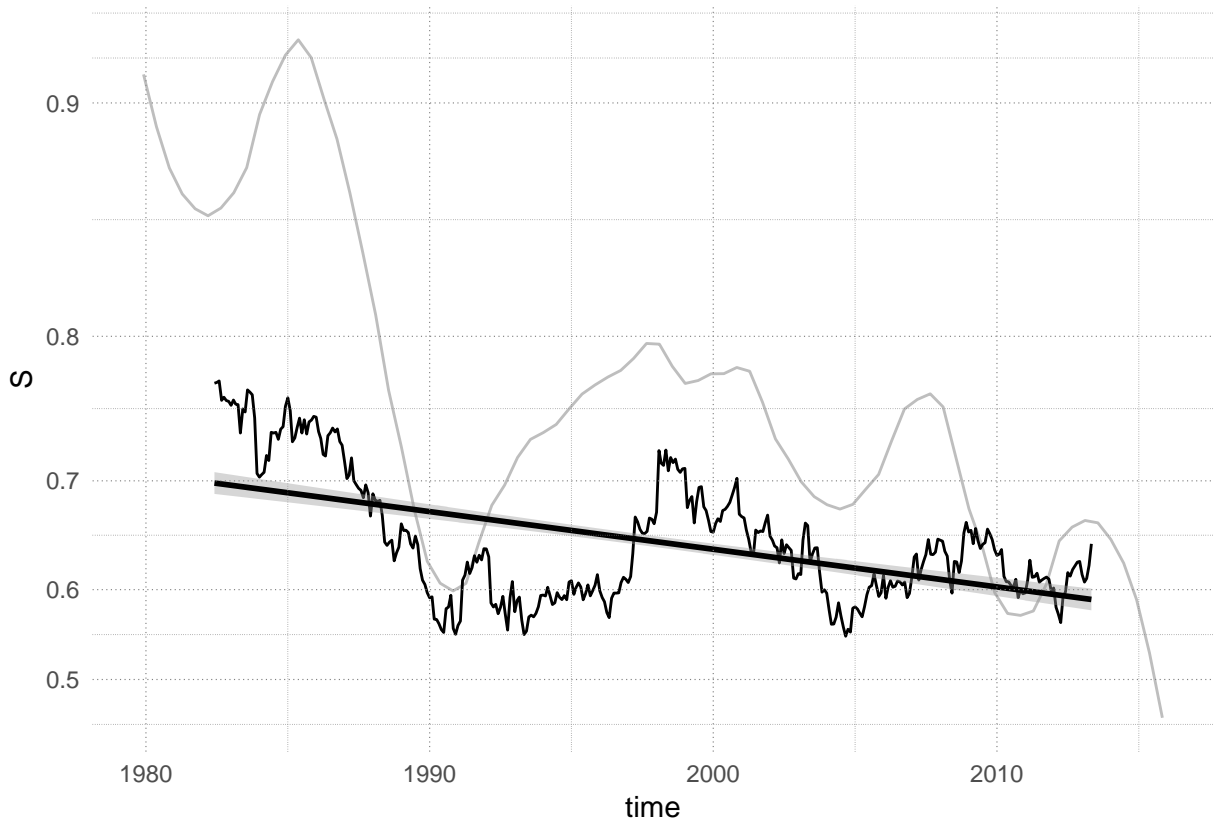


Figure 8: Stationarity (mean wighted correlation) computed in a 5 year rolling window (black) and smoothed correlation (gray).