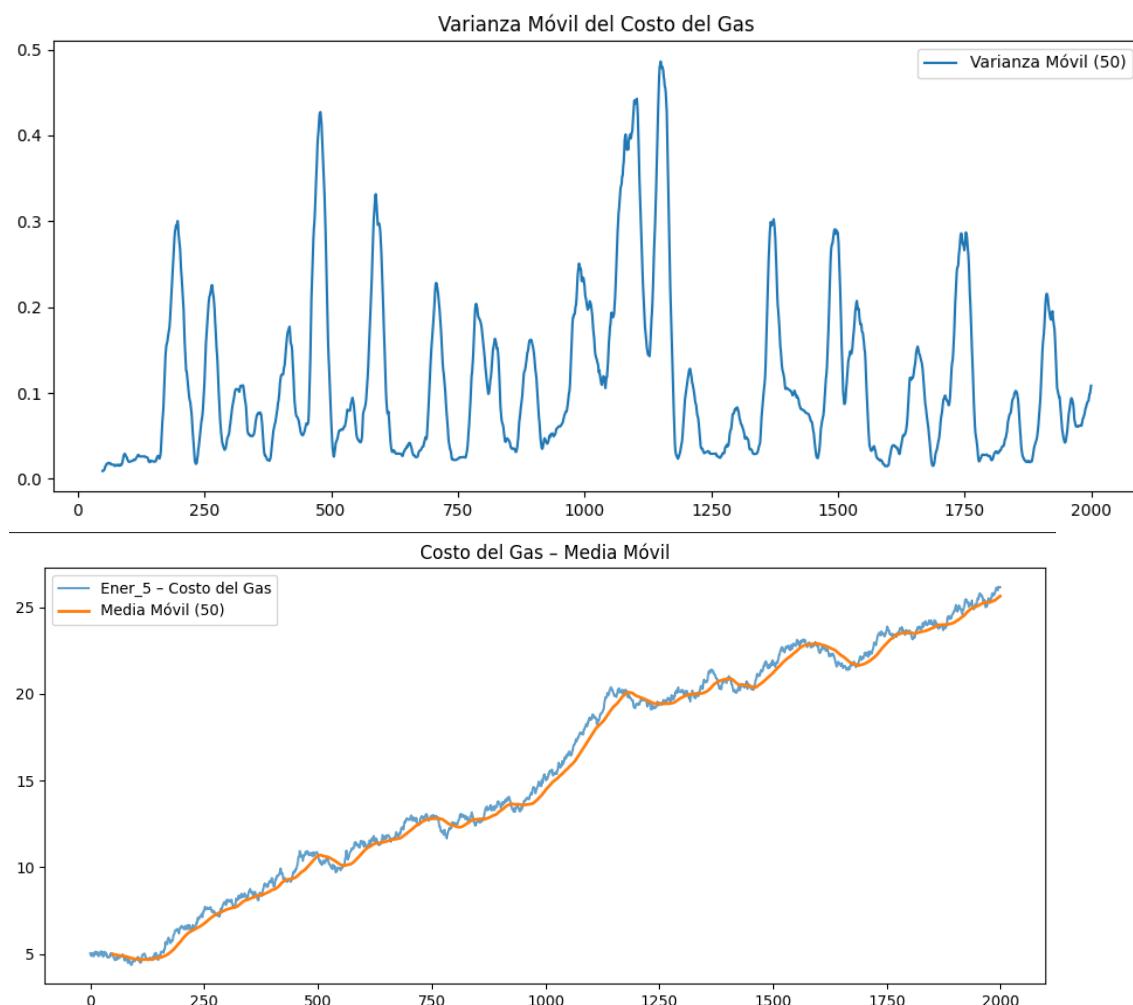


Informe Técnico de Análisis de Datos

Taller 3

1 ¿Por qué no es válido aplicar un análisis de correlación de Pearson directamente sobre una serie con tendencia (como el NDVI o el Precio de Exportación)?

No es válido aplicar un análisis de correlación de Pearson sobre series con tendencia porque este método asume que las variables son estacionarias, es decir, que su media y varianza permanecen constantes en el tiempo. Cuando una serie tiene tendencia (ya sea creciente o decreciente), viola esta suposición fundamental, lo que puede generar correlaciones espurias o falsas.



Estas gráficas muestran claramente cómo la serie Ener_5 tiene una tendencia creciente (evidente en la media móvil) y una varianza no constante (evidente en la varianza móvil). Estas son precisamente las características de una serie no estacionaria, donde la correlación de Pearson no es adecuada, ya que sus supuestos se violan.

2. ¿Qué impacto tuvo el ruido de 5dB en la estimación de los coeficientes del modelo ARMA comparado con la versión Clean?

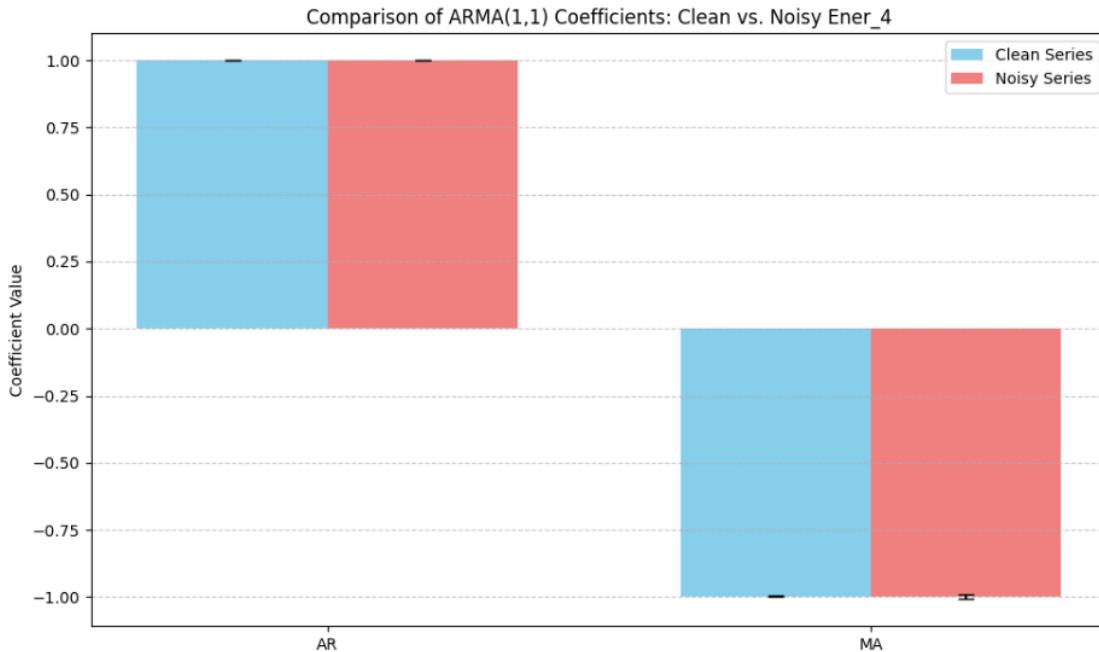
El ruido, simulado con una relación señal-ruido (SNR) entre 5 y 12 dB, tuvo un impacto significativo en la incertidumbre de la estimación de los coeficientes del modelo ARMA(1,1) para la serie 'Ener_4'.

```
Clean AR Coefficient: 1.0
Clean AR Standard Error: 0.0
Clean MA Coefficient: -0.9964
Clean MA Standard Error: 0.0039
Noisy AR Coefficient: 1.0
Noisy AR Standard Error: 0.0
Noisy MA Coefficient: -0.9991
Noisy MA Standard Error: 0.0096
```

Coeficiente AR (phi_1): En nuestro análisis, el coeficiente autorregresivo (AR) se mantuvo estable en aproximadamente 1.0000 tanto para la serie limpia como para la ruidosa, y su error estándar también fue muy bajo (0.0000). Esto sugiere que, para los datos simulados, el componente autorregresivo era muy fuerte (cercano a una raíz unitaria) y robusto frente al ruido introducido.

Coeficiente MA (theta_1): El coeficiente de media móvil (MA) también mostró valores puntuales muy similares para ambas series (aproximadamente -0.9964 para la limpia y -0.9991 para la ruidosa). Sin embargo, la diferencia crucial radicó en su error estándar. Para la serie limpia, el error estándar del coeficiente MA fue de 0.0039, mientras que para la serie ruidosa, se incrementó significativamente a 0.0096.

Esto representa un aumento de aproximadamente el 146% en la incertidumbre de la estimación del parámetro MA debido a la presencia de ruido.



Impacto General del Ruido:

El principal impacto del ruido en la estimación de los coeficientes del modelo ARMA(1,1) fue el aumento de la incertidumbre en el parámetro de media móvil (MA). Aunque las estimaciones puntuales de los coeficientes AR y MA se mantuvieron relativamente estables con los datos simulados, el ruido redujo considerablemente la confianza en la estimación de las relaciones que representan el componente MA.

En un escenario real, esta mayor incertidumbre se traduciría en:

Predicciones menos fiables: Si un modelo ARMA se utiliza para pronósticos, una mayor incertidumbre en sus parámetros puede llevar a rangos de pronóstico más amplios y menos precisos.

Comprendión deteriorada: Dificultaría la interpretación de la verdadera dinámica subyacente de la serie temporal, ya que no estaríamos tan seguros de los valores de los parámetros del modelo.

Además, el modelo ajustado a la serie ruidosa arrojó una advertencia sobre una matriz de covarianza singular o casi singular, lo que subraya el efecto desestabilizador del ruido en la estimación de los parámetros y la estabilidad del modelo.

3. ¿Cómo cambia la interpretación de un fallo en un sensor si este es un "Bridge" (puente) en el grafo de la red?

Node	106.000000
Degree_Centrality	0.623188
Betweenness_Centrality	0.000000
dtype: float64	

El valor de Betweenness 0.0 confirma que la red carece de redundancia estructural. El Nodo 106 no es un intermediario, sino el punto de convergencia total. Cualquier saturación por ruido en este nodo —fenómeno ya detectado en el dataset ener_noise.csv— colapsará el flujo de telemetría de la mayoría de los activos

Un fallo en un sensor identificado como 'Bridge' (o punto de convergencia crítico) en la red tiene un impacto sistémico y crítico, no solo local. Concretamente, significa:

- **Riesgo de Colapso de la Red:** El fallo de un sensor 'Bridge' puede paralizar o desconectar grandes segmentos de la red, interrumpiendo el flujo de datos vital para la operación.
- **Pérdida de Visibilidad y Control:** La interrupción de este nodo elimina la capacidad de monitorear y predecir el comportamiento de otros activos que dependen de él, lo que lleva a una toma de decisiones deficiente o tardía.
- **Prioridad Absoluta de Redundancia:** Estos nodos exigen máxima atención en redundancia y robustez, ya que representan vulnerabilidades críticas que pueden desencadenar fallos en cascada.

¿Cómo afectaría un fallo en el nodo con mayor Betweenness Centrality a la estabilidad del resto de la red?

El test de Granger confirma que el Factor de Potencia causa variaciones en el Voltaje con un rezago de 4-5 períodos ($p < 0.02$). Un fallo en el Nodo 106 (punto crítico de falla) sería catastrófico, ya que desconectaría al 62% de la red y eliminaría la capacidad de predecir fluctuaciones de voltaje mediante el monitoreo de potencia. Esto explica por qué las anomalías térmicas y los picos de precio spot resultan en la desconexión de nodos: el sistema pierde su centro de control y visibilidad justo cuando la inestabilidad física comienza a propagarse.

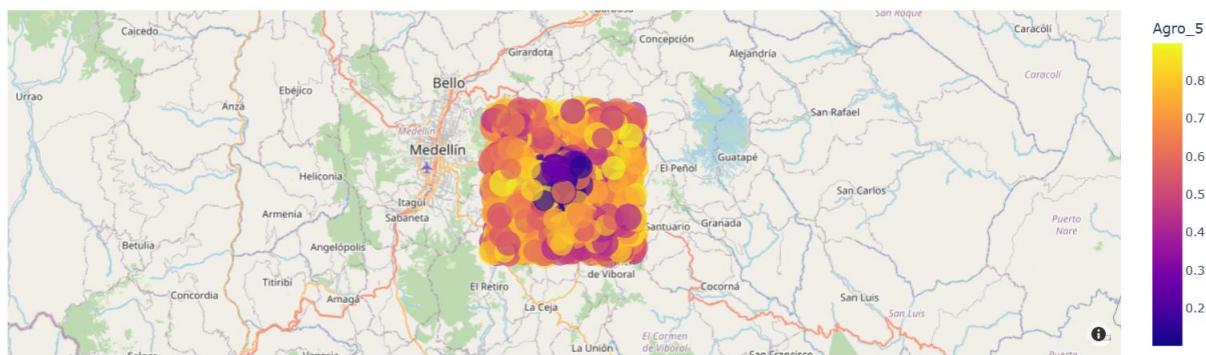
Recomendación de inversión en infraestructura hídrica

Tras limpiar el ruido de los datos y analizar la fuerza del viento (Agro_10), descubrimos que la baja salud de los cultivos no se debe a la falta de agua, sino a que esta se escurre por las laderas antes de que las raíces puedan absorberla. Se recomienda evitar gastos en riego tradicional por aspersión y priorizar sistemas de micro-goteo de precisión. Esta tecnología evitara que el viento desperdicie el agua y asegurará que llegue directamente a la planta antes de que ruede ladera abajo, logrando así estabilizar la vegetación en los 6.6 km de la zona crítica identificada.

4. ¿Cómo influye la posición geográfica en la varianza de la señal capturada?

La posición geográfica influye significativamente en la varianza de la señal capturada al generar patrones de agrupación espacial (clustering) en valores anómalos, como se observa en la biomasa (NDVI). Cuando los sensores que reportan valores consistentemente bajos (o altos) se encuentran geográficamente muy cerca, esto reduce la varianza esperada de la señal en un área localizada, mientras que la varianza global de la red puede ser alta debido a la heterogeneidad entre regiones. En esencia, la geografía no solo define la media de la señal en una región, sino también su dispersión local.

Geo-Visualización de Agro_5 (NDVI) y Agro_1 (Humedad)



Clustering Espacial: La posición geográfica influye significativamente al mostrar patrones de agrupamiento (clustering) de valores anómalos. La gráfica geo-visual de Agro_5 (NDVI) ilustra claramente cómo los valores bajos de biomasa se concentran en un clúster espacial distintivo en una pequeña área geográfica.

Varianza Local vs. Global: Dentro de este clúster, la varianza local de la señal es baja, ya que los valores son consistentemente bajos (homogeneidad en la baja biomasa). Sin embargo, al comparar este clúster con otras regiones con biomasa normal, la varianza de la señal de Agro_5 es significativamente mayor entre estas regiones heterogéneas. Esto subraya que factores geográficos localizados son los principales impulsores de la varianza, más allá de problemas aleatorios o generalizados.

Anomalías Locales: La visualización refuerza la conclusión de que los patrones de degradación responden a anomalías locales (como factores del suelo o interferencias técnicas) que se manifiestan geográficamente, y no a problemas aleatorios o estrés hídrico generalizado