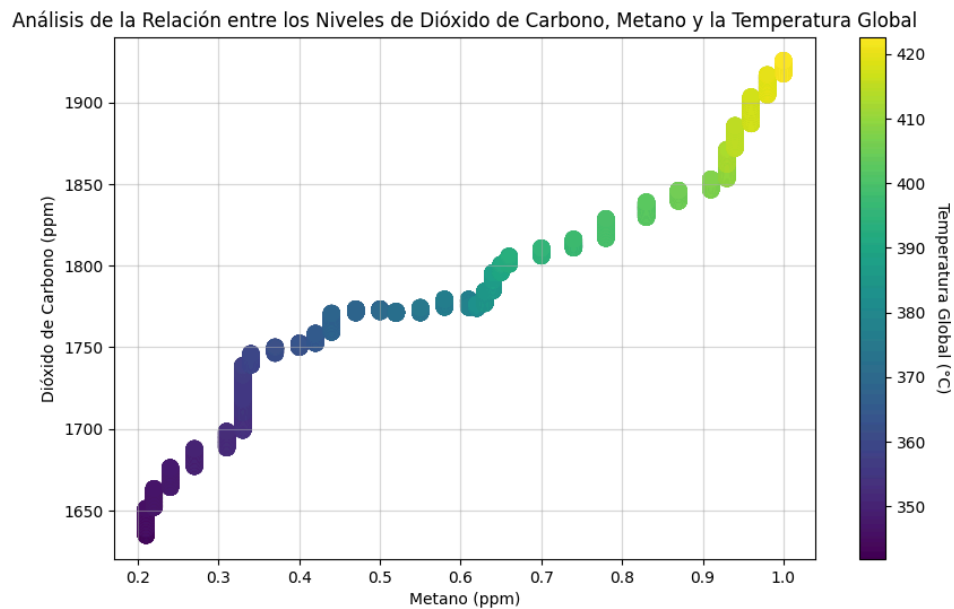


Análisis de Series Temporales de Datos Climáticos: Temperatura Global y Gases de Efecto Invernadero



2025

Cecilia Barozzi

1. Resumen Ejecutivo

El cambio climático es uno de los retos más apremiantes del siglo XXI, impulsado por el incremento de los gases de efecto invernadero (GEI) y sus impactos en el clima global. Este proyecto tiene como objetivo analizar datos históricos sobre Temperatura Global, Metano (CH₄) y Dioxido de Carbono (CO₂), extraídos de la plataforma NASA Climate, para estudiar su comportamiento en el tiempo y las posibles relaciones entre ellos.

A través del análisis de series temporales, se busca:

- Identificar patrones y tendencias a lo largo de los años.
- Detectar anomalías significativas.
- Explorar las correlaciones entre la temperatura global y las concentraciones de CH₄ y CO₂.
- Crear y mejorar modelos de aprendizaje automático para predecir la temperatura global a partir de las variables de los gases CO₂ y metano.

Estos hallazgos no solo permitirán comprender mejor las dinámicas de estos factores, sino también generar evidencia que contribuya a las investigaciones científicas sobre el cambio climático.

2. Metodología

2.1 Descripción de los Datos

Temperatura Global

Se refiere a las anomalías de temperatura promedio del planeta respecto a un período base. Este indicador refleja cambios en el sistema climático global debido al calentamiento global.

Metano (CH₄)

Es un potente gas de efecto invernadero con un impacto en el calentamiento global significativamente mayor que el CO₂ a corto plazo. Se genera principalmente por actividades humanas (agricultura, minería) y procesos naturales.

Dioxido de Carbono (CO₂)

Es el principal gas de efecto invernadero emitido por actividades humanas, como la quema de combustibles fósiles y la deforestación. Su acumulación en la atmósfera es uno de los principales impulsores del cambio climático.

2.2 Fuentes de Datos

Datos de Dioxido de Carbono

- Fuente: Observatorio Mauna Loa y NOAA GML.
- Contacto: Xin Lan (xin.lan@noaa.gov).
- Valores mensuales corregidos para reflejar variaciones estacionales.

Datos de Metano

- Fuente: NOAA GML.
- Reflejan la fracción molar en aire seco.
- Ajustados mediante técnicas estadísticas (Bootstrap y Monte Carlo) para garantizar precisión.

2.3 Relación entre Variables

El CO₂ y el CH₄ son gases de efecto invernadero que contribuyen al calentamiento global al retener el calor en la atmósfera. La temperatura global aumenta como consecuencia directa de la acumulación de estos gases, aunque las interacciones son complejas y están moduladas por factores adicionales como aerosoles, cambios en el albedo terrestre y ciclos naturales climáticos.

2.4 Metas del Proyecto

- Evaluar la calidad y consistencia de los datos climáticos seleccionados.
- Determinar tendencias históricas de las variables (temperatura, CH₄ y CO₂).
- Analizar la relación entre las concentraciones de GEI y la temperatura global.
- Identificar anomalías que puedan asociarse con eventos climáticos específicos.
- Aplicar modelos de aprendizaje automático para la predicción de temperatura global.

3. Orden del análisis

1.Ingesta y Exploración Inicial

Algunos archivos contienen datos combinados en una sola columna, por lo que fue necesario separarlos utilizando delimitadores apropiados para obtener una estructura tabular. Para unificar los datos extraídos de la misma fuente, se utilizó la columna **Year** como referencia, permitiendo alinear las series temporales. Dado que algunos archivos no contenían información desde los mismos años, se recortaron los datos para que todos comenzaran desde el mismo punto temporal.

Pasos realizados:

- Carga de los datasets desde archivos CSV.
 - Revisión del tamaño del dataset (`data.shape`).
 - Exploración de las primeras filas (`data.head()`).
 - Verificación de valores nulos (`data.isnull().sum()`).
 - Identificación de tipos de datos (`data.dtypes`).
 - Revisión y eliminación de valores duplicados.
 - Obtención de estadísticas descriptivas (`data.describe()`).
-

2. Análisis Exploratorio de Datos (EDA)

2.1 Análisis de la Distribución de las Variables

- Visualización de distribuciones (histogramas, gráficos de densidad).
- Cálculo de medidas de tendencia central (media, mediana, moda).
- Evaluación de la dispersión (rango, desviación estándar, varianza).
- Análisis de sesgo y curtosis.
- Identificación de transformaciones necesarias (log, raíz cuadrada, etc.).
- Detección de outliers con boxplots y análisis del rango intercuartil (IQR).

Resultados:

- **Sesgo:** 0.078 (asimetría leve).
- **Curtosis:** -0.30 (distribución ligeramente plana).
- **Transformación recomendada:** Lambda de Box-Cox: 0.179.

2.2 Análisis de Correlaciones y Relaciones entre Variables

- **Correlación de Pearson:** Relación lineal entre variables.
- **Heatmap de correlación:** Visualización de la matriz de correlaciones.
- **Scatter plots:** Exploración de relaciones específicas.

Resultados:

- **Correlación CO₂-Temperatura:** 0.94 (relación positiva fuerte).
- **Correlación Metano-Temperatura:** 0.92 (relación positiva fuerte).

Conclusiones:

- Ambas correlaciones indican una fuerte asociación con la temperatura global.
 - El CO₂ muestra una correlación ligeramente más alta, lo que sugiere un impacto más directo en el calentamiento global a largo plazo.
 - El metano, aunque en menores concentraciones, tiene un alto potencial de calentamiento global.
-

3. Limpieza y Preprocesamiento de Datos

Objetivo: Preparar los datos para el modelado, asegurando calidad y consistencia.

- Manejo de valores nulos.
 - Corrección de outliers.
 - Conversión de tipos de datos.
-

4. Análisis de series temporales

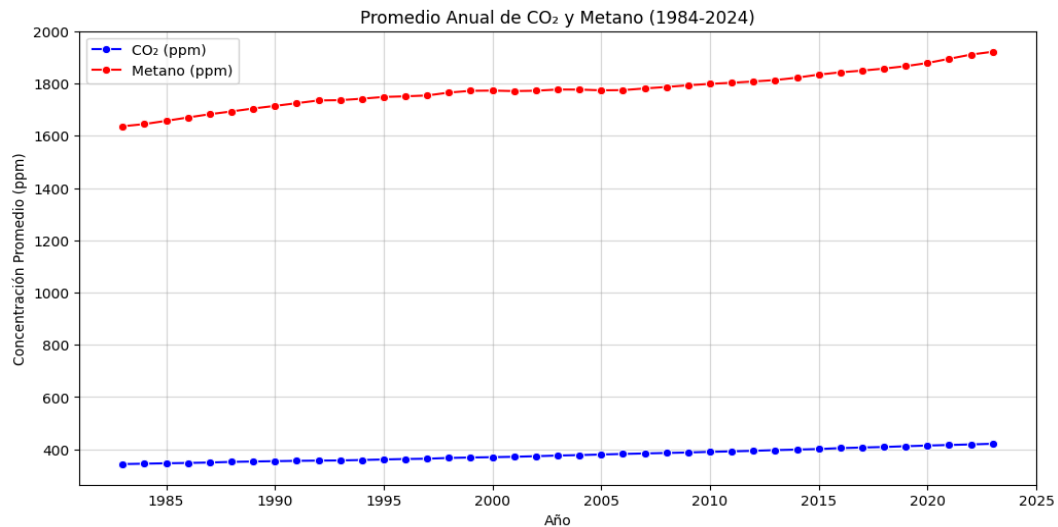
Objetivo: Identificar tendencias, estacionalidad y anomalías.

4.1 Visualización y exploración temporal

Promedio Anual de CO₂ y Metano (1984-2024)

Explorar cómo han cambiado las concentraciones de gases y las temperaturas a lo largo de los años. El promedio (media) proporciona una visión general del comportamiento de la variable. Variables:

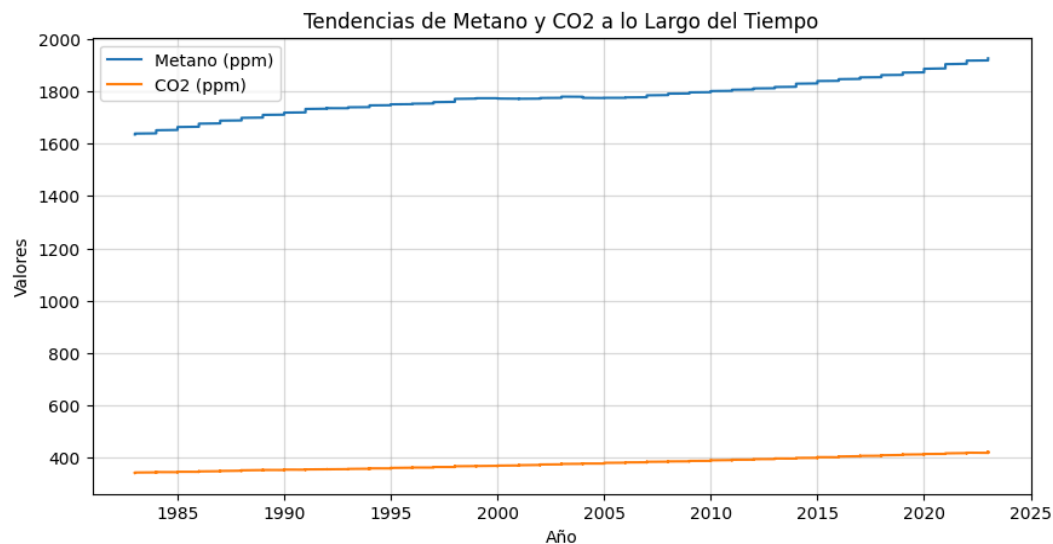
1. average: Promedio mensual de las concentraciones de metano (en partes por millón - ppm) para ese año y mes.
2. monthly average: Promedio mensual de concentraciones de dióxido de carbono (ppm).



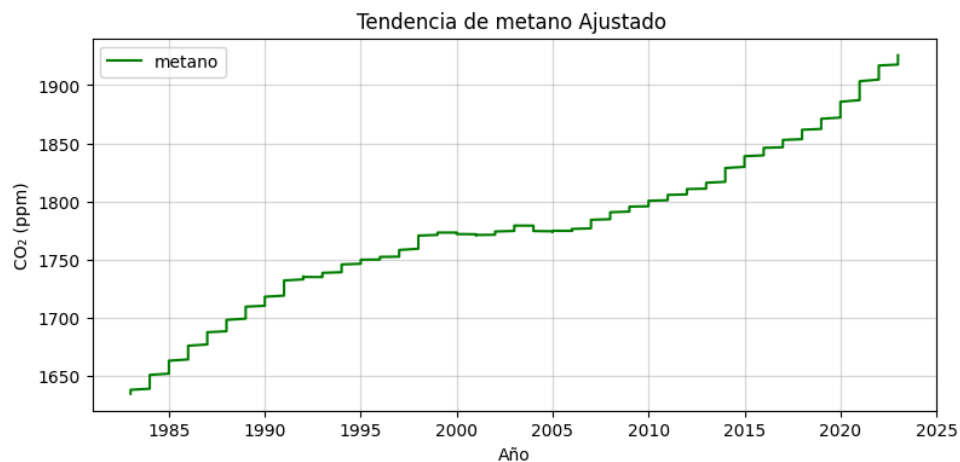
Tendencias de Metano y c02 a lo largo del tiempo

La tendencia muestra la dirección general de los datos a lo largo del tiempo.
variables:

1. trend: Tendencia estimada de las concentraciones de metano después de eliminar los efectos estacionales.
2. de-season alized: Concentraciones ajustadas para eliminar efectos estacionales.

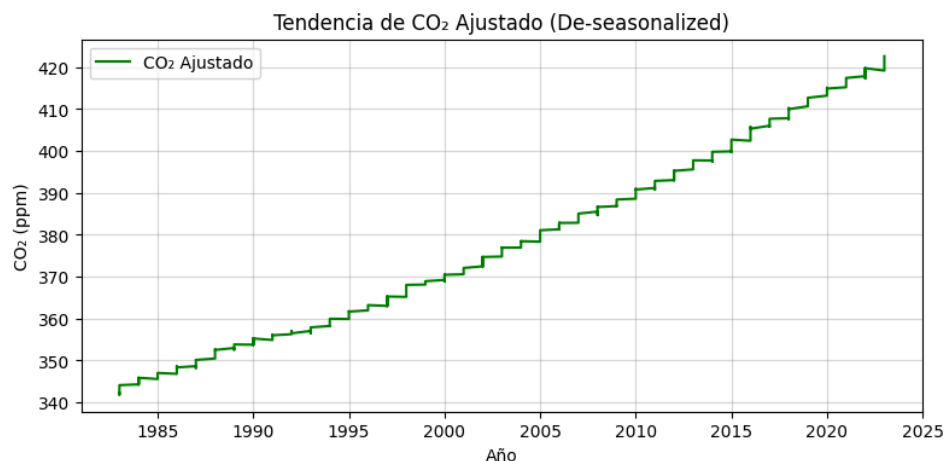


En este gráfico, se comparan las tendencias de metano, CO₂ y los años. La escala de valores del metano es mucho mayor que la del CO₂, lo que hace que visualmente el CO₂ parezca menos pronunciado. A diferencia del gráfico individual de CO₂ o metano, este muestra las tendencias de varias variables juntas, lo que puede afectar la percepción de sus crecimientos individuales.



Tendencia de Metano Ajustado:

Este gráfico muestra la tendencia del metano en ppm a lo largo de los años. La variable trend utilizada en la gráfica sugiere que es un valor suavizado o ajustado de los datos de metano. Se observa un incremento progresivo con algunas mesetas entre 2000 y 2010, pero luego sigue aumentando hasta 2024. Posiblemente, los valores fueron ajustados mediante un método de suavizado o eliminación de estacionalidad.



Tendencia de CO₂ Ajustado (De-seasonalized):

Aquí se muestra la tendencia del CO₂ en ppm sin la variabilidad estacional. Se nota un crecimiento sostenido en la concentración de CO₂ desde 1983 hasta 2023. Al eliminar

la estacionalidad, se observa un patrón más claro y continuo de aumento. Tendencias de Metano, CO₂ y Temperatura Global a lo Largo del Tiempo

Análisis de variabilidad

La desviación estándar ayuda a cuantificar la variabilidad de los datos alrededor de la tendencia o el promedio. En este caso, nos indicará qué tan dispersos son los valores de CO₂ y metano año a año.

Interpretación: CO₂: 22.90 ppm → Las concentraciones de CO₂ tienden a fluctuar, en promedio, ± 22.90 ppm con respecto a su media. Metano: 67.53 ppm → Las concentraciones de metano tienen una variabilidad mayor, oscilando ± 67.53 ppm en torno a su media. ♦ Comparación:

El metano (67.53 ppm) tiene más variabilidad que el CO₂ (22.90 ppm), lo que indica que sus valores presentan mayores fluctuaciones a lo largo del tiempo. Puede sugerir que el metano está más influenciado por factores externos o tiene picos más marcados que el CO₂.

4.2 Cuantificación de variabilidad y crecimiento año tras año

Para cuantificar el crecimiento año tras año se calcula la diferencia entre valores consecutivos en la serie temporal de CO₂ y metano. En términos de series temporales, esto ayuda a comprender si hay un patrón de crecimiento sostenido, aceleraciones o posibles fluctuaciones estacionales.

Los valores de dioxido_diff muestran que el CO₂ varía año tras año, con aumentos y disminuciones. El dioxido_growth (crecimiento porcentual) oscila entre valores positivos y negativos, lo que indica que la tasa de crecimiento del CO₂ no es uniforme. Hay valores negativos, lo que sugiere que en ciertos momentos hubo descensos en la concentración de CO₂.

La columna metano_diff es casi toda 0, lo que sugiere que el metano no varía significativamente año tras año en este dataset. El metano_growth también es 0 en todos los registros, indicando que el metano no muestra variaciones apreciables en este análisis.

Posibles interpretaciones:

✓ El CO₂ presenta una tendencia de crecimiento con fluctuaciones, lo que podría indicar factores estacionales, eventos climáticos o cambios en la actividad humana.

✓ El metano, en cambio, parece mantenerse estable en los valores registrados, lo que podría deberse a una medición menos frecuente o a un crecimiento más uniforme

A continuación vamos a graficar la variabilidad anual de CO₂ y metano. Esto lo haremos a partir de la cuantificación

Los valores dioxido_diff y metano_diff nos permitieron ver cómo cambian las concentraciones cada año. Los valores dioxido_growth y metano_growth nos dieron la tasa de crecimiento, permitiendo detectar tendencias generales. Este análisis es fundamental en series temporales, porque ayuda a determinar si hay tendencias de crecimiento o si el comportamiento es más estacional (fluctuante sin una dirección clara).

Haremos dos gráficos:

1Diferencias anuales (diff): Cuánto cambió la concentración de cada gas de un año a otro.

2Crecimiento relativo (growth): Cambio porcentual año tras año.

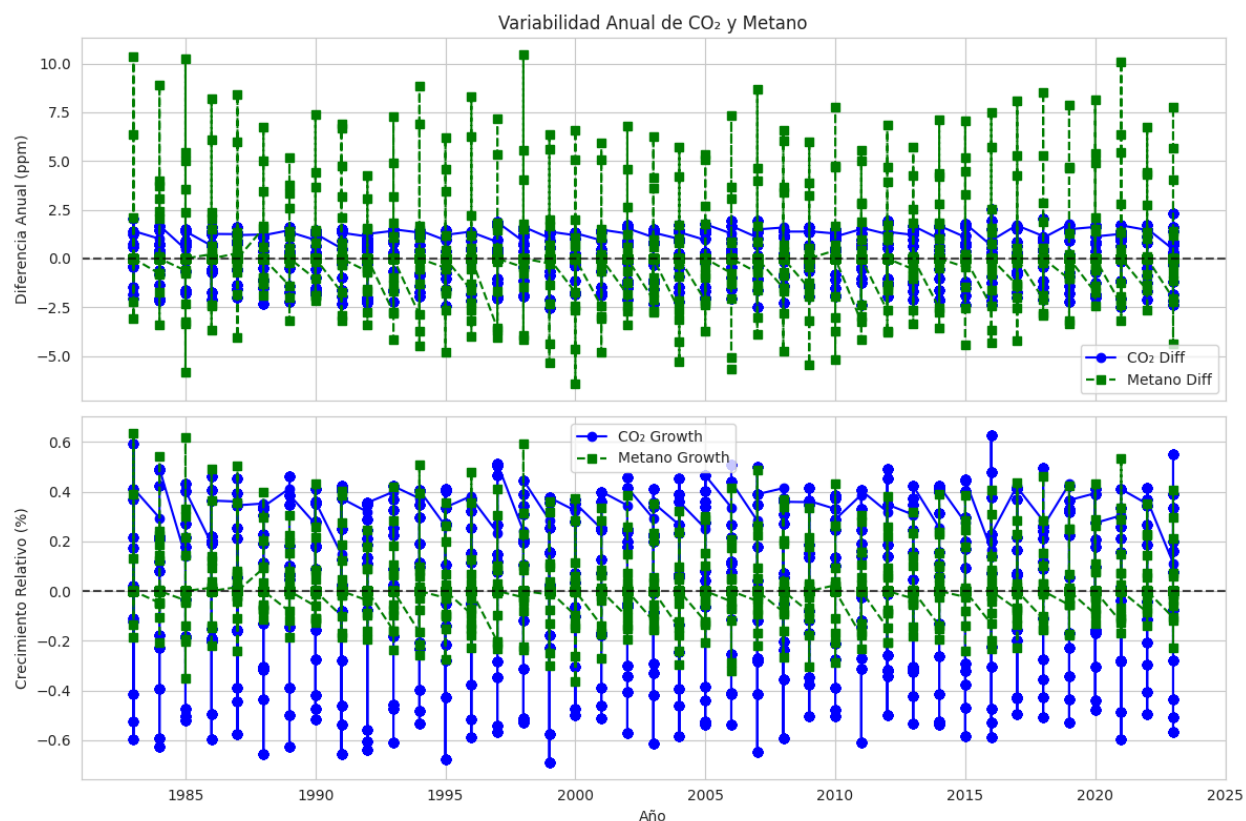


Gráfico superior (Diferencia Anual, diff)

CO₂ (azul): Muestra fluctuaciones moderadas en la variación anual, con valores mayormente positivos. Esto sugiere que el CO₂ ha aumentado cada año, aunque con pequeñas oscilaciones. Metano (verde): Tiene variaciones más extremas, con picos y valles más pronunciados. Esto indica una variabilidad anual mayor que la del CO₂.

Gráfico inferior (Crecimiento Relativo, growth)

CO₂ (azul): Se mantiene en valores positivos en la mayoría de los años, indicando un crecimiento constante en su concentración. Sin embargo, hay algunos años con caídas leves.

Metano (verde): Su crecimiento es más errático, con años de crecimiento y otros de reducción.

Conclusión basada en el gráfico:

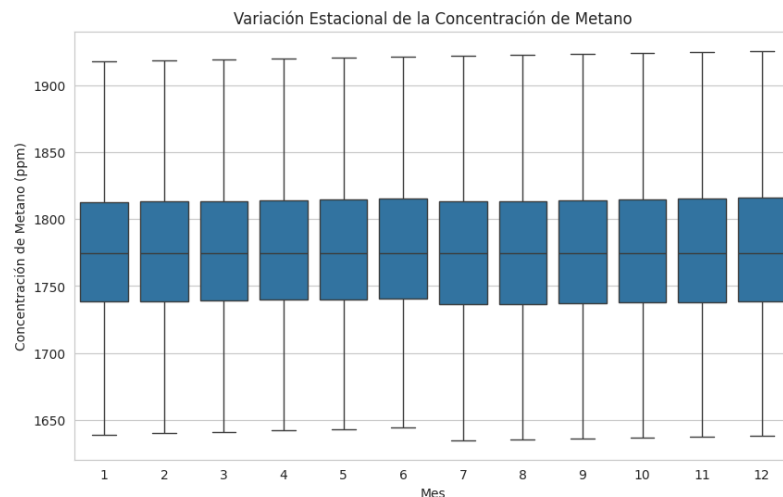
- ♦ El CO₂ muestra una tendencia de crecimiento sostenido, con pequeñas oscilaciones en la variación anual. Su crecimiento relativo rara vez cae por debajo de 0, lo que indica que cada año hay más CO₂ en la atmósfera.
- ♦ El metano es más volátil, con variaciones grandes en diff, lo que sugiere que su concentración fluctúa más de un año a otro. Su crecimiento relativo también tiene valores negativos en varios años, lo que significa que hay períodos donde su concentración ha disminuido.
- ♦ El comportamiento del metano podría estar más influenciado por factores estacionales o eventos específicos (ej. cambios en emisiones agrícolas o industriales).

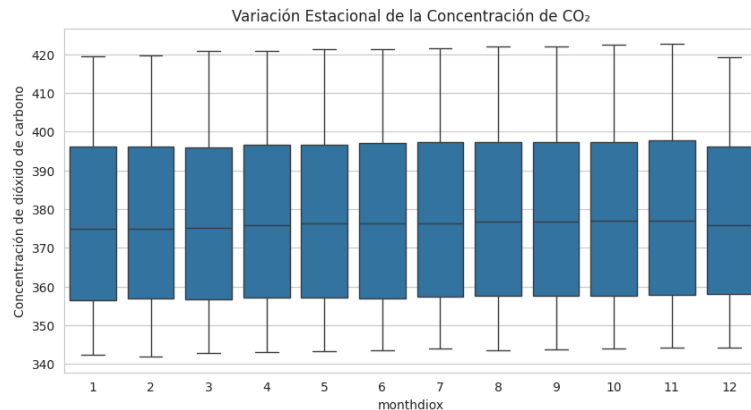
Análisis Estacional metano y c02

Análisis de Variabilidad Estacional (Mensual) de las concentraciones:

Este análisis busca identificar patrones que se repiten anualmente en cada mes.

Variables: tendencia de los datos de c02 y metano





Cada caja muestra la distribución de los datos mensuales con los siguientes elementos:

Mediana (línea en el centro de la caja): Representa el valor central de la distribución mensual.

Rango intercuartílico (caja): Muestra la dispersión de los valores entre el primer y el tercer cuartil (Q1 y Q3).

Bigotes: Indican la extensión de los datos sin considerar valores atípicos.

Valores atípicos (si los hay, como puntos fuera de los bigotes): Representan mediciones inusuales o desviaciones significativas.

Conclusión de los gráficos:

Variabilidad estacional del metano:

La concentración mensual de metano parece mantener una distribución bastante estable a lo largo del año.

No hay cambios estacionales abruptos en la mediana ni en la dispersión de los datos, lo que indica que el metano no tiene una fuerte estacionalidad mensual.

Variabilidad estacional del CO₂:

Similar al metano, la mediana de la concentración de CO₂ es relativamente estable entre los meses. Sin embargo, hay una ligera variabilidad, con algunas diferencias en la dispersión de los datos en ciertos meses.

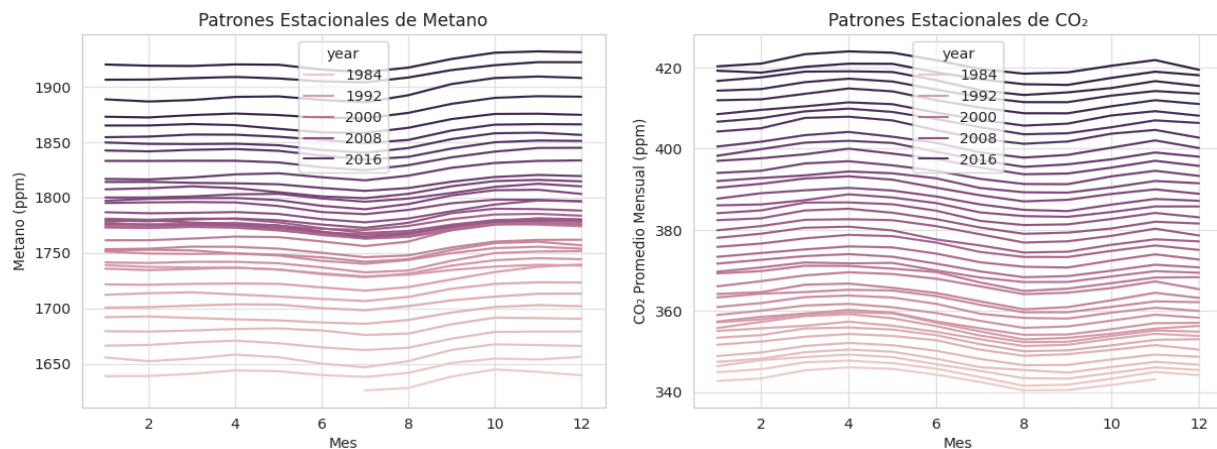
¿Qué significa esto para el análisis estacional?

Si la concentración mensual no muestra fluctuaciones significativas a lo largo del año, podría indicar que no hay una fuerte estacionalidad en los datos.

Para confirmar esto, Se calculan medidas adicionales como la descomposición de series temporales (STL) o aplicar un test de estacionalidad como el de **Dickey-Fuller**. **(más abajo hay mas información)**

Comparación de Estacionalidad entre Décadas

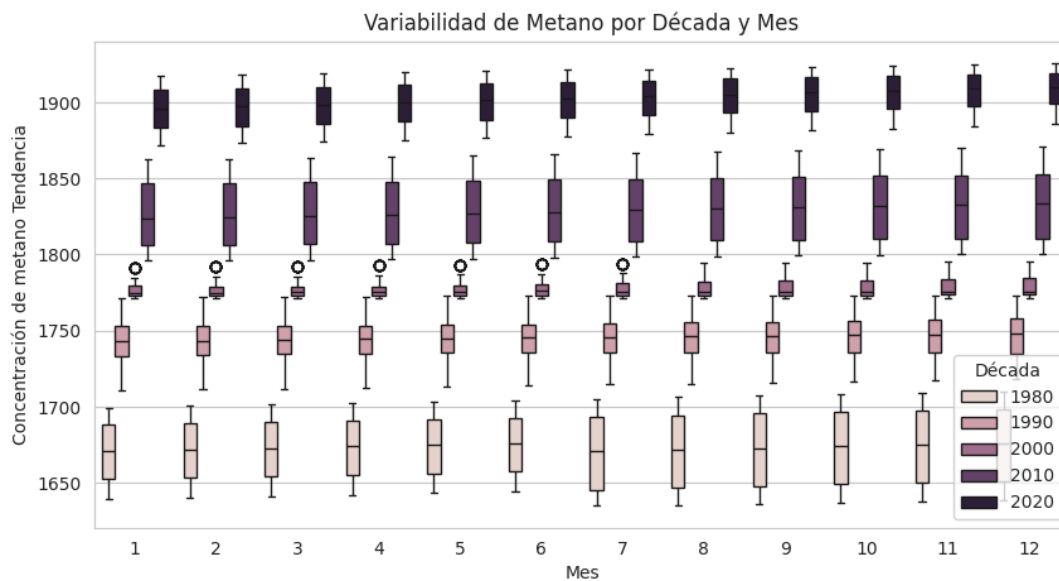
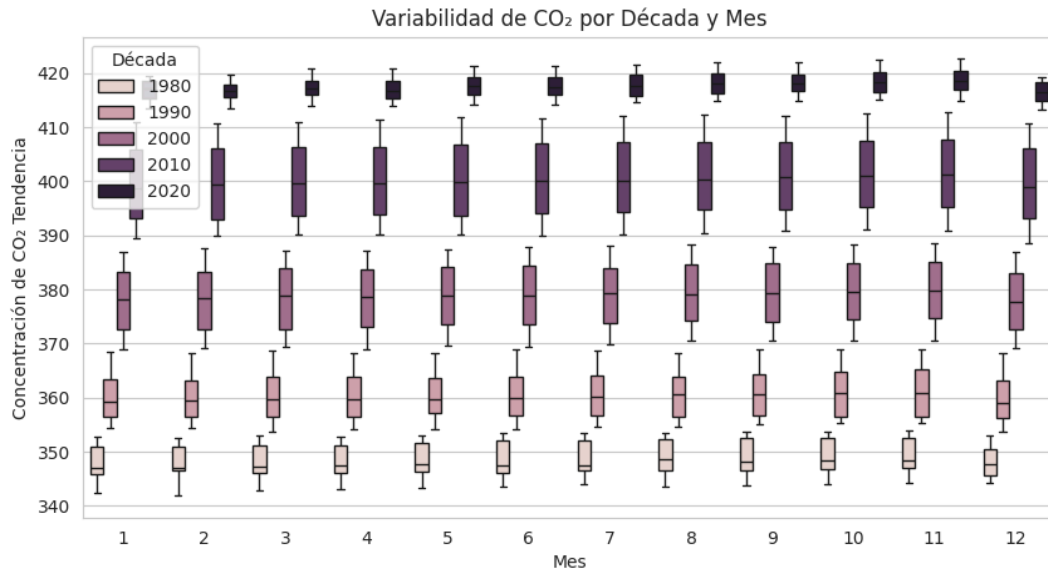
Variables con promedios de metano y c02



Conclusiones del Gráfico de Metano Se observa un patrón estacional en la concentración de metano a lo largo del año, con ciertas fluctuaciones. A medida que avanzan las décadas, los niveles de metano han aumentado progresivamente, lo que sugiere una tendencia creciente similar a la del CO₂. Se nota una variabilidad mensual en los últimos 40 años, con una suave caída a mediados de año y un incremento en los últimos meses. La diferencia entre las curvas de cada década indica que el metano ha crecido a un ritmo sostenido, aunque la variabilidad en los ciclos estacionales se ha mantenido relativamente estable.

Patrones Estacionales de CO₂ Se identifica un patrón estacional repetitivo donde los valores de CO₂ fluctúan a lo largo del año, con picos en ciertos meses y descensos en otros. A pesar de la estacionalidad, hay una tendencia general al alza, con los niveles de CO₂ aumentando progresivamente a lo largo de los años. El comportamiento estacional sugiere que el CO₂ tiene ciclos anuales predecibles, pero con una superposición de una tendencia de crecimiento.

Análisis de los Gráficos de CO₂ y Metano por Década y Mes



Ambos gráficos representan la variabilidad estacional del CO₂ y el Metano en diferentes décadas. La concentración de CO₂ y Metano ha aumentado en cada década, mostrando una tendencia creciente.

¿Cómo ha evolucionado la concentración de CO₂ y Metano?

Ambas concentraciones han aumentado de manera continua en las últimas décadas.

¿En qué décadas ha habido un mayor aumento?

En CO₂, el incremento es notable entre 2000 y 2020. En Metano, el aumento más pronunciado se da a partir de 2000.

¿Se han desplazado los picos estacionales?

No se observa un cambio significativo en los picos estacionales, pero la variabilidad mensual se mantiene. Comparación y conclusión CO₂: Su concentración ha aumentado constantemente,

con una variabilidad estacional que persiste a lo largo de las décadas. Metano: Ha seguido una tendencia similar, con un fuerte incremento en las últimas dos décadas. Conclusión general: El aumento sostenido de ambos gases indica un claro impacto ambiental, reflejando cambios en la emisión y acumulación atmosférica.

Correlación entre gases y temperatura:

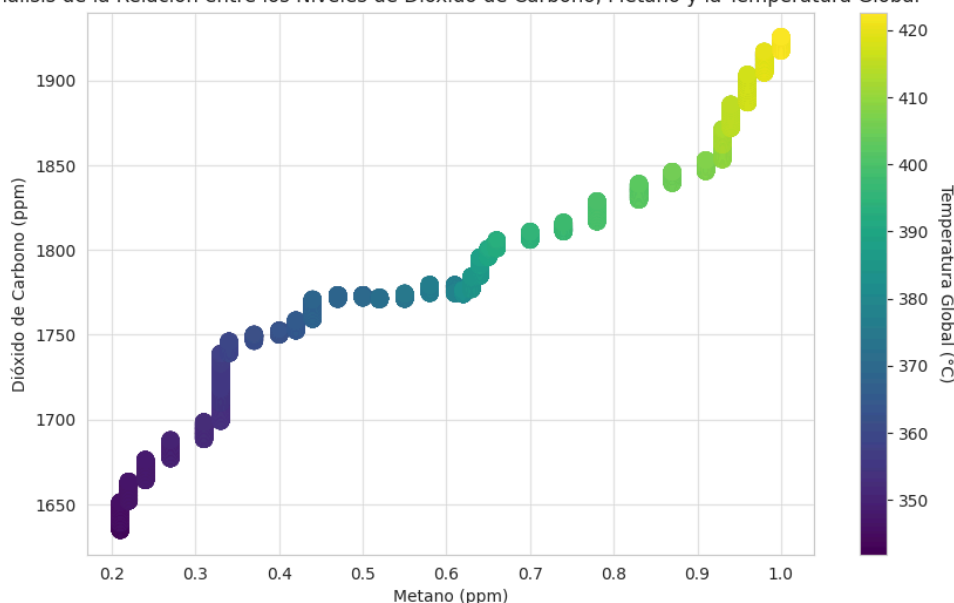
Para la correlación entre gases y temperatura, es más óptimo analizar la tendencia en lugar del promedio. Esto te permitirá ver cómo evolucionan conjuntamente a lo largo del tiempo y si existe una relación consistente.

Analizar cómo el aumento en las concentraciones de metano y CO2 se correlaciona con el incremento en las temperaturas globales. Usando las variables: lowess(5), trend (metano) y de-season alized (CO2).

Relacionar las concentraciones de metano (trend) y dióxido de carbono (de-season alized) con la temperatura global (lowess(5)). Es intuitivo porque metano y CO2 son variables que influyen en el calentamiento global, y el color (representando CO2) añade información útil

Este gráfico tiene sentido porque representa la interacción entre los tres elementos clave en tu análisis (temperatura, metano y CO2). Esto te ayuda a observar si un aumento en metano y CO2 se correlaciona con temperaturas más altas

Análisis de la Relación entre los Niveles de Dióxido de Carbono, Metano y la Temperatura Global



Análisis de la Relación entre los Niveles de Dióxido de Carbono, Metano y la Temperatura Global

Eje X (horizontal): Niveles de metano (en partes por millón, ppm). Eje Y (vertical): Niveles de dióxido de carbono (en ppm). Color de los puntos: Representa la temperatura global (°C). La

barra de color a la derecha indica la correspondencia entre el color y la temperatura.
Interpretación de la Gráfica

Relación entre Metano y Dióxido de Carbono: La gráfica muestra una tendencia general de que a medida que aumentan los niveles de metano, también aumentan los niveles de dióxido de carbono. Esto sugiere una correlación positiva entre ambos gases de efecto invernadero.
Influencia de la Temperatura: El color de los puntos revela que las temperaturas más altas (amarillo) se asocian con niveles más altos de metano y dióxido de carbono. Esto sugiere que la temperatura global está influenciada por las concentraciones de estos gases.

Conclusión

La gráfica sugiere que existe una relación compleja entre el metano, el dióxido de carbono y la temperatura global. El aumento de los niveles de metano parece estar relacionado con el aumento de los niveles de dióxido de carbono, y ambos están asociados con un aumento de la temperatura global.

Consideraciones Adicionales

Causalidad: La gráfica muestra correlación, pero no necesariamente causalidad. Podría haber otros factores que influyen en estas relaciones.

Estos datos respaldan la idea de que los gases de efecto invernadero como el metano y el dióxido de carbono juegan un papel importante en el calentamiento global. La reducción de las emisiones de estos gases podría ser una estrategia efectiva para combatir el cambio climático. En resumen, la gráfica proporciona evidencia visual de una conexión entre los niveles de metano, dióxido de carbono y la temperatura global. Sin embargo, es importante considerar otros factores y realizar investigaciones adicionales para comprender completamente estas relaciones.

Estacionalidad y Estacionariedad en los datos

- ✓ Prueba de Dickey-Fuller: Ver si los datos son estacionarios.
- ✓ Gráficos ACF y PACF: Identificar correlaciones temporales.

Breve explicación :

Estacionalidad: Patrones que se repiten en intervalos regulares (ej. picos anuales). ♦ Si hay estacionalidad: Diferenciación estacional o modelos que la capturen (SARIMA). ♦ Si no hay estacionalidad: No es necesario modificar los datos.

Estacionariedad: La media y varianza no cambian con el tiempo (sin tendencia ni ciclos). ♦ Si no es estacionaria: Diferenciación o transformaciones (log, Box-Cox) si el modelo lo requiere.
♦ Si es estacionaria: No se necesitan cambios.

- ♦ ¿Solo para ML? Sí, la transformación depende del modelo (ej. ARIMA la necesita, Random Forest no).

Prueba de Dickey Fuller, su objetivo es ver si la serie es estacionaria, es decir, si la media, la varianza y la covarianza son constantes en el tiempo.

Si el p-valor es menor a 0.05, la serie es estacionaria y puede usarse directamente en Machine Learning. Si el p-valor es mayor a 0.05, la serie no es estacionaria y necesita transformaciones. como la Diferenciación: Restar el valor anterior del valor actual para eliminar tendencia.

Transformación logarítmica (si la varianza crece con el tiempo). Box-Cox: Para estabilizar la varianza cuando hay heterocedasticidad.

Resultado: Valor p: 0.994350501846125. La serie no es estacionaria (no rechazamos H_0)

Sobre el Sesgo

En este caso el resultado: El sesgo de 0.078 es muy cercano a 0, lo que indica que la distribución de las variables es prácticamente simétrica.

No es necesario hacer transformaciones (log, Box-Cox, etc.), ya que el sesgo es bajo.

4. Modelados

El punto de análisis principal es entender la relación entre los niveles de CO_2 , metano y la temperatura global, y usar esta relación para predecir cambios futuros en la temperatura. Para ello, debes analizar tendencias, correlaciones, causalidad, y construir modelos predictivos. Además, es importante evaluar la distribución de los residuos y el impacto relativo de cada gas.

En primer lugar se optó por el modelo de regresión lineal donde los resultados de las métricas eran buenas pero con otros inconvenientes provenientes de:

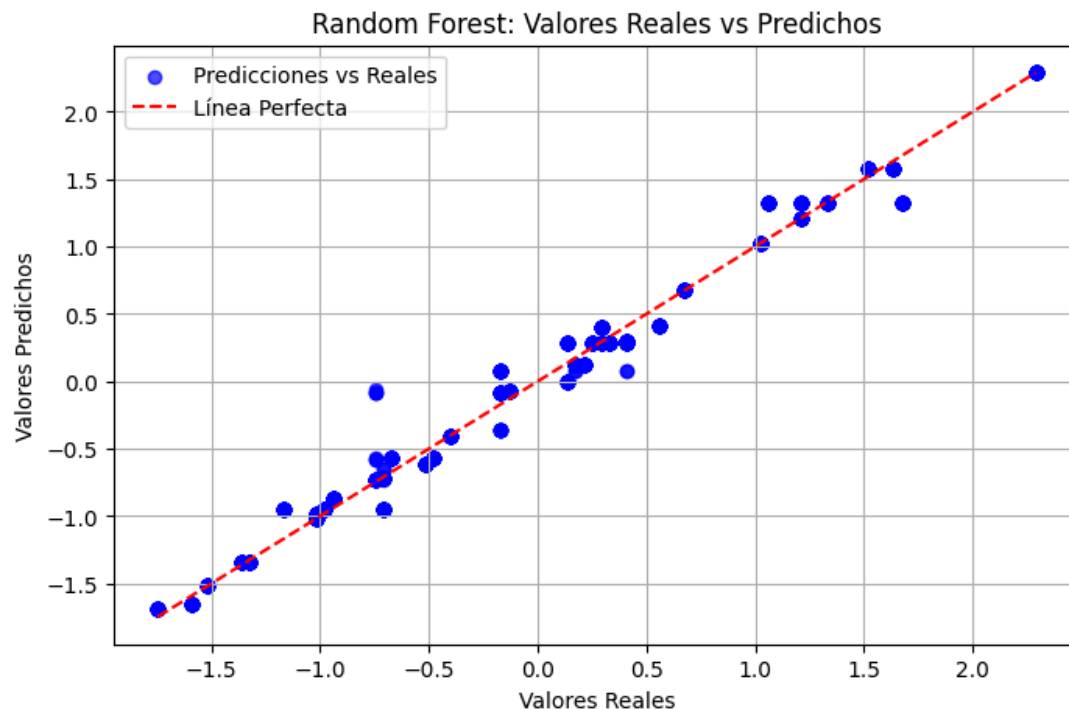
Sesgos: mide la asimetría, un sesgo de menos de 0.5 considera aceptable para muchos análisis estadísticos, pero si se usa un modelo de regresión, es posible que se deba transformar los datos para reducir el sesgo, especialmente si no cumple con otros supuestos como la normalidad de los residuos.

Que en este caso no los estaría cumpliendo. Por lo que se toma la medida primero de transformaciones como la logarítmica, la raíz cuadrada o la de Box-Cox.

En cambio también se podría optar por probar con explorar modelos de machine learning alternativos que sean menos sensibles al sesgo, como los modelos basados en árboles.

El sesgo de 0.078 sugiere una ligera asimetría en la distribución de la variable. Si bien no es extremadamente alto, es importante evaluar si podría afectar el análisis y considerar transformaciones o modelos alternativos si es necesario.

Se mejoró el modelado usando Random Forest



Interpretación de los Resultados

- ① $R^2 = 0.9852$ → Indica que el modelo explica el 98.52% de la variabilidad en la temperatura global. Es un resultado excelente.
- ② $MAE = 0.0847$ → En promedio, la diferencia entre los valores reales y los predichos es de 0.0847 unidades de temperatura.
- ③ $RMSE = 0.1235$ → Muestra la magnitud promedio del error. Cuanto más bajo, mejor.

-Conclusión Final

- ✓ El modelo de Random Forest predice la temperatura global con gran precisión (R^2 alto y errores bajos).
- ✓ Los valores predichos siguen de cerca la tendencia real de la temperatura.
- ✓ La variabilidad de los datos está bien representada, lo que sugiere que no hay sobreajuste extremo.

R^2 promedio en validación cruzada: 0.9845

Desviación estándar: 0.0007

La validación cruzada es una técnica que divide los datos en varios subconjuntos (folds) para entrenar y evaluar el modelo múltiples veces. Esto ayuda a verificar que el modelo no dependa demasiado de un solo conjunto de datos y generalice bien en datos nuevos. EL R^2 0.9845 Indica que el modelo es muy bueno y predice con alta precisión.

La desviación estándar indica qué tanto varían los valores de R^2 entre las diferentes pruebas de validación cruzada. La desviación estándar es 0.0007 (muy baja), significa que el modelo es

estable y generaliza bien, ya que su desempeño casi no cambia entre diferentes conjuntos de validación.

En conclusión, **Random Forest** resultó ser una mejor alternativa en comparación con la **regresión lineal**, ya que maneja mejor la no linealidad de los datos sin necesidad de transformaciones adicionales. Este modelo es una herramienta sólida para predecir cambios futuros en la temperatura global basándose en los niveles de CO₂ y metano.