Clasificación Climática Global mediante PCA y K-Means: Validación Ecológica con Hábitats de Osos

Autores: Iván Fernández Martínez y Sergio Alves Martín

DESCRIPCIÓN BREVE

Este proyecto desarrolla una **clasificación climática global** a partir de datos del **CMIP6**, con el objetivo de identificar **patrones espaciales del clima** y relacionarlos con los **hábitats de distintas especies de osos**.
Se utilizaron tres variables clave —**precipitación, temperatura mínima y máxima**— del periodo **1850–2014**, procesadas y combinadas en un **ensemble multi-modelo**.
Tras aplicar **PCA** para reducir la dimensionalidad, se empleó **K-Means** para agrupar regiones con climas similares.

1. Introducción y objetivos

El estudio del clima a escala global es esencial para entender cómo los cambios en el sistema climático afectan a los ecosistemas y a las especies. Clasificar los distintos tipos de clima del mundo permite identificar patrones comunes, analizar su distribución geográfica y observar cómo varían con el tiempo.

Este proyecto desarrolla una **clasificación climática global** basada en los modelos del **CMIP6 (1850–2014)**, utilizando tres variables clave: **precipitación (pr)**, **temperatura mínima (tasmin)** y **temperatura máxima (tasmax)**. Tras un proceso de preprocesamiento, remallado y cálculo de climatologías mensuales, se generó un **ensemble multi-modelo** que sirvió como base para el análisis estadístico.

Mediante **Análisis de Componentes Principales (PCA)** se redujo la dimensionalidad de los datos y, posteriormente, se aplicó el algoritmo **K-Means** para agrupar regiones del planeta con climas similares. Finalmente, se compararon los resultados con los hábitats de cuatro especies de osos polar, pardo, panda y perezoso, evaluando la correspondencia entre los patrones climáticos y sus condiciones ecológicas.

Objetivos

El objetivo principal del proyecto es crear una clasificación climática global basada en datos del CMIP6 mediante el uso de técnicas estadísticas y de aprendizaje no supervisado, y comprobar su coherencia con los hábitats naturales de diferentes especies.

A partir de este objetivo general, se plantean los siguientes objetivos específicos:

1. Seleccionar y procesar las variables climáticas clave (precipitación, temperatura mínima y máxima) en formato NetCDF para el periodo 1850–2014.
2. Homogeneizar los datos mediante su verificación, remallado y cálculo de climatologías mensuales.
3. Generar un **ensemble multi-modelo**, promediando las climatologías de los distintos modelos del CMIP6.
4. Aplicar el **Análisis de Componentes Principales (PCA)** para reducir la dimensionalidad del conjunto de datos conservando la mayor parte de la información.
5. Implementar el algoritmo **K-Means** para identificar regiones climáticamente homogéneas y determinar el número óptimo de clústeres mediante el método del codo.
6. Visualizar y comparar los resultados con los hábitats naturales de los osos polar, pardo, panda y perezoso, analizando la coherencia bioclimática entre ambos conjuntos de información.

2. Metodología

2.1. Selección de Variables Climáticas

La selección de variables se diseñó para representar equilibradamente los dos componentes principales del sistema climático: la **energía térmica** y el **balance hídrico**. Se eligieron tres variables fundamentales precipitación (pr), temperatura mínima (tasmin) y temperatura máxima (tasmax), todas ellas con resolución mensual para el periodo 1850–2014.

- **Precipitación (pr):** mide la cantidad de agua que cae sobre una superficie en forma de lluvia o nieve. Es clave para identificar regiones áridas, templadas o húmedas.
- **Temperatura mínima (tasmin):** refleja las condiciones térmicas nocturnas, asociadas con la pérdida de calor y el enfriamiento superficial.
- **Temperatura máxima (tasmax):** representa el calor diurno y el nivel de radiación solar absorbido por la superficie terrestre.

Estas tres variables permiten captar los contrastes climáticos más relevantes a escala global y son ampliamente utilizadas en estudios bioclimáticos. Además, su disponibilidad homogénea en los modelos del CMIP6 facilitó su comparación y combinación.

2.2. Fuentes y preparación de los datos

2.2.1. Modelos Climáticos Utilizados

El análisis se basó en datos procedentes de **ocho modelos globales del proyecto internacional CMIP6**, considerado el estándar de referencia en estudios climáticos. Los modelos fueron seleccionados por disponer de información completa y continua de las tres variables en el periodo histórico 1850–2014.

Los datos se descargaron en formato **NetCDF (.nc)**, ampliamente empleado para almacenar información geoespacial multidimensional (latitud, longitud, tiempo y variable).

2.2.2. Organización y estructura de los datos

Los datos originales se organizaron en una carpeta de trabajo denominada data_raw, donde cada modelo incluía las tres variables seleccionadas. Cada archivo NetCDF mantiene la misma estructura general:

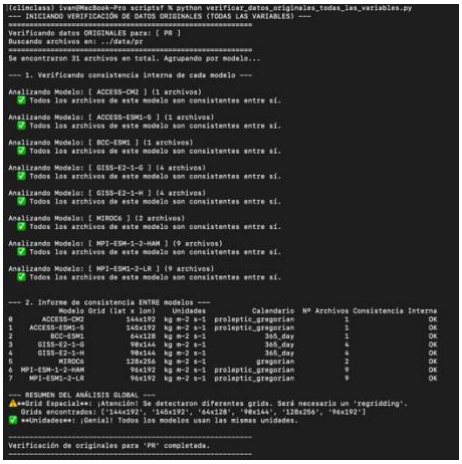
- **Dimensiones espaciales:** latitud y longitud, que definen la cuadrícula global.
- **Dimensión temporal:** serie mensual entre 1850 y 2014.
- **Variable climática:** una de las tres seleccionadas, expresada en sus unidades físicas originales.

2.2.3. Procesamiento y homogeneización de datos

El preprocesamiento se desarrolló a través de una secuencia de **scripts en Python**, cada uno enfocado en una función específica del flujo de trabajo. El objetivo general fue **homogeneizar** y **preparar** la información de los modelos CMIP6, garantizando su compatibilidad espacial, temporal y de unidades antes del análisis estadístico.

Verificación de los datos originales

Script: verificar_datos_originales_todas_las_variables.py



Se comprobó la integridad de los archivos NetCDF para asegurar que:

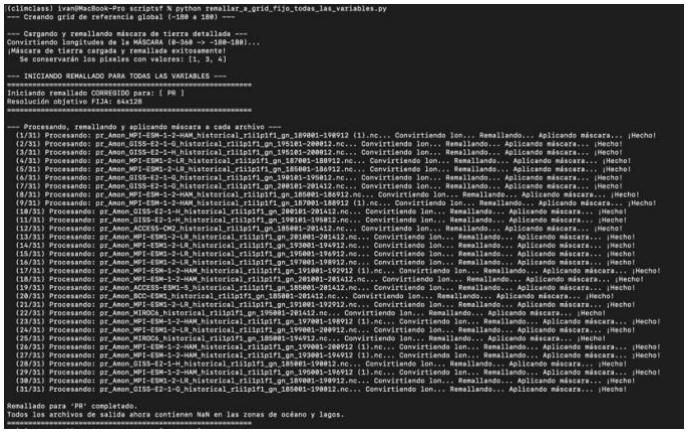
- Cubrieran el periodo 1850–2014 completo.
- Compartieran el mismo sistema de coordenadas.
- No contuvieran valores nulos ni errores de lectura.

Aunque no genera nuevos archivos, este paso fue esencial para detectar incoherencias y garantizar la fiabilidad del conjunto previo al remallado.

Remallado a una cuadrícula común

Script: remallar_a_grid_fijo_todas_las_variables.py

Dado que los modelos climáticos tienen resoluciones espaciales distintas, se realizó una **interpolación a una malla uniforme de 128×64 puntos**, lo más fácil era reducir todas a la más pequeña, asegurando que cada celda representara una superficie comparable. Durante este proceso se aplicó una **máscara terrestre (landsea.nc)**, eliminando las zonas oceánicas y conservando solo las áreas continentales y polares.

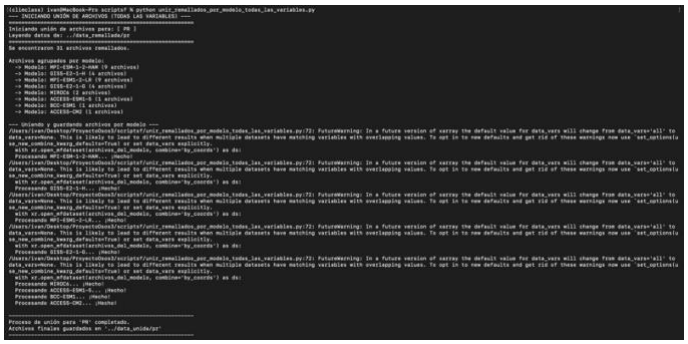


Resultado:
Archivos NetCDF remallados y estandarizados, almacenados en la carpeta data_remallada.

Unificación de archivos por modelo

Script: unir_remallados_por_modelo_todas_las_variables.py

Una vez remallados, los datos de cada modelo se encontraban en archivos separados por variable.
Este script **fusionó los archivos** en un único bloque por modelo y variable, verificando la coherencia temporal y espacial.



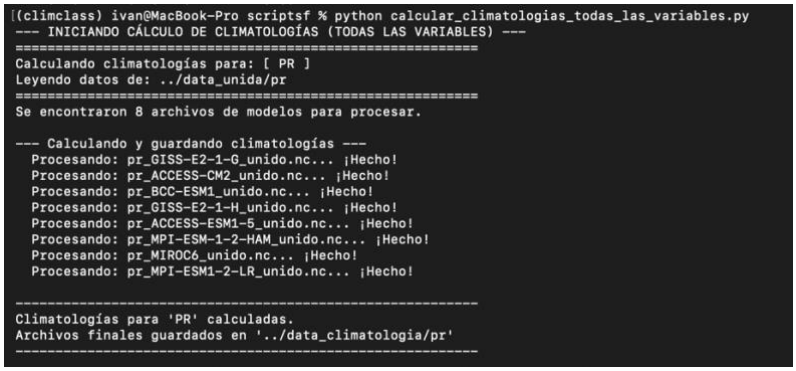
Resultado:
Archivos combinados y organizados en la carpeta data_unida.

Cálculo de climatologías mensuales

Script: calcular_climatologias_todas_las_variables.py

Se calcularon las **climatologías mensuales** (promedios de cada mes entre 1850 y 2014), condensando 160 años de datos en 12 valores medios representativos.

- En **precipitación (pr)**, se observaron patrones estacionales y contrastes árido–húmedo.
- En **tasmin** y **tasmax**, se identificaron gradientes térmicos y la oscilación media anual.



Resultado:
Archivos NetCDF con 12 capas mensuales, guardados en la carpeta data_climatologias.

Creación del ensemble multi-modelo

Script: crear_ensemble_todas_las_variables.py

Se generó un **ensemble multi-modelo**, calculando el promedio de las climatologías de los diez modelos. Este procedimiento redujo sesgos individuales y proporcionó una visión más equilibrada del clima global.

Archivos generados:

- ensemble_pr.nc – Precipitación media global
- ensemble_tasmin.nc – Temperatura mínima media
- ensemble_tasmax.nc – Temperatura máxima media

Resultado:
Archivos listos para el análisis estadístico posterior, almacenados en data_ensemble.

Resumen del preprocesamiento

El proceso completo produjo tres conjuntos de datos **homogéneos, remallados, verificados y promediados**, que representan el clima medio global en términos de temperatura y precipitación.
Estos archivos sirvieron como base para aplicar las técnicas de reducción de dimensionalidad y clasificación climática descritas a continuación.

2.3. Análisis PCA y clasificación K-Means

Con los archivos *ensemble* finales se llevó a cabo el **análisis estadístico multivariante**, orientado a reducir la dimensionalidad de los datos y clasificar las zonas del planeta según su similitud climática.
Para ello se combinaron dos técnicas complementarias: el **Análisis de Componentes Principales (PCA)** y el **agrupamiento no supervisado K-Means**, implementadas mediante scripts en Python.

Aplicación del Análisis de Componentes Principales (PCA)

Script: aplicar_pca.py

El PCA se aplicó a los tres archivos (ensemble_pr.nc, ensemble_tasmin.nc, ensemble_tasmax.nc), con el fin de identificar las combinaciones lineales que mejor explicaran la variabilidad climática global.
El script:

- Extrajo y normalizó los valores de cada variable.
- Combinó los datos en una matriz de características espaciales.
- Aplicó la descomposición PCA con *scikit-learn*.

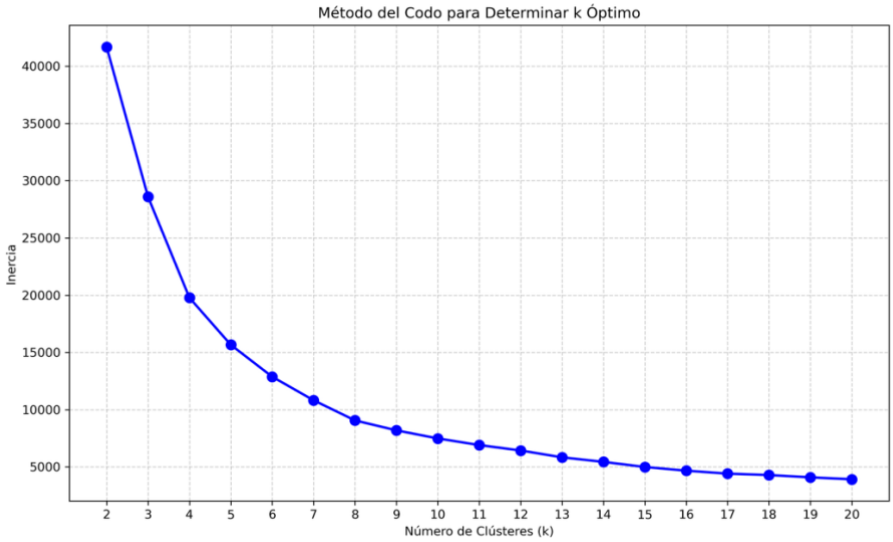
Se obtuvieron las componentes principales y la varianza explicada acumulada, que permitió conservar el **95 % de la información original**.
Los resultados se guardaron en componentes_principales.nc, junto con el modelo pca_model.joblib.

El joblib guarda un modelo ya entrenado para no tener que volver a calcularlo.

Determinación del número óptimo de clústeres (Método del codo)

Script: calcular_y_guardar_codo.py

Para definir el número adecuado de clústeres se utilizó el **método del codo**, ejecutando el algoritmo K-Means para valores de k entre 2 y 10.
El punto de inflexión de la curva de inercia determinó que **k = 6** era el valor más apropiado.
Este resultado se almacenó en k_optimo.txt.



Clasificación climática global mediante K-Means

Script ejecutado: generar_mapa_kmeans.py

Una vez identificado el número óptimo de clústeres mediante el método del codo (k = 6), se aplicó el algoritmo K-Means sobre los datos transformados por el PCA. Este procedimiento permitió agrupar cada punto geográfico del planeta en una de las seis categorías climáticas identificadas según su similitud en las variables de temperatura y precipitación.

Durante la ejecución se generaron mapas de clasificación para distintos valores de *k* (de 5 a 10), lo que facilitó una comparación visual de la coherencia espacial de las particiones. Los resultados se exportaron en formato PNG (mapa_clasificacion_k5.png, mapa_clasificacion_k6.png, etc.) y en formato NetCDF, garantizando la trazabilidad de los datos.

El mapa correspondiente a *k = 6* (mapa_clasificacion_k6.png) fue inicialmente seleccionado como la clasificación climática global principal. Este valor representaba un equilibrio entre simplicidad y diferenciación, permitiendo distinguir los grandes tipos climáticos del planeta como las zonas polares, templadas, áridas, tropicales, montañosas y ecuatoriales de manera coherente con las observaciones climatológicas conocidas.

Resultado:

- Archivos gráficos: mapa_clasificacion_k5.png a mapa_clasificacion_k10.png y scatter_clasificacion_k5.png a scatter_clasificacion_k10.png
- Archivo .nc: mapa_clasificacion_k[num].nc

Exploración alternativa: clasificación con nueve clústeres

Script: nueve_clusters.py

Aunque el método del codo determinó que *k = 6* era el número óptimo desde un punto de vista estadístico, se consideró relevante realizar una segunda clasificación con diferentes valores de k, llegando a la conclusión de que el número más optimo y que mejor definía los hábitats era k= 9, para explorar patrones climáticos de mayor detalle. Este análisis adicional buscó comprobar si un número superior de grupos permitía capturar mejor las transiciones entre climas o las particularidades de ciertas regiones.

El script nueve_clusters.py siguió la misma metodología que el anterior, aplicando el algoritmo K-Means sobre los mismos datos PCA, pero fijando manualmente el número de clústeres en nueve. El resultado se guardó como mapa_clasificacion_k9.nc y su representación gráfica (mapa_clasificacion_k9.png) mostró una segmentación más rica y precisa del planeta.

En esta nueva clasificación, se observaron subdivisiones dentro de zonas previamente homogéneas. Por ejemplo:

- En las regiones templadas, el modelo separó áreas oceánicas y continentales.
- En las zonas áridas, distinguió entre desiertos cálidos y fríos.
- En los trópicos, identificó diferencias entre selvas húmedas y sabanas secas.

Estos resultados, aunque más complejos, mostraron una correspondencia espacial más ajustada. Por tanto, el mapa con *k = 9* fue interpretado como una **versión refinada del modelo climático global**, con mayor poder descriptivo a nivel biogeográfico.

Resultado:

- Archivo gráfico: mapa_clasificacion_k9.png.
- Archivo .nc: mapa_clasificacion_k9.nc

Validación bioclimática: comparación con hábitats de osos

Script: analizar_y_mapear_habitats.py

Se realizó una validación ecológica superponiendo los mapas de clasificación con las distribuciones de cuatro especies:

- **Oso polar**
- **Oso pardo**
- **Oso perezoso**
- **Oso panda**

La comparación mostró **alta coherencia espacial** entre los clústeres y las regiones de presencia de cada especie:
el oso polar en climas fríos y secos, el pardo en zonas templadas, el perezoso en áreas cálidas y húmedas mientras que el panda es un animal que se concentra únicamente en una zona montañosa de China.

Conclusión del análisis estadístico

La combinación de PCA y K-Means permitió identificar una tipología climática global robusta y reproducible.

- El modelo con *k = 6* proporcionó una visión sintética de los principales tipos de clima.
- El modelo con *k = 9* ofreció un mayor nivel de detalle y coherencia ecológica.

Ambos enfoques complementan la comprensión del clima global, y su integración representa una herramienta útil para futuras investigaciones sobre la relación entre clima y biodiversidad.

3. Resultados y visualizaciones

3.1. Representación de clústeres en el espacio PCA

ras aplicar el **Análisis de Componentes Principales (PCA)** sobre las variables climáticas —temperatura mínima, temperatura máxima y precipitación—, se representaron los resultados del algoritmo **K-Means** en el plano definido por las dos primeras componentes principales, que concentran la mayor parte de la variabilidad térmica e hídrica global.

En la **Figura 1 (k = 6)**, cada punto representa una celda del conjunto de datos climáticos proyectada en el espacio PCA, y los colores indican los seis clústeres obtenidos. Se observa una separación clara entre grupos: los valores situados en el extremo izquierdo corresponden a **climas fríos y secos**, mientras que los de la derecha reflejan **climas cálidos y húmedos**, lo que demuestra que el modelo distingue adecuadamente los principales contrastes térmicos del planeta.

Posteriormente, se exploró un escenario alternativo con **nueve clústeres (Figura 2)**, que permite una **segmentación más detallada**, diferenciando zonas de transición entre climas templados y tropicales. Aunque el método del codo señaló **k = 6** como el número óptimo por su equilibrio entre compacidad y simplicidad, el modelo con **k = 9** mostró una **mayor coherencia visual y geográfica** en los mapas, por lo que ambos escenarios se mantuvieron para su análisis comparativo posterior.

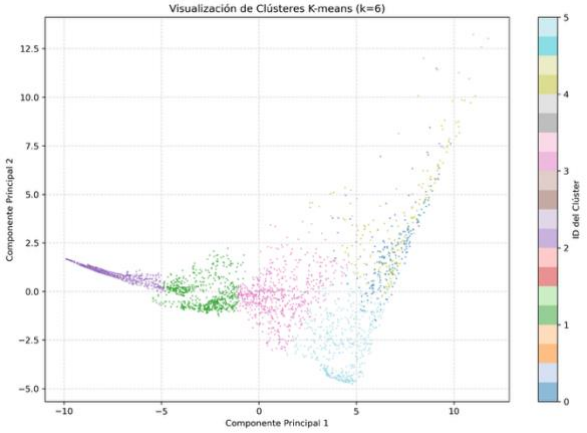


Figura 1

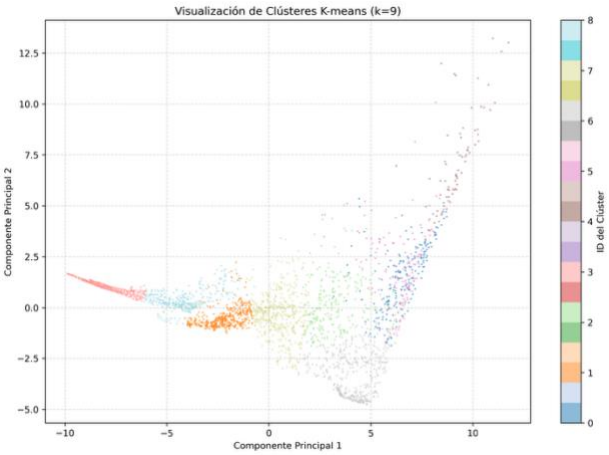


Figura 2

3.2. Mapas de clasificación climática

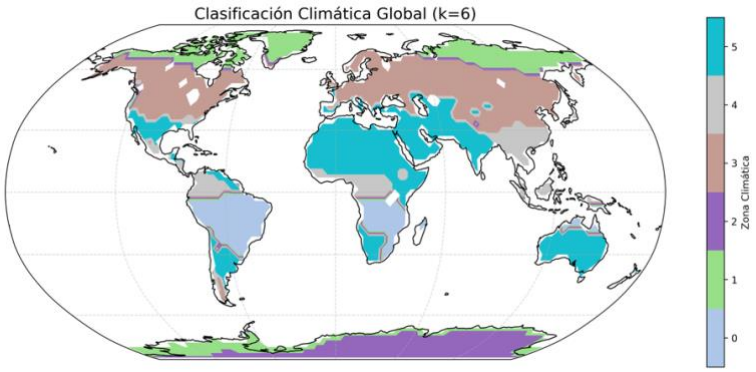
Una vez completado el análisis K-Means, se generaron mapas globales para distintos valores de *k*.
Estos mapas permiten comparar visualmente la segmentación del planeta según las variables climáticas analizadas, evaluando cómo varía el detalle de la clasificación al aumentar el número de clústeres.

Clasificación con k = 6

El modelo con **seis clústeres** genera una visión general simplificada del clima mundial. Los colores agrupan amplias zonas climáticas del planeta, de la siguiente manera:

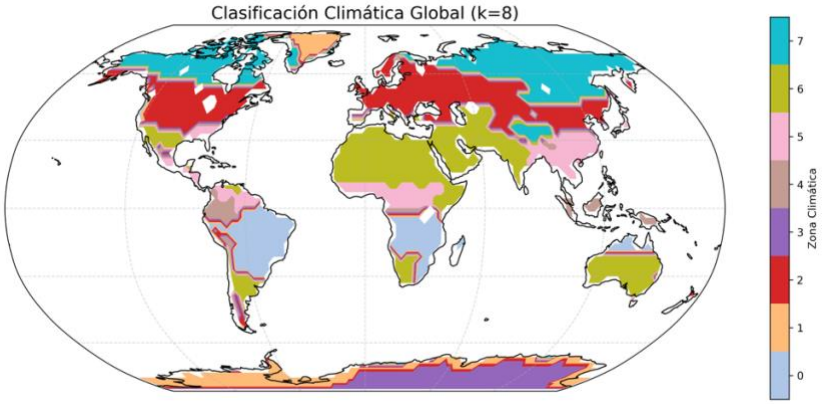
- **Verde claro**: zonas **frías boreales y subpolares**, correspondientes al norte de Canadá, Escandinavia y Siberia. Estas regiones presentan bajas temperaturas anuales y precipitaciones moderadas.
- **Marrón rojizo**: climas **templados continentales**, característicos del interior de Eurasia y Norteamérica, con marcadas diferencias entre invierno y verano.
- **Celeste**: climas **áridos y semiáridos**, que abarcan el norte de África, Oriente Medio, Asia central y zonas del interior de Australia.
- **Gris y azul claro**: climas **tropicales y ecuatoriales**, como la Amazonía, el África subsahariana o Australia, caracterizados por altas temperaturas y fuertes precipitaciones.
- **Violeta**: climas **fríos antárticos y subantárticos**, donde predominan las bajas temperaturas extremas durante todo el año.

Aunque este mapa capta correctamente los grandes contrastes latitudinales, **fusiona bajo un mismo color zonas climáticamente distintas** (por ejemplo, selvas húmedas y sabanas), lo que reduce el nivel de detalle y dificulta el análisis ecológico más fino.



Clasificación con k = 8

Al aumentar a **k = 8**, se observa una **mayor diferenciación de las zonas templadas y tropicales**, comenzando a distinguirse los climas continentales, áridos y monzónicos. Sin embargo, la división aún presenta cierta homogeneidad dentro de grandes continentes, lo que sugiere que un número algo mayor de clústeres puede capturar mejor la variabilidad regional.



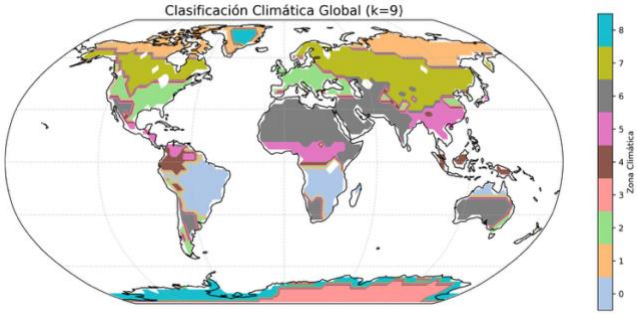
Clasificación con k = 9

Al incrementar el número de clústeres a **nueve**, se obtiene una segmentación **más detallada y realista**, que separa con mayor precisión las transiciones climáticas. En este caso, los colores reflejan zonas climáticas más específicas:

- El **azul claro (0)** abarca **la mayor parte de Sudamérica, el sur de África y el norte de Australia**, zonas de clima **templado-cálido húmedo**, con estaciones bien marcadas y lluvias moderadas.
- El **amarillo (1)** se localiza en **el norte de Canadá y Siberia**, representando climas **fríos y continentales**, con inviernos largos y veranos cortos.
- El **verde claro (2)** cubre **el sur y centro de Europa y la zona central de Estados Unidos**, reflejando un clima **templado húmedo o oceánico**, con temperaturas moderadas y precipitaciones regulares.
- El **salmón (3)** aparece en **el este del Polo Sur**, asociado a condiciones **frías polares**, con temperaturas bajo cero la mayor parte del año.
- El **marrón (4)** domina **Indonesia y regiones ecuatoriales del sudeste asiático**, caracterizando un clima **tropical húmedo**, cálido todo el año y con fuertes precipitaciones.
- El **rosa (5)** también se concentra **en torno al ecuador**, especialmente en **África central y el norte de Sudamérica**, representando un clima **tropical cálido y muy húmedo**.
- El **gris (6)** ocupa **el norte de África, la India, el este de África y el centro de Australia**, indicando zonas **áridas o semiáridas**, con escasas lluvias y altas temperaturas.

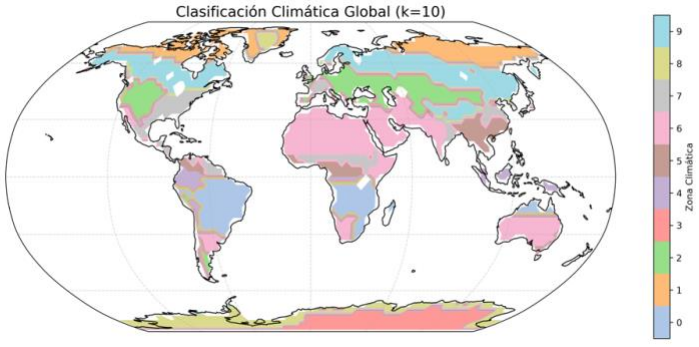
- La verde oliva **(7)** se extiende por **el norte de América, Escandinavia y el norte de Asia**, identificando climas **fríos boreales o subpolares**, con inviernos largos y veranos muy breves.
- Por último, el **celeste (8)** aparece en **el Polo Sur y el centro de Groenlandia**, donde predominan condiciones **polares extremas**, con frío permanente y ausencia casi total de vegetación.

El modelo con **k=9** permite **distinguir mejor las variaciones térmicas y de humedad** dentro de los grandes biomas. Por ejemplo, separa los desiertos cálidos de los templados, y diferencia entre selvas tropicales húmedas y zonas monzónicas. Además, esta mayor resolución climática se traduce en **una mejor correspondencia con los hábitats reales de las especies de osos**, lo que respalda su elección como clasificación principal para el análisis posterior.



Clasificación con k = 10

Finalmente, el mapa con **k = 10** introduce divisiones adicionales en áreas tropicales y subtropicales, pero sin aportar mejoras significativas a nivel interpretativo, lo que sugiere un leve sobreajuste respecto a la estructura climática real.

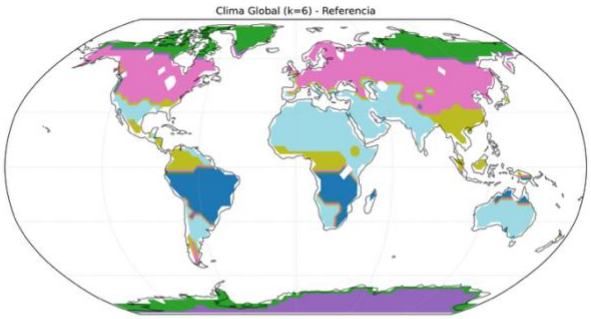


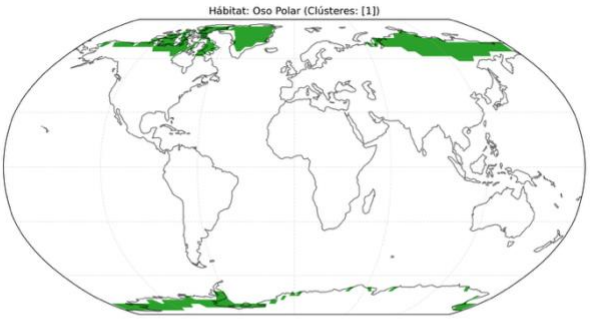
En conjunto, **el modelo con k = 9 logra el equilibrio más adecuado entre detalle y estabilidad**, por lo que fue seleccionado como referencia principal para los análisis comparativos con los hábitats de los osos.

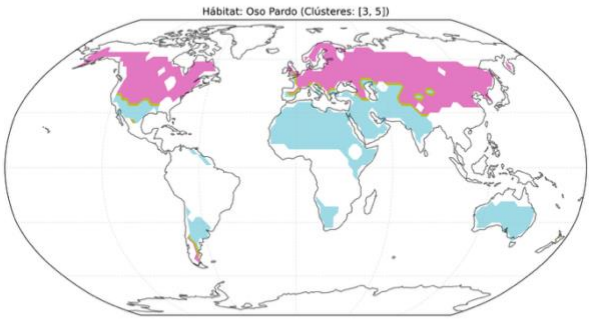
3.3. Validación mediante hábitats de osos

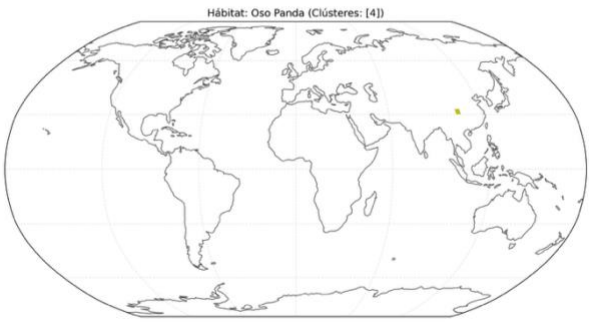
Para evaluar la validez ecológica de la clasificación climática obtenida, se compararon los resultados de los modelos de **k = 6** y **k = 9** con los **hábitats naturales de cuatro especies de osos**: el **oso polar**, el **oso pardo**, el **oso perezoso** y el **oso panda**. El objetivo fue comprobar hasta qué punto los clústeres climáticos derivados del análisis PCA + K-Means se corresponden con las condiciones ambientales reales que definen la distribución de estas especies.

Clasificación con k = 6







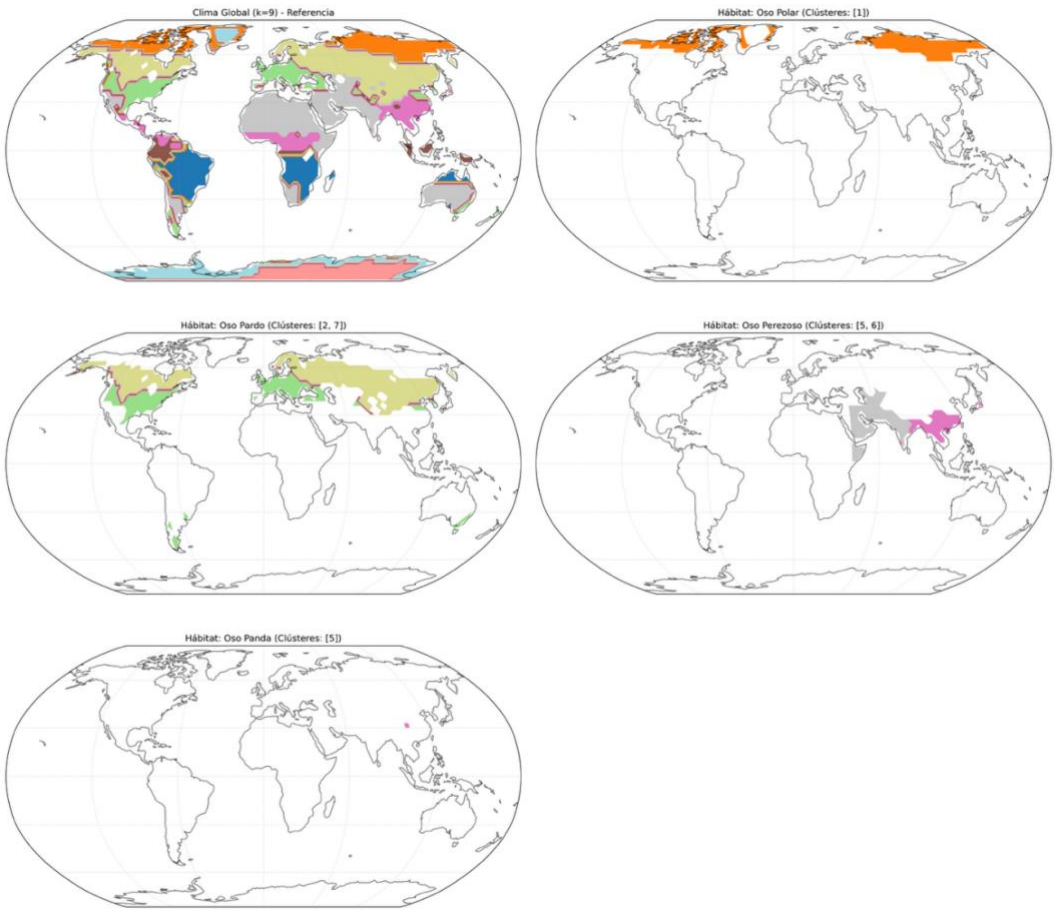


El modelo con **seis clústeres** ofrece una visión general de la estructura climática global, pero la segmentación es demasiado amplia para captar diferencias térmicas y de precipitación más sutiles. Esto genera **agrupaciones climáticas excesivamente generalistas**, que en algunos casos no reflejan correctamente la realidad ecológica.

- **Oso polar:** se asocia de manera coherente con el clúster **verde**, correspondiente a **regiones frías polares y subárticas**. Su distribución en el mapa coincide con el Ártico, Groenlandia y Siberia septentrional, zonas efectivamente habitadas por la especie, pero aparece de forma incorrecta en la Antártida debido a que comparten temperaturas muy frías.
- **Oso pardo:** aparece vinculado principalmente a los clústeres **rosado y azul claro**, que representan **climas templados húmedos y fríos continentales**. En Europa, Asia y América del Norte la correspondencia es razonable, pero el modelo **extiende el clúster azul claro hacia regiones donde la especie no existe**, como el **norte de África, Australia, África tropical e incluso zonas del cono sur americano**.
  Esta anomalía se debe a que el clúster azul claro agrupa **zonas con temperaturas medias moderadas, pero con precipitaciones muy dispares**, mezclando **climas templados húmedos** con **regiones áridas o semiáridas**. En consecuencia, el modelo confunde grandes desiertos cálidos (como el Sahara o el centro de Australia) con regiones templadas reales, creando **falsas coincidencias ecológicas**.
- **Oso perezoso:** se asocia con los clústeres **amarillo y celeste**, que abarcan **regiones tropicales y subtropicales húmedas del sur de Asia**, incluyendo la India y Sri Lanka. Aunque la correspondencia es razonable, el modelo también extiende el clúster hacia zonas del sudeste asiático donde la especie no está presente, lo que indica un exceso de generalización.
- **Oso panda:** se sitúa dentro del clúster **grisáceo**, vinculado a **zonas templadas húmedas de montaña**. Su localización en el suroeste de China coincide con su hábitat real, aunque el clúster también cubre otras zonas montañosas de Europa y América, lo que muestra que el modelo no diferencia microclimas tan específicos.

En conjunto, el modelo de **k=6** presenta una **visión climática global coherente**, pero su nivel de generalización **reduce la precisión ecológica**. Los principales problemas son la **confusión entre zonas templadas y áridas**, la **presencia artificial del clúster azul claro en regiones sin correlato biológico (como Australia o África tropical)** y la **pérdida de detalle en hábitats especializados**. Estos errores se explican porque, con pocos clústeres, K-Means tiende a agrupar zonas con valores medios similares, ignorando la variabilidad estacional o local.

Clasificación con k = 9



Al aumentar el número de clústeres a **nueve**, el modelo ofrece una **segmentación climática más detallada y ecológicamente coherente**. Esta mayor resolución permite diferenciar con más precisión las variaciones regionales en temperatura y precipitación, mejorando la correspondencia entre climas y hábitats reales.

- **Oso polar:** se asocia al clúster **naranja**, que representa **climas árticos y subárticos**. La delimitación coincide de forma precisa con su hábitat natural en las regiones costeras del Ártico, incluyendo Groenlandia, Canadá y Siberia y ahora sí que evita la Antártida.
- **Oso pardo:** se vincula principalmente con los clústeres **verde oliva y amarillo**, que corresponden a **climas templados húmedos y fríos continentales**. Estos se distribuyen por América del Norte, Europa y Asia, reflejando fielmente su área de ocupación y **corrigiendo los errores del modelo anterior**, al no extender su rango a África ni a zonas áridas.
- **Oso perezoso:** ocupa los clústeres **gris y rosa**, que representan **regiones tropicales húmedas y monzónicas** del sur de Asia. La separación entre áreas húmedas y semiáridas mejora sustancialmente la correspondencia con su distribución real.

- **Oso panda:** se restringe al clúster **rosa pálido**, asociado a **climas templados húmedos de montaña** en el suroeste de China, coincidiendo casi exactamente con su hábitat natural en las montañas de Sichuan.

En conjunto, el modelo de **k=9** presenta **una coherencia espacial y ecológica superior** a la del modelo de k=6. Permite identificar con mayor detalle las zonas climáticas relevantes para cada especie y **reduce los solapamientos irreales**, especialmente en regiones áridas.
Mientras que k=6 simplifica excesivamente el mosaico climático mundial, el modelo con nueve clústeres **logra capturar la complejidad real del sistema climático y su relación con los hábitats animales**, validando su uso como la opción más representativa en este estudio.

Además, se incorporó un **filtro geográfico** basado en las coordenadas de latitud y longitud disponibles gracias a la máscara de datos. Este procedimiento permitió **restringir el área de visualización** de cada clúster a las zonas donde realmente se distribuye cada especie de oso, evitando extensiones artificiales en regiones no habitadas. En la práctica, el filtro actúa limitando el rango de longitudes y latitudes que se representan en el mapa para cada tipo de oso, de modo que el clúster climático seleccionado se **adecúa mejor al hábitat real** de la especie.
Esta corrección mejora notablemente la **coherencia ecológica y espacial** del modelo, ya que los resultados reflejan con mayor precisión la relación entre las condiciones climáticas y la distribución geográfica de cada oso, ofreciendo una representación más fiel de la realidad bioclimática.

4. Interpretación y discusión

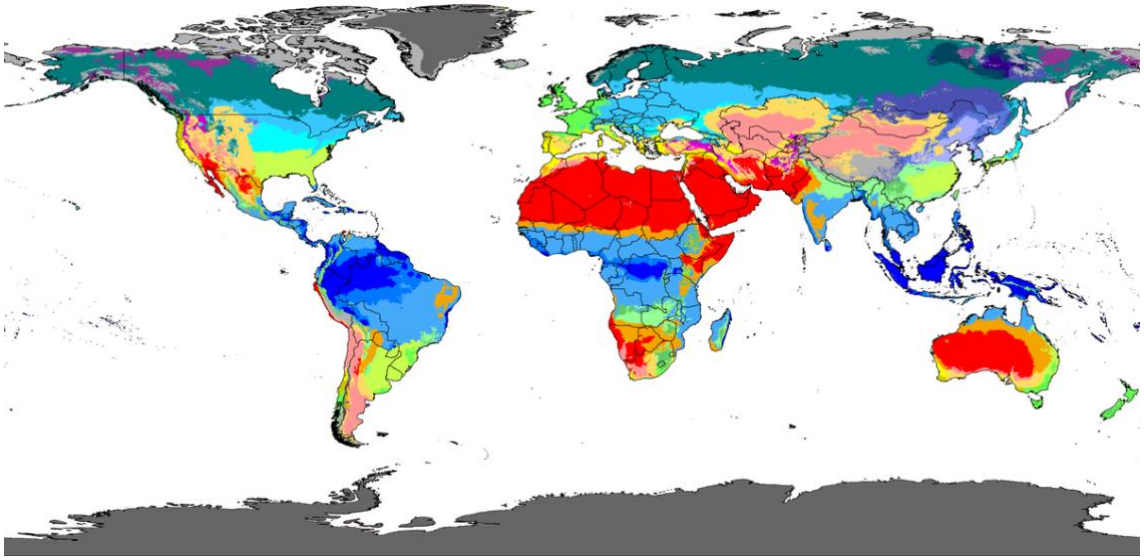
4.1 Correspondencia entre climas y hábitats

Como se observó en el apartado anterior, los resultados de clasificación muestran una clara relación entre los clústeres climáticos y las áreas de distribución de las especies de osos. En particular, el modelo con **k = 9** refleja con mayor precisión las condiciones térmicas e hídricas reales:

- El **oso polar** se concentra en los clústeres fríos y secos del Ártico.
- El **oso pardo** ocupa zonas templadas y húmedas del hemisferio norte.
- El **oso perezoso** aparece en regiones tropicales cálidas y húmedas del sur de Asia.
- El **oso panda** se localiza en climas templados de montaña del suroeste chino.

En cambio, el modelo de **k = 6**, aunque captó bien los grandes contrastes globales, tiende a simplificar en exceso. Por ejemplo, asigna al **oso pardo** áreas impropias como el norte de África o gran parte de Australia, que en la realidad presentan condiciones áridas. Esto evidencia que un número bajo de clústeres puede ocultar diferencias bioclimáticas relevantes.

4.2Comparativa con Köppen Geiger



Los resultados del modelo **PCA + K-Means** se parecen bastante a las clasificaciones climáticas tradicionales, como la de **Köppen-Geiger**. En ambos casos se distinguen claramente las grandes zonas del planeta: **polares, templadas, tropicales y áridas**. El modelo con **k = 9** ofrece un nivel de detalle similar al de Köppen, aunque con **límites más marcados**, sobre todo en las zonas tropicales y semiáridas.
En general, la comparación demuestra que la clasificación es **coherente y válida**, aunque se basa solo en variables climáticas y no considera factores biogeográficos.

4.3 Limitaciones

Durante el proyecto surgieron varias dificultades técnicas y metodológicas. La principal fue el **exceso de ruido** en los primeros mapas generados con cinco variables climáticas (pr, tasmin, tasmax, hurs y evspsbl). Las dos últimas **humedad relativa** y **evaporación** mostraron inconsistencias espaciales, lo que nos generaba mapas con mucho ruido por lo que finalmente se limitó el análisis a tres variables más estables: **precipitación** y **temperaturas mínima y máxima**.

También se detectaron **problemas con la proyección y las longitudes** de los mapas, ya que la **máscara terrestre** cortaba parte de la representación global. Esto obligó a reajustar el remallado y los límites espaciales para asegurar una cobertura continua.

Otra dificultad fue el error asociado al parámetro **time_bnds** durante el remallado, que anulaba la lectura de algunos archivos. Se resolvió **omitiendo dicho parámetro**, aunque con una ligera pérdida de precisión temporal.

En el análisis estadístico, como ya hemos dicho en contadas ocasiones a lo largo de este informe, el método del codo indicó **k = 6** como número óptimo de clústeres, pero el resultado más coherente desde el punto de vista ecológico se obtuvo con **k = 9**, que ofreció una mejor diferenciación de climas y hábitats.

5. Conclusiones y mejoras futuras

El presente trabajo ha permitido desarrollar una **clasificación climática global** basada en datos del **CMIP6**, integrando técnicas estadísticas de **Análisis de Componentes Principales (PCA)** y **agrupamiento K-Means**.
A partir de tres variables fundamentales **precipitación**, **temperatura mínima** y **temperatura máxima** se obtuvieron patrones climáticos coherentes que reflejan con notable realismo los contrastes térmicos e hídricos del planeta.

Los resultados demostraron que la combinación de PCA y K-Means es **una herramienta eficaz para resumir y clasificar grandes volúmenes de datos climáticos**, produciendo mapas interpretables y reproducibles.
Además, la **comparación con los hábitats de las especies de osos** evidenció una clara correspondencia entre los clústeres obtenidos y las zonas bioclimáticas donde viven estas especies, validando el enfoque empleado.

No obstante, el estudio también puso de manifiesto algunas **limitaciones** relacionadas con la selección de variables, la resolución espacial de los modelos y el número de clústeres óptimo.
Por ello, de cara a trabajos futuros se plantea:

- Incluir **más variables climáticas** (como humedad del suelo, radiación o presión atmosférica).
- Probar **métodos de clasificación alternativos** (como DBSCAN o modelos de mezcla gaussiana) para capturar estructuras no lineales.

En conjunto, este proyecto constituye una **base sólida y escalable para el análisis bioclimático global**, demostrando el potencial de la integración entre **modelización climática y análisis de datos** en la comprensión de los patrones que determinan la distribución de la biodiversidad.