

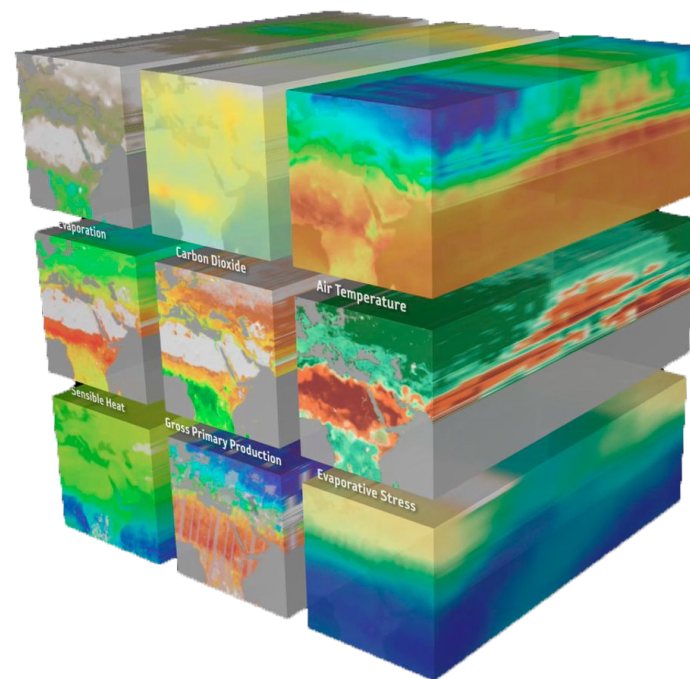


**Maximiliano Garavito Chtefan
Juan Pablo Moreno Rios
Maria Fernanda Navas Burgos
Liceth Yaneth Rozo Quintana**

Simulación Digital

**Detección de eventos
climáticos anormales
multivariados**

Justificación



Los climas extremos son una consecuencia del cambio climático y ponen en peligro la vida de las personas. Para poder predecir cuando pueden ocurrir, cómo mitigarlos, cómo minimizar el daño, generar medidas de respuesta eficaces y ayudar a que se recupere el ecosistema es necesario poder extraer dichos eventos y entender la dinámica de las interacciones de variables oceanicas, terrestres y atmosfericas que los conforman.

Temas aplicados

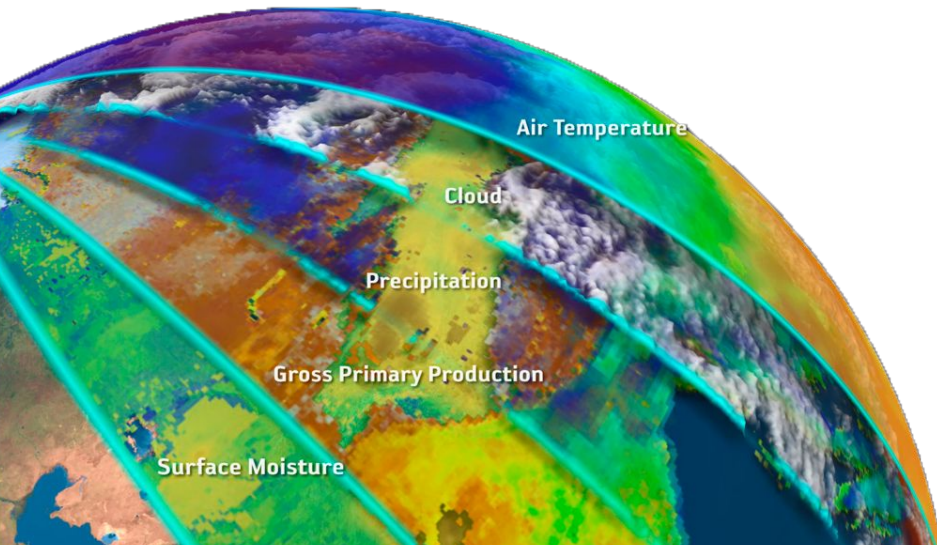
Fourier Transform

Signal Processing

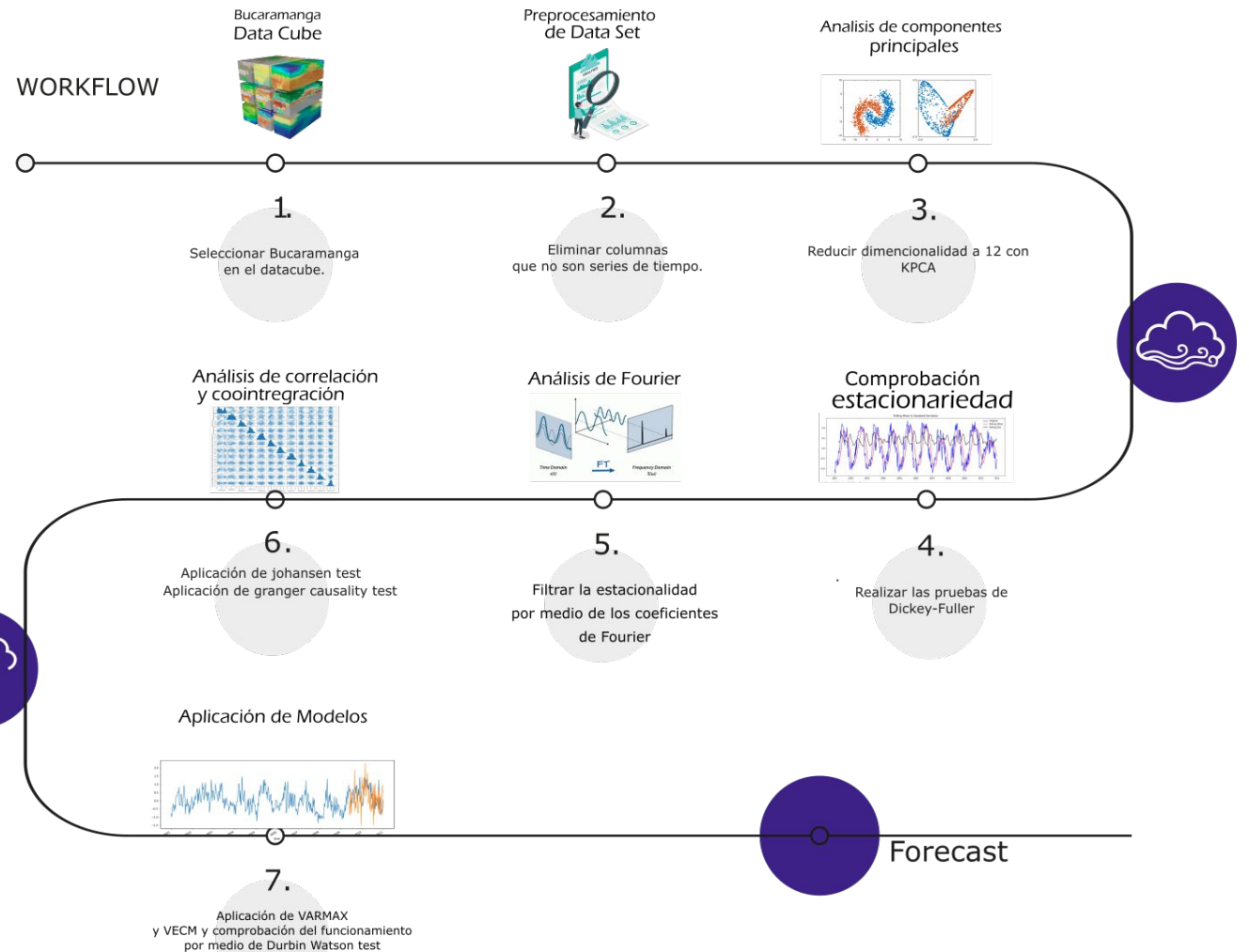
KernelPCA

VARMAX

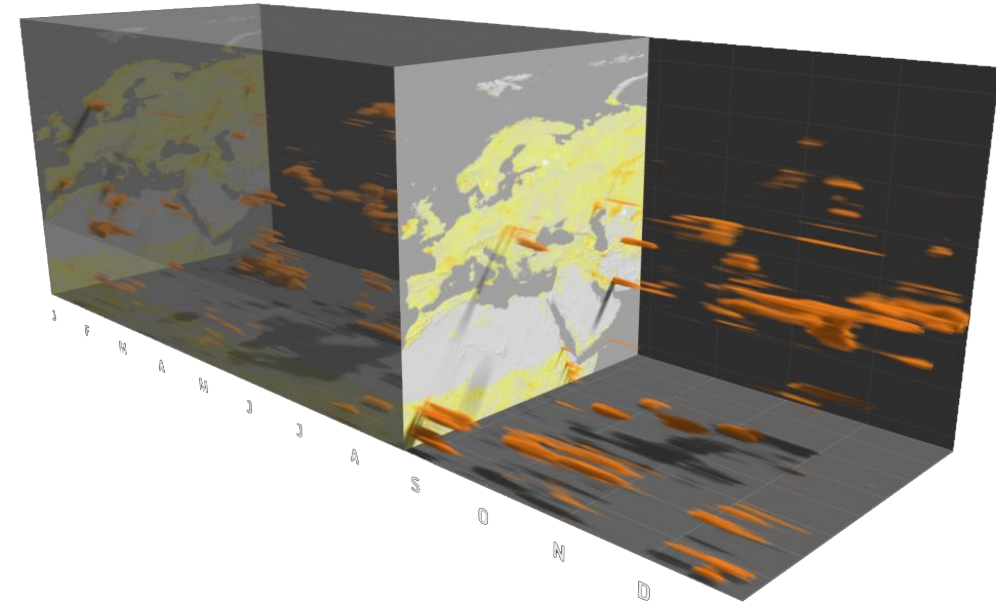
Frequency Analysis



Workflow

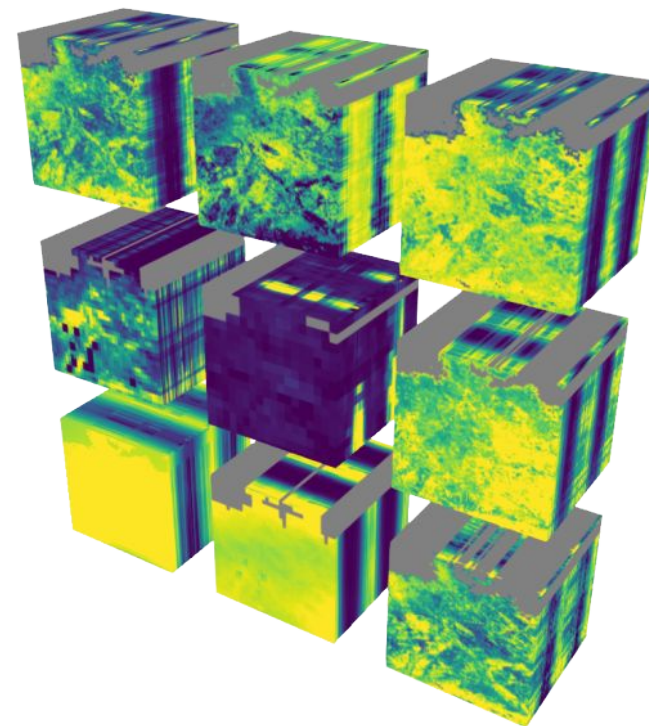


DataSet



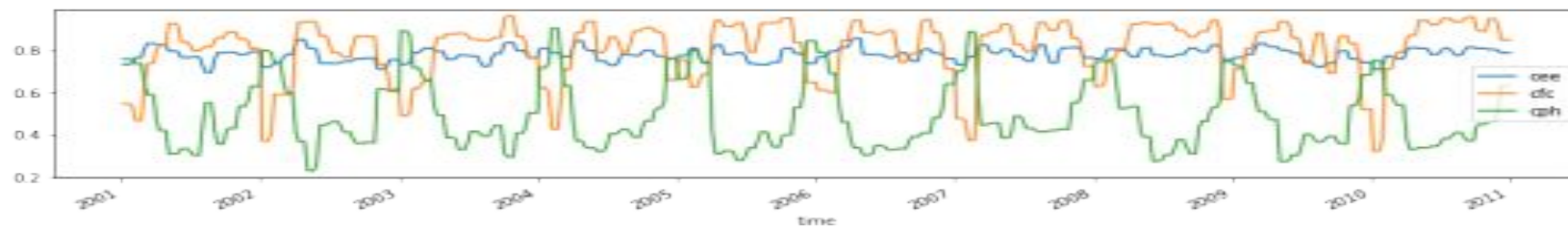
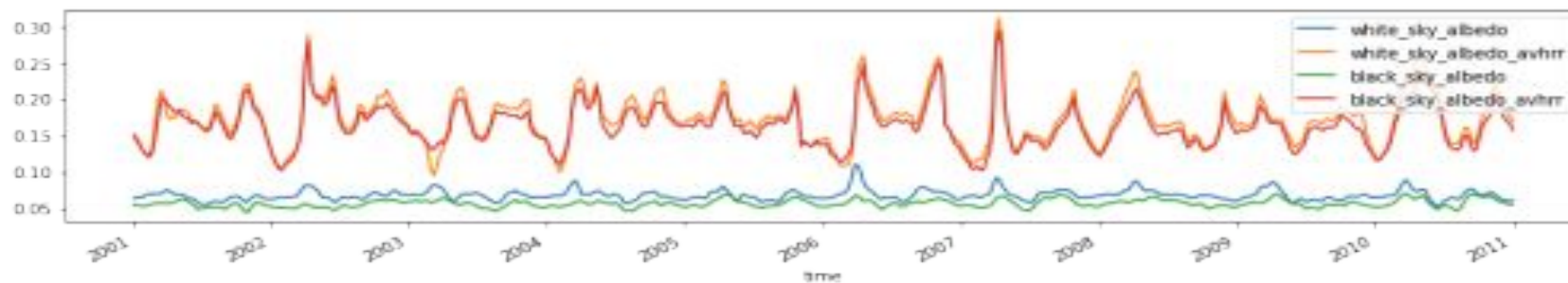
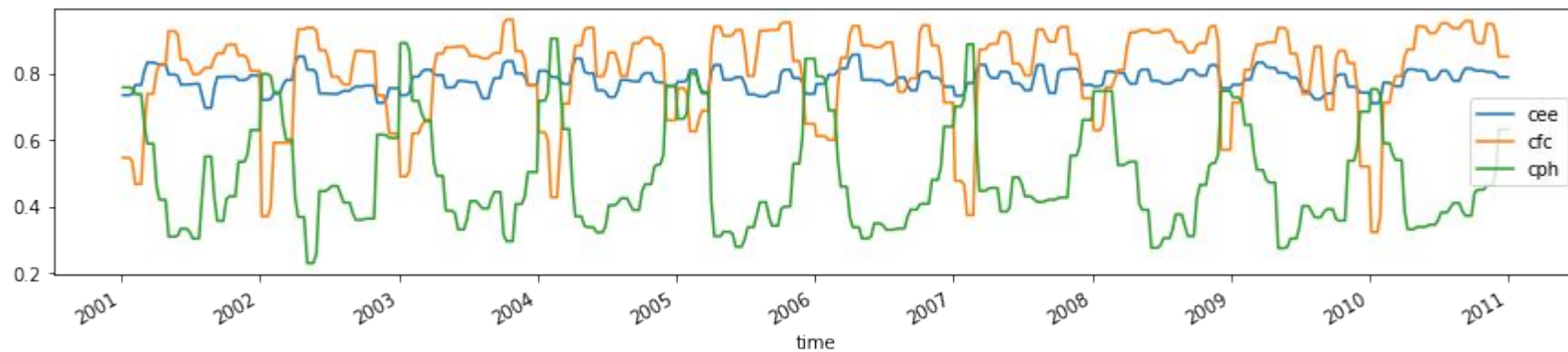
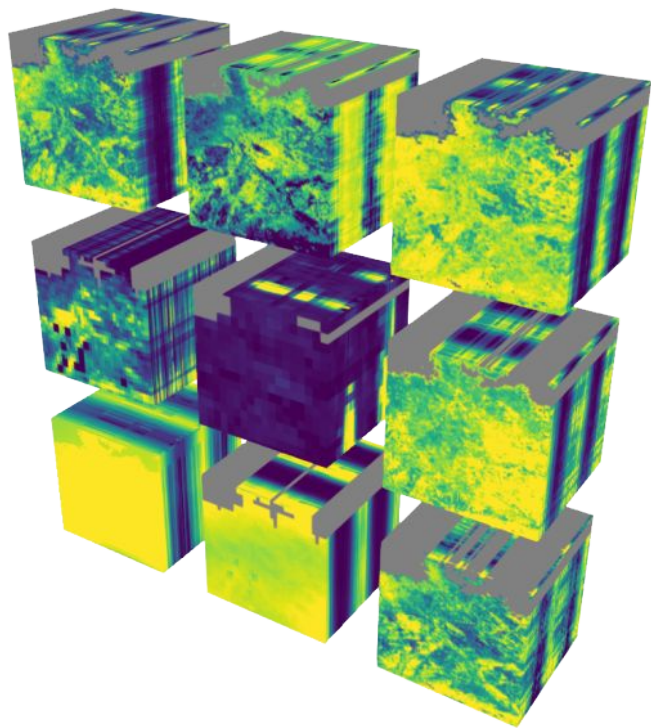
El dataset usado fue extraído del Earth System Data Lab, el cual es un proyecto de la Agencia Espacial Europea que consiste en mantener diferentes datacubes globales para uso público. En nuestro proyecto usamos el datacube global versión 2.0.0 optimizado temporalmente con una resolución de 0.25 grados, del cual extrajimos la serie de tiempo multivariada asociada a las coordenadas espaciales de Bucaramanga

Procesamiento

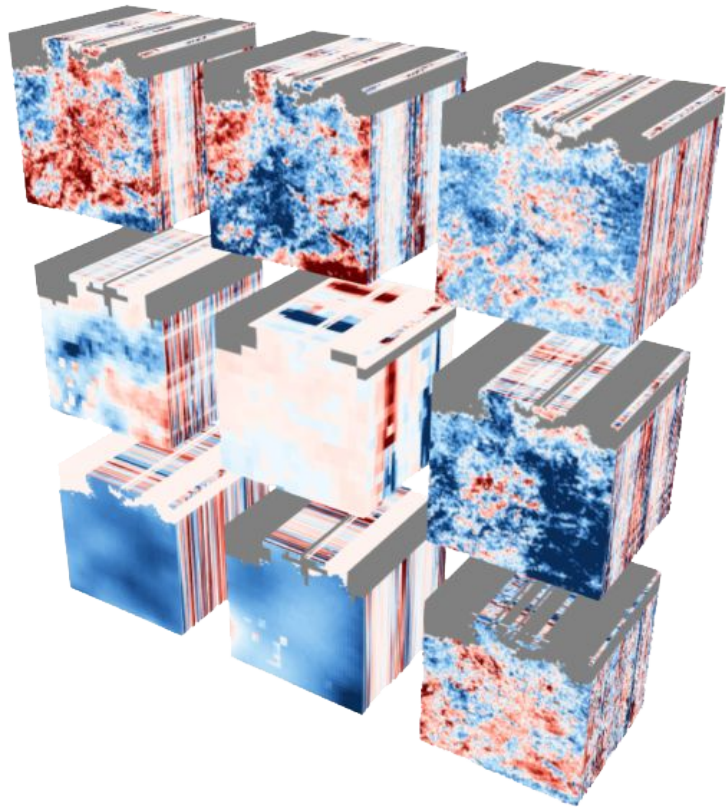


Se importan los datos al notebook, Se realiza una graficación de los datos de las series de tiempo resultantes, para observar posibles correlaciones. Se observan 29 posibles variables para procesar.

Procesamiento

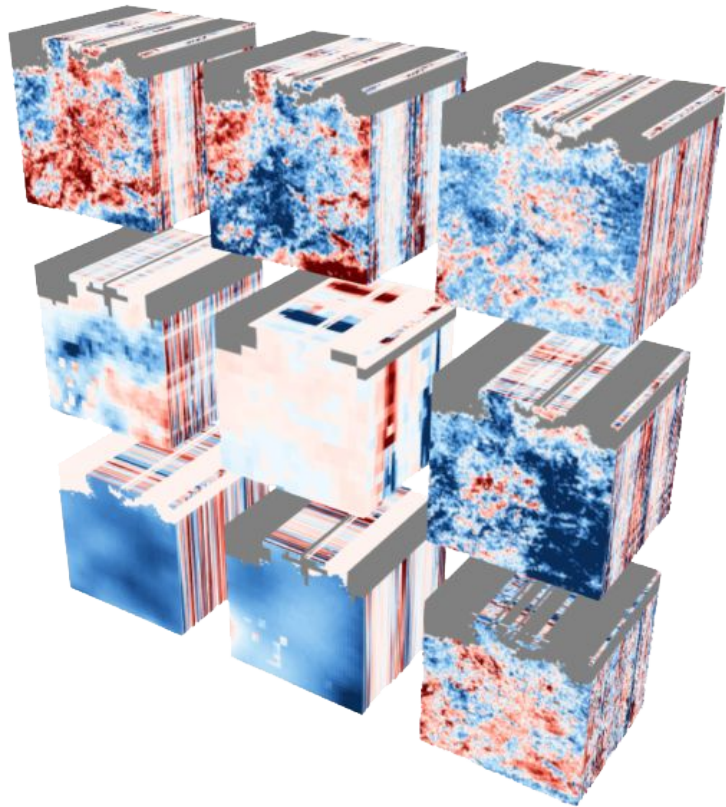


Kernel-PCA



Se observa una correlación fuerte entre los datos, y se procede a la reducción de dimensionalidad con KPCA. Ya que en python solo se pueden trabajar 12 series de tiempo simultáneamente, se aplicó un KPCA para reducir el dataset a los 12 componentes mas importantes

Kernel-PCA



Se observa una correlación fuerte entre los datos, y se procede a la reducción de dimensionalidad con KPCA. Ya que en python solo se pueden trabajar 12 series de tiempo simultáneamente, se aplicó un KPCA para reducir el dataset a los 12 componentes mas importantes

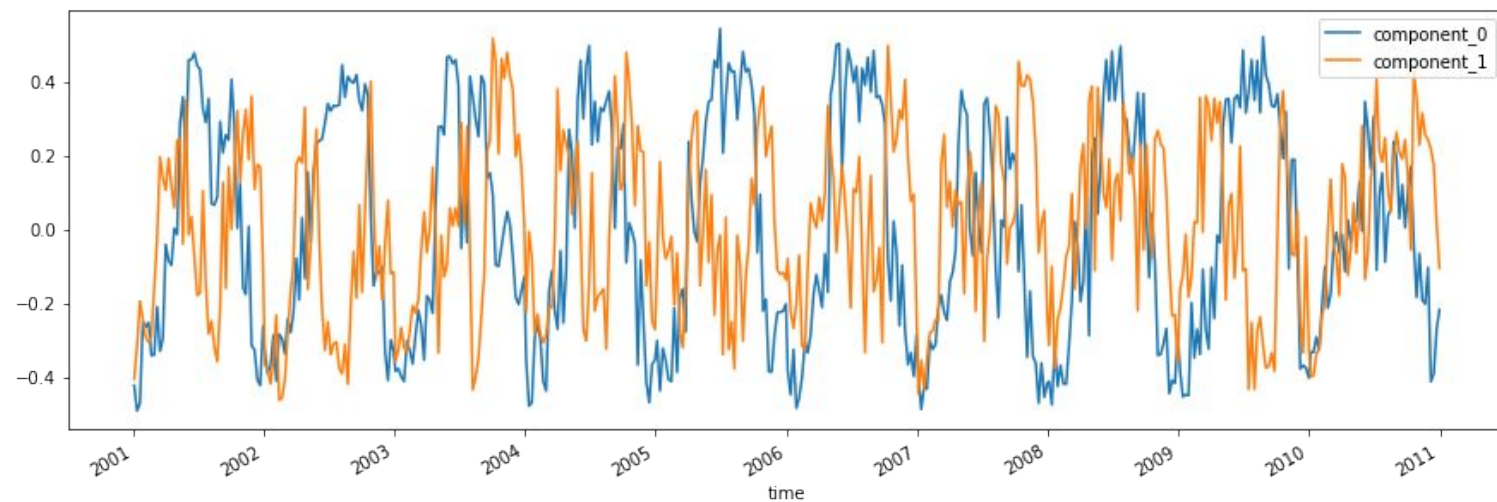
Time Series

Cointegración

Cuando dos o más series de tiempo están cointegradas, significa que tienen una relación estadísticamente significativa. Por lo cual se realiza el test de cointegración de Soren Johanssen. Si existe una combinación lineal de series de tiempo que tienen un orden de integración menor que la serie individual, se dice que la colección de series está cointegrada.

Los resultados de este test indican que las 12 series de tiempo están correlacionadas, lo cual es bueno, ya que la base que nos permite aplicar el modelo de predicción VARMAX, es que exista correlación entre dos o mas series de tiempo.

Time Series

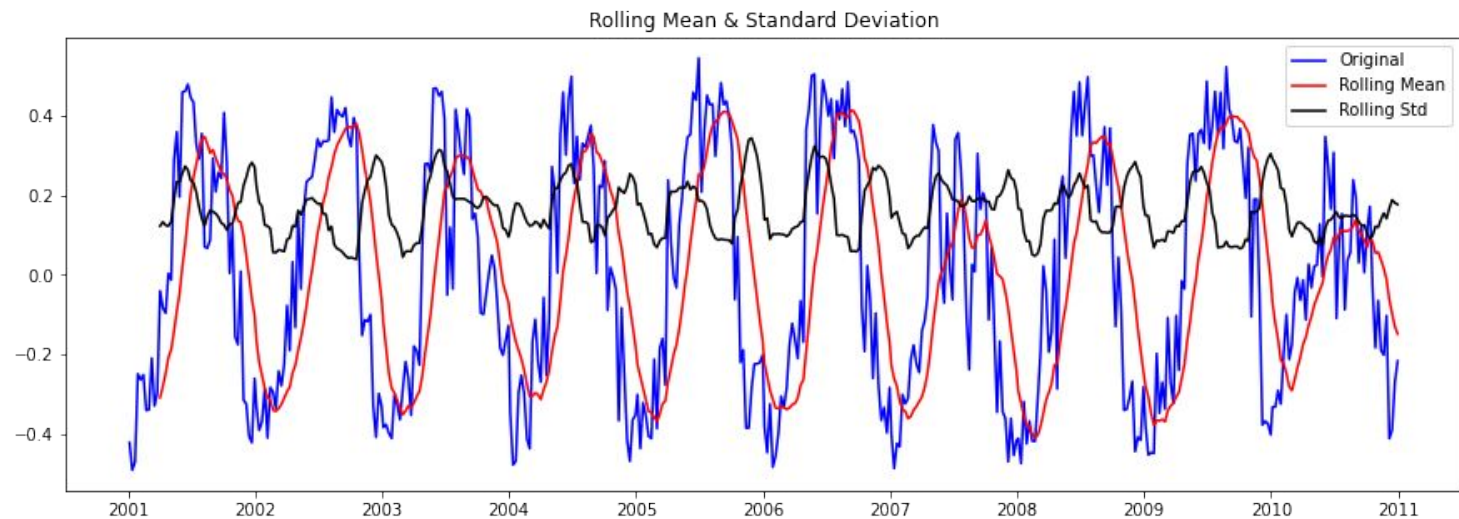


Se procede a graficar los componentes para comparar si hay correlación, y en algunos se nota a simple vista

Time Series

Estacionareidad

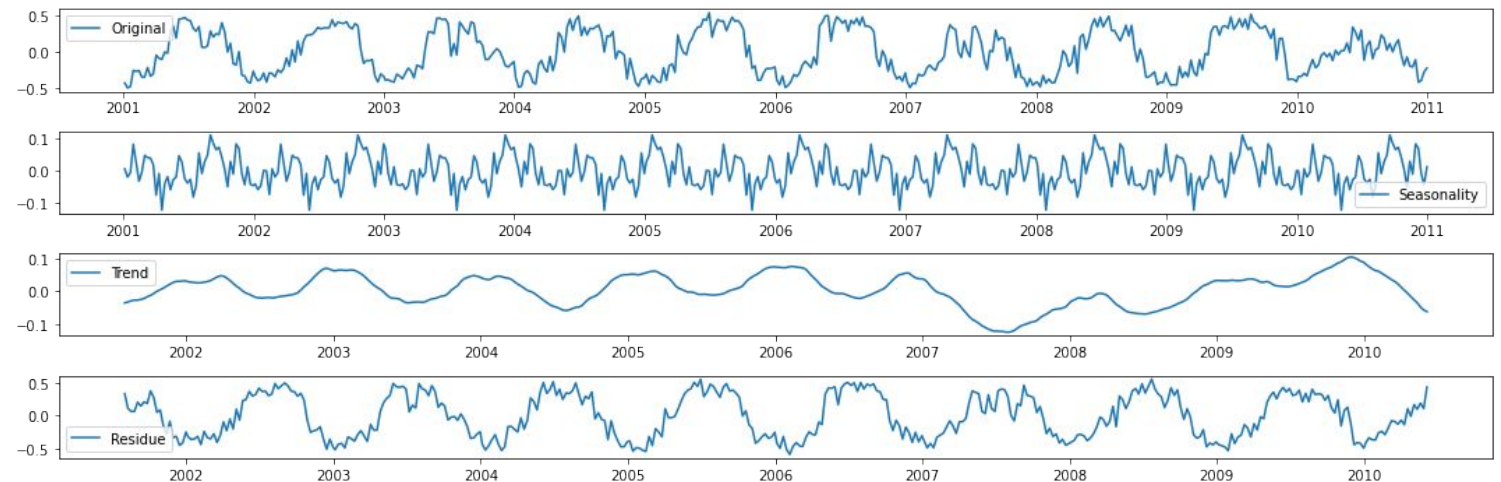
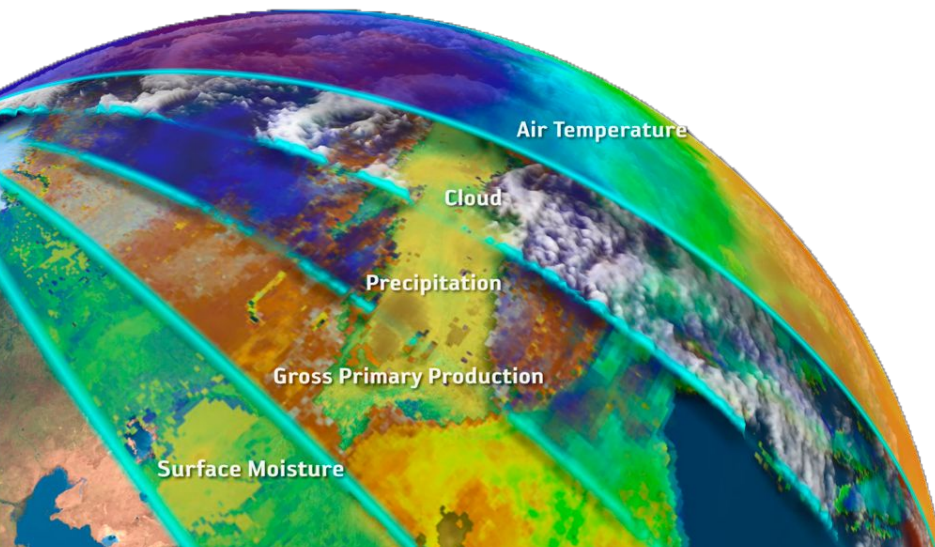
Luego se procede a realizar la prueba de estacionareidad de Dickey-Fuller, la cual nos dirá si la distribución y los parámetros de la serie no varían con el tiempo



Frequency Analysis

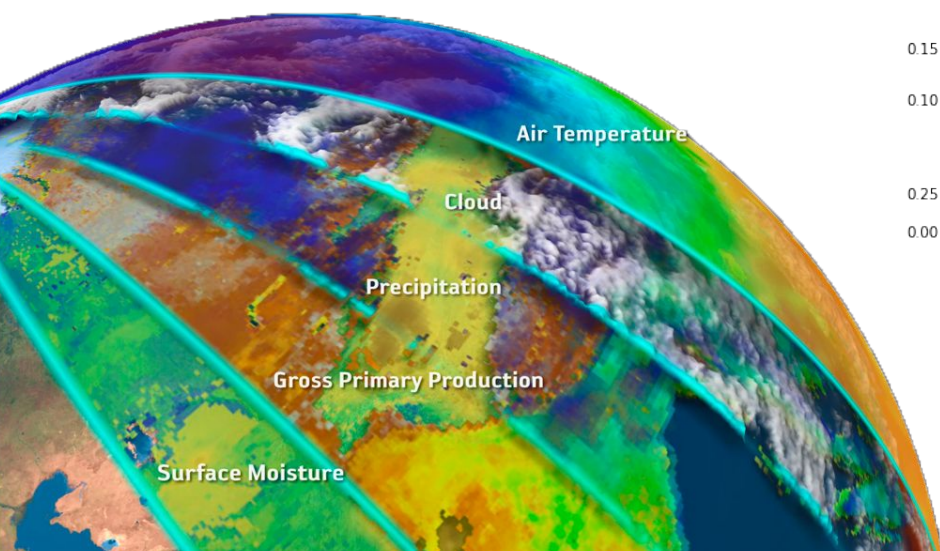
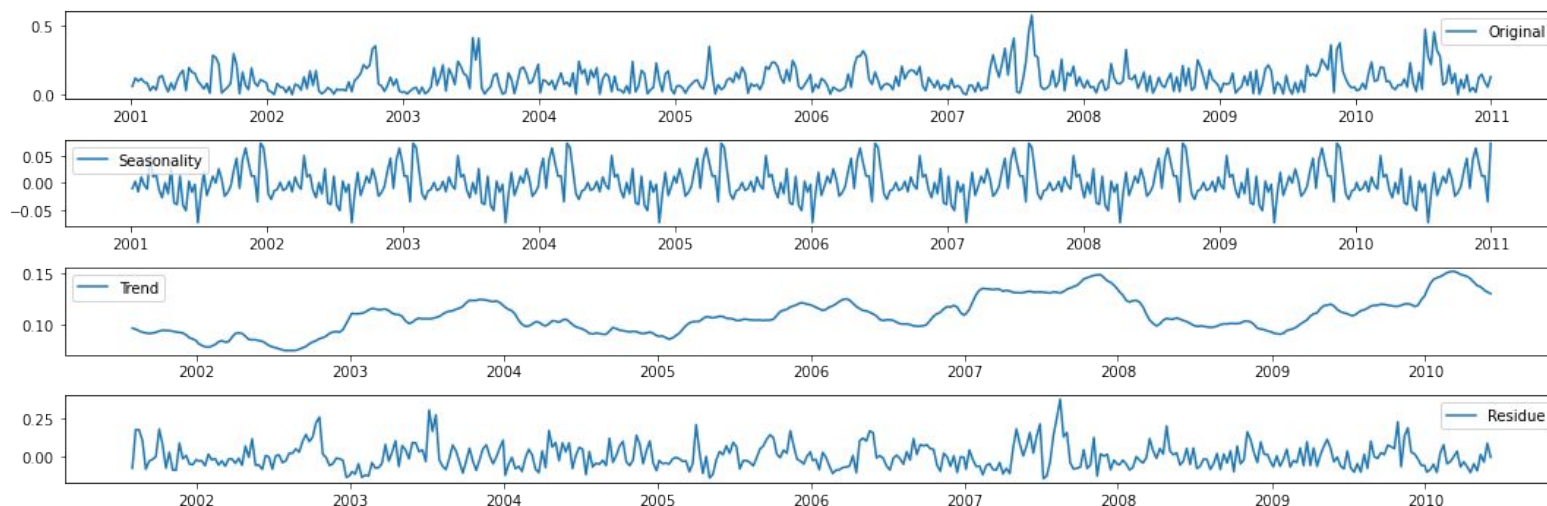
Estacionalidad

Se realiza descomposición estacional para eliminar trends y suavizar las series, pero aún así en los residuos se puede notar estacionalidad, por lo cual se debe transformar las series al espacio de frecuencias y eliminar las dominantes, para lo cual se analiza con Fourier

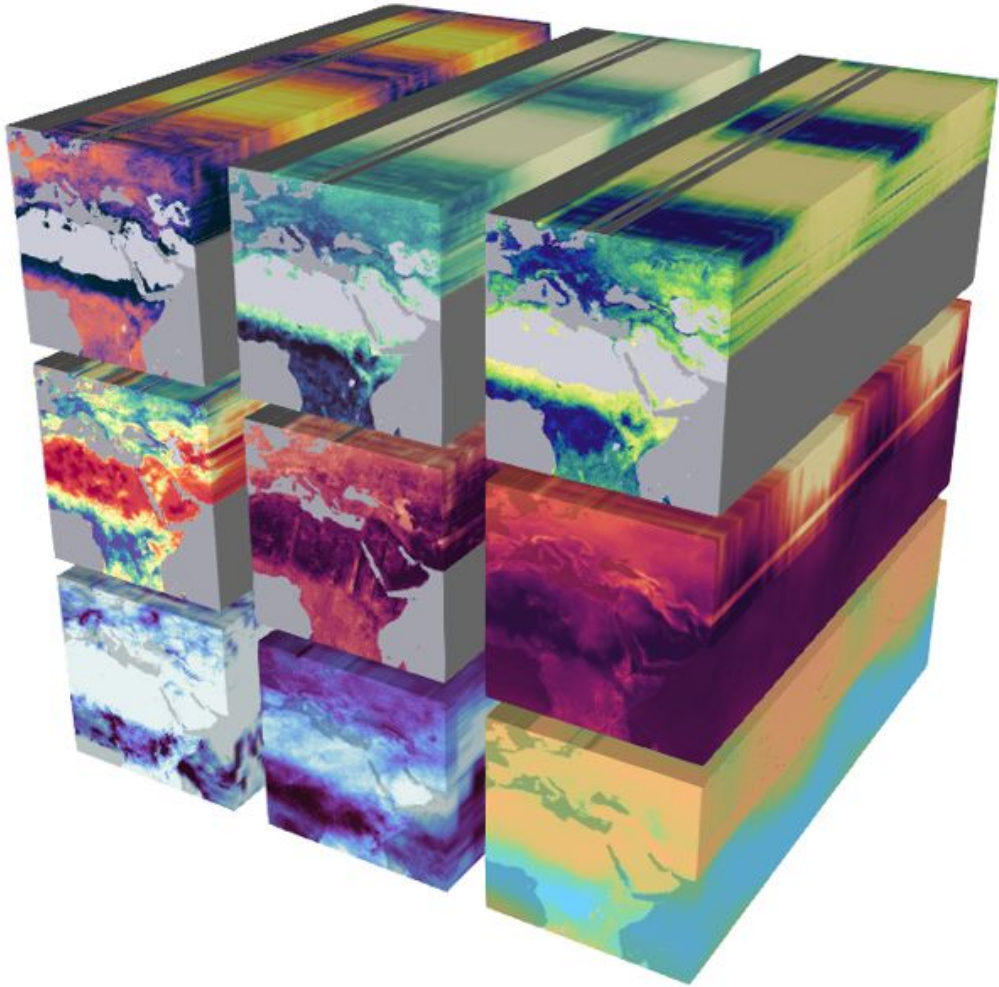


Fourier Transform

Al transformar con Fourier el comportamiento de las series dejan de tener un componente estacional fuerte, lo cual nos permite seguir el proyecto. A estas series transformada se les hacen las pruebas de Dickey-Fuller nuevamente para revisar si son estacionarias



VARMAX

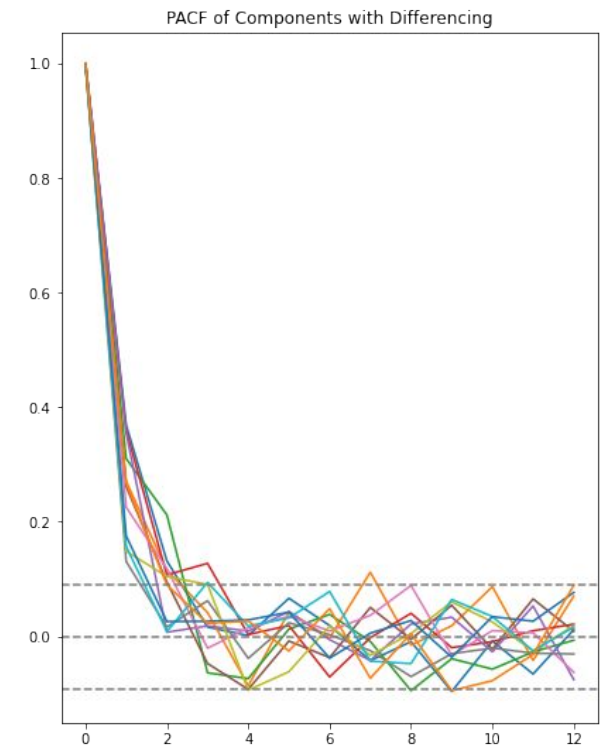
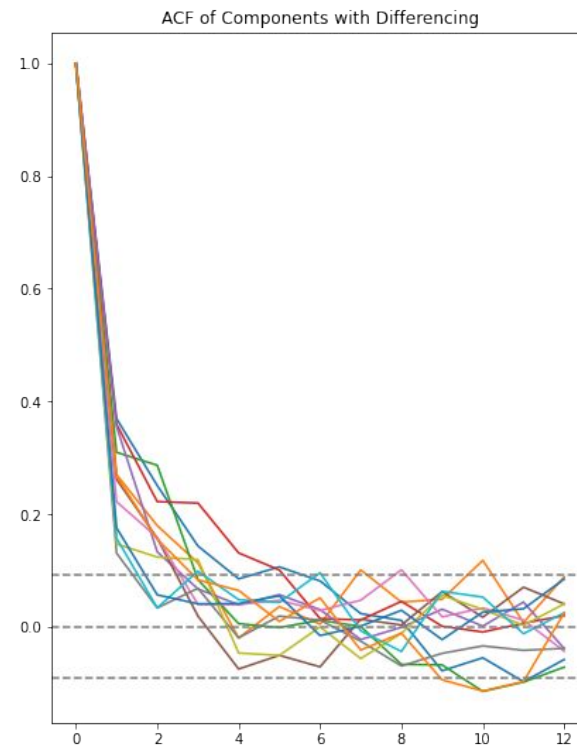


Dada una serie de tiempo multivariante, el procedimiento VARMAX estima los parámetros del modelo y genera pronósticos asociados con procesos de media móvil autorregresiva vectorial con modelos de regresores exógenos.

VARMAX

ACF & PACF

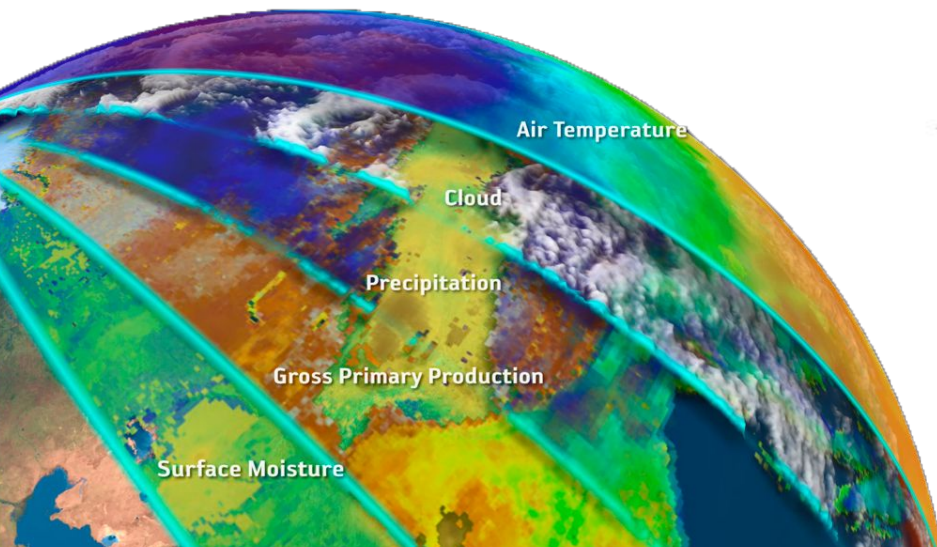
Se realizan ACF y PACF para hallar los parámetros óptimos AR y MA.



Estadístico de Durbin-Watson

Por último antes de predecir observaremos si hay correlación serial entre los residuos. Esta correlación quiere decir que hay un patrón de comportamiento que el modelo no explica en su totalidad

La prueba del estadístico de Durbin-Watson, el cual es un valor entre 0 y 4, nos dice que si el valor tiende a 2, no hay correlación entre residuos



$$DW = \frac{\sum_{t=2}^T ((e_t - e_{t-1})^2)}{\sum_{t=1}^T e_t^2}$$

```
component_0 : 1.91
component_1 : 1.93
component_2 : 1.9
component_3 : 1.91
component_4 : 1.9
component_5 : 1.96
component_6 : 1.93
component_7 : 2.02
component_8 : 1.98
component_9 : 2.0
component_10 : 1.96
component_11 : 1.89
```

Referencias

Jason Brownlee . (.). A Gentle Introduction to Markov Chain Monte Carlo for Probability. ., de . Sitio web:

<https://machinelearningmastery.com/markov-chain-monte-carlo-for-probability/>

Markowitz's Efficient Frontier in Python. ., de . Sitio web:

<https://medium.com/python-data/efficient-frontier-in-python-34b0c3043314>

Will Koehrsen. (.). Markov Chain Monte Carlo in Python. ., de . Sitio web:

<https://towardsdatascience.com/markov-chain-monte-carlo-in-python-44f7e609beg8>

Zachary M. Jones. (.). Monte-Carlo Methods for Prediction Functions. ., de . Sitio web:

<https://cran.r-project.org/web/packages/mmpf/vignettes/mmpf.html>

Milan Flach¹ , Fabian Gans¹ , Alexander Brenning^{2,4} , Joachim Denzler^{3,4,5} , Markus Reichstein^{1,4,5} , Erik Rodner^{3,4} , Sebastian Bathiany⁶ , Paul Bodesheim¹ , Yanira Guanche^{3,4} , Sebastian Sippel¹ , and Miguel D. Mahecha. (.). Multivariate anomaly detection for Earth observations: a comparison of algorithms and feature extraction techniques. ., de .

Sitio web: <https://www.earth-syst-dynam.net/8/677/2017/esd-8-677-2017.pdf>

