

# Interpolation of discontinuous optical data using Fourier series and Whittaker method: a toolbox written in python

## Interpolación de datos ópticos discontinuos mediante series de Fourier y Método Whittaker: un toolbox escrito en python

Humberto Cruz-Sanabria<sup>1</sup>, María Guadalupe Sánchez<sup>1</sup>, Santiago Belda<sup>2</sup>, Juan Pablo Rivera-Caicedo<sup>3</sup>, and Daniel Fajardo-Delgado<sup>1</sup>

<sup>1</sup> *TecNM, Instituto Tecnológico de Cd. Guzmán,*  
Departamento de Sistemas y Computación,  
Av. Tecnológico 100, Col. Centro, Cd. Guzmán, Jal., México, 49000.  
{m19291006,maria.sc1, daniel.fd}@cdguzman.tecnm.mx

<sup>2</sup> *Universidad de Valencia, Laboratorio de Procesado de Imágenes,*  
Catedrático José Beltrán 2, Paterna, Valencia, Spain, 46980  
santiago.belda@uv.es

<sup>3</sup> *Universidad Autónoma de Nayarit, CONACYT - Secretaría de Investigación y Posgrado,*  
Boulevard Tepic-Xalisco 325, Col. Centro, Tepic, Nay., México, 63155.  
jprivera@uan.edu.mx

### Abstract

The use of time series using optical data has been of great help in the study of the spatio-temporal variability of agricultural crops on a global scale. A recurring problem of optical data is that it is spatially and temporally discontinuous, with missing values due to a multitude of factors such as adverse weather conditions, low temporal resolution, sensor failures, among others. As a consequence, data gap filling solutions are needed for accurate crop monitoring. This article presents a standalone image processing toolbox written in python. This allows to perform different tasks in time series in Sentinel-2 products, which include: (1) generating time series with continuous spatial data from discontinuous data using two interpolation methods: Fourier and Whittaker series, (2) detecting parameters phenological (beginning of season, highest peak and end of season, that is, key stages of crop growth) and (3) Reconstructing maps of some spectral index of a day for which there is no data. The toolbox has been tested in an area cultivated with sugar cane, and the results show that Whittaker method best interpolates the data. This toolbox will bene-

fit farmers for better monitoring of sugarcane cultivation, considering all the days of the development cycle and expressed through a spectral index.

**Keywords**— Toolbox, Interpolation, Fourier, Whittaker

### Resumen

El uso de series temporales utilizando datos ópticos ha sido de gran ayuda en el estudio de la variabilidad espacio-temporal de cultivos agrícolas a escala global. Un problema recurrente en el uso de datos ópticos es que son espacial y temporalmente discontinuos, con valores perdidos debido a multitud de factores como puede ser condiciones climáticas adversas, baja resolución temporal, fallos en el sensor, entre otras. Como consecuencia, se necesitan soluciones de relleno de huecos de datos para un monitoreo preciso de los cultivos. En el presente artículo se expone un toolbox de procesamiento de imágenes escrito en python. Éste permite realizar diferentes tareas en series temporales en productos Sentinel-2, las cuales incluyen: (1) generar series temporales con datos espacialmente continuos a partir de datos discontinuos utilizando dos métodos de interpolación: Series

de Fourier y Whittaker, (2) detectar parámetros fenológicos (inicio de temporada, pico más alto y fin de la temporada, es decir etapas clave de crecimiento del cultivo) y (3) Reconstruir mapas de algún índice espectral de un día del cual no se tengan datos. El toolbox ha sido probado en una zona cultivada con caña de azúcar, y los resultados muestran que el método Whittaker interpola mejor los datos. Este toolbox beneficiará a los agricultores para un mejor monitoreo del cultivo de la caña de azúcar, considerando todos los días del ciclo de desarrollo y expresado a través de un índice espectral.

**Palabras clave**— Toolbox, Interpolación, Fourier, Whittaker

## I. Introducción

El estudio de la variabilidad espacio-temporal de la tierra ha sido un área de interés en la teledetección [1]. Específicamente en la agricultura, es un requisito clave para estudiar los cambios morfológicos que sufren las plantas en su ciclo de vida, así como también comprender como éstas responden a su entorno [2].

El uso de series temporales utilizando datos ópticos ha sido de gran ayuda al monitoreo a escala global de los cambios estacionales de áreas con vegetación a través de sus propiedades espectrales.

Generalmente estos datos ópticos son obtenidos a través de sensores sobre plataformas espaciales como los satélites, pues tienen la ventaja de que las áreas geográficas de interés se revisan en un ciclo regular, además de que algunos ofrecen el acceso a los datos sensados de manera gratuita.

Un ejemplo de ello es la misión Sentinel-2 del programa COPERNICUS de la Agencia Espacial Europea (ESA, por sus siglas en inglés) [3], la cual consiste en una constelación de dos satélites en órbita polar, en fases de 180° entre sí, con un tiempo de revisita de 5 días sobre un mismo punto, usando el sensor Multispectral Instrument (MSI, por sus siglas en inglés), el cual es capaz de capturar 13 bandas del espectro electromagnético a 10m, 20m y 60m de resolución [4].

Uno de los principales problemas de los datos ópticos de teledetección es que son espacial y temporalmente discontinuos, con valores perdidos debido a multitud de factores como puede ser condiciones climáticas adversas, baja resolución temporal, fallos en el sensor, entre otras [5]. Como consecuencia, se necesitan soluciones de relleno de huecos de datos para un monitoreo preciso del cultivo [1].

Actualmente existen toolboxes que implementan diversas técnicas de interpolación que permiten rellenar huecos en series temporales (Series de Fourier, Whittaker, k-nearest neighbors regression, entre otros).

Algunos ejemplos de estas toolboxes son: DATimeS [1] escrita en MATLAB y que utiliza diversas técnicas de aprendizaje automático para el llenado de huecos y la detección de tendencias de fenología de la vegetación; TIMESTATS [6] escrita en Interactive Data Language (IDL) y distribuida libremente con la máquina virtual IDL, permite el análisis de archivos de datos de teledetección multitemporales, como MODIS, AVHRR, MERIS y SPOT-Vegetation, mediante métodos paramétricos y no paramétricos para detección de tendencias; BeeBox [7] es una herramienta basada en la web con un enfoque en la apicultura mediante el análisis de patrones de crecimiento en recursos florales usando MODIS y otros datos; EO Time Series Viewer [8] es un complemento QGIS de código abierto y gratuito para visualizar y etiquetar datos de series de tiempo de observación de la tierra basados en rásters.

En el presente artículo se expone un toolbox de procesamiento de imágenes escrito en python 3.8. Este toolbox permite realizar diferentes tareas en series temporales en productos Sentinel-2 y funciona como un complemento para el Sentinel Application Platform (SNAP) [9].

Las tareas del toolbox incluyen: (1) generar series temporales con datos espacialmente continuos a partir de datos discontinuos utilizando dos métodos de interpolación: Series de Fourier y método de Whittaker, (2) detectar parámetros fenológicos (inicio de temporada, pico más alto y fin de la temporada, es decir etapas clave de crecimiento del cultivo) y (3) reconstruir mapas de algún índice espectral de un día del cual no se tengan datos.

El artículo está estructurado de la siguiente manera: en la sección II se describen los datos utilizados del Sentinel-2 para realizar las pruebas, el área de estudio y el índice utilizado para el monitoreo, así mismo se describen las especificaciones de la misión Sentinel-2. En la sección III se detalla cada uno de los procesos del toolbox, así como también los métodos de relleno de huecos donde no se tienen datos. Finalmente, en las secciones IV y V se muestran los resultados obtenidos y las conclusiones de las pruebas realizadas, respectivamente.

## II. Materiales

### II.1. Área de estudio

El área de estudio está ubicada en el municipio del Rincón, Jalisco, México con coordenadas (19° 40'01.4" N, 103° 24'42.6" W). El terreno se encuentra sembrado con la variedad de caña de azúcar ATEMex 96-40 y los datos registrados corresponden al valor del índice de área foliar (LAI, por sus siglas en inglés) para el ciclo de corte resoca 1, que inició en febrero de 2019 y terminó en marzo de 2020.

## II.2. Datos de estudio

Para este trabajo se utilizaron los datos del sensor MSI, que esta abordo de la misión Sentinel-2 de la Agencia Espacial Europea (ESA, por sus siglas en inglés), esta misión consiste de una constelación de dos satélites en órbita polar, en fases de 180° entre sí. En la Fig. 1 se muestra el diseño de ambos satélites. El objetivo de la misión es monitorear la variabilidad en las condiciones de la superficie terrestre, con su gran ancho de franja y alto tiempo de revisión (10 días en el ecuador con un satélite y 5 días con dos satélites en condiciones sin nubes, lo que da como resultado 2-3 días a mediados de latitudes). La Tabla 1 muestra el resumen de la resolución espectral y espacial del sensor MSI ofrece datos en 13 bandas distintas del espectro electromagnético, las imágenes capturadas tienen una resolución de 10, 20 y 60 metros por píxel [5]. Los datos fueron descargados del sitio web de Sentinels Scientific Data Hub de la ESA, con un nivel de nubosidad menor al 50 %, esto debido a que es el umbral en donde las imágenes disponibles no presentaban nubes en el área de estudio. En total se obtuvieron 54 imágenes que se capturaron del satélite durante el periodo de evaluación del cultivo y que corresponden a un porcentaje de nubosidad menor al 50 %.

**Tabla 1:** Resolución de las bandas del Sentinel-2

Banda	Intervalos de longitud de onda	Resolución de píxel
1	443 nm	60 m
2	540 nm	10 m
3	560 nm	10 m
4	665 nm	10 m
5	705 nm	20 m
6	740 nm	20 m
7	783 nm	20 m
8	842 nm	10 m
8a	965 nm	20 m
9	940 nm	60 m
10	1375 nm	60 m
11	1610 nm	20 m
12	2190 nm	20 m



**Figura 1:** Sentinel-2

Fuente: ESA

## II.3. Índice de área foliar (LAI)

LAI (Índice de área foliar, por sus siglas en inglés) es un parámetro utilizado para estudiar los procesos fisiológicos en ecosistemas forestales, tales como pronosticar el crecimiento, la productividad o el intercambio gaseoso. LAI hace referencia al conjunto de masa foliar de una cobertura vegetal y, por tanto, a su capacidad de fotosíntesis relacionado con el contenido de pigmentos [10]. La estimación de este índice se realizó mediante el software SNAP utilizando la función *Biophysical Processor*. El LAI se obtiene por medio de una red neuronal de tres capas con la siguiente topología: capa de entrada de 11 neuronas (8 bandas y 3 ángulos de la geometría de observación e iluminación), la capa oculta con 5 neuronas y función de transferencia tipo tangente sigmoideal, como capa de salida una neurona con una función de activación lineal. La ecuación 1 muestra la función de transferencia.

$$f(x) = \frac{1}{1 - e^{-x}} \quad (1)$$

## II.4. Métricas de calidad

Para evaluar la metodología propuesta, se ha utilizado dos métricas de calidad, error cuadrático medio (MSE, por sus siglas en inglés) y el índice de similitud estructural (SSIM, por sus siglas en inglés), cuyas ecuaciones se describen en 2 y 3 [11], respectivamente.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (2)$$

donde  $n$  = número de datos,  $Y_i$  = valores observados,  $\hat{Y}_i$  = los valores predichos. Un valor menor indica que hay mayor similitud entre  $Y_i$  y  $\hat{Y}_i$ .

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (3)$$

Un valor de SSIM cercano a uno es mejor para esta métrica.

## III. Metodología

En la Fig. 2 se presenta la metodología general implementada del toolbox.

A continuación, se describe cada uno de los procesos del método propuesto.

### III.1. Lectura de mapas

El primer paso en la metodología es la lectura de los mapas del índice espectral a utilizar. Algunos de ellos son, el índice de área foliar (LAI), el Índice de Vegetación

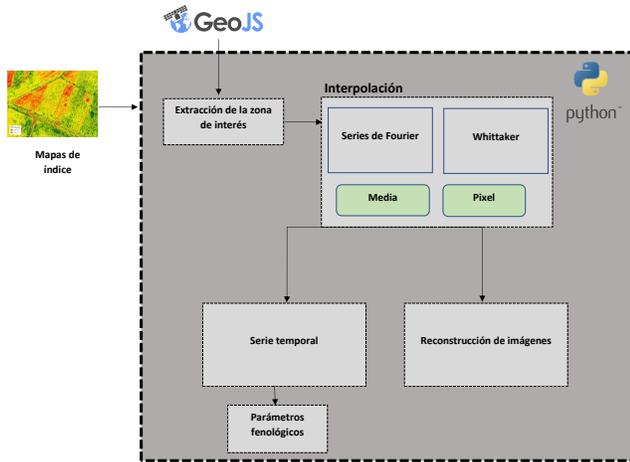


Figura 2: Metodología general empleada

Perpendicular (PVI, por sus siglas en inglés) [12], el Índice de Vegetación de diferencia normalizada Verde (GNDVI, por sus siglas en inglés) [13]. Sin embargo, para este estudio solo se ha considerado el LAI. Se ha utilizado el formato *tif* para el tratamiento de las imágenes, la lectura del formato se realiza mediante la API Rasterio y el manejo de los píxeles mediante *numpy* de *python*.

### III.2. Extracción de la zona de interés

Una vez que las imágenes se leen, se utiliza un polígono en formato *shapefile* para extraer los píxeles que solo comprenden la zona de interés, este proceso se realiza utilizando la librería *fiona* de *python*. El polígono se puede generar mediante la herramienta en línea *GeoJSON*.

### III.3. Interpolación

Una vez que se extrae la zona de interés de las imágenes, se utiliza dos métodos de interpolación: Series de Fourier y Whittaker, los cuales se explican a continuación.

#### III.3.1. Series de Fourier

Las series de Fourier es la descomposición de una señal periódica en términos de señales básicas (senos y cosenos) cuyas frecuencias son múltiplos de la señal original [14]. La serie de Fourier de una función periódica  $f(x)$  de periodo  $T$ , también conocida como señal, definida en un intervalo de longitud  $T$  esta dada por la ecuación. 4.

$$\frac{a_0}{2} \sum_{n=1}^{\infty} (a_n \cos(n\omega_0 x) + b_n \sin(n\omega_0 x)), \quad (4)$$

$$\text{donde } \omega_0 = (2\pi)/T, a_0 = \frac{1}{T} \int_T f(x) dx,$$

$$a_n = \frac{1}{T} \int_T f(x) \cos(n\omega_0 x) dx,$$

$$b_n = \frac{1}{T} \int_T f(x) \sin(n\omega_0 x) dx.$$

Para efectuar la regresión de la serie temporal con este método, se proporciona el grado de la serie de fourier y posteriormente se podrá efectuar la regresión a través de la media aritmética o utilizando algún píxel de los mapas.

#### III.3.2. Método Whittaker

Este método fue desarrollado por Eilers [15], se trata de un método de ajuste flexible dado por la Ec III.3.2.

$$O(z) = (y - z)^T W_0 (y - z) + \lambda_s z^T D_s^T D_s z \quad (5)$$

En el presente trabajo se ha elegido un valor de  $\lambda$  de 500 de manera arbitraria. Para efectuar este método de regresión el usuario solamente deberá proporcionar si desea que la serie sea en base a la media o por píxel.

#### III.4. Serie temporal

Una vez efectuada la regresión por uno de los métodos de interpolación, se presenta una gráfica donde se visualiza la serie temporal con los datos originales y la serie temporal generada con el método seleccionado, esta se genera utilizando *matplotlib* de *python*.

#### III.5. Parámetros fenológicos

Adicionalmente, el toolbox permite obtener los parámetros fenológicos (del cultivo a través del índice) de inicio de temporada, pico más alto y final de la temporada.

#### III.6. Reconstrucción de imágenes

Este proceso permite generar mapas a partir de alguno de los dos métodos de interpolación (Whittaker y Series de Fourier) de algún día del cual no se tenga datos.

## IV. Resultados

En esta sección se muestran los resultados obtenidos al aplicar cada uno de los procesos del toolbox al área de estudio.

En las Figs. 3 y 4 se muestran los resultados obtenidos al aplicar Series de Fourier con grado cuatro y el método Whittaker a la serie temporal. En ambos casos el eje de las  $x$  representa los días del ciclo de vida del cultivo y el eje  $y$  el valor del índice LAI. Se puede observar que se presenta la serie temporal original con cada dato registrado marcado con un punto, y la serie aplicando el método de interpolación especificado.

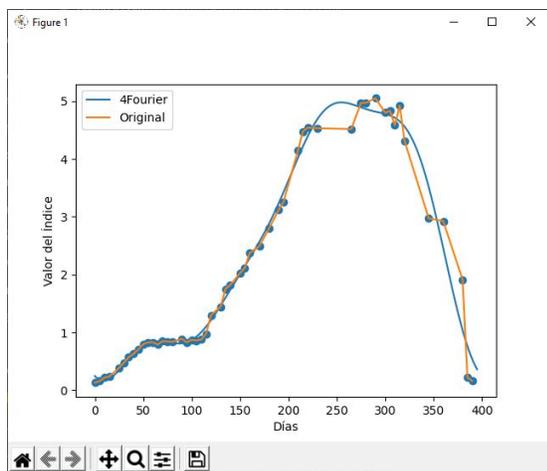


Figura 3: Interpolación: Serie de fourier de grado 4 efectuada sobre la media del mapa

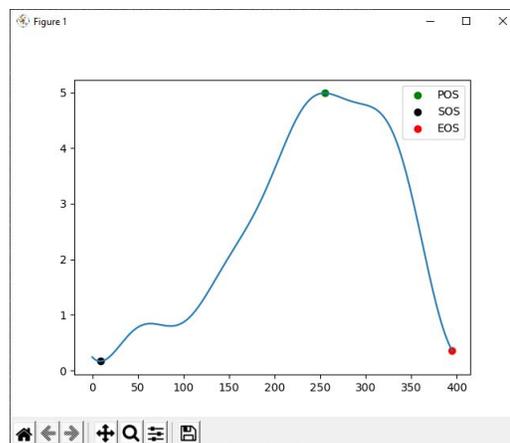


Figura 5: Parámetros fenológicos

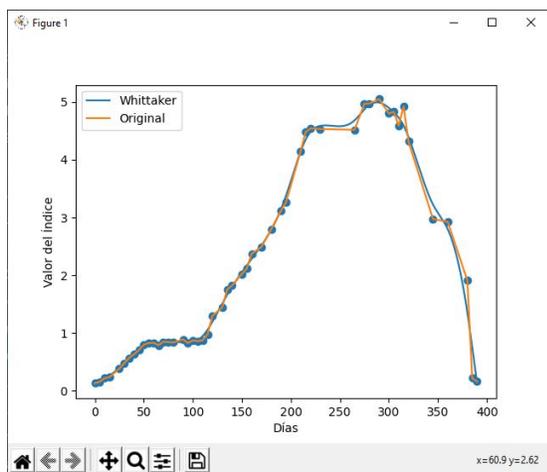


Figura 4: Interpolación: Método de whittaker efectuado sobre la media del mapa

Se puede observar que el método de Whittaker se ajusta más a los datos reales de la serie temporal que el método de Series de Fourier. Sin embargo, hay que considerar la calidad de los datos capturados, pues éstos a pesar de tener una etapa de pre-procesamiento, aun pueden presentar ruido que puede afectar el valor del índice.

#### IV.1. Parámetros Fenológicos

Los resultados para este proceso se muestran en la Fig. 5. El inicio de la temporada esta plasmado en la serie por el punto (SOS), el pico más alto esta plasmado en la serie por el punto (POS) y el final de la temporada esta plasmado en la serie por el punto (EOS).

#### IV.2. Reconstrucción de imágenes

Con el objetivo de probar la precisión de cada método de interpolación, se optó por eliminar un dato  $x$  en la serie original y de esta manera estimar el mapa para el índice LAI de ese día. El día elegido aleatoriamente fue el 145 de 395 días (que es la duración del ciclo del cultivo de la caña de azúcar). En la Fig. 6 se muestra el dato eliminado dentro de la serie temporal.

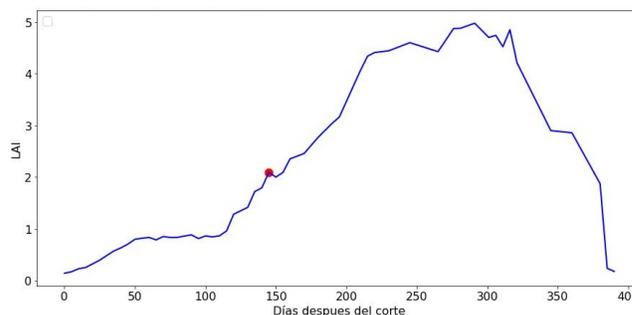


Figura 6: Día 145 de la serie eliminado

El resultado para cada método de interpolación se encuentra en las Figs. 7 y 8. En la Tabla 2 se muestra una comparativa de los métodos de interpolación utilizando las métricas MSE y SSIM. Como puede observarse, el método Whittaker es el que mejor se comporta logrando un MSE de 0.0177 y un SSIM de 0.9884, siendo éste el que mejor se ajustó a los valores reales del mapa LAI en el cultivo de caña de azúcar.

Tabla 2: Comparativa entre ambos métodos.

	Serie de Fourier	Whittaker
MSE	0.0293	0.0177
SSIM	0.9797	0.9884

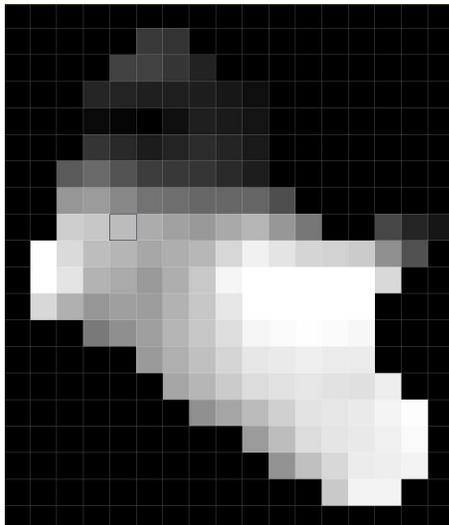


Figura 7: Interpolación de imagen para el día 145 utilizando serie de fourier de grado 4

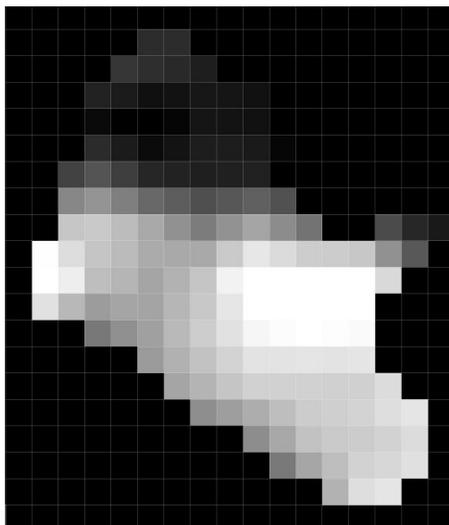


Figura 8: Interpolación de imagen para el día 145 utilizando el método de whittaker

## V. Conclusiones

El toolbox presentado en este artículo, es un conjunto de herramientas que permite el análisis de series temporales escrito en python 3.8, el cual reúne dos métodos de interpolación (Series de Fourier y Whittaker), las cuales son: (1) generar series temporales con datos espacialmente continuos a partir de datos discontinuos, (2) detectar parámetros fenológicos (inicio de temporada, pico mas alto y fin de la temporada) y (3) reconstruir mapas de algún índice espectral de un día del cual no se tengan datos. Con el fin de probar el toolbox, se seleccionó el cultivo de caña de azúcar sobre el cual se efectuaron las tres herramientas incluidas en el software. Los resultados

obtenidos fueron satisfactorios pues para cada uno de las herramientas, las técnicas de interpolación utilizadas se ajustaron satisfactoriamente a los datos discontinuos de la serie temporal del cultivo de caña de azúcar, además de lograr identificar los parámetros fenológicos de éste. Como trabajo futuro se plantea implementar mas técnicas de interpolación e incluir análisis de tendencias de series temporales, además de incluir técnicas de extrapolación como puede ser un modelo autorregresivo integrado de promedio móvil o ARIMA.

## Agradecimientos

Gracias al programa de apoyo de becas nacionales del CONACYT a través del numero de beca 005412.

## Referencias

- [1] Santiago Belda y col. «DATimeS: A machine learning time series GUI toolbox for gap-filling and vegetation phenology trends detection». En: *Environmental Modelling & Software* 127 (2020), pág. 104666.
- [2] Lauricio Endres y col. «Morphological changes recorded in different phenophases of sugarcane plants subjected to water stress in tropical field conditions». En: *Australian Journal of Crop Science* 12.7 (2018), págs. 1041-1050.
- [3] S Jutz y MP Milagro-Pérez. «Copernicus: the European Earth Observation programme». En: *Revista de Teledetección* 56 (2020).
- [4] Matthias Drusch y col. «Sentinel-2: ESA's optical high-resolution mission for GMES operational services». En: *Remote sensing of Environment* 120 (2012), págs. 25-36.
- [5] Sivasathivel Kandasamy y col. «A comparison of methods for smoothing and gap filling time series of remote sensing observations—application to MODIS LAI products». En: *Biogeosciences* 10.6 (2013), págs. 4055-4071.
- [6] Thomas Udelhoven. «TimeStats: A software tool for the retrieval of temporal patterns from global satellite archives». En: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4.2 (2010), págs. 310-317.
- [7] Jonathan Arundel y col. «A web-based application for beekeepers to visualise patterns of growth in floral resources using MODIS data». En: *Environmental Modelling & Software* 83 (2016), págs. 116-125.

- [8] Benjamin Jakimow y col. «Visualizing and labeling dense multi-sensor earth observation time series: The EO Time Series Viewer». En: *Environmental Modelling & Software* 125 (2020), pág. 104631.
- [9] Marco Zuhlke y col. «SNAP (sentinel application platform) and the ESA sentinel 3 toolbox». En: *Sentinel-3 for Science Workshop*. Vol. 734. 2015, pág. 21.
- [10] Rafael M<sup>a</sup> Navarro-Cerrillo. «Evaluación de variables biofísicas mediante teledetección en pinares mediterráneos». En: *Ambienta* 105 (2013), págs. 78-88.
- [11] Gabriel Prieto Renieblas y col. «Structural similarity index family for image quality assessment in radiological images». En: *Journal of medical imaging* 4.3 (2017), pág. 035501.
- [12] Charles R Perry Jr y Lyle F Lautenschlager. «Functional equivalence of spectral vegetation indices». En: *Remote sensing of environment* 14.1-3 (1984), págs. 169-182.
- [13] Anatoly A Gitelson, Yoram J Kaufman y Mark N Merzlyak. «Use of a green channel in remote sensing of global vegetation from EOS-MODIS». En: *Remote sensing of Environment* 58.3 (1996), págs. 289-298.
- [14] Cornelius Lanczos y John Boyd. *Discourse on Fourier series*. SIAM, 2016.
- [15] Paul HC Eilers. «A perfect smoother». En: *Analytical chemistry* 75.14 (2003), págs. 3631-3636.