

Valero-Jorge, A.; Casterad, M.; Alcalá, J. TSGenerator: librería R de código abierto para el procesado integrado de productos fenológicos HR-VPP de Copernicus

# TSGenerator: librería R de código abierto para el procesado integrado de productos fenológicos HR-VPP de Copernicus

Valero-Jorge, Alexey <sup>1</sup> Casterad, M<sup>a</sup> Auxiliadora <sup>1</sup> Alcalá, José-Tomás <sup>2</sup>

<sup>1</sup> Dpto. de Sistemas Agrícolas, Forestales y Medio Ambiente (Unidad asociada a EEAD-CSIC Suelos y Riegos), Centro de Investigación y Tecnologías Agroalimentaria de Aragón, España

<sup>2</sup> Dpto. de Métodos Estadísticos, Instituto Universitario de Investigación en Matemáticas y Aplicaciones (IUMA), Universidad de Zaragoza, España

ORCID: Valero-Jorge 0000-0002-5993-7346 Casterad 0000-0003-4458-6966 Alcalá 0000-0001-7549-8825

Correspondencia: [avalero@cita-aragon.es](mailto:avalero@cita-aragon.es) [acasterad@cita-aragon.es](mailto:acasterad@cita-aragon.es) [jtalcala@unizar.es](mailto:jtalcala@unizar.es)

## RESUMEN

El conjunto de productos *High Resolution Vegetation Phenology and Productivity* del *Copernicus Land Monitoring Service* proporciona series temporales de índices de vegetación y parámetros fenológicos derivados de Sentinel-2, con creciente aplicación en agricultura y monitorización de ecosistemas. Sin embargo, su explotación operativa requiere flujos complejos de descarga, filtrado de calidad y análisis temporal que limita su uso por usuarios no expertos. Este trabajo presenta TSGenerator, una librería R de código abierto que integra el procesamiento de productos HR-VPP, incluyendo descarga, preprocesamiento, extracción espacial y análisis de calidad e imputación, junto con aplicaciones Shiny orientadas a usuarios sin experiencia en programación. Su funcionalidad se valida mediante la monitorización de 1809 parcelas agrícolas durante el periodo 2018–2023, mostrando alta eficiencia en preprocesamiento y extracción espacial en equipos convencionales, siendo la descarga el principal cuello de botella. En el estudio se detecta una elevada variabilidad en la completitud de las series, así como patrones estacionales no aleatorios de datos ausentes, poniendo de relieve la importancia del análisis de calidad. TSGenerator reduce barreras técnicas y facilita el uso operativo de productos HR-VPP en aplicaciones agrícolas y ambientales..


**Palabras clave:** *Sentinel-2; NDVI; series temporales, programación, cultivos*


Fecha de recepción: 5 febrero 2026 · Fecha de aceptación: 5 febrero 2026

# TSGenerator: librería R de código abierto para el procesamiento integrado de productos fenológicos HR-VPP de Copernicus


Valero-Jorge, Alexey <sup>(1)</sup>, Casterad, M<sup>a</sup> Auxiliadora <sup>(1)</sup>, Alcalá, José-Tomás <sup>(2)</sup>

<sup>(1)</sup> Dpto. de Sistemas Agrícolas, Forestales y Medio Ambiente (Unidad asociada a EEAD-CSIC Suelos y Riegos), Centro de Investigación y Tecnologías Agroalimentaria de Aragón, España.

 0000-0002-5993-7346, avalero@cita-aragon.es.

 0000-0003-4458-6966, acasterad@cita-aragon.es.

<sup>(2)</sup> Dpto. de Métodos Estadísticos, Instituto Universitario de Investigación en Matemáticas y Aplicaciones (IUMA), Universidad de Zaragoza, España.

 0000-0001-7549-8825, jtalcala@unizar.es.

**Resumen:** El conjunto de productos *High Resolution Vegetation Phenology and Productivity* del *Copernicus Land Monitoring Service* proporciona series temporales de índices de vegetación y parámetros fenológicos derivados de Sentinel-2, con creciente aplicación en agricultura y monitorización de ecosistemas. Sin embargo, su explotación operativa requiere flujos complejos de descarga, filtrado de calidad y análisis temporal que limita su uso por usuarios no expertos. Este trabajo presenta *TSGenerator*, una librería R de código abierto que integra el procesamiento de productos HR-VPP, incluyendo descarga, preprocesamiento, extracción espacial y análisis de calidad e imputación, junto con aplicaciones *Shiny* orientadas a usuarios sin experiencia en programación. Su funcionalidad se valida mediante la monitorización de 1809 parcelas agrícolas durante el periodo 2018–2023, mostrando alta eficiencia en preprocesamiento y extracción espacial en equipos convencionales, siendo la descarga el principal cuello de botella. En el estudio se detecta una elevada variabilidad en la completitud de las series, así como patrones estacionales no aleatorios de datos ausentes, poniendo de relieve la importancia del análisis de calidad. *TSGenerator* reduce barreras técnicas y facilita el uso operativo de productos HR-VPP en aplicaciones agrícolas y ambientales.

**Palabras clave:** Sentinel-2; NDVI; series temporales, programación, cultivos

## ***TSGenerator: open-source R library for integrated processing of Copernicus HR-VPP phenological products***

**Abstract:** *The High Resolution Vegetation Phenology and Productivity product suite from the Copernicus Land Monitoring Service provides time series of vegetation indices and phenological parameters derived from Sentinel-2, with growing applications in agriculture and ecosystem monitoring. However, its operational use requires complex workflows involving data download, quality filtering, and temporal analysis, which limits its use by non-expert users. This paper presents TSGenerator, an open-source R library that integrates HR-VPP product processing, including downloading, preprocessing, spatial extraction, quality analysis, and imputation, along with Shiny applications designed for users with no programming experience. Its functionality is validated by monitoring 1,809 agricultural plots during the 2018–2023 period, demonstrating high efficiency in preprocessing and spatial extraction on conventional hardware, with downloading being the main bottleneck. The study detects high variability in the completeness of the time series, as well as non-random seasonal patterns of missing data, highlighting the importance of quality analysis. TSGenerator reduces technical barriers and facilitates the operational use of HR-VPP products in agricultural and environmental applications.*

**Keywords:** *Sentinel-2; NDVI; time series, programming, crops*

## 1. INTRODUCCIÓN

Los productos *High Resolution Vegetation Phenology and Productivity* (HR-VPP) del *Copernicus Land Monitoring Service* (CLMS), disponibles en la plataforma WEKEO—de acceso libre previa autenticación y

disponibilidad condicionada por los tiempos de procesamiento del servicio—ofrecen series temporales de alta resolución de índices de la vegetación (NDVI, LAI, fAPAR, PPI) y parámetros fenológicos derivados de Sentinel-2, esenciales para aplicaciones en agricultura de precisión, monitorización de ecosistemas y

evaluación de impactos del cambio climático en España y Europa (CLMS, 2022). Sin embargo, su procesado implica desafíos significativos en cuanto a la descarga autenticada vía API de acceso a datos armonizados (HDA API WEKEO), el filtrado de píxeles de baja calidad mediante el indicador de calidad QFLAG2, el manejo de grandes volúmenes de datos y una extracción espacial de datos eficiente, lo que restringe su uso por usuarios con experiencia limitada en programación.

El paquete oficial *hda* desarrollado por WEKEO/EEA proporciona acceso nativo en R a la HDA API para descarga directa de productos CLMS, incluyendo HR-VPP, pero carece de flujos integrados específicos para limpieza, extracción temporal e interfaces accesibles para usuarios no expertos.

*TSGenerator* resuelve estas limitaciones mediante un paquete R de código abierto que integra descarga (actualmente vía *reticulate* y *hda* Python, con posible migración futura a *hda*), descarga de datos, preprocesamiento, extracción de series temporales y análisis de calidad e imputación, complementado con aplicaciones Shiny (Chang *et al.*, 2021) interactivas para un uso intuitivo sin programación avanzada. El objetivo de este trabajo es presentar *TSGenerator*, demostrar su funcionalidad con un caso práctico agrícola y promover su adopción en la comunidad de la teledetección.

## 2. EL PAQUETE TSGenerator

### 2.1. Arquitectura del paquete

*TSGenerator* se estructura en cuatro módulos funcionales interconectados que implementan un flujo de trabajo completo para el procesamiento de productos HR-VPP. El módulo de **descarga** (Tabla 1), permite búsquedas parametrizadas por producto: índices de vegetación incluido el producto de calidad QFLAG2, las trayectorias estacionales del PPI y los diferentes parámetros fenológicos; extensión espacial y rango temporal. La autenticación exigida para acceder a los datos se gestiona con el protocolo de autorización OAuth2 con la HDA API mediante la integración de la librería Python *hda* a través del paquete R *reticulate*.

El módulo de **preprocesamiento** implementa el análisis automatizado de la banda QFLAG2, identificando píxeles afectados, siguiendo las especificaciones técnicas del producto (CLMS, 2022). Los usuarios pueden definir configuraciones predeterminadas correspondientes a las clases de cielo despejado, proximidad de nubes, proximidad de sombras de nubes, proximidad de nubes extendidas y proximidad de sombras de nubes extendidas.

**Tabla 1.** Principales funciones del paquete TSGenerator.

Módulos	Funciones principales
Descarga	Download.VI; Download.VPP; Download.STPPI
Preprocesamiento	QFLAG2.Mask; get.Stack; renames.VI; renames.VPP
Extracción espacial	get.series.median; get.series.mean; get.series.VPP
Análisis de calidad e imputación	TSImpute; GAM.missing; general.Quality

El módulo de **extracción espacial** se utiliza para extraer series temporales de geometrías poligonales. La extracción implementa estrategias eficientes de agregación espacial mediante estadísticos descriptivos (media, mediana, desviación estándar) y gestión optimizada de memoria para *datasets* de gran escala. Para ello, utiliza las capacidades vectoriales y las funcionalidades ráster de las librerías *sf* (Pebesma, 2018) y *terra* (Hijmans, 2023).

Finalmente, el módulo de **análisis de calidad e imputación** calcula métricas cuantitativas de completitud temporal (porcentaje de observaciones válidas, distribución de gaps), identifica patrones *outliers* mediante un Modelo Aditivo Generalizado (GAM, ver fórmula en Valero-Jorge *et al.*, 2025). Como método de imputación, se ofrece el Filtro de Kalman, ampliamente usado en series temporales.

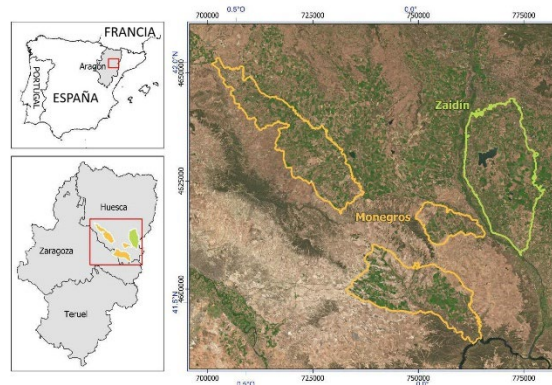
### 2.2. Interfaz gráfica Shiny

Complementariamente a las funciones programáticas, *TSGenerator* incorpora 8 aplicaciones *Shiny* que eliminan la barrera de entrada para usuarios sin experiencia en programación R. Las 3 aplicaciones de descarga permiten configurar consultas HDA mediante formularios interactivos y gestionar descargas masivas con seguimiento de progreso. Las 4 aplicaciones de preprocesamiento conforman un módulo completo para el preprocesamiento interactivo de las series temporales de imágenes descargadas. Y la aplicación GAM que permite realizar el análisis y visualización de patrones ausentes en las series temporales de índices espectrales extraídas. Todas estas aplicaciones pueden ser lanzadas utilizando la función *runTSapp*.

## 3. CASO DE ESTUDIO: MONITORIZACIÓN DE CULTIVOS HERBÁCEOS

### 3.1. Procedimiento

Para validar la funcionalidad de *TSGenerator*, se implementó un caso de uso en agricultura sobre 1809 parcelas agrícolas de cultivos herbáceos—cebada, trigo, guisante, maíz bajo sistema de mono y doble cultivo y girasol de doble cultivo—ubicadas en las zonas de riego de Monegros y Zaidín, en Aragón, España (Figura 1).



**Figura 1.** Ubicación geográfica del área de estudio.

Los pasos seguidos para la monitorización se esquematizan en la Figura 2. Primero, se descargaron

los productos NDVI, la banda de calidad QFLAG2 y parámetros fenológicos correspondientes al período 2018-2023, cubriendo la extensión territorial de todas las parcelas, mediante consultas parametrizadas a la API HDA. Luego se aplicó una máscara conservadora sobre QFLAG2 (clases cielo despejado, QFLAG = 1).

Las series temporales de NDVI se extrajeron agregando espacialmente la mediana por parcela, tras filtrar previamente los píxeles de baja calidad mediante QFLAG2. Se asegura así que la agregación se realiza únicamente sobre los píxeles válidos disponibles en cada fecha. Finalmente, se evaluó la calidad temporal de las series y los patrones de datos ausentes.

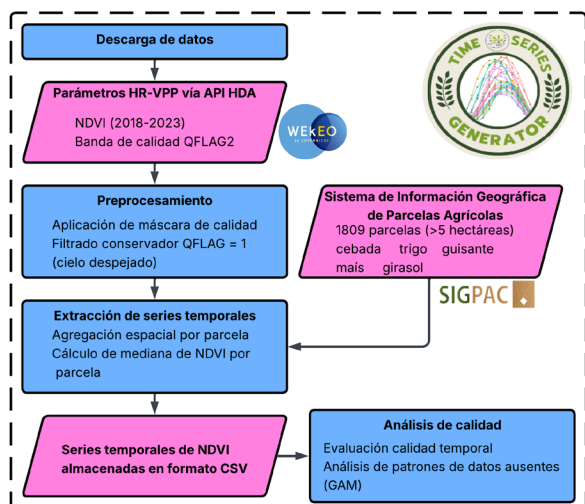


Figura 2. Flujo de procesos desarrollado.

Los cultivos presentes en cada parcela se identificaron utilizando registros anónimos de las declaraciones de la Política Agrícola Común proporcionadas por el Gobierno de Aragón. La información espacial y la delimitación geográfica de las parcelas se obtuvieron del Sistema de Información Geográfica de Parcelas Agrícolas (SIGPAC) en formato *shapefile*. Es importante resaltar que el objetivo del caso de uso propuesto no es presentar aplicaciones complicadas y análisis profundos, sino ejemplos sencillos y didácticos de cómo implementar el paquete *TSGenerator*. Todo el procesamiento computacional se llevó a cabo en un ordenador equipado con un procesador Intel Core i9-12900 de 12ª generación (16 núcleos y 24 hilos), 32 GB de memoria RAM y unidad de almacenamiento SSD de 500 GB.

### 3.2. Eficiencia operacional

La adquisición autenticada de 252 productos Sentinel-2 (NDVI y QFLAG2) para la escena 30TYM (período 2018–2023) a través de la HDA API de WEKEO tomó 8 horas y 24 minutos en total. El proceso es dominado por factores de red (ancho de banda del servidor y conexión del usuario), con un volumen aproximado de 80–90 GB, convirtiéndolo en una tarea intensiva en transferencia de datos. En equipos con hardware inferior, el impacto en el tiempo total sería mínimo, ya que el cuello de botella radica en la latencia y rendimiento de la red, no en las capacidades locales de CPU o RAM.

El preprocesamiento de los 252 productos (incluyendo enmascaramiento y renombrado de bandas) tomó

aproximadamente 3 horas y 42 minutos en total, con un uso de memoria pico inferior a 8 GB incluso al manejar paquete de imágenes multianuales. Esta optimización permite procesar grandes volúmenes sin requerir hardware especializado, y el módulo automatiza la selección de clases de calidad, reduciendo errores manuales y tiempo de ejecución en comparación con flujos *ad-hoc*.

El tiempo total para extraer series NDVI correspondientes a 6 años, para todas las parcelas fue de 54 minutos, con un consumo eficiente de recursos gracias al procesamiento por “trozos” y la gestión optimizada de memoria. La función *get.series.median* implementa paralelismo implícito, lo que le permite escalar adecuadamente a conjuntos de mayor tamaño. Este rendimiento se debe principalmente a dicho paralelismo y a una gestión eficiente de la memoria, más que al hardware específico utilizado, lo que sugiere una buena escalabilidad en equipos convencionales. En comparación con enfoques manuales como los utilizados en QGIS, *TSGenerator* reduce el tiempo en un factor de 5–10 al integrar todo el flujo en un solo paquete. Alternativamente, para análisis exploratorios sobre subconjuntos de parcelas, podrían emplearse enfoques basados en servicios remotos que eviten la descarga completa de imágenes.

### 3.3. Calidad de series temporales

El análisis de las series de NDVI mostró una alta variabilidad de calidad temporal tanto interanual como entre cultivos en la disponibilidad de observaciones válidas tras aplicar un QFLAG2 conservador (Figura 3).

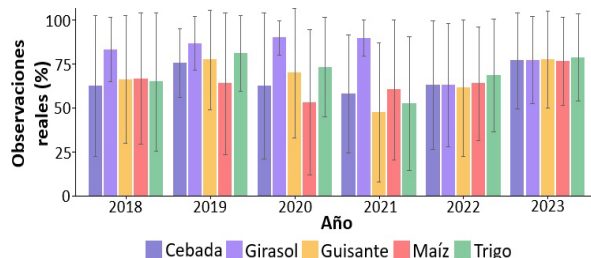


Figura 3. Porcentaje medio anual de observaciones disponibles.

El año 2023 registró el mayor porcentaje de observaciones disponibles (76,6–78,5 %) para todos los cultivos, constituyéndose como un año de baja cobertura nubosa, en general aceptable para la construcción de series temporales. Por el contrario, 2021 presentó los valores más bajos (47,6–89,5 %), lo que refleja condiciones meteorológicas adversas caracterizadas por mayor nubosidad persistente, que limitaron significativamente la adquisición de datos ópticos de calidad en la región de estudio, como es habitual en años con patrones climáticos más inestables en el noreste peninsular.

### 3.4. Patrones de datos ausentes

Los GAM ajustados para la probabilidad de datos ausentes en las series temporales de NDVI mostraron patrones estacionales significativos ( $p < 0,05$ ) en todos los cultivos, evidenciando una variación no lineal a lo largo del año (Figura 4). Al agrupar por ciclo fenológico,

los cultivos de invierno (cebada, trigo y guisante) presentaron una varianza explicada media del 7,0 % (rango 6,5–7,6 %), reflejando un componente estacional detectable pero débil, dominado por la alta variabilidad aleatoria inherente a los datos de calidad óptica de Sentinel-2. En contraste, los cultivos de verano (maíz y girasol) alcanzaron una varianza explicada media del 19,0 % (rango 17,2–20,8 %), indicando patrones de ausencias más estructurados y predecibles, probablemente asociados a periodos de nubosidad persistente durante sus fases críticas de desarrollo. En ambos grupos, las curvas GAM confirman que la ausencia de datos no es aleatoria, con máximos notorios en invierno-primavera y otoño, lo que subraya la necesidad de imputación robusta en esos periodos clave para garantizar series temporales fiables en monitorización fenológica agrícola. Estos resultados son coherentes con trabajos que destacan la necesidad de considerar la estructura temporal en imputación y análisis fenológicos (Simões *et al.*, 2021; Kong *et al.*, 2022 y Valero-Jorge *et al.*, 2025).

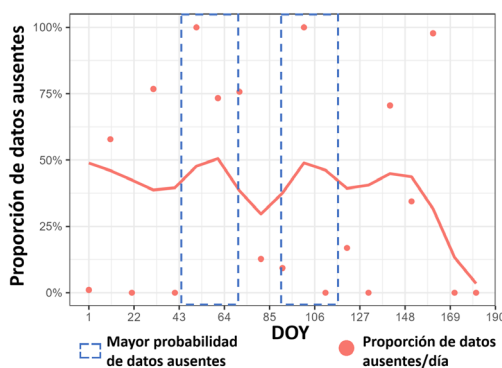


Figura 4. Probabilidad de datos ausentes en cebada de 2018.

#### 4. CONTRIBUCIÓN Y CONCLUSIONES

*TSGenerator* constituye una solución de código abierto para el procesamiento integrado de productos HR-VPP, con beneficios operacionales. Integra en un único paquete R las etapas principales del procesamiento de productos HR-VPP, reduciendo la dependencia de herramientas externas y facilitando el uso de estos datos por usuarios sin mucha experiencia en programación. Los resultados operacionales evidencian que la descarga es el principal cuello de botella, condicionado por la red, en coherencia con lo descrito previamente para flujos Sentinel-2 por Ranghetti *et al.* (2020).

El preprocesamiento, la extracción espacial y el análisis temporal son computacionalmente eficientes en equipos convencionales. El caso de estudio demuestra tiempos de ejecución compatibles con aplicaciones de monitorización a nivel de parcela y valida la utilidad del análisis de calidad para identificar patrones de gaps relevantes para series fenológicas. La relevancia del módulo de análisis de calidad se confirma al evidenciarse la marcada variabilidad interanual y entre cultivos en la completitud de las series, particularmente en 2021.

La disponibilidad de aplicaciones *Shiny* amplía el alcance de la herramienta respecto a otras del ecosistema R, facilitando la adopción operacional de HR-VPP por parte de usuarios no programadores.

El desarrollo futuro se orienta a integrar *hdar* y agregar otros métodos de imputación convencionales, reforzando la relevancia del paquete para el análisis fenológico basado en teledetección. *TSGenerator* está disponible actualmente en **GitHub** (<https://github.com/avalero92/TSGenerator>) y **Zenodo** (<https://zenodo.org/records/14936909>).

#### 5. AGRADECIMIENTOS

Este trabajo es parte del proyecto LAIKcA, I+D+i PID2021-124029OR-I00, financiado por MICIU/AEI/10.13039/501100011033 y FEDER/UE. El primer autor agradece la ayuda PRE2022-102328 financiada por MICIU/AEI/10.13039/501100011033 y por el FSE+.

#### 6. REFERENCIAS

- Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2021). shiny: Web Application Framework for R. R package version 1.7.1. <https://CRAN.R-project.org/package=shiny>
- Copernicus Land Monitoring Service. (2022). Product user manual of seasonal trajectories and VPP parameters: High Resolution Vegetation Phenology and Productivity (HR-VPP) (Issue 2.5). European Environment Agency. <https://land.copernicus.eu/en/technical-library/product-user-manual-of-seasonal-trajectories/@@download/file>
- European Environment Agency. (s.f.). *hdar*: R client for the Harmonised Data Access (HDA) API of WEKEO [Paquete de software R]. CRAN. Recuperado el 24 de diciembre de 2025, de <https://cran.r-project.org/package=hdar>.
- Hijmans, R. J. (2023). terra: Spatial data analysis. CRAN: *Contributed packages*. <https://CRAN.R-project.org/package=terra>
- Kong, D. (2020). phenofit: A state-of-the-art vegetation phenology extraction package (Versión 0.3.9) [R software package]. Zenodo. <https://doi.org/10.5281/zenodo.3605560>.
- Pebesma, E. (2018). Simple features for R: standardized support for spatial vector data. <https://doi.org/10.32614/RJ-2018-009>
- Ranghetti, L., Boschetti, M., Nutini, F., & Busetto, L. (2020). "sen2r": An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. *Computers & Geosciences*, 139, 104473. <https://doi.org/10.1016/j.cageo.2020.104473>
- Simões, R., Camara, G., Queiroz, G., Souza, F., Andrade, P. R., Santos, L., Carvalho, A., & Ferreira, K. (2021). Satellite image time series analysis for big Earth observation data. *Remote Sensing*, 13(13), Article 2428. <https://doi.org/10.3390/rs13132428>
- Valero-Jorge, A.; Casterad, M.A.; Alcalá, J.-T. Evaluating the Influence of Missing Data from the Crop Vegetation Index Time Series on Copernicus HR-VPP Phenological Products. *Eng. Proc.* 2025, 94, 4. <https://doi.org/10.3390/engproc2025094004>