Congresos UEx. Actas de Congresos. ISSN: 3101-7177,
vol. 2, 2026, XXI Congreso de la Asociación Española de Teledetección
DOI: https://doi.org/10.17398/3101-7177.2.271

Procesamiento de datos Sentinel-2 en entornos de Big Data: evaluación de Google Earth Engine frente a procesamiento local para la generación de índices de vegetación aplicados a predicción de incendios

Jiménez Romanillos, José [1] **Martínez García, Francisco Manuel** [2] **García Ramajo, Jorge Juan** [2] **Quirós, Elia** [2]

[1] Junta de Extremadura, España

[2] Universidad de Extremadura

ORCID: Jiménez Romanillos 0009-0009-0612-4390 Martínez García 0000-0002-0862-9933 García Ramajo 0009-0000-7296-3082 Quirós 0000-0002-8429-045X

Correspondencia: jose.jimenez@juntaex.es fmmgarcia@unex.es jjgarciar@unex.es equiros@unex.es

RESUMEN

La disponibilidad masiva de imágenes Sentinel-2 ha impulsado el uso de entornos Big Data para aplicaciones ambientales, planteando la necesidad de evaluar su eficiencia y coherencia metodológica. Este trabajo compara la generación de series temporales de índices de vegetación mediante Google Earth Engine (GEE) y procesamiento local. Se utilizaron imágenes Sentinel-2 L2A entre el 29 de marzo y el 13 de septiembre de 2019, implementando un flujo de trabajo equivalente que incluyó filtrado espacial y temporal, enmascarado de nubes, cálculo de índices espectrales y composición temporal. La comparación se realizó mediante métricas de rendimiento computacional, almacenamiento, automatización y consistencia radiométrica. Los resultados muestran diferencias relevantes en eficiencia y escalabilidad, con mejor rendimiento de GEE en términos de automatización y procesamiento distribuido. Las discrepancias radiométricas fueron reducidas y no afectan de forma significativa la consistencia de los índices.

Palabras clave: *Sentinel-2, Google Earth Engine, incendios forestales, Big Data*

Fecha de recepción: 19 febrero 2026 · **Fecha de aceptación:** 19 febrero 2026

Procesamiento de datos Sentinel-2 en entornos de Big Data: evaluación de Google Earth Engine frente a procesamiento local para la generación de índices de vegetación aplicados a predicción de incendios

José Jiménez-Romanillos [(1)], Francisco Manuel Martínez García [(2)], Jorge Juan García-Ramajo [(2)], Elia Quirós [(2)]

[(1)] Junta de Extremadura.

0009-0009-0612-4390, jose.jimenez@juntaex.es

[(2)] Universidad de Extremadura.

0000-0002-0862-9933, fmmgarcia@unex.es ; 0009-0000-7296-3082, jjgarciar@unex.es

0000-0002-8429-045X, equiros@unex.es

Resumen: La disponibilidad masiva de imágenes Sentinel-2 ha impulsado el uso de entornos Big Data para aplicaciones ambientales, planteando la necesidad de evaluar su eficiencia y coherencia metodológica. Este trabajo compara la generación de series temporales de índices de vegetación mediante *Google Earth Engine* (GEE) y procesamiento local. Se utilizaron imágenes Sentinel-2 L2A entre el 29 de marzo y el 13 de septiembre de 2019, implementando un flujo de trabajo equivalente que incluyó filtrado espacial y temporal, enmascarado de nubes, cálculo de índices espectrales y composición temporal. La comparación se realizó mediante métricas de rendimiento computacional, almacenamiento, automatización y consistencia radiométrica. Los resultados muestran diferencias relevantes en eficiencia y escalabilidad, con mejor rendimiento de GEE en términos de automatización y procesamiento distribuido. Las discrepancias radiométricas fueron reducidas y no afectan de forma significativa la consistencia de los índices.

Palabras clave: Sentinel-2, Google Earth Engine, incendios forestales, Big Data

Sentinel-2 Data Processing in Big Data Environments: Evaluation of Google Earth Engine versus local processing for Vegetation Index Generation Applied to Wildfire Prediction

***Abstract**: The large availability of Sentinel-2 imagery has promoted the use of Big Data environments for environmental applications, requiring evaluation of their efficiency and methodological consistency. This study compares the generation of vegetation index time series using Google Earth Engine (GEE) and local processing. It focuses on Sentinel-2 L2A imagery acquired between March 29 and September 13, 2019, implementing an equivalent workflow that included spatial and temporal filtering, cloud masking, computation of spectral indices, and temporal compositing. The assessment considered computational performance, storage requirements, automation level, and radiometric consistency. Results indicate notable differences in efficiency and scalability, with GEE showing superior automation and distributed processing capabilities. Radiometric discrepancies among platforms were minimal and did not significantly affect radiometric consistency.*

***Keywords**: Sentinel-2, Google Earth Engine, wildfire modeling, Big Data*

1. INTRODUCCIÓN

El incremento exponencial de datos procedentes de los programas de observación de la tierra ha consolidado el paradigma *Big Data* en el ámbito de la teledetección. La misión Sentinel-2, con imágenes multiespectrales de hasta 10 m de resolución espacial, 13 bandas y una frecuencia de revisita de cinco días en condiciones nominales, genera volúmenes masivos de información (Berra *et al*., 2024; Drusch *et al*., 2012). Esta disponibilidad masiva de datos plantea nuevos retos técnicos y metodológicos, especialmente en lo que respecta a la capacidad de almacenamiento, la escalabilidad del procesamiento y la eficiencia computacional.

En este contexto han cobrado protagonismo las plataformas en la nube basadas en el enfoque *compute-*

to-data, donde el procesamiento se realiza junto al dato, evitando su descarga masiva (Berra *et al*., 2024). Entre ellas destacan *Google Earth Engine* (GEE) y *EO Browser*, que permiten trabajar directamente sobre colecciones completas de imágenes. En particular, GEE ofrece un entorno de programación basado en *JavaScript* o *Python* con capacidad de procesamiento distribuido a escala planetaria (Gorelick *et al*., 2017; Saad El Imanni *et al*., 2022).

Sin embargo, el procesamiento local continúa siendo una alternativa relevante, especialmente en entornos donde se requiere control completo sobre el flujo de trabajo, reproducibilidad estricta o integración directa con pipelines de análisis avanzados (Yang *et al*., 2024). El enfoque tradicional *data-to-compute* implica la descarga de las escenas, su almacenamiento local y el procesamiento mediante herramientas como GDAL, SNAP o librerías científicas en *Python*, lo que puede suponer elevados costes de almacenamiento y tiempos de procesamiento cuando se manejan series temporales extensas.

La generación de índices espectrales derivados, principalmente índices de vegetación y humedad, constituye un paso crítico en los estudios ambientales. Su cálculo a gran escala exige recursos computacionales significativos y, al mismo tiempo, una elevada consistencia radiométrica y reproducibilidad (Saad El Imanni *et al*., 2022), ya que de ello depende la fiabilidad de análisis posteriores en ámbitos como la gestión ambiental o la planificación territorial.

A pesar de la creciente adopción de plataformas *cloud*, existe aún una limitada evaluación comparativa entre entornos de procesamiento en términos de rendimiento computacional, volumen de almacenamiento requerido, reproducibilidad y consistencia radiométrica de los productos generados. El objetivo de este trabajo es evaluar comparativamente dos enfoques de procesamiento (GEE y procesamiento local) para la generación de series temporales de índices de vegetación derivados de Sentinel-2. A través del análisis de métricas de rendimiento, eficiencia de almacenamiento, automatización, reproducibilidad y coherencia numérica de los índices generados.

2. MATERIAL Y MÉTODOS

2.1. Área de estudio y datos

El estudio se centró en la descarga y análisis de imágenes Sentinel-2 L2A correspondientes a la tesela 29TPH (Figura 1) ubicada al noroeste de la Península Ibérica (Lugo), para el periodo comprendido entre el 29 de marzo y el 13 de septiembre de 2019 donde ocurrieron tres incendios con una superficie superior a 100 ha. La resolución final de análisis fue de 10 m tras el remuestreo de todas las bandas.

2.2. Índices de vegetación

Se calcularon los siguientes índices de vigor y actividad fotosintética, como el NDVI (Índice de Vegetación de Diferencia Normalizada); GNDVI (Índice de Vegetación de Diferencia Normalizada Verde); SAVI (Índice de Vegetación Ajustado al Suelo); MSAVI2 (Índice de Vegetación Ajustado al Suelo Modificado 2) y el índice NDI45 (Índice de Diferencia Normalizada 4-5), así como indicadores de humedad del combustible (NDMI (Índice de Humedad de Diferencia Normalizada); MSI (Índice de Estrés Hídrico). Además, se incorporaron variables estructurales biofísicas como el CI (Índice de Clorofila) y BI (Índice de brillo). Todos estos índices espectrales están ampliamente contrastados en el desarrollo de trabajos técnicos y científicos en teledetección, lo que respalda su idoneidad para el análisis.

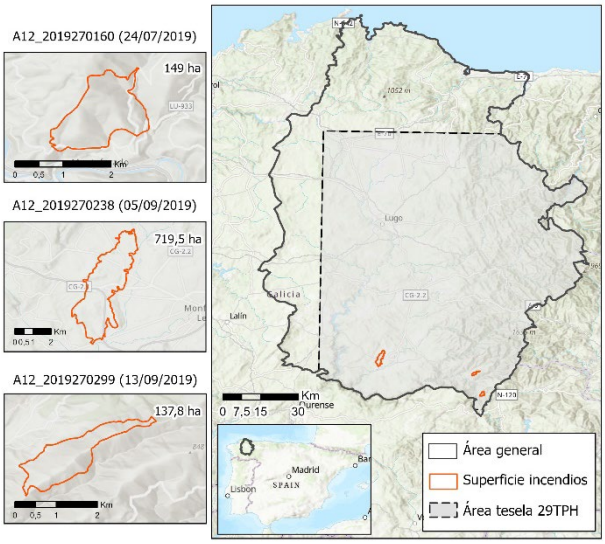

Figura 1. Área de estudio que recoge los incendios estudiados.

2.3. Flujo de procesamiento

Se diseñó un flujo de trabajo equivalente en ambas plataformas:

- Filtrado espacial y temporal. Se seleccionaron las fechas a descargar indicando el inicio y el final, y las coordenadas en formato EPSG:4326.
- Enmascarado de nubes. Se seleccionaron las imágenes con menos de un 20% de nubes, y dentro de estas imágenes se descartaron los píxeles que sean nieve, nubes y *thin cirrus*, de acuerdo con la *Scene Classification Layer* (SCL) de Sentinel-2 L2A.
- Remuestreo de imágenes. Se realizó un remuestreo con el objetivo de mantener una resolución de 10 m.
- Cálculo de índices. A partir de las bandas de cada imagen, se calcularon todos los índices deseados.
- Composición temporal. Para cada índice se generó una serie temporal de tres meses anteriores a la fecha del incendio, a partir de la cual se calcularon valores medios agregados para intervalos de 7, 14 y 30 días.

2.4. Características del procesamiento local

Para el procesamiento local se utilizaron 13 imágenes Sentinel-2 L2A correspondientes a la tesela de estudio, previamente filtradas para excluir aquellas con un porcentaje de nubes superior al 20%. Estas se

procesaron en un equipo con CPU Intel Xeon Gold 5218 @ 2.30 GHz y 128 GB de RAM a 2666 MHz, con almacenamiento distribuido en tres discos de 444 GB, 3.63 TB y 3.63 TB. El procesamiento se realizó mediante el uso de las librerías de *Python* geopandas, pandas, rasterio, xarray, rioxarray y esa_snappy, configurando las operaciones para mantener una resolución de 10 m y asegurar equivalencia con el flujo de trabajo implementado en GEE.

2.5. Métricas de evaluación

La comparación entre plataformas se realizó considerando cuatro dimensiones principales: rendimiento computacional, eficiencia de almacenamiento, consistencia radiométrica y operatividad del flujo de trabajo.

2.5.1. Rendimiento computacional

Se registró el tiempo total de procesamiento (*end-to-end*) desde la ejecución inicial hasta la obtención del producto final (raster o tabla de variables). Cuando fue posible, se desglosó por etapas: filtrado de colección, enmascarado de nubes, cálculo de índices, composición temporal y exportación.

Para permitir la comparación entre plataformas y escenarios con diferente tamaño de área o número de fechas, se definió un tiempo normalizado (Eq. 1):

$$T_{norm} = \frac{T_{total}}{(Área/100) \cdot (Fechas/100) \cdot Índices} \quad (1)$$

donde Área se expresa en km² y Fechas corresponde al número de composiciones temporales generadas. Esta métrica permite comparar eficiencia relativa independientemente de la escala del experimento.

2.5.2. Eficiencia de almacenamiento y transferencia

Se cuantificó el tamaño total de los productos generados (GB), diferenciando entre datos intermedios y salida final. En el caso del procesamiento local, se consideró además el volumen total descargado de escenas Sentinel-2. Se evaluó la eficiencia de almacenamiento en términos de tamaño por fecha y por km² procesado.

2.5.3. Consistencia radiométrica

Para evaluar la coherencia numérica entre plataformas, se compararon los valores de los índices calculados sobre las mismas fechas y áreas, tomando como referencia los índices de procesamiento local. Se emplearon métricas de error estándar:

- Error cuadrático medio (RMSE)
- Sesgo medio (bias)
- Coeficiente de correlación de Pearson (r)

Las comparaciones se realizaron a partir de estadísticas agregadas del área de estudio, garantizando la equivalencia espacial y temporal entre ambos enfoques. Asimismo, el análisis incluyó una evaluación a nivel de píxel, comparando para cada fecha los valores correspondientes a las mismas ubicaciones espaciales en ambas plataformas.

2.5.4. Operatividad y reproducibilidad

Se evaluaron cualitativamente aspectos relacionados con:

- Posibilidad de reproducir exactamente el proceso mediante scripting.
- Limitaciones operativas (cuotas, tamaño máximo de exportación, concurrencia).

Estas dimensiones permiten valorar no solo la eficiencia técnica, sino también la idoneidad de cada entorno para flujos de trabajo *Big Data* orientados a modelado predictivo.

3. RESULTADOS Y DISCUSIÓN

3.1. Rendimiento computacional

El tiempo total de procesamiento fue de 21 min en GEE y 76 min en local; en este último caso, 54 min correspondieron al procesamiento (cálculo de índices, reproyección, composición y exportación) y 22 min a la descarga y preparación de las imágenes (Tabla 1). El tiempo normalizado (Eq. 1) mostró una mayor eficiencia relativa en GEE, con una reducción del 72.4% respecto al procesamiento local. Cabe señalar que estos resultados se obtuvieron bajo una licencia no comercial. En entornos de producción, con cuotas de procesamiento comercial, la mayor capacidad de paralelización permitiría reducir los tiempos de ejecución y exportación.

Tabla 1. Comparación de rendimiento computacional.

Plataforma	Tiempo (min)	Tiempo relativo	Tamaño salida (MB)	Espacio ocupado (GB)
GEE	21	2.6602	27.7	0
Local	76	9.6274	29.9	68.6

3.2. Almacenamiento y transferencia de datos

El volumen total generado como salida final, es decir, los CSV finales, fue de 27.7 MB (GEE) y 29.9 MB (local). En el caso del procesamiento local, el volumen total descargado de escenas Sentinel-2 junto a su procesamiento ascendió a 68.6 GB, lo que incrementó significativamente las necesidades de almacenamiento. La descarga previa consistió en un total de 10.3 GB, mientras que las imágenes procesadas ocuparon un total de 58.3 GB.

Las diferencias observadas se relacionan con:

- Nivel de la máscara a la hora de descartar nubes, nieve y *thin cirrus*.
- Enmascarado de imágenes a partir de porcentaje de nubes.
- Nivel de compresión aplicado.
- Generación o no de productos intermedios.

Desde una perspectiva Big Data, el enfoque *compute-to-data* (GEE) reduce sustancialmente el movimiento de datos, mientras que el enfoque *data-to-compute* (local) implica mayores costes de transferencia y almacenamiento.

3.3. Consistencia radiométrica

Los resultados numéricos entre plataformas mostraron una alta concordancia (Tabla 3) , lo que se refleja en la coincidencia de sus distribuciones (Figura 2). El índice con mayor discrepancia fue MSI (RMSE = 0.11982),

mientras que BI presentó la mayor similitud con un RMSE de 0.00721.

Tabla 2. Métricas comparando ambas fuentes.

Índice	RMSE	Sesgo medio	Pearson
BI	0.007212	0.000295	0.976184
CI	0.038827	-0.000699	0.970586
GNDVI	0.041317	0.002066	0.963288
MSAVI2	0.045119	0.006784	0.970575
MSI	0.119822	0.001228	0.941064
NDI45	0.060906	0.003336	0.934917
NDMI	0.042988	0.001011	0.975850
NDVI	0.048766	0.001137	0.969170

En la Figura 2 se muestran los valores obtenidos para cada índice, apreciándose ligeras diferencias entre ambos procesamientos. Estas se relacionan con la construcción de mosaicos, el enmascarado de nubes, el tratamiento de píxeles parcialmente cubiertos, los ajustes radiométricos de GEE y las estrategias de reproyección, pero no afectaron la coherencia espacial ni la interpretación ambiental de los índices. Estos resultados muestran que GEE es más eficiente que el procesado local, especialmente con grandes volúmenes de datos.

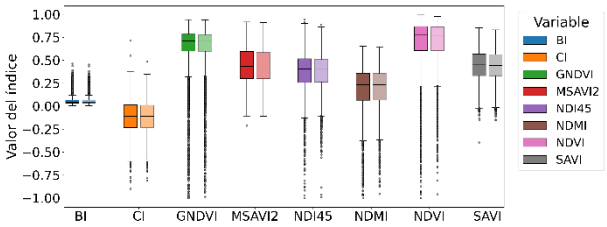

Figura 2. Boxplot comparativo por índice.

4. CONCLUSIONES

Este trabajo compara el procesamiento de datos Sentinel-2 mediante GEE y procesamiento local para la generación de índices de vegetación aplicados al modelado de incendios forestales. Las principales conclusiones son:

GEE destaca por su elevada eficiencia computacional y su capacidad de automatización, lo que lo convierte en una herramienta especialmente adecuada para flujos de trabajo basados en *Big Data* e integrados en pipelines de *machine learning*.

Por su parte, el procesamiento en local ofrece un control total sobre el flujo de trabajo y favorece la reproducibilidad de los análisis. No obstante, implica mayores requerimientos de almacenamiento y tiempos de descarga, lo que puede constituir una limitación relevante en estudios con grandes volúmenes de datos o amplias áreas de estudio.

Por tanto, la elección de la plataforma debe fundamentarse en factores como el volumen de datos, la necesidad de escalabilidad, la infraestructura disponible y el grado de automatización requerido. Para análisis a gran escala y la generación sistemática de variables destinadas a modelado predictivo, los entornos *cloud* con capacidades avanzadas de *scripting* se perfilan como la opción más eficiente.

5. AGRADECIMIENTOS

Cofinanciado por la Unión Europea a través del Programa Interreg VI-A España-Portugal (POCTEP) 2021-2027.

”Redes de alertas tempranas, para la teledetección de riesgos derivados del cambio climático, por satélites de observación de la tierra para respuesta de protección civil (RAT_EOS_PC)”.

6. REFERENCIAS

Berra, E. F., Fontana, D. C., Yin, F., & Breunig, F. M. (2024). Harmonized Landsat and Sentinel-2 Data with Google Earth Engine. *Remote Sensing*, *16*(15). https://doi.org/10.3390/rs16152695

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA’s Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, *120*. https://doi.org/10.1016/j.rse.2011.11.026

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*. https://doi.org/10.1016/j.rse.2017.06.031

Saad El Imanni, H., El Harti, A., & El Iysaouy, L. (2022). Wheat Yield Estimation Using Remote Sensing Indices Derived from Sentinel-2 Time Series and Google Earth Engine in a Highly Fragmented and Heterogeneous Agricultural Region. *Agronomy*, *12*(11). https://doi.org/10.3390/agronomy12112853

Yang, L., He, W., Qiang, X., Zheng, J., & Huang, F. (2024). Research on remote sensing image storage management and a fast visualization system based on cloud computing technology. *Multimedia Tools and Applications*, *83*(21). https://doi.org/10.1007/s11042-023-17858-6