

Consideraciones sobre la colección 2 de Landsat

Xavier Pons, Cristina Cea, Óscar González-Guerrero, Jordi Cristóbal

Grumets Research Group. Departament de Geografia, Edifici B. Universitat Autònoma de Barcelona.
08193 Bellaterra, Catalunya, España. xavier.pons@uab.cat; cristina.cea@uab.cat; oscar.gonzalez.guerrero@uab.cat;
jordi.cristobal@uab.cat

Resumen: Este trabajo presenta algunas consideraciones sobre la colección 2 de Landsat, iniciada en 2020. El esfuerzo del Servicio Geológico de Estados Unidos (USGS) por proporcionar una serie coherente desde varios puntos de vista (geométrico, radiométrico, de metadatos, etc) para un volumen de datos inmenso, ligado a una historia de varias décadas, con desafíos científicos y técnicos constantes es, sin duda alguna, encomiable. En el artículo se destacan tanto aspectos de buenas prácticas (p.ej., la elección del formato de distribución o el marcado de píxeles saturados), de inconsistencias radiométricas (en zonas de solape de escenas, entre imágenes de fechas cercanas, etc), como de decisiones que no ayudan a la comunidad de usuarios (como la retirada de productos menos procesados, la no inclusión de productos de nivel 2 para los datos de los tres primeros satélites de la serie, la inclusión de metadatos que pueden llevar a confusión, la inconsistencia de valores para sintatos [NoData], etc). También se interroga sobre la enorme diferencia en radiancias respecto a los datos procesados por ESA (CEOS), y se aporta luz y justificación a la decisión de cambiar el sentido de los tradicionales DN, que ha conllevado que los conversores a radiancias ahora cambien a lo largo del año; asimismo, se ofrecen alternativas para algún aspecto problemático del tratamiento de los datos del infrarrojo térmico. El trabajo pretende ayudar a otros usuarios, así como contribuir al debate sobre las mejores prácticas en el procesamiento de imágenes de teledetección.

Palabras clave: Landsat, colección-2, radiancia, reflectancia TOA, reflectancia BOA.

English Title: Considerations on Landsat Collection 2

Abstract: This paper presents some considerations about Landsat Collection 2, started in 2020. The USGS effort to provide a coherent series from various points of view (geometric, radiometric, metadata, etc) for a long term and extensive volume of data with constant scientific and technical challenges is, without a doubt, admirable. The article highlights both aspects of good practices (p.ej., the choice of the distribution format or the marking of saturated pixels), as well as radiometric inconsistencies (in areas of scene overlap, close data images, etc) and decisions that pose difficulties for the user community (such as the lower processing level products termination, the exclusion of level 2 products for the first three Landsat satellites, the inclusion of metadata that can lead to confusion, the inconsistency of NoData values, etc). It also questions the enormous difference in radiances regarding data processed by ESA (CEOS), and sheds light and justification on the decision to change the meaning of the traditional DN, resulting in radiance conversion factors shifts throughout the year. Moreover, some alternatives are offered for some problematic aspects of thermal infrared data processing. The work aims to help other users, as well as contribute to the debate on best practices in Remote Sensing image processing.

Keywords: Landsat, collection-2, radiance, TOA reflectance, BOA reflectance.

1. INTRODUCCIÓN

Desde sus inicios, los datos Landsat han sido accesibles en una gran variedad de soportes (principalmente cintas CCT, CD-ROM y descarga desde Internet) y formatos (como CEOS en las imágenes distribuidas durante años gracias a convenios entre el USGS y la ESA, GeoTIFF, etc). En las últimas décadas, el USGS ha realizado diversos esfuerzos para proporcionar una serie coherente desde varios puntos de vista (geométrico, radiométrico, de metadatos, etc), dado que han sido usuales las distribuciones con inconsistencias importantes (p.ej., Cristóbal *et al.*, 2004; Chen *et al.*, 2021). Durante estos años han sido frecuentes, asimismo, las vacilaciones de diversa índole, incluyendo aspectos tan básicos como los formatos de metadatos (actualmente en un fichero MTL), o la codificación de los sintatos (NoData) como 0 o 255 en los sensores de resolución radiométrica de 8 bits.

La Colección 2 (Crawford *et al.*, 2023), C2 en adelante, es el segundo esfuerzo importante de reprocesamiento en

el archivo Landsat, iniciado en 2020 (la Colección 1, C1, se desarrolló en 2016) que ha conseguido varias mejoras gracias a la aplicación de avances diversos en el procesamiento de datos, el desarrollo de algoritmos y las capacidades de acceso y distribución de datos. La C2 contiene datos de nivel 1 de Landsat (Lan en adelante cuando se aluda a numeración de satélites) 1-9 así como productos científicos de nivel 2 y 3 para los Lan 4-9.

En este artículo se abordarán las consideraciones de interés (principalmente radiométricas) o que no resultan claras en las páginas web del USGS (páginas accesibles desde <https://www.usgs.gov/Lan-missions/landsat-collection-2>) ni en el reciente artículo de Crawford *et al.* (2023) principalmente en relación con los datos del espectro solar, pero también con alguna consideración sobre el infrarrojo térmico. El artículo se complementa con algunas reflexiones finales relativas a metadatos, geometría, etc, de las imágenes.

Para una comparativa C1 vs C2 puede acudirse a <https://d9-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/s3fs-public/atoms/files/Landsat-C1vsC2-2021-0430-LMWS.pdf>.

Modestamente creemos que, dada la importancia de la serie histórica Landsat, estas reflexiones pueden ser de utilidad a su amplia comunidad de usuarios, así como fomentar el debate sobre las mejores prácticas en el procesamiento de imágenes de teledetección.

2. RADIOMETRÍA

2.1. Los DN ya no significan lo clásico

En la C2 se ha continuado con la estrategia de la C1 de modificar los DN originalmente captados por el sensor a unos nuevos valores, denominados Qcal, por “*Quantized and calibrated standard product pixel values*”. En efecto, anteriormente, y aún hoy para la mayoría de sensores de características similares, las radiancias (L, típicamente en $W/(m^2 \cdot sr)$, o en $W/(m^2 \cdot sr \cdot \mu m)$ si se proporciona la magnitud espectral, L_λ) captadas por el sensor se escalaban a niveles digitales (DN) a través de una función lineal conocida, por lo que se podía proporcionar también la función lineal inversa, de pendiente a y ordenada en el origen b , para recuperar los datos de L mediante $L = a \cdot DN + b$; habitualmente a y b se mantenían constantes a lo largo del tiempo, hasta la “próxima recalibración”. En las colecciones Landsat esto ya no es así: se reescriben los DN a Qcal para poder disponer de conversores desde Qcal a reflectancias exoatmosféricas o “*Top of the Atmosphere*”, TOA (R_{TOA} en adelante) en el producto de nivel 1 (que contine Qcal de nivel 1) o a nivel del terreno o “*Bottom of the Atmosphere*”, BOA (R_{BOA}) (con Qcal de nivel 2); los conversores, REFLECTANCE_MULT_BAND_# y REFLECTANCE_ADD_BAND_# [# es el número de la banda espectral] (documentados en los ficheros MTL), se ajustan para que sean los mismos siempre a lo largo del tiempo e incluso para todas las bandas espectrales. Debe notarse que los conversores se encuentran documentados en grupos: los conversores a TOA en LEVEL1_RADIOMETRIC_RESCALING, y los a BOA en LEVEL2_SURFACE_REFLECTANCE_PARAMETERS. La servitud de esta estrategia es que dejan de ser constantes entre fechas los conversores a L (RADIANCE_MULT_BAND_# y RADIANCE_ADD_BAND_#). Para comprender este cambio nos centraremos en el producto de nivel 1, siendo el razonamiento análogo para el producto de nivel 2. Partiremos de la fórmula clásica de $R_{TOA} = \pi \cdot L \cdot d^2 / E_0$ (p. ej., de la Eq. 1 de Pons *et al.* (2014), ignorando, como hace el USGS, el coseno del ángulo cenital solar, asumiendo que las transmitancias τ son iguales a 1 y no aplicando ninguna corrección atmosférica [$L_a = 0$]), donde d es la distancia Tierra-Sol en unidades astronómicas y E_0 es la irradiancia solar exoatmosférica, típicamente en W/m^2 . Los datos así calculados (que valdrán entre $[0,1]$ teóricamente), se escalan entre el valor mínimo y máximo conveniente, de modo que obtenemos los Qcal a través de una nueva transformación lineal: $Qcal = R_{TOA} \cdot m + n$. Aislando R_{TOA} se deducen los conversores que nos permitirán pasar de los Qcal a R_{TOA} :

$$\begin{aligned} \text{REFLECTANCE_MULT_BAND_}\# &= 1/m, \\ \text{REFLECTANCE_ADD_BAND_}\# &= -n/m. \end{aligned}$$

A partir de estos Qcal también es posible pasar a L calculando una nueva pendiente, y una nueva ordenada en el origen, pero que serán dependientes de d^2 , o sea cambiantes en cada fecha, lo que no ocurre en la mayoría de sensores de teledetección:

$$Qcal = R_{TOA} \cdot m + n = (\pi \cdot L \cdot d^2 / E_0) \cdot m + n$$

De donde:

$$L = E_0 \cdot (Qcal - n) / (m \cdot \pi \cdot d^2) =$$

$$[E_0 / (m \cdot \pi \cdot d^2)] \cdot Qcal - n / (m \cdot \pi \cdot d^2) =$$

$$(1/d^2) \cdot [E_0 / (m \cdot \pi)] \cdot Qcal - (1/d^2) \cdot n / (m \cdot \pi)$$

Con lo que:

$$\text{RADIANCE_MULT_BAND_}\# = (1/d^2) \cdot [E_0 / (m \cdot \pi)]$$

$$\text{RADIANCE_ADD_BAND_}\# = (1/d^2) \cdot n / (m \cdot \pi)$$

Obsérvese, pues, que tanto la pendiente como la ordenada en el origen para obtener radiancias han pasado a tener un cambio cíclico anual, que depende de la fecha puesto que es función de d^2 , como lo demuestra la Figura 1, que además refleja la conocida variación del cuadrado de la distancia Tierra-Sol, en alrededor de $\pm 3.3\%$ a lo largo del ciclo anual.

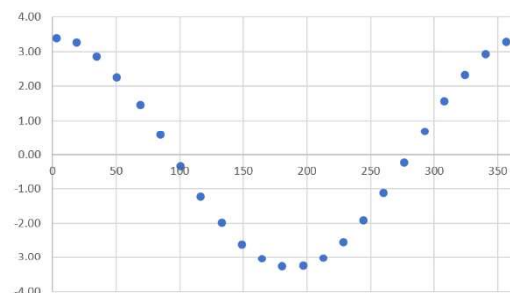


Figura 1. Oscilación (en %) de la pendiente de conversión a L a lo largo del año (ejemplo para la banda 4 de 24 imágenes entre el 3-1 y el 21-12 de 2022 en la escena 197-031).

2.2. Tratamiento no coherente en las zonas solapadas entre unidades de distribución

Si se toma un par de imágenes consecutivas, como las 197 030 y 031 del WRS-2, y se analiza la zona de solape, es frecuente encontrar píxeles con valores radiométricos distintos entre una y otra escena ya en el producto de nivel 1, que arrastrará esta incoherencia a todos los posteriores productos. Por ej., en la fecha 10/06/2004 encontramos diferencias en los píxeles situados en las coordenadas longit.,latit. siguientes: 2.468796,42.510419 (nieve), 2.519669,42.495735 (nubes), 2.102762,42.571561 (agua), 2.550657,42.499102 (vegetación de coníferas) y 2.873622,42.416712 (zona urbana). Algunas de estas diferencias son mayores que un DN y en más de una región espectral a la vez.

2.3. Discrepancias de reflectancia aceptables en el nivel 1, excesivas en el nivel 2

Cuando se analizan píxeles de una misma zona captados con pocos días de diferencia y en el mismo año, sin evidencias de cambios bruscos de cubierta, cabría esperar que, dado que los ángulos de iluminación y observación son altamente similares, las discrepancias sólo deberían obedecer a causas atmosféricas. Si tomamos un píxel

invariante (totalmente ocupado por cemento, horizontal) (lon,lat: 2.226218, 41.410353, escena 197 031) las diferencias llegan a 1.9 unidades de % de R_{TOA} en el IRc entre el 5 de dic. y el 3 de enero, aceptables por su naturaleza TOA; sin embargo, llegamos a diferencias de 6.2 unidades de % de R_{BOA} en el producto de nivel 2, contrariamente a lo esperable. Por otra parte, y como puede verse en la Tabla 1, si, dado el carácter invariante de dicho punto, calculamos sus diferencias extremas R_{TOA} y R_{BOA} en las bandas Azul, Verde, Rojo, Infrarrojo cercano e Infrarrojo de onda corta de $1.6 \mu\text{m}$ y $2.2 \mu\text{m}$ a lo largo del año obtenemos valores que difieren hasta 16.1 (IRoc2) en R_{TOA} , con un promedio de 13.0 entre todas las bandas, y hasta 6.7 (IRc) en R_{BOA} , con un promedio de 4.7. Estas discrepancias, juntamente con las citadas en el subapartado 2.6 de detalles adicionales, nos llevan a reflexionar sobre la conveniencia de usar, alternativamente al producto oficial, métodos de ajuste atmosférico apoyados en zonas pseudoinvariantes (Pons *et al.*, 2014), o en datos de radiometría de campo representativa gracias al uso de UAV (Padró *et al.*, 2018).

Tabla 1. Rango de variación de las R_{TOA} y R_{BOA} para un píxel invariante (cemento) en 3 momentos del mismo año.

	TOA	BOA
A	-10.2	-2.8
V	-10.7	-3.9
R	-11.4	-5.0
IRc	-14.5	-6.7
IRoc1	-15.3	-5.6
IRoc2	-16.1	-4.3
Media:	-13.0	-4.7
Mediana:	-12.9	-4.7

2.4. Enormes diferencias con la radiometría de productos ESA-CEOS

Resulta muy atractiva la tarea de homogeneización radiométrica entre sensores de la serie Landsat que se ha realizado hasta llegar a la C2. Sin embargo, si se acude a los productos ESA-CEOS y se comparan, simplemente, los valores de L de las mismas imágenes (idénticas fechas), es posible detectar enormes diferencias. Véase la Tabla 2. La pregunta, obviamente, es: ¿están esas diferencias justificadas por la homogeneización a lo largo del tiempo que ha realizado el USGS? Dar respuesta va más allá de las posibilidades de este artículo, pero es necesario señalar este potencial problema.

2.5. Problemas en la estimación de la temperatura de la superficie

Actualmente se usan datos provenientes de ASTER Global Emissivity Dataset (GED) para la estimación de la emisividad necesaria para el cálculo de la temperatura de la superficie (T_s). Diversos problemas en su correcta estimación (*blockiness artifact* debido a la interpolación del vecino más cercano en el producto GED, la pérdida de datos de la serie temporal o la anomalía de ajuste de la vegetación en la emisividad) provocan que no se pueda estimar la T_s en ciertos píxeles o imágenes del producto C2 ST. Para solucionar este problema, se pueden usar otros algoritmos para el cálculo de la emisividad de la vegetación o suelos desnudos con los datos del espectro solar de Landsat, como por

ejemplo el propuesto por Sobrino *et al.* (2008). Otras problemáticas pueden consultarse en <https://www.usgs.gov/landsat-missions/landsat-collection-2-known-issues#ST>.

Tabla 2. Medianas de L (unidades: $W/(m^2 \cdot sr \cdot \mu\text{m})$) en las bandas Azul, Verde, Rojo, Infrarrojo cercano e Infr. de onda corta de $1.6 \mu\text{m}$ y $2.2 \mu\text{m}$, para una gran zona de 7526.24 km^2 de la escena 197 031 en las fechas indicadas. Se proporcionan los valores en imágenes ESA (CEOS) y Earth Explorer (EE), su diferencia y el porcentaje respecto a ESA (CEOS).

	2004 02 19				2004 07 28			
	CEOS	EE	EE-CEOS	%vsCEOS	CEOS	EE	EE-CEOS	%vsCEOS
A	28.62	38.30	9.69	33.84	42.47	57.45	14.98	35.26
V	20.70	27.57	6.87	33.21	33.62	43.50	9.88	29.39
R	14.92	19.71	4.79	32.13	23.78	31.19	7.41	31.16
IRc	29.45	32.65	3.20	10.88	62.03	68.57	6.54	10.54
IRoc1	3.68	4.32	0.65	17.63	7.76	9.14	1.38	17.78
IRoc2	0.75	0.90	0.15	20.31	1.54	1.82	0.28	18.14

	2002 04 26				2004 06 10			
	CEOS	EE	EE-CEOS	%vsCEOS	CEOS	EE	EE-CEOS	%vsCEOS
A	61.55	65.87	4.32	7.02	38.86	52.09	13.23	34.05
V	47.92	97.09	49.17	102.60	30.10	40.61	10.51	34.91
R	34.16	65.65	31.48	92.15	19.75	27.02	7.26	36.78
IRc	63.72	61.56	-2.16	-3.38	67.73	74.70	6.97	10.29
IRoc1	8.57	8.90	0.32	3.76	7.54	8.66	1.11	14.75
IRoc2	1.71	2.93	1.22	71.37	1.31	17.48	16.17	1232.80

2.6. Detalles adicionales

Como novedad en la C2 cabe destacar que se distribuyen sendas capas que indican los píxeles saturados, lo cual es interesante para poderlos descartar en aproximaciones radiométricas finas. También cabe destacar que en <https://www.usgs.gov/landsat-missions/landsat-collection-2-known-issues#SR> se enumeran varios problemas relativos al algoritmo de corrección radiométrica, cuya discusión va más allá de este breve artículo. En todo caso, resulta decepcionante que todavía se confunda, como se reconoce en la fuente indicada, la codificación de los sintatos (NoData) con valores radiométricos cero, lo cual nada tiene que ver con la objetiva dificultad de la corrección radiométrica en determinadas situaciones.

3. OTROS ASPECTOS

3.1. Empobrecimiento de las opciones de distribución

Las plataformas de distribución del USGS han dejado de ofrecer los productos clásicos más cercanos a la captación en bruto (*raw*). Es cierto que los usuarios de Landsat no disponían prácticamente nunca de datos totalmente en bruto, pero los productos menos procesados permitían mejoras en la georreferenciación y, tanto o más importante, escapar a la interpolación bicúbica (el método ciertamente menos respetuoso con la radiometría original, algo especialmente discutible en aplicaciones de cartografía temática categórica, así como en la obtención de parámetros biofísicos). En otras ocasiones, nos encontramos con imágenes muy mal georreferenciadas (*p.ej.*, LT05_L1GS_19803_1_20060927_20200831_02_T2 en la Península Ibérica, así como en numerosas imágenes de Groenlandia, debido a la poca calidad del MDE utilizado); dichas imágenes “simplemente” se etiquetan como “*tiers 2*” para avisar, aunque la corrección correcta es perfectamente posible (nosotros mismos lo hicimos con esta imagen hace años, y afortunadamente la conservamos, pero ahora nadie podrá puesto que ya no está a disposición la imagen original). Finalmente, en países donde la proyección UTM no es la oficial, o se usa

un dátum sensiblemente diferente al WGS84 para una resolución espacial como la Landsat, el remuestreo de la imagen causa problemas indeseables; si esta afirmación no fuera cierta no leeríamos, en relación al hecho que los productos de nivel 2 de la mayor parte de los Estados Unidos de América están en proyección cónica de Albers: “*Raw Landsat data are calibrated and reprojected directly into the [...] Albers projection, which means [they] [...] are not resampled twice*”.

3.2. Capas de consignación de calidad

Numerosos metadatos de calidad (QA) se siguen ofreciendo en la C2. Sin embargo, el formato continúa siendo no trivial para muchos usuarios debido a la mezcla de bits con distinto sentido en un mismo byte y al hecho que en algunos casos existen propiedades que exigen la consideración de más de un bit a la vez. Es por ello por lo que algunos programas informáticos han optado por módulos *ad hoc* (ArcGIS: <https://www.usgs.gov/landsat-missions/landsat-quality-assessment-arcgis-toolbox>) o por la separación en bandas específicas para cada propiedad (MiraMon: Pons, 2024).

3.3. Serie temporal incompleta en el nivel 2

Los productos de nivel 2 no están disponibles para Lan 1-3 (<https://www.usgs.gov/landsat-missions/landsat-collection-2> y Crawford *et al.*, 2023), lo cual conlleva una importante pérdida de información en la serie temporal durante los diversos años que transcurren entre Lan 1 (1972) y Lan 4 (1982).

3.4. COG directo para geoservicios

La C2 se distribuye en Cloud-Optimized GeoTIFF (COG), lo cual es un acierto puesto que esta reciente especificación del OGC (Masó, 2023) permite ofrecer las imágenes desde servidores sin necesidad de preparación previa, pero sin comprometer la lectura de los mismos ficheros desde entornos locales.

3.5. Metadatos que dan lugar a confusión

En el producto de nivel 2 se incluyen coeficientes de conversión a L y a R_{TOA} (véase 2.1). Si un usuario reflexiona sobre su sentido puede deducir que en ningún caso son aplicables a la imagen a la cual acompañan (debido a la no reversibilidad, con simples coeficientes, del proceso de escritura de los nuevos Qcal proporcionales a las R_{BOA}), pero su presencia nos parece altamente confusa y propensa a que se cometan errores, al tiempo que nada didáctica en un entorno docente: no es fácil entender que, en el mismo fichero, haya dos valores para la misma clave REFLECTANCE_MULT_BAND_4=, o un valor RADIANCE_MULT_BAND_4= que no debe aplicarse.

4. AGRADECIMIENTOS

Los autores desean agradecer al USGS su política de datos abiertos, así como el esfuerzo para seguir mejorando la serie temporal Landsat de imágenes de Teledetección. Este trabajo se ha realizado en parte gracias a la ayuda

SGR de Grupos de Investigación Consolidados y de Calidad de la Generalitat de Catalunya (SGR 2021 núm. 00554) y al hecho que X. P. es perceptor de una ICREA Academia Excellence in Research Grant (2023–2027).

5. BIBLIOGRAFÍA

- Crawford, C.J., Roy, D.P., [32 autores más] ..., Zahn, S. 2023. The 50-Year Landsat Collection 2 Archive. *Science of Remote Sensing*, 8, 100103. <https://doi.org/10.1016/j.srs.2023.100103>
- Chen, F., Wang, C., Zhang, Y., Yi, Z., Fan, Q., Liu, L., Song, Y. 2021. Inconsistency among Landsat Sensors in Land Surface Mapping: A Comprehensive Investigation Based on Simulation. *Remote Sensing*, 13(7), 1383. <https://doi.org/10.3390/rs13071383>
- Cristóbal J., Pons X. and Serra P. 2004. Sobre el uso operativo de Landsat-7 ETM+ en Europa. *Revista de Teledetección* 21, 55-59. <http://www.aet.org.es/revistas/revista21/AET21-11.pdf>
- Masó, J. 2023. *OGC Cloud Optimized GeoTIFF Standard*. Open Geospatial Consortium. <http://www.opengis.net/doc/is/COG/1.0>
- Padró, J.C, Muñoz, F.J., Ávila, L.A., Pesquer, L., Pons, X. 2018. Radiometric Correction of Landsat-8 and Sentinel-2A Scenes Using Drone Imagery in Synergy with Field Spectroradiometry. *Remote Sensing*, 10, 1687. <https://doi.org/10.3390/rs10111687>
- Pons, X., Pesquer, L., Cristóbal, J., González-Guerrero, O. 2014. Automatic and improved radiometric correction of Landsat imagery using reference values from MODIS surface reflectance images. *International Journal of Applied Earth Observation and Geoinformation*, 33, 243-254. <https://doi.org/10.1016/j.jag.2014.06.002>
- Pons, X. 2024. *MiraMon. Sistema de Información Geográfica y software de Teledetección*. Bellaterra: CREAM. https://www.miramon.cat/Index_es.htm
- Sobrino, J.A., Jiménez-Muñoz, J.C., Soria, G., Romaguera, M., Guanter, L., Moreno, J. 2008. Land Surface Emissivity Retrieval from Different VNIR and TIR Sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2), 316-327. <https://doi.org/10.1109/TGRS.2007.904834>