Facilitating advanced Sentinel-2 analysis through a simplified computation of Nadir BRDF Adjusted Reflectance

David Montero[1,2,3], Miguel D. Mahecha[1,2,3,4], César Aybar[5], Clemens Mosig[1,2,4], Sebastian Wieneke[1,2]

[1] Remote Sensing Centre for Earth System Research (RSC4Earth), Leipzig University, Germany
– (david.montero, miguel.mahecha, clemens.mosig, sebastian.wieneke)@uni-leipzig.de
[2] Institute of Earth System Science & Remote Sensing, Leipzig University, Germany
[3] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Germany
[4] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Germany
[5] Image Processing Laboratory, Universitat de València, Spain – cesar.aybar@uv.es

Keywords: Sentinel-2, BRDF, STAC, Earth System Data Cubes, Python.

Abstract

The Sentinel-2 (S2) mission from the European Space Agency's Copernicus program provides essential data for Earth surface analysis. Its Level-2A products deliver high-to-medium resolution (10-60 m) surface reflectance (SR) data through the MultiSpectral Instrument (MSI). To enhance the accuracy and comparability of SR data, adjustments simulating a nadir viewing perspective are essential. These corrections address the anisotropic nature of SR and the variability in sun and observation angles, ensuring consistent image comparisons over time and under different conditions. The c-factor method, a simple yet effective algorithm, adjusts observed S2 SR by using the MODIS BRDF model to achieve Nadir BRDF Adjusted Reflectance (NBAR). Despite the straightforward application of the c-factor to individual images, a cohesive Python framework for its application across multiple S2 images and Earth System Data Cubes (ESDCs) from cloud-stored data has been lacking. Here we introduce sen2nbar, a Python package crafted to convert S2 SR data to NBAR, supporting both individual images and ESDCs derived from cloud-stored data. This package simplifies the conversion of S2 SR data to NBAR via a single function, organized into modules for efficient process management. By facilitating NBAR conversion for both SAFE files and ESDCs from SpatioTemporal Asset Catalogs (STAC), sen2nbar is developed as a flexible tool that can handle diverse data format requirements. We anticipate that sen2nbar will considerably contribute to the standardization and harmonization of S2 data, offering a robust solution for a diverse range of users across various applications. sen2nbar is an open-source tool available at `https://github.com/ESDS-Leipzig/sen2nbar`.

1. Introduction

The Sentinel-2 (S2) mission from the Copernicus program of the European Space Agency (ESA), comprising the Sentinel-2A (S2A) and Sentinel-2B (S2B) satellites, is equipped with the MultiSpectral Instrument (MSI) sensor (Drusch et al., 2012). This sensor is designed to capture data across various spectral bands with high-to-medium spatial resolutions (Table 1). The Level-2A product from S2 provides surface reflectance (SR) data, which are indispensable for detailed Earth surface analysis. This product has been employed in diverse applications, examples include the estimation of carbon fluxes (Pabon-Moreno et al., 2022), the investigation of vegetation dynamics using spectral indices (Montero et al., 2023a), and Land Use and Land Cover (LULC) products generation via Artificial Intelligence (AI) models (Brown et al., 2022). The utility of S2 SR data is significantly enhanced through the incorporation into Earth System Data Cubes (ESDCs, Mahecha et al., 2020; Montero et al., 2023b), which offer an organized spatiotemporal framework, allowing for simplified multitemporal analyses.

While S2 SR data is widely used, its viewing angle ($\pm 10.3°$) and field of view ($20.6°$) amplify Bidirectional Reflectance Distribution Function (BRDF) effects due to SR anisotropy (Roy et al., 2017b). To minimize BRDF effects, adjustments simulating a nadir viewing perspective are needed (Roy et al., 2016). These corrections address the directional effects arising from the anisotropic nature of SR and the variability of sunlight and satellite viewing angles. Such adjustments are crucial for the

consistent comparison of images taken at different times and under various sensor acquisition conditions. This is particularly important for the processing and analysis of analysis-ready ESDCs, which are increasingly utilized due to their organized spatiotemporal structure and the simplicity of generating them from cloud-stored data (Montero et al., 2023b).

The spectral parameters derived from the MODIS BRDF model facilitate the computation of directional reflectance across any specified sensor viewing and solar angles (Roy et al., 2008). Using this framework, Roy et al. (2008, 2016) introduced the c-factor method, a straightforward approach for adjusting Landsat SR data by applying the MODIS spectral BRDF model parameters. This adjustment yields Nadir BRDF Adjusted Reflectance (NBAR) by multiplying the observed Landsat SR with the ratio of reflectances predicted by the MODIS BRDF model for both the observed Landsat SR and a standard nadir view under fixed solar zenith conditions. Roy et al. (2017b) and Roy et al. (2017a) extended this methodology for multiple S2 spectral bands. Despite the simplicity of the c-factor method for individual S2 images, a unified Python framework for applying this conversion uniformly across multiple images, especially for ESDCs derived from SpatioTemporal Asset Catalogs (STAC), is missing so far.

This paper presents sen2nbar, an open-source Python package designed to facilitate converting S2 SR data to NBAR using the c-factor method. sen2nbar is engineered to support both automatic NBAR conversions through a single function, accommodating SAFE files and ESDCs derived from STAC. The

Name	Band	Resolution (m)	Wavelength (nm)		f_{iso}	f_{geo}	f_{vol}	Reference
			Sentinel-2A	Sentinel-2B				
Blue	B02	10	496.6	492.1	0.0774	0.0079	0.0372	Roy et al. (2017b)
Green	B03	10	560.0	559.0	0.1306	0.0178	0.0580	Roy et al. (2017b)
Red	B04	10	664.5	665.0	0.1690	0.0227	0.0574	Roy et al. (2017b)
Red Edge 1	B05	20	703.9	703.8	0.2085	0.0256	0.0845	Roy et al. (2017a)
Red Edge 2	B06	20	740.2	739.1	0.2316	0.0273	0.1003	Roy et al. (2017a)
Red Edge 3	B07	20	782.5	779.7	0.2599	0.0294	0.1197	Roy et al. (2017a)
NIR	B08	10	835.1	833.0	0.3093	0.0330	0.1535	Roy et al. (2017b)
SWIR 1	B11	20	1613.7	1610.4	0.3430	0.0453	0.1154	Roy et al. (2017b)
SWIR 2	B12	20	2202.4	2185.7	0.2658	0.0387	0.0639	Roy et al. (2017b)

Table 1. BRDF spectral parameters for the Sentinel-2 bands.

document is organized into the following sections: Section 2 presents the sen2nbar framework, elaborating on the modules that simplify the processing steps; Section 3 illustrates the application of sen2nbar, providing practical examples; Section 4 explores the limitations of sen2nbar as well as its potential, with a particular emphasis on AI; and Section 5 summarizes our conclusions.

2. The sen2nbar Python package

The sen2nbar package, developed in Python, facilitates the conversion of S2 L2A Surface Reflectance values into NBAR values through a single function. To achieve this higher level, the package is structured into several modules, ensuring a methodical conversion process leveraging multidimensional arrays via xarray (Hoyer and Hamman, 2017) and numpy (Harris et al., 2020):

2.1 The axioms module

This module stores the predefined MODIS BRDF spectral model parameters, namely $f_{\text{iso}}(\lambda)$, $f_{\text{vol}}(\lambda)$, and $f_{\text{geo}}(\lambda)$ (as shown in Table 1), in addition to preserving the values corresponding to the spatial resolution of each spectral band. In accordance to the methodology proposed by Roy et al. (2017a,b), our implementation encompasses all spectral bands except for the Aerosols (B01) and the narrow NIR (B8A) bands. Consequently, the NBAR values are not computed for B01 and B8A bands. Also note that for the S2 Red Edge bands, the BRDF spectral model parameters were determined through a process of linear interpolation between the red and NIR MODIS BRDF spectral model parameters (Roy et al., 2017a).

2.2 The metadata module

This module is tasked with the extraction of two critical sets of data from the scene's metadata files (MTD_MSIL2A.xml and MTD_TL.xml, Figure 1a), essential for the computation and harmonization of NBAR values: firstly, it extracts the sun and sensor viewing angles; secondly, it determines the processing baseline of the scene. The acquisition of this information is conducted directly from the metadata file when dealing with SAFE files[1]. Alternatively, when operating with ESDCs constructed via stackstac or cubo (Montero et al., 2024), the module retrieves the requisite data from the corresponding metadata STAC asset within a specific item in a STAC catalog using the requests library. This dual-path approach ensures flexibility and adaptability in processing various data sources for NBAR computation and harmonization.

---

[1] https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/data-formats

2.2.1 Solar and sensor viewing angles: The extraction process for solar and sensor viewing angles in degrees involves the aggregation of detector-specific data across each spectral band into a unified 4-dimensional array $\mathbf{A} \in \mathbb{R}^{|\Lambda| \times |\Theta| \times H \times W}$. This array is structured with dimensions $|\Lambda| = |B| + 1$, where B are the bands undergoing conversion (with an additional index allocated for solar information) and Λ is the set of bands added to the solar component; $|\Theta| = 2$ (the number of extracted angles, i.e. $\Theta = \{\text{zenith, azimuth}\}$); and $H = W = 23$, reflecting the spatial resolution of this information (5 km). The assignment of spatial coordinate values within this grid uses the Upper Left (UL) X (ULX) and Y (ULY) coordinates, derived from the tile's geocoding information.

2.2.2 Processing Baseline: The Processing Baseline (PB) of the scene is extracted and converted into a floating-point numerical format. This PB version serves as a criterion for initiating a harmonization procedure for scenes possessing a PB version of 4.00 or greater.

2.3 The kernels module

A kernel-based BRDF model integrates various scattering mechanisms as a linear combination of distinct scattering modes, termed kernels (Lucht et al., 2000). The Ross-Li BRDF model (Section 2.4), also known as the Ross-Thick/Li-Sparse Reciprocal BRDF model, which is the standard for MODIS BRDF (Lucht et al., 2000), is composed of three components: the isotropic scattering parameter, the radiative transfer-type volumetric scattering kernel (Volumetric Kernel K_{vol}), and geometric-optical surface scattering kernel (Geometric Kernel K_{geo}). The objective of this module is to calculate the volumetric and geometric kernels. The computation of these kernels adheres to the mathematical formulations presented by Lucht et al. (2000). Here we use the solar zenith angle $\theta = \mathbf{A}_{\{\text{sun, zenith}\}}$, the view zenith angle $\vartheta = \mathbf{A}_{\{B, \text{zenith}\}}$, and the view-sun relative azimuth angle $\phi = \mathbf{A}_{\{\text{sun, azimuth}\}} - \mathbf{A}_{\{B, \text{azimuth}\}}$ (all angles expressed in radians). Note that $\theta \in \mathbb{R}^{H \times W}$ (since it is a 2-dimensional array representing the solar zenith angle) while $\vartheta, \phi \in \mathbb{R}^{|B| \times H \times W}$ (since they are 3-dimensional arrays representing the zenith and relative azimuth angles for all bands). This implies that $K_{\text{vol}}(\theta, \vartheta, \phi), K_{\text{geo}}(\theta, \vartheta, \phi) \in \mathbb{R}^{|B| \times H \times W}$:

2.3.1 Volumetric Kernel: The Volumetric Kernel K_{vol} in the Ross-Li BRDF model is represented by the RossThick kernel (Roujean et al., 1992), which is based on the radiative transfer theory presented by Ross (1981):

$$K_{\text{vol}}(\theta, \vartheta, \phi) = \frac{(\pi/2 - \xi)\cos \xi + \sin \xi}{\cos \theta + \cos \vartheta} - \frac{\pi}{4} \quad (1)$$

Where $\cos \xi = \cos \theta \cos \vartheta + \sin \theta \sin \vartheta \cos \phi$.
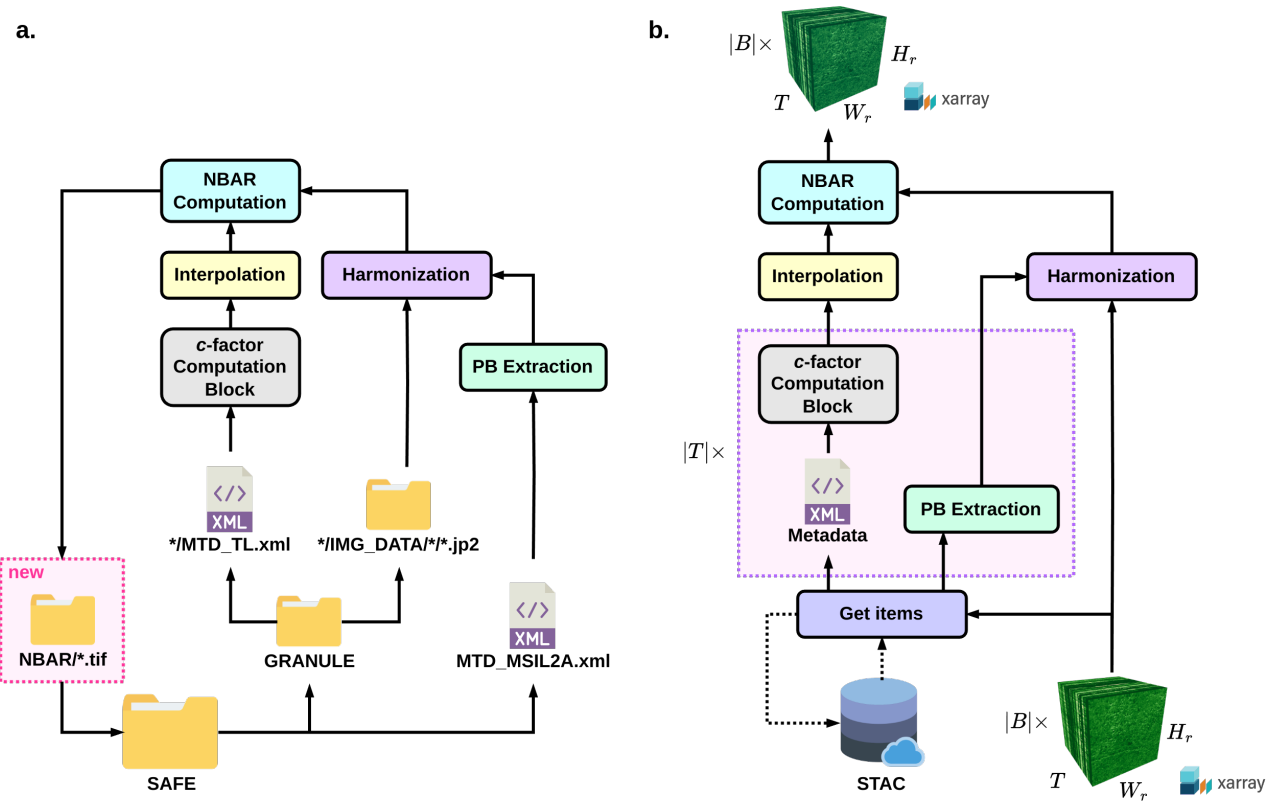
Figure 1. Flowcharts depicting the process of computing NBAR for SAFE files (a) and ESDCs (b). In (a), NBAR values are calculated for each spectral band and subsequently stored either as GeoTIFF or Cloud Optimized GeoTIFF (COG) within a newly created folder titled "NBAR" inside the original SAFE file's directory. For ESDCs, depicted in (b), the NBAR computation results in the generation of a new ESDC, represented as an xarray object. The dashed arrows linking the STAC object in (b) signify that this procedure is automated, eliminating the need for users to manually extract items from the STAC.

2.3.2 Geometric Kernel: The Geometric Kernel K_{geo} in the Ross-Li BRDF model is represented by the LiSparse kernel (Wanner et al., 1995), which is based on the geometric-optical mutual shadowing BRDF model by Li and Strahler (1992):

$$K_{\text{geo}}(\theta, \vartheta, \phi) = O - \sec \theta' - \sec \vartheta'$$
$$+ \frac{1}{2}(1 + \cos \xi') \sec \theta' \sec \vartheta' \qquad (2)$$

where

$$O = \frac{1}{\pi}(t - \sin t \cos t)(\sec \theta' + \sec \vartheta') \qquad (3)$$

$$\cos t = \frac{h}{b} \frac{\sqrt{D^2 + (\tan \theta' \tan \vartheta' \sin \phi)^2}}{\sec \theta' + \sec \vartheta'} \qquad (4)$$

$$D = \sqrt{\tan^2 \theta' + \tan^2 \vartheta' - 2 \tan \theta' \tan \vartheta' \cos \phi} \qquad (5)$$

$$\cos \xi' = \cos \theta' \cos \vartheta' + \sin \theta' \sin \vartheta' \cos \phi \qquad (6)$$

$$\theta' = \tan^{-1}\left(\frac{b}{r} \tan \theta\right) \qquad (7)$$

$$\vartheta' = \tan^{-1}\left(\frac{b}{r} \tan \vartheta\right) \qquad (8)$$

Here, the $\cos t$ term is constrained to [-1,1], $h/b = 2$, and $b/r = 1$ (Lucht et al., 2000).

2.4 The brdf module

This module is dedicated to the computation of the Ross-Li BRDF model, leveraging the scattering kernels sourced from the kernels module and the BRDF spectral parameters obtained from the axioms module, in accordance with the mathematical framework outlined by Lucht et al. (2000). Prior to computation, the BRDF spectral model parameters are transformed into xarray.DataArray objects such that $f_{\text{iso}}, f_{\text{vol}}, f_{\text{geo}} \in \mathbb{R}^{|B|}$. This enhances the efficiency and scalability of the BRDF model computation. Consequently, this computation of the Ross-Li BRDF model outputs values as a multidimensional array $\text{BRDF}(\theta, \vartheta, \phi) \in \mathbb{R}^{|B| \times H \times W}$:

$$\text{BRDF}(\theta, \vartheta, \phi) = f_{\text{iso}} + f_{\text{vol}} K_{\text{vol}}(\theta, \vartheta, \phi)$$
$$+ f_{\text{geo}} K_{\text{geo}}(\theta, \vartheta, \phi) \qquad (9)$$

2.5 The c_factor module

This module is tasked with calculating the c-factor, adhering to the methodology delineated by Roy et al. (2008). The c-factor is instrumental in modifying the reflectance values to align with any given viewing or solar geometry, as outlined by Roy et al. (2016). To standardize the reflectance measurements to a nadir viewing zenith angle, the value of the viewing zenith angle, ϑ, is predetermined to be 0. Note that $c(\theta, \vartheta, \phi) \in \mathbb{R}^{|B| \times H \times W}$:
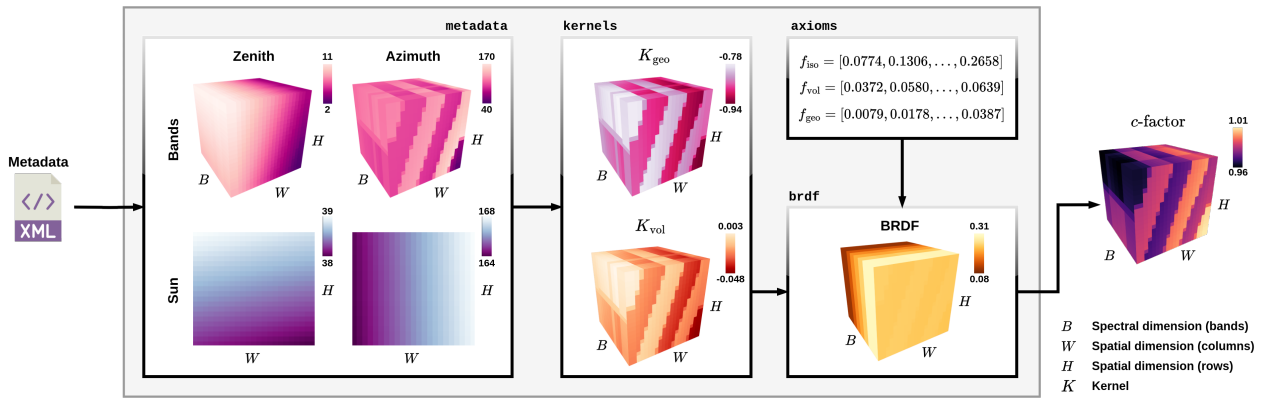
Figure 2. Diagram illustrating the c-factor computation block. The arrays depicted in this block serve as examples derived from a single metadata file, with the understanding that values could vary across different files.

$$c(\theta, \vartheta, \phi) = \frac{\text{BRDF}(\theta, 0, \phi)}{\text{BRDF}(\theta, \vartheta, \phi)} \quad (10)$$

Furthermore, this module integrates information and functionalities from preceding modules to directly derive the c-factor from a metadata file, establishing a comprehensive function specifically tailored for SAFE files. Additionally, recognizing the accessibility of metadata files through STAC items, a separate function has been developed to compute the c-factor directly from STAC items. This latter function is notably equipped with the capability to reproject the c-factor to a designated CRS when required. This reprojection feature is essential for ensuring the alignment and consistency of ESDCs created from scenes that originate in varying CRS, thereby facilitating their integration and analysis within a unified spatial framework.

2.6 The `nbar` module

This module is designed to compute NBAR values, tailored to the format of the input data, whether they are SAFE files or ESDCs constructed via `stackstac` or `cubo`:

2.6.1 SAFE files: For SAFE files (Figure 1a), the spatial resolution is band-specific, denoted as $\lambda \in B$, leading to the necessity of computing NBAR values for each λ independently. Let $\rho_\lambda \in \mathbb{R}^{H_\lambda \times W_\lambda}$ represent the SR of band λ, with H_λ and W_λ indicating its spatial dimensions. Initially, metadata files are used to extract the PB $\in \mathbb{R}_+$ and to compute the c-factor $c(\theta, \vartheta, \phi) \in \mathbb{R}^{|B| \times H \times W}$. The PB is then applied to harmonize ρ_λ if required:

$$\rho*_\lambda = \begin{cases} \rho_\lambda - 1000, & \text{if PB} \geq 4, \\ \rho_\lambda, & \text{otherwise} \end{cases} \quad (11)$$

Here, $\rho*_\lambda$ signifies the harmonized SR values. Subsequently, the NBAR values are computed. It's important to note that the c-factor is derived using the original spatial resolution of the angle information (5 km), which makes the computation faster compared to performing angle interpolation beforehand. Therefore, $c(\theta, \vartheta, \phi_\lambda)$ undergoes bilinear interpolation, denoted by the function I such that $I : \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{H_\lambda \times W_\lambda}$, to align with the original resolution of band λ. The NBAR values are then generated as follows:

$$\text{NBAR}_\lambda = I(c(\theta, \vartheta_\lambda, \phi_\lambda)) \times \rho*_\lambda \quad (12)$$

Resulting NBAR images are subsequently exported individually into the same SAFE file directory, available in either COG or GeoTIFF formats.

2.6.2 Earth System Data Cubes: For ESDCs constructed from STAC (Figure 1b), the spatial resolution is predefined, enabling the generation of NBAR values for the entire ESDC via operations on multidimensional arrays. We denoted $\rho_T \in \mathbb{R}^{|T| \times |B| \times H_r \times W_r}$ as the SR of the ESDC, where T represents the temporal dimension items, and H_r and W_r are its spatial dimensions at resolution r, alongside the bounding box (BBox) of the ESDC. Initially, the module accesses the STAC, retrieving metadata files and PB values for each item $\tau \in T$. For each τ, the c-factor $c_\tau(\theta_\tau, \vartheta_\tau, \phi_\tau)$ is calculated and subsequently concatenated along the temporal dimension such that $c_T = [c_\tau(\theta_\tau, \vartheta_\tau, \phi_\tau)]_{\tau \in T}$, resulting in $c_T \in \mathbb{R}^{|T| \times |B| \times H \times W}$. The PB values for each τ are also concatenated as $\text{PB}_T \in \mathbb{R}^{|T|}$. This array facilitates the harmonization of ρ_T:

$$\rho*_T = \left[\begin{cases} \rho_\tau - 1000, & \text{if PB}_\tau \geq 4, \\ \rho_\tau, & \text{otherwise} \end{cases} \right]_{\tau \in T} \quad (13)$$

Here, $\rho*_T$ represents the harmonized SR values of the ESDC. Following this, NBAR values are calculated. Given that the c-factor is derived from the original spatial resolution of angle information, it undergoes bilinear interpolation through function I such that $I : \mathbb{R}^{|T| \times |B| \times H \times W} \rightarrow \mathbb{R}^{|T| \times |B| \times H_r \times W_r}$, aligning with the predetermined resolution and BBox of $\rho*_T$. This process enables NBAR computation for ESDCs with uniform spatial dimensions ($H_r = W_r$, as in cubes generated via `cubo`) and for those with non-uniform dimensions ($H_r \neq W_r$, as in cubes generated via `stackstac`). Note that I only interpolates the spatial dimensions H and W. The NBAR values are then derived as follows:

$$\text{NBAR}_T = I(c_T) \times \rho*_T \quad (14)$$

3. Showcase

In a case study designed to showcase the capabilities of the `sen2nbar` package in handling multidimensional data, we util-
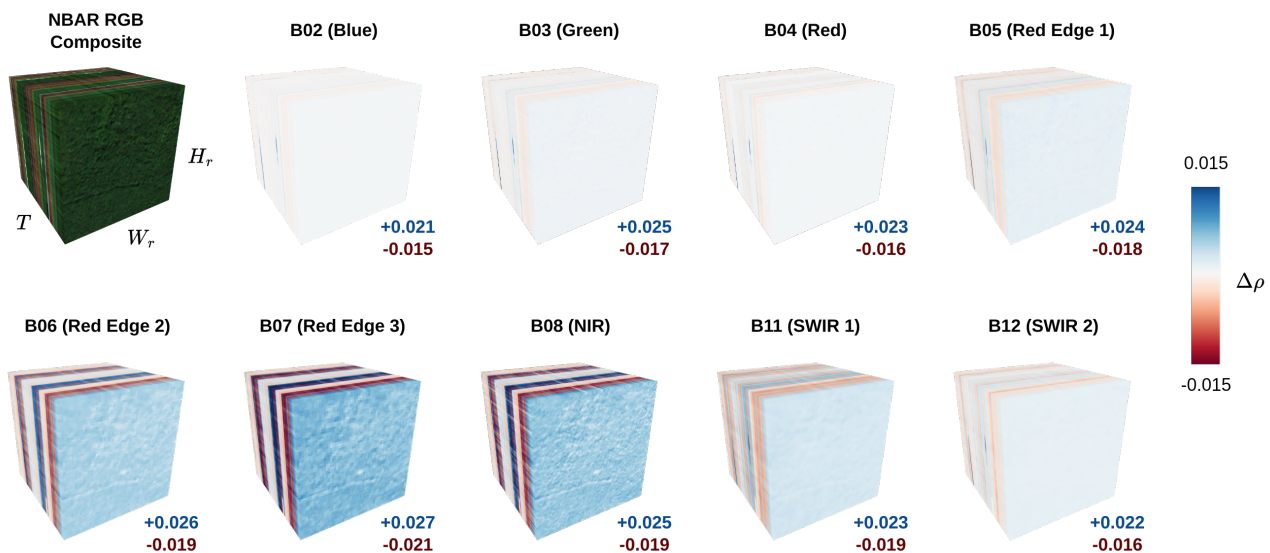
Figure 3. ESDCs illustrating the difference between NBAR and SR values. The initial ESDC presents an RGB composite utilizing visible bands to create a true color reference image. The maximum (blue) and minimum (red) $\Delta\rho$ values for each band are indicated in the lower-right corner of the corresponding ESDC.

ized an ESDC of S2 L2A data generated for the Hainich Natural National Park area, proximate to the DE-Hai Eddy Covariance tower (Knohl et al., 2003), a deciduous broadleaf forest in Germany, using cubo (Montero et al., 2024). The ESDC was crafted with a uniform resolution of 10 meters ($r = 10$) across all bands, featuring an edge size of 128 pixels, which translates to $H_r = W_r = 128$. This ESDC covers the period from 2021-01-01 to 2023-12-31, focusing on images with less than 10% cloud coverage for analysis, encompassing a total of 90 images ($|T| = 90$). It includes all S2 bands with the exception of B01 and B8A. Data from both satellites (S2A and S2B) were incorporated.

Utilizing the aforementioned ESDC, NBAR values were computed, followed by the calculation of the difference:

$$\Delta\rho = \text{NBAR}_T - \rho*_T \tag{15}$$

This metric shows the variations in reflectance values when transitioning from SR to NBAR data, particularly within the spatiotemporal framework of the ESDC. Figure 3 illustrates the $\Delta\rho_\lambda$ for the ESDCs. Spatially, while the reflectance scales according to the c-factor and shows minimal variance owing to the confined area of study, the temporal variations are more pronounced, underscoring the significance of NBAR for accurate multitemporal analysis. This effect is especially marked in the Red Edge-NIR region, where discrepancies between values are more substantial compared to other bands. However, it's important to note that the maximum and minimum $\Delta\rho$ values exhibit similar ranges across all bands, highlighting the impact of NBAR adjustments.

Reflectance changes, as highlighted previously, can greatly influence the derivation of various products. To illustrate this effect, we computed four widely used vegetation indices for both the NBAR_T and $\rho*_T$: the Normalized Difference Vegetation Index (NDVI, Rouse et al., 1974), the Near-Infrared Reflectance of Vegetation (NIRv, Badgley et al., 2017), the Kernel NDVI (kNDVI, Camps-Valls et al., 2021), and the Inverted Red

Edge Chlorophyll Index (IRECI, Frampton et al., 2013). Additionally, the difference between the indices derived from NBAR values and those from SR values was calculated as:

$$\Delta\psi = \psi_{\text{NBAR}} - \psi_{\rho*} \tag{16}$$

Here, ψ represents the index in question. Figure 4 displays the $\Delta\psi$ for the ESDCs. It is noted that the impact on NDVI is minimal, but it considerably increases for the other indices, particularly for kNDVI and IRECI, where the largest absolute changes observed reach values of 0.051 and 0.087, respectively. This variance underlines that the influence of reflectance adjustments on derived indices is index-dependent and can be substantial enough to potentially skew analysis reliant on these derived features.

4. Discussion

Here we present various factors that either constrain or enhance the capabilities of sen2nbar, focusing on data providers and scalability. As we discuss below, this approach enhances the possibilities for big data computations, workflow generation towards analysis-ready data cubes, and therefore eases the application of novel artificial intelligence (AI) methods.

4.1 The Impact of Metadata Availability

sen2nbar is fully operational with SAFE files, which are widely accessible as they are publicly available (including metadata files). When it comes to ESDCs, sen2nbar is compatible with those generated from STAC, particularly from public collections. A prime example of this compatibility is with the S2 L2A collection from the Microsoft Planetary Computer STAC, which possesses the required metadata files for transforming SR to NBAR values. Conversely, sen2nbar does not support the S2 L2A collection from the Element84 STAC due to the absence of necessary metadata files. Similarly, while tools like cubo facilitate the creation of ESDCs from GEE, the lack of metadata files in this context precludes the calculation of NBAR
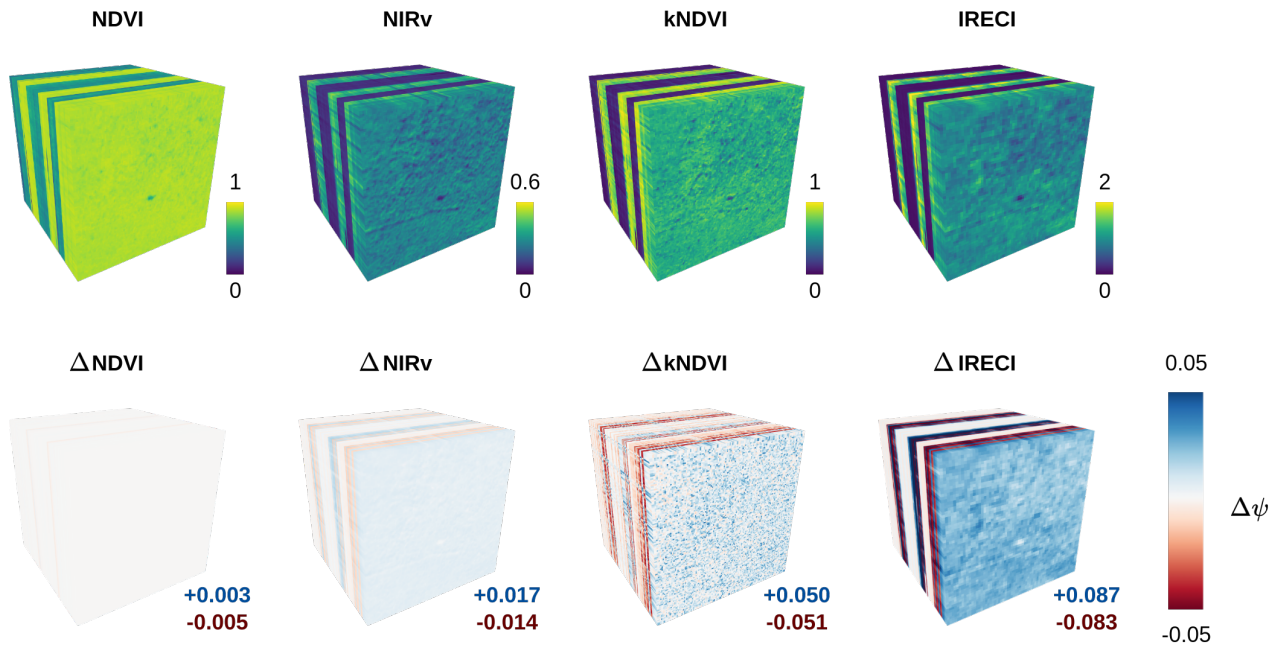
Figure 4. ESDCs illustrating the difference between indices calculated using NBAR and SR values. The first row presents the four calculated indices using the SR values. The second row presents the $\Delta\psi$ values for each ESDC. The maximum (green) and minimum (brown) $\Delta\psi$ values for each index are indicated in the lower-right corner of the corresponding ESDC.

values. Future versions of `sen2nbar` may overcome these limitations by enabling the interconnection of STAC assets (leveraging data from one STAC provider while sourcing metadata from another). This approach, however, depends on the availability of identical scenes across different collections.

4.2 Parallel Processing and Lazy Evaluation

`sen2nbar` employs multidimensional arrays through `xarray` and `numpy`, enabling the execution of various array operations. In the context of ESDCs, the outcome of NBAR processing is another multidimensional array mirroring the dimensions of the input array. This architecture facilitates the acceleration of computations by exploiting lazy evaluations and parallel processing when necessary. Specifically, when an ESDC is generated via `stackstac` or `cubo`, the cube initially remains unevaluated, and its actual creation occurs only upon data retrieval. This setup allows the NBAR processing to be integrated into the lazy evaluation chain, as `xarray` accommodates `dask` arrays (Rocklin, 2015), ensuring that the NBAR computation is conducted concurrently with cube creation. Moreover, if an ESDC is pre-existing, either stored in memory or on disk, it can be divided into chunks for processing. `sen2nbar` is then capable of executing the NBAR transformation in a lazy manner on these chunks. This approach is particularly effective with inherently chunked data formats, such as `zarr`, which can be processed lazily and re-written to disk in chunks, enhancing efficiency.

4.3 Refining Multidimensional Dataset Production

`sen2nbar` can enhance the accuracy of S2 data and derived features, supporting a wide range of applications, from individual snapshot analyses to comprehensive multitemporal studies. This extends to datasets generated from S2 data. This improvement will be largely attributed to the streamlined generation of ESDCs via STAC coupled with the data accuracy refine-

ment process offered by `sen2nbar`. This conversion process ensures interoperability across the temporal dimension for all kind of applications. As a result, `sen2nbar` streamlines the processing of multidimensional datasets like FluxnetEO (Walther et al., 2022), which originates from MODIS NBAR data designed for vegetation monitoring but is adaptable to S2 data, or BigEarthNet (Sumbul et al., 2019), derived from S2 patches aimed at land cover classification benchmarks. Therefore, the `sen2nbar` package emerges as a crucial tool in remote sensing, improving the accuracy and utility of data for a broad spectrum of applications.

4.4 Artificial Intelligence

The ultimate impact of `sen2nbar` is its potential to greatly improve AI models in remote sensing. With the introduction of Transformer architectures (Vaswani et al., 2017), AI has seen the development of new models that surpass traditional architectures across a variety of tasks, including Convolutional Neural Networks (CNNs) and Long-Short Term Memory (LSTM) networks. In remote sensing, the adoption of Transformer models has notably advanced, especially with pretrained models such as SITS-Former (Yuan et al., 2022). These foundation models, capable of learning from unlabeled datasets, are instrumental for various specific applications. Notable examples include Prithvi-100M (Jakubik et al., 2023) and SpectralGPT (Hong et al., 2023), which are trained on Harmonized Landsat Sentinel (HLS, Claverie et al., 2018) data and BigEarthNet, respectively. As these models form the basis for diverse applications, the role of `sen2nbar` in improving the quality of input datasets becomes invaluable. By ensuring that these models have access to clean and accurate data, particularly in a multitemporal context, `sen2nbar` will greatly contribute to the improvement of AI models' performance in remote sensing applications.

5. Conclusions

In this paper, we introduced `sen2nbar`, an open-source Python package crafted to facilitate the computation of Nadir BRDF Adjusted Reflectance (NBAR) values for Sentinel-2 L2A data. By enabling the computation of NBAR values through a single function, `sen2nbar` greatly simplifies the process, demonstrating its adaptability and compatibility with SAFE files as well as Earth System Data Cubes (ESDCs) derived from STAC through `stackstac` or `cubo`. We foresee `sen2nbar` playing an instrumental role in enhancing the accuracy of remote sensing data analyses within Earth system research. Its utility is especially pronounced in spatio-temporal analysis and in strengthening advanced AI-driven tasks via refined Sentinel-2 datasets.

Acknowledgements

References

Badgley, G., Field, C. B., Berry, J. A., 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. *Science Advances*, 3(3). http://dx.doi.org/10.1126/sciadv.1602244.

Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V. J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A. M., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1). http://dx.doi.org/10.1038/s41597-022-01307-4.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, A., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L., Jung, M., Gamon, J. A., Reichstein, M., Running, S. W., 2021. A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, 7(9). http://dx.doi.org/10.1126/sciadv.abc7447.

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, 145–161. http://dx.doi.org/10.1016/j.rse.2018.09.002.

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, 25–36. http://dx.doi.org/10.1016/j.rse.2011.11.026.

Frampton, W. J., Dash, J., Watmough, G., Milton, E. J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 83–92. http://dx.doi.org/10.1016/j.isprsjprs.2013.04.007.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T. E., 2020. Array programming with NumPy. *Nature*, 585(7825), 357–362. http://dx.doi.org/10.1038/s41586-020-2649-2.

Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., Plaza, A., Gamba, P., Benediktsson, J. A., Chanussot, J., 2023. SpectralGPT: Spectral Remote Sensing Foundation Model. https://arxiv.org/abs/2311.07113.

Hoyer, S., Hamman, J., 2017. xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software*, 5(1), 10. http://dx.doi.org/10.5334/jors.148.

Jakubik, J., Roy, S., Phillips, C. E., Fraccaro, P., Godwin, D., Zadrozny, B., Szwarcman, D., Gomes, C., Nyirjesy, G., Edwards, B., Kimura, D., Simumba, N., Chu, L., Mukkavilli, S. K., Lambhate, D., Das, K., Bangalore, R., Oliveira, D., Muszynski, M., Ankur, K., Ramasubramanian, M., Gurung, I., Khallaghi, S., Hanxi, L., Cecil, M., Ahmadi, M., Kordi, F., Alemohammad, H., Maskey, M., Ganti, R., Weldemariam, K., Ramachandran, R., 2023. Foundation Models for Generalist Geospatial Artificial Intelligence. https://arxiv.org/abs/2310.18660.

Knohl, A., Schulze, E.-D., Kolle, O., Buchmann, N., 2003. Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany. *Agricultural and Forest Meteorology*, 118(3–4), 151–167. http://dx.doi.org/10.1016/S0168-1923(03)00115-1.

Li, X., Strahler, A., 1992. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 276–292. http://dx.doi.org/10.1109/36.134078.

Lucht, W., Schaaf, C., Strahler, A., 2000. An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing*, 38(2), 977–998. http://dx.doi.org/10.1109/36.841980.

Mahecha, M. D., Gans, F., Brandt, G., Christiansen, R., Cornell, S. E., Fomferra, N., Kraemer, G., Peters, J., Bodesheim, P., Camps-Valls, G., Donges, J. F., Dorigo, W., Estupinan-Suarez, L. M., Gutierrez-Velez, V. H., Gutwin, M., Jung, M., Londoño, M. C., Miralles, D. G., Papastefanou, P., Reichstein, M., 2020. Earth system data cubes unravel global multivariate dynamics. *Earth System Dynamics*, 11(1), 201–234. http://dx.doi.org/10.5194/esd-11-201-2020.

Montero, D., Aybar, C., Ji, C., Kraemer, G., Söchting, M., Teber, K., Mahecha, M. D., 2024. On-Demand Earth System Data Cubes. https://arxiv.org/abs/2404.13105.

Montero, D., Aybar, C., Mahecha, M. D., Martinuzzi, F., Söchting, M., Wieneke, S., 2023a. A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research. *Scientific Data*, 10(1). http://dx.doi.org/10.1038/s41597-023-02096-0.

Montero, D., Kraemer, G., Anghelea, A., Aybar, C., Brandt, G., Camps-Valls, G., Cremer, F., Flik, I., Gans, F., Habershon, S., Ji, C., Kattenborn, T., Martínez-Ferrer, L., Martinuzzi, F., Reinhardt, M., Söchting, M., Teber, K., Mahecha, M., 2023b. Data Cubes for Earth System Research: Challenges Ahead. http://dx.doi.org/10.31223/X58M2V.

Pabon-Moreno, D. E., Migliavacca, M., Reichstein, M., Mahecha, M. D., 2022. On the Potential of Sentinel-2 for Estimating Gross Primary Production. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–12. http://dx.doi.org/10.1109/TGRS.2022.3152272.

Rocklin, M., 2015. Dask: Parallel computation with blocked algorithms and task scheduling. *Proceedings of the Python in Science Conference*, SciPy, SciPy.

Ross, J., 1981. *The radiation regime and architecture of plant stands*. Springer Netherlands.

Roujean, J., Leroy, M., Deschamps, P., 1992. A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *Journal of Geophysical Research: Atmospheres*, 97(D18), 20455–20468. http://dx.doi.org/10.1029/92JD01411.

Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. Technical report, NASA.

Roy, D., Li, Z., Zhang, H., 2017a. Adjustment of Sentinel-2 Multi-Spectral Instrument (MSI) Red-Edge Band Reflectance to Nadir BRDF Adjusted Reflectance (NBAR) and Quantification of Red-Edge Band BRDF Effects. *Remote Sensing*, 9(12), 1325. http://dx.doi.org/10.3390/rs9121325.

Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., Lindquist, E., 2008. Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sensing of Environment*, 112(6), 3112–3130. http://dx.doi.org/10.1016/j.rse.2008.03.009.

Roy, D. P., Li, J., Zhang, H. K., Yan, L., Huang, H., Li, Z., 2017b. Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 199, 25–38. http://dx.doi.org/10.1016/j.rse.2017.06.019.

Roy, D., Zhang, H., Ju, J., Gomez-Dans, J., Lewis, P., Schaaf, C., Sun, Q., Li, J., Huang, H., Kovalskyy, V., 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 176, 255–271. http://dx.doi.org/10.1016/j.rse.2016.01.023.

Söchting, M., Mahecha, M. D., Montero, D., Scheuermann, G., 2024. Lexcube: Interactive Visualization of Large Earth System Data Cubes. *IEEE Computer Graphics and Applications*, 44(1), 25–37. http://dx.doi.org/10.1109/MCG.2023.3321989.

Sumbul, G., Charfuelan, M., Demir, B., Markl, V., 2019. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. https://arxiv.org/abs/1706.03762.

Walther, S., Besnard, S., Nelson, J. A., El-Madany, T. S., Migliavacca, M., Weber, U., Carvalhais, N., Ermida, S. L., Brümmer, C., Schrader, F., Prokushkin, A. S., Panov, A. V., Jung, M., 2022. Technical note: A view from space on global flux towers by MODIS and Landsat: the FluxnetEO data set. *Biogeosciences*, 19(11), 2805–2840. http://dx.doi.org/10.5194/bg-19-2805-2022.

Wanner, W., Li, X., Strahler, A. H., 1995. On the derivation of kernels for kernel-driven models of bidirectional reflectance. *Journal of Geophysical Research: Atmospheres*, 100(D10), 21077–21089. http://dx.doi.org/10.1029/95JD02371.

Yuan, Y., Lin, L., Liu, Q., Hang, R., Zhou, Z.-G., 2022. SITS-Former: A pre-trained spatio-spectral-temporal representation model for Sentinel-2 time series classification. *International Journal of Applied Earth Observation and Geoinformation*, 106, 102651. http://dx.doi.org/10.1016/j.jag.2021.102651.