



OPEN

DATA DESCRIPTOR

A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research

David Montero¹✉, César Aybar^{ID 2,3}, Miguel D. Mahecha^{ID 1,4,5}, Francesco Martinuzzi^{ID 1,5}, Maximilian Söchting^{ID 1,6} & Sebastian Wieneke¹

Spectral Indices derived from multispectral remote sensing products are extensively used to monitor Earth system dynamics (e.g. vegetation dynamics, water bodies, fire regimes). The rapid increase of proposed spectral indices led to a high demand for catalogues of spectral indices and tools for their computation. However, most of these resources are either closed-source, outdated, unconnected to a catalogue or lacking a common Application Programming Interface (API). Here we present “Awesome Spectral Indices” (ASI), a standardized catalogue of spectral indices for Earth system research. ASI provides a comprehensive machine readable catalogue of spectral indices, which is linked to a Python library. ASI delivers a broad set of attributes for each spectral index, including names, formulas, and source references. The catalogue can be extended by the user community, ensuring that ASI remains current and enabling a wider range of scientific applications. Furthermore, the Python library enables the application of the catalogue to real-world data and thereby facilitates the efficient use of remote sensing resources in multiple Earth system domains.

Background & Summary

The constant monitoring of surface processes is a prerequisite for understanding Earth System’s dynamics and functioning^{1–6}. The big expansion of freely available remote sensing resources^{6–8} opens unprecedented opportunities for exploring interactions between the Earth’s surface and electromagnetic radiation. As new sensors come in ever higher spatial, temporal, and spectral resolutions, we constantly gain better insights into different Earth sub-systems. Surface materials have characteristic spectral signatures due to their physical properties and their interactions with electromagnetic radiation. As environmental effects can modify these interactions, spectral signatures give some insights into different surface processes, e.g., vegetation states⁹, urban expansion^{5,10}, fires, and burned areas^{11,12}. However, often one has to combine specific spectral regions of interest (i.e., spectral bands) in order to reduce unwanted (i.e., confounding effects) when studying a certain phenomenon or material^{13–16}. Such combinations of bands, named spectral indices, supply condensed information on specific sub-systems, materials, or processes. For different application domains, multiple proposals for ideal spectral indices exist, and the field is constantly developing and testing novel spectral indices.

The most common application domain of spectral indices is the terrestrial biosphere, with a special focus on vegetation monitoring via Vegetation Indices (VI). Typically, VIs use spectral information (i.e., reflectance) from the visible and near-infrared (NIR) regions of the electromagnetic spectrum, exploiting the chlorophyll absorption bands and plant structural features for assessing vegetation states¹⁷. Along these lines, multiple VIs have been created, and the total number of indices has been growing steadily through time^{17,18}. The most renowned and applied VI is the Normalized Difference Vegetation Index (NDVI)¹³, computed as the normalized

¹Remote Sensing Centre for Earth System Research (RSC4Earth), Leipzig University, Leipzig, 04103, Germany.

²Image Processing Laboratory, University of Valencia, Valencia, 46980, Spain. ³High Mountain Ecosystem Research Group, National University of San Marcos, 15081, Lima, Peru. ⁴Helmholtz Centre for Environmental Research, Leipzig, 04318, Germany. ⁵Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig, 04105, Germany. ⁶Image and Signal Processing Group, Leipzig University, Leipzig, 04109, Germany.

✉e-mail: david.montero@uni-leipzig.de

difference between NIR and red reflectances, and generally used to assess vegetation greenness in space and time. Variants have also been developed by incorporating additional correction factors, such as the canopy background adjustment parameter in the Soil Adjusted Vegetation Index (SAVI)¹⁴, which reduces soil brightness effects. Some spectral indices have even been designed to serve as correction factors for scattering effects in Solar Induced Fluorescence (SIF), e.g., Fluorescence Correction Vegetation Index (FCVI)¹⁹ and Hyperspectral NIRv (NIRvH)²⁰.

The goal of developing VIs specifically suited for a more accurate mapping of dynamic ecosystem processes, such as gross primary production (GPP, i.e., gross carbon uptake by plants)⁶, has led to a particularly wide variety of indices. The NIR Reflectance of Vegetation (NIRv)²¹ as well as the generalized Kernel NDVI (kNDVI)²² are examples of novel indices that aim to serve as a proxy for GPP. New sensors are capable of retrieving the reflectance in the red edge, which is highly sensitive to chlorophyll content due to the chlorophyll absorption shift to a longer wavelength. This led to a new set of indices around these bands. For instance, the Medium Resolution Imaging Spectrometer (MERIS) Terrestrial Chlorophyll Index (MTCI)²³ was created from the red and red edge bands of the sensor on board the Envisat satellite and served as a proxy for estimating the Red Edge Position, also known as the Red Edge Point (REP), which is used for estimating canopy chlorophyll content. In particular, the Multispectral Instrument (MSI) on board Sentinel-2 satellites enables the computation of red edge indices, such as the Inverted Red Edge Chlorophyll Index (IRECI) as well as the Sentinel-2 REP (S2REP)²⁴.

Indices for other application domains also contribute to the growing list of spectral indices. Water Indices (WI) generally use reflectance from the green, NIR and Shortwave Infrared (SWIR) bands to discriminate water bodies from other land cover types such as soil, vegetation, or constructed surfaces²⁵. One of the most used WIs is the Normalized Difference Water Index (NDWI)¹⁶, computed as the normalized difference between green and NIR reflectances for delineating water bodies. Variants of this index, such as the Modified NDWI (MNDWI)²⁶, have been created by replacing the NIR for SWIR reflectance, particularly reducing the influence of the noise produced by built-up areas. The number of WIs is also increasing due to the launching of new satellites and enhanced sensors. The Multi-Band Water Index (MBWI)²⁷ was developed for images of the Operational Land Imager (OLI) sensor of Landsat-8, while the Sentinel-2 Water Index (S2WI)²⁸ was created by taking advantage of the red edge bands in Sentinel-2 MSI imagery.

Modifications of existent indices as well as the development of new sensors with additional bands or different spectral characteristics are the main reasons for the growth of new spectral indices in other application domains. Besides WIs, indices for identifying burned areas and active fires are being expanded. For instance, the Burned Area Index (BAI)²⁹, one of the most used indices for detecting fire-affected areas, initially developed for the Advanced Very-High-Resolution Radiometer (AVHRR) instrument, has recently been modified to match Sentinel-2 MSI spectral bands (BAIS2)³⁰. BAIS2 uses red edge bands as well as the narrow NIR band. Another example is snow and ice cover indices. The Normalized Difference Snow Index (NDSI)¹⁵ was developed for the Landsat's Thematic Mapper (TM) and the Moderate Resolution Imaging Spectroradiometer (MODIS) instruments using the green and SWIR bands, while one of the newest indices, the Snow Water Index (SWI)³¹, developed for Landsat-8 OLI and Sentinel-2 MSI instruments, additionally use the NIR band aiming to discriminate better between snow cover and water bodies. This is also the case of Urban and Soil Indices. The Urban Index (UI)³², developed to map urban areas and investigate urban density and socio-economic variables, was constructed using images from the Landsat TM instrument. Contemporaneous indices such as the Dry Built-Up index (DBI)³³ use the Thermal Infrared (TIR) band from the Landsat OLI instrument to improve the discrimination of built-up areas and bare soil in dry climates. Likewise, Soil Indices like the Bareness Index (BaI)³⁴ were created for detecting the degree of bareness on the surface using bands from Landsat TM. By using bands from Landsat OLI, new indices such as the Modified Bare Soil Index (MBI)³⁵ and the Enhanced MBI (EMBI)⁵ have been developed, which particularly suppress built-up areas when detecting bare soil areas.

The steady growth of spectral indices across application domains inevitably calls for a central repository, gallery, or catalogue, where the remote sensing community can explore the vast amount of indices for specific applications or analyses. Arguably, one of the most comprehensive catalogues of spectral indices is the Index DataBase (IDB)^{36,37}, which has condensed several indices in a relational data model, specially VIs, as well as multiple attributes such as their formula, acronym, original name, used bands for different sensors and a collection of articles that cite most of the indices. The IDB includes a JavaScript Object Notation (JSON) Application Programming Interface (API) for querying spectral indices, which can be used upon request. Specifically for VIs, the review article by Xue *et al.*, (2017)¹⁸ presents a comprehensive review of several VIs, including their formulas and references. Proprietary software, such as ArcGIS and ENVI, count with their own spectral indices gallery (see <https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/indices-gallery.htm> and <https://www.3harrisgeospatial.com/docs/alphabeticallyspectralindices.html>, respectively). While they exhibit formulas, descriptions, and references, and can be accessed freely online, they require a license to compute them. There are few open-source projects providing tools to calculate spectral indices. Examples include the RadiometricIndices application from Orfeo Toolbox³⁸ (see https://www.orfeo-toolbox.org/CookBook/Applications/app_RadiometricIndices.html); the Vegetation Indices tools from SAGA-GIS³⁹ (see https://saga-gis.sourceforge.io/saga_tool_doc/8.4.1/imagery_tools.html); the spectralIndices function from RStoolbox (see <https://bleutner.github.io/RStoolbox/rstbx-docu/spectralIndices.html>), which is written in C and works with raster objects in R; the nightmares package, which also leverages the computation of spectral indices inside R (see <https://cran.r-project.org/web/packages/nightmares/index.html>); and the multispectral module from xarray-spatial (see <https://xarray-spatial.org/reference/multispectral.html>), which works with dask for xarray.DataArray objects⁴⁰ in Python. However, all the above mentioned open-source projects implement outdated sets of hard-coded indices that are not connected to a standardized open community catalogue.

Despite the fact that numerous resources have been developed for querying and computing spectral indices, multiple barriers still abridge the remote sensing community from exploiting them: The first obstacle is the outdated proprietary-dependent set of indices. The lack of novel spectral indices prevents users from making the most of spectral features that are offered by today's instruments. This hurdle grows bigger as multiple resources have a limited amount of indices in their catalogues, depriving users not just from novel, but also from traditional, as well as rare spectral indices. As the number of indices continues to increase, the remote sensing community requires an easy way to implement indices on their own on demand. The second obstacle is that spectral indices toolboxes are not connected to an existent open-source catalogue and indices are usually hard-coded inside specific modules. While this is not a major issue for open-source resources where the community can contribute by adding new indices, it is a stumbling block for non-developers who need a quick way to compute spectral indices that are not included yet. Vice versa, catalogues that are not linked to a toolbox are not entirely harnessing their potential by providing an application that the community may use to query and compute the included spectral indices.

The aim of this paper is to present “Awesome Spectral Indices” (ASI), an open community catalogue for handling spectral indices in Earth System research that addresses the research gap described above. ASI comprises a spectral indices catalogue, which is connected to a Python library. The catalogue of spectral indices is a free-to-access online open-source repository that follows a basic key-value structure and a simple standard for spectral indices in different application domains. The catalogue's straightforward architecture allows the community to request or add new indices as they become available. The Python library, named `spyindex`, uses the catalogue to create an user-friendly module for querying and automatically computing spectral indices for multiple Python objects (e.g. arrays and data frames). ASI leverages the data model structure as well as a Python library for easy management of spectral indices for their use in the analysis of Earth's surface processes using remote sensing data streams.

Methods

Catalogue design. *Data model.* The catalogue is designed as a simple nested key-value model (Fig. 2b) that stores attributes for multiple spectral indices. For simplicity, the set of spectral indices (represented by the acronyms as string objects) is defined as $\Psi = \{\psi_1, \dots, \psi_n\}$, where ψ is a spectral index and n is the number of indices in the list. The set of application domains is defined as $O = \{\ll\text{vegetation}\gg, \ll\text{water}\gg, \ll\text{burn}\gg, \ll\text{snow}\gg, \ll\text{urban}\gg, \ll\text{soil}\gg, \ll\text{radar}\gg, \ll\text{kernel}\gg\}$. The catalogue is formally defined as $C = \{(\psi_1, A_{\psi_1}), \dots, (\psi_n, A_{\psi_n})\}$, where C is the catalogue, ψ is a spectral index and $\psi \in \Psi$, A_{ψ} is the set of attributes of ψ and n is the number of indices in C . Each index ψ is the primary key of the model and the corresponding value A_{ψ} is a set of 9 attributes represented by a secondary key-value model (Fig. 2c) such that

$$A_{\psi} = \left\{ \begin{array}{l} (\ll\text{short_name}\gg, \psi) \\ (\ll\text{long_name}\gg, \eta_{\psi}) \\ (\ll\text{application_domain}\gg, o_{\psi}) \\ (\ll\text{formula}\gg, E_{\psi}) \\ (\ll\text{bands}\gg, V_{\psi}) \\ (\ll\text{platforms}\gg, Z_{\psi}) \\ (\ll\text{reference}\gg, r_{\psi}) \\ (\ll\text{date_of_addition}\gg, d_{\psi}) \\ (\ll\text{contributor}\gg, c_{\psi}) \end{array} \right\} \quad (1)$$

where ψ is the short name (i.e., acronym) of the index and $\psi \in \Psi$, η_{ψ} is the long name (i.e., original name) of the index, o_{ψ} is the application domain of the index and $o_{\psi} \in O$, E_{ψ} is the expression of the index, V_{ψ} is the set of operands in E_{ψ} and $V_{\psi} \subseteq V$ (see Expressions section), Z_{ψ} is the set of platforms with the required bands for the index computation, r_{ψ} is the URL or DOI of the reference of the index, d_{ψ} is the date of addition to C , and c_{ψ} is the URL of the contributor's GitHub profile.

Contributors must fill all attributes of an index except V_{ψ} and Z_{ψ} , which are extracted automatically after E_{ψ} is validated. The set of platforms Z_{ψ} is filled up taking V_{ψ} into account. If a platform $z \in Z$ counts with the required bands for a spectral index computation, that is, $V_{\psi} \subseteq V_z$, the platform z is added to Z_{ψ} . As of version 0.4.0, the set of platforms Z is $\{\ll\text{MODIS}\gg, \ll\text{Landsat-TM}\gg, \ll\text{Landsat-ETM+}\gg, \ll\text{Landsat-OLI}\gg, \ll\text{Sentinel-2}\gg, \ll\text{Sentinel-1 (Dual HH-HV)}\gg, \ll\text{Sentinel-1 (Dual VV-VH)}\gg, \ll\text{Planet-Fusion}\gg\}$. Note that platforms with the same sensors (or spectral bands) were grouped (e.g. $\ll\text{Landsat-TM}\gg$ refers to the group of the Landsat 4 and 5 satellites).

Expressions. An expression E is a string object that can be parsed as a representation of an operation (e.g. $\ll a + b \gg$ is an expression representing the $a + b$ mathematical operation, where $\ll a \gg$ and $\ll b \gg$ are operands, and $\ll + \gg$ is an operator). Given this, a well defined valid expression can be easily evaluated in any programming language, leveraging the need to re-write each spectral index for all languages where spectral indices can be implemented.

To establish a standard expression evaluation model, all possible expression operands V and operators G in spectral indices must follow a standard naming. The complete set of standards is given by $S = V \cup G$. Given this,

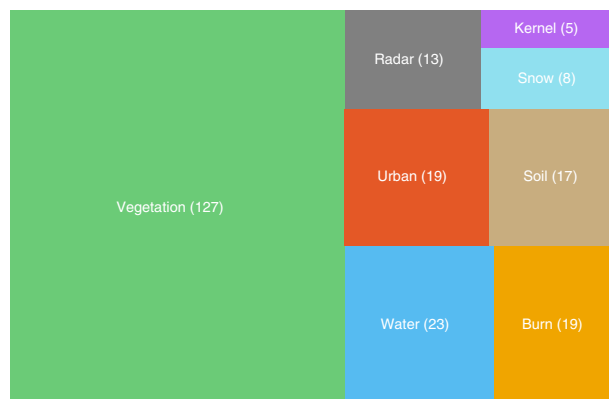
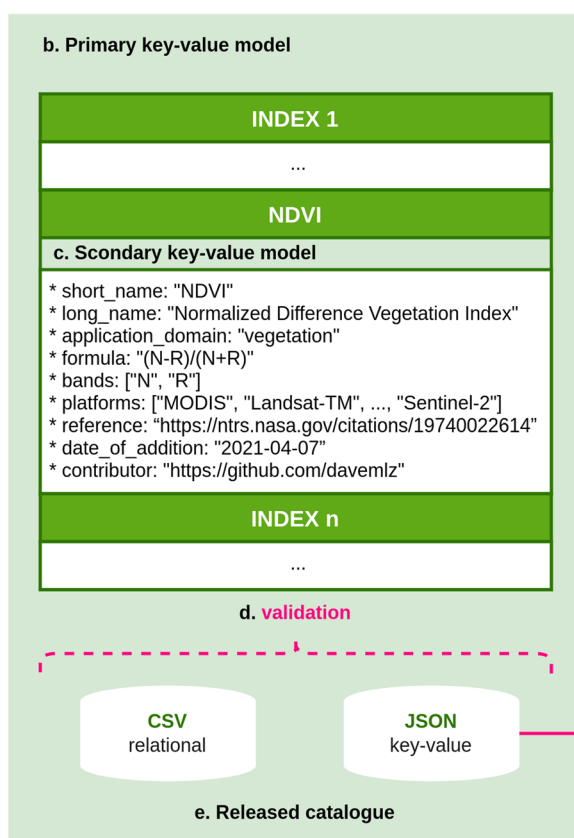


Fig. 1 Treemap of the count of spectral indices per application domain (as of version 0.4.0). The number of indices is reported inside parentheses.

a. Catalogue



f. Python library: spyindex

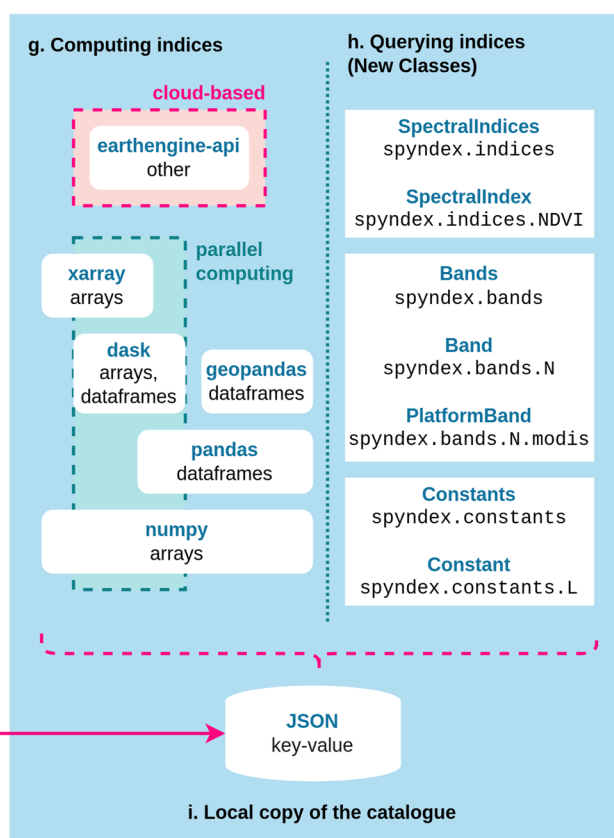


Fig. 2 A simplified diagram of the structure and workflow supported by ASI. (a) Simplified representation of the nested key-value data model, (b) primary key-value model representing indices and the set of attributes, (c) secondary key-value model representing the values for each attribute, (d) validation process of the catalogue before releasing, (e) final formats of the released catalogue, (f) simplified representation of the Python library, (g) libraries and data types suitable for spectral indices computation, (h) new classes with usage examples for querying spectral indices as well as their attributes and additional parameters, and (i) local copy of the JSON file used as base for the library.

bands, additional parameters, band dependent parameters, and common operators require a well defined standard. G was created from the standard operators $\ll + \gg$, $\ll - \gg$, $\ll * \gg$, and \ll / \gg , that were used for addition, subtraction, multiplication and division operations, respectively. The $\ll ** \gg$ operator was used for exponentiation operations instead of $\ll ^ \gg$ to make it compatible with Python. V was created from the bands of popular

Module	Class	Type	Computing
Python Built-in	int	Primitive	Serial
	float	Primitive	Serial
numpy	int	Primitive	Serial
	float	Primitive	Serial
	ndarray	Array	Serial
xarray	Dataset	Array	Serial
	DataArray	Array	Serial
pandas	Series	Array	Serial
	DataFrame	Tabular	Serial
geopandas	Series	Array	Serial
	GeoDataFrame	Tabular	Serial
earthengine-api	Image	Other	Parallel
	Number	Other	Parallel
dask	Array	Array	Parallel
	DataFrame	Tabular	Parallel

Table 1. Supported Python objects.

multispectral sensors in satellite remote sensing platforms, therefore, each standard name corresponds to a band in one of these platforms (Table 2). The standard names were kept as short as possible, representing them in most of cases as a single character in capital letters to avoid large string expressions. When more than one band for the same region is encountered, a consecutive number is added at the end of the standard name. For Radar indices, the standards «HH», «HV», «VV», and «VH» were created for the corresponding polarization mode no matter the wavelength (e.g. C-band of the C-SAR instrument on Sentinel-1 satellites).

This standardization accounts for spectral indices that do not require any additional parameters, such as the NDVI:

$$\text{NDVI} = (\rho_{\text{NIR}} - \rho_{\text{red}}) / (\rho_{\text{NIR}} + \rho_{\text{red}}) \quad (2)$$

where ρ_i is the reflectance in band i . Equation 2 is represented by $E_{\text{NDVI}} = \ll(N - R)/(N + R)\gg$ using S , where $V_{\text{NDVI}} = \{\ll N \gg, \ll R \gg\}$. Nevertheless, additional parameters must also be standardized for spectral indices that require them. These parameters are standardized according to the notation used in the original index, keeping the standard names as similar as possible to the original parameters no matter the length, and using the name of greek letters when encountered, e.g., the weighting coefficient α in the Wide Dynamic Range Vegetation Index (WDRVI)⁴¹ is represented by the standard name «alpha» (Table 3). Band dependent parameters, such as band wavelengths λ_i and kernels $K(a, b)$ were standardized by using the common names «lambdaA» and «kAB», where «lambdaA» is the wavelength of band A and «kAB» is a kernel function of A and B (being A and B band reflectances or other additional parameters).

By using these additional standard names, spectral indices like the Hyperspectral Near-Infrared Reflectance of Vegetation (NIRvH2)²⁰ are also accounted for. Given this, the equation:

$$\text{NIRvH2} = \rho_{\text{NIR}} - \rho_{\text{red}} - k \times (\lambda_{\text{NIR}} - \lambda_{\text{red}}) \quad (3)$$

where k is the soil slope between red and NIR and λ_i is the wavelength of band i , is represented by the expression $E_{\text{NIRvH2}} = \ll N - R - k * (\text{lambdaN} - \text{lambdaR}) \gg$, where $V_{\text{NIRvH2}} = \{\ll N \gg, \ll R \gg, \ll k \gg, \ll \text{lambdaN} \gg, \ll \text{lambdaR} \gg\}$. The «N» and «R» characters in the «lambdaN» and «lambdaR» expressions have no conflicts with the ρ_{NIR} and ρ_{red} standards since the complete expression is parsed as a single variable as long as it doesn't have any spaces or operators. This principle also applies to generalized kernel indices such as the kNDVI:

$$\text{kNDVI} = \frac{K(\rho_{\text{NIR}}, \rho_{\text{NIR}}) - K(\rho_{\text{NIR}}, \rho_{\text{red}})}{K(\rho_{\text{NIR}}, \rho_{\text{NIR}}) + K(\rho_{\text{NIR}}, \rho_{\text{red}})} \quad (4)$$

where $K(\rho_i, \rho_j)$ is a kernel function of the reflectances in bands i and j . Equation 4 is represented by the expression $E_{\text{kNDVI}} = \ll (\text{kNN} - \text{kNR}) / (\text{kNN} + \text{kNR}) \gg$ using S , where $V_{\text{kNDVI}} = \{\ll \text{kNN} \gg, \ll \text{kNR} \gg\}$. This standard has no conflicts with the standard of raw reflectances and allows the optimization of generalized kernel indices by enabling the free choice of the kernel function as well as their parameters.

Release. After validation (see Technical Validation), all indices are exported as a JSON file by keeping the original nested key-value model and the GitHub repository is updated. Additionally, the JSON file is used to convert the nested key-value model to a relational model exported as a CSV file. This model permits the use of the catalogue as a tabular structure that can be easily read as a data frame datatype in many programming languages (e.g. R and Julia). Furthermore, it can be loaded as a table using a Relational Database Management System (RDBMS) such as PostgreSQL, allowing the use of the domain specific SQL language for querying the

Band	Standard	Landsat			Sentinel		Terra/Aqua
		TM	ETM+	OLI	MSI-2A	MSI-2B	MODIS
Aerosols	A			1 (440.0)	1 (442.7)	1 (442.3)	
Blue	B	1 (485.0)	1 (485.0)	2 (480.0)	2 (492.4)	2 (492.1)	3 (469.0)
Green 1	G1						11 (531.0)
Green	G	2 (560.0)	2 (560.0)	3 (560.0)	3 (559.8)	3 (559.0)	4 (555.0)
Yellow	Y						
Red	R	3 (660.0)	3 (660.0)	4 (655.0)	4 (664.6)	4 (665.0)	1 (645.0)
Red Edge 1	RE1				5 (704.1)	5 (703.8)	
Red Edge 2	RE2				6 (740.5)	6 (739.1)	
Red Edge 3	RE3				7 (782.8)	7 (779.7)	
Near Infrared	N	4 (830.0)	4 (835.0)	5 (865.0)	8 (832.8)	8 (833.0)	2 (858.5)
Near Infrared 2	N2				8A (864.7)	8A (864.0)	
Water Vapour	WV				9 (945.1)	9 (943.2)	
Short-wave Infrared 1	S1	5 (1650.0)	5 (1650.0)	6 (1610.0)	11 (1613.7)	11 (1610.4)	6 (1640.0)
Short-wave Infrared 2	S2	7 (2215.0)	7 (2220.0)	7 (2200.0)	12 (2202.4)	12 (2185.7)	7 (2130.0)
Thermal Infrared	T	6 (11450.0)	6 (11450.0)				
Thermal Infrared 1	T1			10 (10895.0)			
Thermal Infrared 2	T2			11 (12005.0)			

Table 2. Standard band naming and the corresponding band number for different platforms. Central wavelengths in nm are specified for each platform in parenthesis.

Parameter	Standard	Spectral Indices
Gain Factor	g	EVI
Canopy Background Adjustment	L	EVI, EVI2, MNLI, SARVI, SAVI, SAVIT
First Coefficient for the aerosol resistance term	C1	EVI, kEVI
Second Coefficient for the aerosol resistance term	C2	EVI, kEVI
c correction factor	cexp	OCVI
n exponent for the SR power operation	nexp	GDVI
α weighting coefficient	alpha	BWDRVI, NDPI, WDRVI
β calibration parameter	beta	NDSI _{ns}
γ weighting coefficient	gamma	ARVI
ω weighting coefficient	omega	MBWI
$f(\Delta)$ adjustment factor for terrain shadows distortion	fdelta	SEVI
Soil line slope	sla	ATSAVI, SAVI2, TSAVI, WDWI
Soil line intercept	slb	ATSAVI, SAVI2, TSAVI
Photosynthetically Active Radiation	PAR	NIRvP
Soil slope parameter	k	NIRvH2
Center wavelength of band A (nm)	lambdaA	DVI+, NDGI, NIRvH2
Kernel function of A and B	kAB	kEVI, kNDVI, kRVI, kVARI

Table 3. Standard additional parameters naming and the spectral indices where they are used.

catalogue, and even as a spreadsheet in proprietary software such as Microsoft Excel. After being updated, a new version of the GitHub repository is officially released. All sources are matched to their relevant indices in the main data files using the «reference» attribute, and cited here^{5,9–16,19–35,41–166}.

Python library design. `spyndex` is an open-source Python package hosted in GitHub that works as the Python library for ASI (Fig. 2f). This package allows users to query and compute spectral indices from the catalogue. The spectral indices are automatically updated in the `spyndex` repository once they are released by keeping a local copy of the JSON file provided by the catalogue itself (Fig. 2i). This allows users to access and compute spectral indices even if they don't have an internet connection. Nevertheless, users can still use the most recent version of the catalogue for spectral indices computation if an internet connection is provided.

Querying spectral indices and parameters. The local copy of the JSON file is used to create a new class named `SpectralIndex`. This class stores the attributes of each index in the catalogue and saves them as its own attributes. Additionally, this class counts with a method for computing the index by passing the required parameters as a dictionary object or as keywords arguments. By using this class, each index in the local copy of the

catalogue is transformed to a `SpectralIndex` object that can be accessed and computed directly. All transformed indices are then stored in a `SpectralIndices` object, which is constructed by inheritance from the `Box` class¹⁶⁷, allowing indices querying by using dot notation and text auto-completion (Fig. 2h).

Standards for band reflectances also have a new specific class for querying information relative to different satellite platforms and their specifications. The `Band` class is created to store the `«standard»` and `«long_name»` attributes, which users can access from the specific band. Additionally, the attributes `«sentinel2a»`, `«sentinel2b»`, `«landsat4»`, `«landsat5»`, `«landsat7»`, `«landsat8»`, `«landsat9»` and `«modis»` are also created for each band, containing information on the band name, band number, center wavelength (nm) and bandwidth (nm) for each specific platform in a class named `PlatformBand`. The `Band` class is useful for the computation of indices that require the specific platform band wavelengths λ_i such as the NIRvH2 and the Normalized Difference Greenness Index (NDGI)¹⁵⁹. All bands are then stored in a `Bands` object, also inherited from the `Box` class to allow the use of dot notation and auto-completion (Fig. 2h).

Additional parameters (Table 3) that also follow a standard naming are included in a new class named `Constant`. This class stores the `«standard»`, `«long_name»` and `«default»` attributes, representing the standard name, the original name and the default value of the parameter. Furthermore, specific kernel parameters were also included here for the calculation of kernel indices. These kernel parameters include the length-scale parameter σ for the Radial Basis Function (RBF) Kernel K_{RBF} (Eq. 7), and the polynomial degree p and the trade-off parameter c in the Polynomial Kernel K_{poly} (Eq. 6). By using this class, users can decide whether to use the default values for the additional parameters, or to optimize them by using any other value or a compatible Python object. This becomes handy when computing indices that require additional parameters such as the Enhanced Vegetation Index (EVI)⁹⁰ and the Optimized Chlorophyll Vegetation Index (OCVI)¹⁴¹, among others. All parameters are stored in a `Constants` object inherited from a `Box` class for their access using dot notation and text auto-completion (Fig. 2h).

Computing spectral indices. In Python, a single string expression can be evaluated as $f(E, P)$, where f is the evaluation function, E is the expression to evaluate, and P is a dictionary of local variables that serve as inputs for undefined variables in E . Since $E_\psi \in A_\psi$ is a valid expression that was curated in the catalogue validation process, it can be used without any modification to compute a spectral index by being evaluated. This evaluation is performed by passing the required parameters V_ψ for the formula computation as interoperable Python objects that support mathematical overloaded operators. This means that the evaluation process is successful for objects with mathematical overloaded operators that perform the actual mathematical operations (e.g. a `«+»` operator performs an addition for two integers), but it will fail for objects with mathematical overloaded operators that do not perform the actual mathematical operations (e.g. a `«+»` operator performs a concatenation for two lists).

By this definition, spectral indices can be computed by taking advantage of the expression evaluation process in Python. A single spectral index ψ can then be evaluated by $f(E_\psi, P_\psi)$, where E_ψ is the expression of the spectral index ψ and P_ψ is the dictionary of parameters required to evaluate the index expression such that $P_\psi = \{(v_1, \rho_1), \dots, (v_n, \rho_n)\}$, where $v_i \in V_\psi$ and ρ_i is the reflectance of band i as a compatible Python object. For example, to calculate the NDVI (Eq. 2) it is required to compute $f(E_{\text{NDVI}}, P_{\text{NDVI}})$, where $E_{\text{NDVI}} = \langle\langle N - R \rangle\rangle / \langle\langle N + R \rangle\rangle$, and $P_{\text{NDVI}} = \{(\langle\langle R \rangle\rangle, \rho_{\text{red}}), (\langle\langle N \rangle\rangle, \rho_{\text{NIR}})\}$. Here, P_{NDVI} is a dictionary where the standard names `«N»` and `«R»` are keys and their values ρ_i can be either single values or n -dimensional arrays.

This computation process is leveraged by the function `computeIndex`, which takes the acronym of the index ψ and the dictionary of parameters P_ψ for the index computation as inputs. By doing this, `computeIndex` works on a higher level by computing $g(\psi, P_\psi)$ instead of $f(E_\psi, P_\psi)$, which leverages the need of writing the complete expression by just using the index acronym. E_ψ is taken from the `«formula»` attribute in the local copy of the catalogue, although it can be requested directly from the online version in the updated catalogue. P_ψ is passed by the user and must contain the same keys as V_ψ . If this is not complied, the function will raise an error. In addition, the dictionary of parameters is interchangeable with keyword arguments for a more fluid scripting.

Multiple indices can be computed at once by passing a list of index acronyms $\Psi_s \subseteq \Psi$ to the function. In this case, the dictionary of parameters is now $P_{\Psi_s} = \bigcup_{i=1}^n P_{\psi_i}$, where $\psi_i \in \Psi_s$, and n is the number of indices in Ψ_s . The function can then be computed as $g(\Psi_s, P_{\Psi_s})$, which is easier and more efficient than computing $g(\psi_i, P_{\psi_i})$ for all $\psi_i \in \Psi_s$. For example, in order to compute the Chlorophyll Index Red Edge (CI_{RE})⁷⁸, where $P_{\text{CI}_{\text{RE}}} = \{(\langle\langle \text{RE1} \rangle\rangle, \rho_{\text{RE1}}), (\langle\langle N \rangle\rangle, \rho_{\text{NIR}})\}$, and the NDVI, where $P_{\text{NDVI}} = \{(\langle\langle R \rangle\rangle, \rho_{\text{red}}), (\langle\langle N \rangle\rangle, \rho_{\text{NIR}})\}$, the required dictionary of parameters is now passed as $P_{\text{CI}_{\text{RE}}} \cup P_{\text{NDVI}} = \{(\langle\langle R \rangle\rangle, \rho_{\text{red}}), (\langle\langle \text{RE1} \rangle\rangle, \rho_{\text{RE1}}), (\langle\langle N \rangle\rangle, \rho_{\text{NIR}})\}$, and both indices can be computed at once by passing $\Psi_s = \{\langle\langle \text{CI}_{\text{RE}} \rangle\rangle, \langle\langle \text{NDVI} \rangle\rangle\}$.

Generalized kernel indices can also be computed using `computeIndex`. However, to allow the optimization of kernel parameters, the `computeKernel` was created. This function permits users to compute a kernel $K(a, b)$, where K is a kernel function, and a and b can be whether reflectances or additional parameters. As an illustrative example, the kNDVI computation requires $P_{\text{kNDVI}} = \{(\langle\langle \text{kNN} \rangle\rangle, K(\rho_{\text{NIR}}, \rho_{\text{NIR}})), (\langle\langle \text{kNR} \rangle\rangle, K(\rho_{\text{NIR}}, \rho_{\text{red}}))\}$, where K can be any kernel function computed using the `computeKernel` method. Three kernel functions are implemented in `spyndex`. These kernels are (1) the Linear Kernel:

$$K_{\text{linear}}(a, b) = ab \quad (5)$$

(2) the Polynomial Kernel:

$$K_{\text{poly}}(a, b) = (ab + c)^p \quad (6)$$

where c is a trade-off parameter between higher and lower order terms in the polynomial and p is the polynomial degree, and (3) the RBF Kernel:

$$K_{\text{RBF}}(a, b) = \exp\left(-\frac{(a - b)^2}{2\sigma^2}\right) \quad (7)$$

where σ is the length-scale parameter. The K_{RBF} is suggested for the kNDVI computation with $\sigma = 0.5(\rho_{\text{NIR}} + \rho_{\text{red}})^{22}$. However, the σ parameter can be chosen by the user in `computeKernel`, allowing the optimization of K_{RBF} . This principle also applies for the c and p parameters in K_{poly} .

Likewise, this is useful for the creation and optimization of generalized kernel indices. Since all indices can be kernelized²², the `computeKernel` function can be used for the required bands in kernelized indices that are being created and are not part of the catalogue yet. As an illustrative example, the Green Normalized Difference Vegetation Index (GNDVI)⁷⁴ can be kernelized as:

$$\text{kGNDVI} = \frac{K(\rho_{\text{NIR}}, \rho_{\text{NIR}}) - K(\rho_{\text{NIR}}, \rho_{\text{green}})}{K(\rho_{\text{NIR}}, \rho_{\text{NIR}}) + K(\rho_{\text{NIR}}, \rho_{\text{green}})} \quad (8)$$

and all kernels in this equation can be easily computed by passing the corresponding ρ_{NIR} and ρ_{green} reflectances to `computeKernel` and selecting the desired kernel function with the required parameters for optimization.

Data Records

The ASI Catalogue (Fig. 2a) is a curated growing collection of currently 231 spectral indices (as of version 0.4.0) widely used in remote sensing. The catalogue is divided into 8 groups, which mostly represent specific application domains (Fig. 1), namely: vegetation, water, burn, snow, urban, radar, soil, and kernel indices. Note that radar and kernel indices reflect specific methodological approaches. For instance, indices from any Earth system application domain could be converted to generalized kernel indices²². However, we decided to create a group in its own right for indices of this kind since their mathematical structure involves a free parameterization.

The complete set of spectral indices per application domain (as of version 0.4.0) in JSON and CSV formats is available online via Zenodo¹⁶⁸. The JSON file stores the catalogue using a key-value model, with the acronym of each index as the key and the set of 9 attributes as the value. The CSV file stores the catalogue as a relational model, with each row representing an index and the 9 attributes as columns. The data types for each attribute are listed in Table 4. Note that by modifying the original key-value model in the CSV file, the `<<bands>>` and `<<platforms>>` attributes lose their array structure and additional string management must be done to recover it while keeping the relational model.

In addition to the set of spectral indices, two JSON files containing multiple attributes for the standard of each band as well as the additional parameters are included in the Zenodo repository. The JSON file for the bands includes the `<<long_name>>`, `<<short_name>>`, `<<platforms>>`, `<<min_wavelength>>`, and `<<max_wavelength>>` attributes. These attributes describe each one of the bands, including their spectral range and the set of platforms where the band is found. Furthermore, it includes the `<<common_name>>` attribute, which creates a direct mapping of the ASI standard to the Electro-Optical Extension Specification (see <https://github.com/stac-extensions/eo/>) for SpatioTemporal Asset Catalogues (STAC). The JSON file for the additional parameters includes the `<<default>>`, `<<description>>`, and `<<short_name>>` attributes. These attributes describe each one of the additional parameters, including the default value of the parameter in the original paper where it is found.

Technical Validation

Catalogue validation. The living catalogue is continuously updated by the main developers based on rigorous literature reviews. Users can request the addition of new indices by using issues or pull requests. All spectral indices are validated using `pydantic`¹⁶⁹ (Fig. 2d), a Python package for data validation using type annotations. For this validation step, all indices must comply with the following constraints before being added to the catalogue: (1) All attributes must be filled with a string datatype, (2) ψ must be written in capital letters if possible and must not contain spaces, (3) E_ψ must be a valid expression and must use items from the set of standards S (see Expressions section), (4) d_ψ must be in “YYYY-MM-DD” format, (5) c_ψ must be a valid GitHub profile URL or an email, and (6) $o_\psi \in O$. If the input indices do not comply with the above mentioned constraints the system raises an error and the catalogue is not updated until corrected. In the case of source references, the main developers review the reference provided by the contributor of an index. If the reference is not the original source of the index (e.g., a published paper, a poster, or grey literature), the index is not added to the catalogue until the original source is provided.

Compatibility with python objects. The expression evaluation taking mathematical overloaded operators into account makes it possible for different types of Python objects to be compatible with `spyndex` (Table 1). First, the primitive numerical Python built-in types `int` (integers) and `float` (floating points) are compatible with mathematical overloaded operators and thus compatible with `spyndex`. This primitive compatibility allows most Python objects that are built on top of them to be compatible with `spyndex`. Second, the primitive data

Attribute	Data type
short_name	string
long_name	string
application_domain	string
formula	string
bands	array (string)
platforms	array (string)
reference	string
date_of_addition	datetime
contributor	string

Table 4. Spectral indices attributes and their data types. Data types of array items are presented in parenthesis.

types in the `numpy` package¹⁷⁰ (e.g., `numpy.int` and `numpy.float`) are also compatible with mathematical overloaded operators and therefore most Python objects built on top of `numpy` are also compatible with `spyndex` (Fig. 2g).

This compatibility with the primitives of Python and `numpy` makes `spyndex` compatible with the most used objects: arrays and data frames. The primitive array objects can be created through `numpy` and they can be n -dimensional. These primitive arrays are the source of data arrays in the `xarray` package⁴⁰, which allows the creation and manipulation of labelled data arrays and datasets. This is extremely useful for images with 3 dimensions {columns, rows, bands}, or data cubes with 4 dimensions {columns, rows, bands, time}. By using `xarray.DataArray` objects as inputs in a spectral index calculation, the result will be another `xarray.DataArray` with the same dimensions as the input arrays. If more than one index is calculated at the same time, by default, the result is a `xarray.DataArray` with an extra dimension named «index» and the acronyms of the indices as labels.

Additionally, `spyndex` is also compatible with data frames in the `pandas` package¹⁷¹, which allows the creation and manipulation of tabular data, and `geopandas`¹⁷², which adds a spatial dimension to tabular data in `pandas`. Usually, the input data consist of `pandas.Series` or `geopandas.Series`, and the result would be another series with the same length when one index is computed. When multiple indices are computed, by default, the result is a `pandas.DataFrame` where each column corresponds to a spectral index.

This compatibility enables the direct use of SpatioTemporal Asset Catalogues (STAC) for spectral indices computation with `spyndex`, allowing the querying of satellite imagery from a STAC provider (e.g., Planetary Computer) and the direct computation of spectral indices for the requested images according to the data type in which the data is stored. Additionally, `spyndex` also extends the compatibility to Google Earth Engine (GEE) objects for its Python API⁸. GEE is a cloud-based geoprocessing service that counts with a vast catalogue of satellite imagery that users can process online. Originally, GEE objects, such as `ee.Image` and `ee.Number`, are not compatible with mathematical overloaded operators. However, they are extended through the `eemont` package¹⁷³. This permits the `spyndex` compatibility with GEE objects for large spectral indices computations in the cloud.

Other objects can also be compatible with `spyndex` if mathematical overloaded operators are implemented for the object's class. This is useful for users that aim to create their own classes or modules. By emulating numeric types, the created classes can serve as inputs for objects that are compatible for spectral indices computation. If the new classes are created by inheritance of classes that already have overloaded operators, the new classes are automatically compatible with `spyndex` (e.g., a new class named `Signature` that inherits from `numpy.ndarray` and that represents a spectral signature can be used to compute spectral indices with `spyndex`).

Usage Notes

The potential of the ASI catalogue can be exploited to boost computations of spectral indices from real-world data using parallel computing and multidimensional data management through the `spyndex` Python library. Nevertheless, the machine readable catalogue can be used in further programming languages with a special potential for Earth system research. Furthermore, advanced applications involving hyperspectral indices or Machine Learning (ML) pipelines can also be exploited with ASI.

Python library usage examples. As the `spyndex` Python library works with any type of object that is compatible with overloaded operators, primitive types can be used to compute one or multiple spectral indices in a simple way. This is also true for more complex structures such as `numpy.ndarray` or `pandas.Series` objects (e.g. Figure 3). Nevertheless, real world applications usually require complex multidimensional objects that store several amounts of remote sensing data. Figure 4 shows an example where a Sentinel-2 L2A data cube for the year 2018 is retrieved as a `xarray.DataArray` given a specific pair of coordinates in the Hainich National Park, Germany, followed by the computation of NDVI, NIRv, and kNDVI.

Parallel computing. Remote sensing datasets are usually large, and calculating spectral indices can be computationally expensive. For example, a single Sentinel-2 L2A image may have a data size of around 800 MB, which can increase by resampling the 20 and 60 m bands to 10 m, the highest spatial resolution of the MSI sensor. A multitemporal study using just Sentinel-2 imagery can easily overpass 1 TB of data, and appropriate scalability

```

import spyndex
import numpy as np
import pandas as pd

# Using floats
idx = spyndex.computeIndex(
    index=["NDVI", "SAVI"],
    N=0.68,
    R=0.09,
    L=0.5
)

# Using numpy
idx_numpy = spyndex.computeIndex(
    index=["NDVI", "SAVI"],
    N=np.random.normal(0.6, 0.10, 10000),
    R=np.random.normal(0.1, 0.05, 10000),
    L=0.5
)

# Using pandas
df = pd.DataFrame(
    dict(
        NIR=np.random.normal(0.6, 0.10, 10000),
        Red=np.random.normal(0.1, 0.05, 10000)
    )
)

idx_pandas = spyndex.computeIndex(
    index=["NDVI", "SAVI"],
    N=df["NIR"],
    R=df["Red"],
    L=0.5
)

```

Fig. 3 Example code for computing NDVI and SAVI for float, numpy.ndarray, and pandas.Series objects. Both indices can be computed at the same time with the same input values. Furthermore, additional parameters such as the Canopy Background Adjustment L can be passed as a single constant value. Note that spyndex returns an object of the same type as the input objects.

is required to avoid delayed processing. Computing spectral indices by using GEE objects can leverage this to scale the satellite imagery processing, which is a transparent process for users. In the specific case of the GEE JavaScript API, users may use `spectral`¹⁷⁴, a module that extends the ASI catalogue to the Code Editor. However, when using Python, users may prefer to use arrays and data frames to have greater processing flexibility for remote sensing data. This is also useful when Machine Learning (ML) or Deep Learning (DL) modelling techniques are going to be applied.

To leverage this scalability, `dask` objects can be used in combination with `numpy`, `xarray` and `pandas`. `dask` is a library for parallel computing in Python that extends the arrays and data frame objects by using chunks of data in combination with a dynamic task scheduling. Since it extends the already known arrays and data frames, `dask` classes are also compatible with `spyndex` (Table 1), allowing their calculation in parallel computing instead of the usual serial computing. By using `dask`, spectral indices computation can be done in parallel and multiple indices computation at once becomes an advantage over serial computing of spectral indices (Fig. 5).

Computing spectral indices for multidimensional arrays. As Earth System Data Cubes (ESDC) are gaining strength in Earth system research², spectral indices cubes can be added as new derived variables in them. Taking the vegetation application domain as example, here we present how ASI can be used for investigating the spatiotemporal spectral response of vegetation by using spectral indices at the ecosystem scale in the Hainich National Park, a Broadleaf Deciduous Forest (DBF) in central Germany¹⁷⁵. The `spyndex` Python library was used to compute several spectral indices for Sentinel-2 imagery. To exploit the potential of data cubes for Earth system research, we will describe this example by extending the concept of ESDC to Earth System Mini Cubes

```

import spyndex
import cubo
import xarray as xr
import numpy as np

da = cubo.create(
    lat=51.0792,
    lon=10.4521,
    collection="sentinel-2-l2a",
    bands=["B04", "B08"],
    start_date="2018-01-01",
    end_date="2018-12-31",
    edge_size=350,
    resolution=10,
    query={"eo:cloud_cover": {"lt": 2}}
).where(lambda x: x > 0, other=np.nan)

da = da.groupby("time").mean()

idx = spyndex.computeIndex(
    index=["NDVI", "NIRv", "kNDVI"],
    N=da.sel(band="B08"),
    R=da.sel(band="B04"),
    kNN=1.0,
    kNR=spyndex.computeKernel(
        kernel="RBF",
        a=da.sel(band="B08"),
        b=da.sel(band="B04"),
        sigma=da.sel(band=["B08", "B04"]).mean("band").median("time")
    )
)

```

Fig. 4 Example code for computing NDVI, NIRv, and kNDVI for a xarray.DataArray object. First, a Sentinel-2 L2A data cube is retrieved as a data cube via the cubo Python package. This data cube is a chunked xarray. DataArray with 4 dimensions: <<bands>>, <<time>>, <<x>>, and <<y>>. The <<bands>> dimension contains the labeled data for NIR and red surface reflectance values, therefore, it is used to select the values for each operand in the computation. Furthermore, this example displays how to compute a kernel for kernel indices such as the kNDVI. Note that the <<kNN>> operand was set to 1.0 since K_{RBF} was the selected kernel and $K_{\text{RBF}}(x, x) = 1.0$.

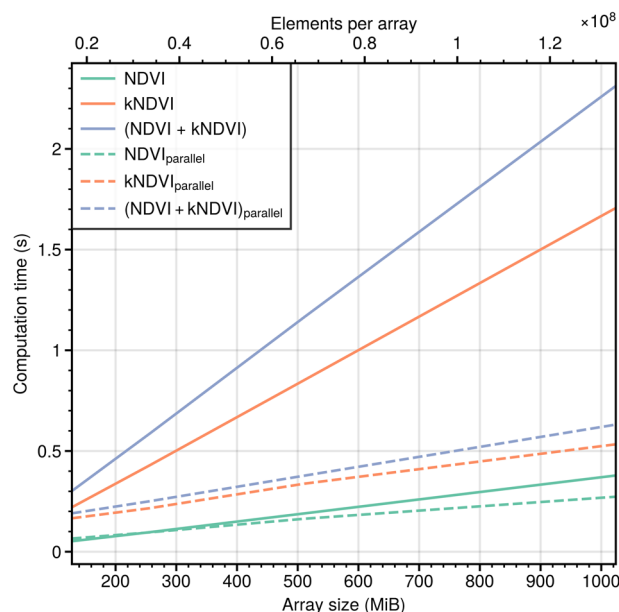


Fig. 5 Computation time for NDVI, kNDVI and both indices at the same time (NDVI + kNDVI) for different array sizes using randomly generated array objects with chunks of 64 MiB. Performance was evaluated for serial (bold lines) and parallel (dashed lines) computing.

(ESMC), i.e., data cubes with a small spatial extent, but with a higher spatiotemporal resolution than global data cubes).

Sentinel-2 L2A MSI data from the Hainich National Park was retrieved from the Microsoft's Planetary Computer platform. All images were centered at the DE-Hai Eddy Covariance (EC) Tower¹⁷⁶ from the Integrated Carbon Observation System (ICOS) and stored as a ESMC $\mathcal{C}_{S_2}(L)$ in zarr format, where L is the set of

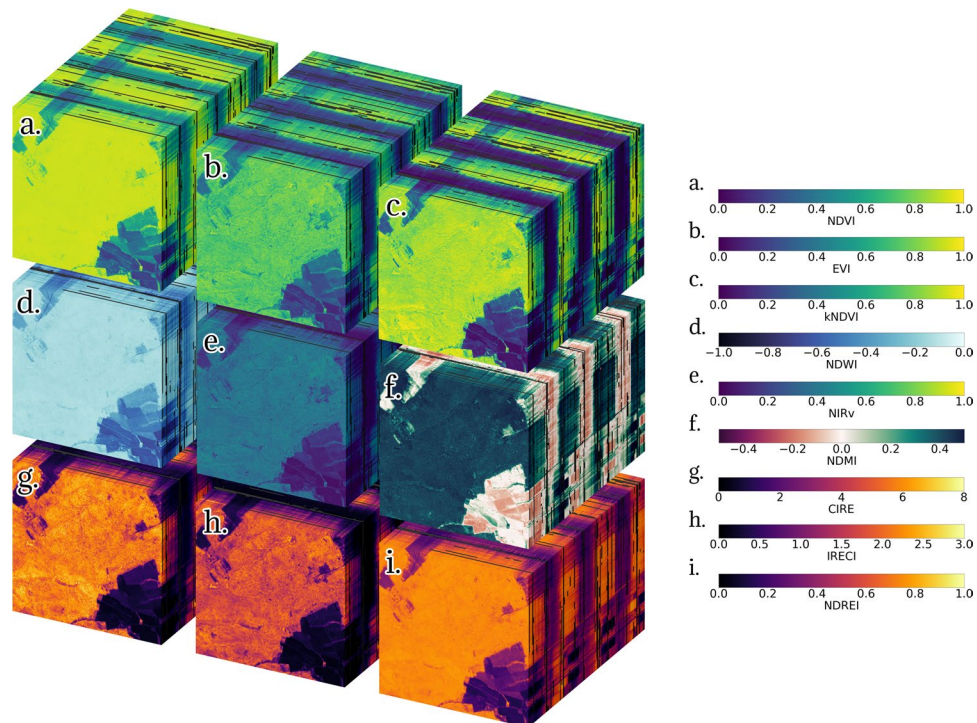


Fig. 6 Spectral indices mini cubes computed from Sentinel-2 displaying a 2.56 km radius around the DE-Hai site. The cubes have dimensions $\{x = 512, y = 512, \text{time} = 195\}$. The initial date is 2018-07-09 (front) and the final date is 2021-09-26 (back). The displayed indices are (a) NDVI, (b) EVI, (c) kNDVI, (d) NDWI, (e) NIRv, (f) NDMI, (g) CIRE, (h) IRECI, and (i) NDREI. For an interactive variant of the image visit <https://www.lexcube.org/>.

dimensions of \mathcal{C}_{S_2} , and $L = \{\text{time}, \text{band}, x, y\}$. The ESMC was preprocessed as follows: (1) the cube was cropped to a spatial extent of a 2.56 km radius around the EC tower, (2) the surface reflectance was adjusted to nadir BRDF reflectance using the *c*-factor approach^{177–179}, (3) all bands were resampled to 10 m using the nearest neighbour algorithm, (4) clouds and cloud shadows were masked using the KappaMask L2A model¹⁸⁰, and (5) snow was masked using the Fmask algorithm^{181,182}. The temporal resolution of the dataset is 1-day and covers the period 2016–2021. However, since the revisit time of the Sentinel-2 constellation only allows for an image every 2–5 days, most of the time steps are missing.

The mini cube $\mathcal{C}_{S_2}(L)$ was used to compute multiple spectral indices such that it was mapped to a spectral indices mini cube $\mathcal{C}_{S_2}(R)$, where R are the set of output dimensions, and $R = \{\text{time}, \text{index}, x, y\}$. Since not all indices can be computed for Sentinel-2, a subset $\Psi_{S_2} \subset \Psi$ was selected such that $\llcorner \text{Sentinel-2} \gg \in Z_\psi$, where $\psi \in \Psi$. $\mathcal{C}_{S_2}(L)$ was loaded as a chunked `xarray.DataArray` object and the spectral indices Ψ_{S_2} were computed in parallel using `spynex` and `dask`. The additional optimizable parameters in spectral indices that required them were set to their default value according to the literature and the kernel indices were computed using $K_{\text{RBF}}(\rho_i, \rho_j)$ (Eq. 7) with

$$\sigma = \text{median}_{\substack{\{x,y\} \\ \{\text{time}, x, y\}}} \left(\frac{\rho_i + \rho_j}{2} \right) \quad (9)$$

where ρ is the spatiotemporal univariate mini cube of bands i and j . Note that for σ the temporal dimension was reduced by mapping the median function across each pixel of the cube^{22,183}. This applies not only for the kNDVI, but also for the remaining kernel indices.

For this document, we selected 9 spectral indices that use different bands from the visible, Red Edge, NIR and SWIR spectrum: (1) NDVI¹³, (2) EVI⁹⁰, (3) kNDVI²², (4) NDWI¹⁶, (5) NIRv²¹, (6) Normalized Difference Moisture Index (NDMI)¹⁴⁸, (7) CIRE⁷⁸, (8) IRECI²⁴, and (9) Normalized Difference Red Edge Index (NDREI)⁷². The complete set of spectral indices ψ_{S_2} can be explored with the Leipzig Explorer of Earth Data Cubes (Lexcube, see <https://www.lexcube.org/>). Figure 6 shows the visualization of the 9 spectral indices for the DE-Hai site. Since all indices were computed from \mathcal{C}_{S_2} , the shape of the x , y , and time dimensions are maintained for each index. As the time dimension was no regular in the original cube, the timesteps are not regular in the spectral indices cubes either. Note that there are visible gaps due to the cloud-, clouds shadow- and snow-masking for all cubes (i.e., NaN values are kept as they are when indices are computed).

This example demonstrates that multiple spectral indices can be computed for high-resolution data in multi-dimensions using the Python library of ASI. While this example shows how to compute indices for Sentinel-2

data, it can be naturally extended to coarser resolution data like MODIS or to finer resolution data like satellite imagery from Planet constellations. This shows the flexibility of ASI according to the study scale. Furthermore, the potential increases with the number of indices that can be computed. Here we presented 9 spectral indices representing combinations from all spectral regions available for the Sentinel-2 platform, but hundreds more can be computed from the same bands. Additionally, indices from different application domains can be applied to each other. In this case, WIs were computed for a vegetation application domain since the water absorption bands may give meaningful insights for monitoring vegetation states.

Spectral indices and platform capabilities. As shown before, the ASI catalogue is suitable for platforms that work on high spatial resolution datasets with a rich spectral resolution such as Sentinel-2. While this is true for the platform that was tested here, it can also serve multiple former platforms that were not mentioned, such as Landsat 1-5 with the MultiSpectral Scanner (MSS), the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), MODIS, or even hyperspectral platforms like Hyperion EO-1. The use of these satellite platforms power up analyses from the global to the ecosystem scale. Nevertheless, these scales might not be sufficient according to the problem-specific objective of an Earth system study (e.g., sub-pixel sampling, recurring time-sampling, higher spectral requirements). As an example, when looking at plots with a limited spatial extent (e.g., $<100\text{ m}^2$, which is the mixed area for each Sentinel-2 pixel), using airborne or drone imagery is one of the most useful approaches to take. While the spatial resolution can be significantly enhanced by using these platforms (e.g., from meters to sub-centimetres), the spectral resolution may decrease as commercial RGB cameras are used in most cases, leaving most of the spectral indices out of reach according to the platform specifications. However, spectral indices in the visible spectrum can still be computed for drone missions with RGB cameras since the ASI catalogue includes several spectral indices of this kind. The subset of spectral indices in the visible spectrum can be easily retrieved as $\Psi_{\text{VIS}} \subset \Psi$ such that $V_{\psi} \subseteq V_{\text{VIS}}$, where $V_{\text{VIS}} = \{\langle\langle B \rangle\rangle, \langle\langle G \rangle\rangle, \langle\langle R \rangle\rangle\}$, e.g., Excess Green Index (ExG)¹⁴⁹, Visible Atmospherically Resistant Index (VARI)⁷⁷, Green Leaf Index (GLI)¹⁰⁶, etc.

Computing hyperspectral and narrow-band indices. As a matter of simplicity, the ASI catalogue currently consists of broad-band spectral indices. Some narrow-band spectral indices as well as hyperspectral indices were coerced to broad-band indices if the specific wavelength was close to a common spectral band in popular multispectral sensors (Table 2). This rules out the need of the actual reflectance in the specific wavelength, but leaves a portion of narrow-band and hyperspectral indices out of the catalogue. Some hyperspectral indices have different variants that use slightly different wavelengths in their formulas and become completely omitted. One example is the Simple Ratio (SR), which has multiple variants that can be addressed one by one when exploiting the Red Edge reflectance of vegetation^{55,95,184,185}. This also applies to complete application domains, as is the case of geology indices, which generally use several SWIR narrow-bands of ASTER for lithologic mapping¹⁸⁶. However, users can still use data from specific wavelengths by providing them in P_{ψ} when computing a spectral index that can be used as a hyperspectral index, such as the NDVI or the Ratio Vegetation Index (RVI). This is extensible to coerced hyperspectral indices in the catalogue, such as the Structure Insensitive Pigment Index (SIPI)¹¹⁸ or the Transformed Chlorophyll Absorption in Reflectance Index (TCARI)⁸³, where users can pass the actual data from specific wavelengths as parameters for the index computation.

Feeding machine learning pipelines. As data driven techniques are gaining strength in Earth system research, the ASI catalogue can power up ML models. For instance, classical ML approaches are regularly used for estimating bio-physical as well as bio-chemical variables from local to global scales^{183,187–189}. This includes linear regression models as well as non-linear approaches such as ensemble methods (e.g. Random Forest, Gradient Boosting, etc.), kernel methods (e.g. Support Vector Regression, SVR, Gaussian Processes, GP, etc.)^{190,191}, and probabilistic methods (e.g. Naive Bayes). As Python is highly used for ML (e.g., using packages such as `scikit-learn`¹⁹²; `gpy`, see <http://sheffieldml.github.io/GPy>; `XGBoost`, see <https://xgboost.ai/>; etc.), the `spyndex` Python library fits in the pipeline by supporting several Python objects that are often used for these methods (e.g., arrays and data frames). This means that spectral indices can be directly calculated over the dataset of interest and included as spectral features that can be even further pre-processed (e.g., computing textural features from spectral indices in 2-dimensional arrays) before being fed into a ML model. This principle applies for other pre-processing approaches in feature engineering such as dimensionality reduction. Since several spectral indices can be computed for the same input bands, a new high-dimensional feature space can be derived. As multiple indices can be highly correlated, this feature space can be reduced to a latent one, either linearly or non-linearly.

Since multiple data driven approaches also include DL techniques, ASI can also be included in DL pipelines inside Python. As spectral indices can be computed for n -dimensional arrays, multiple types of architectures can be used for modelling a certain variable of interest or classifying land covers. For example, a fully connected neural network can be fed with tensors that contains several spectral indices of interest for Land Use Land Change (LULC) pixel-based mapping or GPP modelling^{188,189}. However, an enhanced potential of DL can be achieved by considering the spatial and temporal dimensions when we talk about remote sensing. The spatial domain is covered by using Convolutional Neural Networks (CNN) that can be fed with tensors from a 3-dimensional array. We can also grip time with Recurrent Neural Networks (RNN) architectures by adding a temporal dimension¹⁹³. Full potential can be described by taking the spatiotemporal dimension into consideration with 4-dimensional arrays¹⁹⁴. This can be effectively achieved by using different frameworks in Python, e.g., `tensorflow`¹⁹⁵ and `pytorch`¹⁹⁶, and exploiting the potential of ASI for handling spectral indices with n -dimensional data cubes.

Using ASI outside python. The Python library is designed to manage multiple Python objects, which makes it a great tool for spectral indices computation on data frames with `pandas`, or data cubes by using labelled data arrays with `xarray`. This is extremely important for Earth system research as many variables are usually distributed along a regular spatiotemporal grid. Remote sensing data structures can then be easily transformed to fit these grids and spectral indices can be computed directly. In practice, this can be done in multiple programming languages and platforms, so their specific features as well as third-party libraries can be exploited. Given this, the ASI catalogue was already used to create a JavaScript module for the GEE Code Editor¹⁷⁴, which is extensively used in Earth system research. Nevertheless, two additionally and particularly relevant programming languages to use are R and Julia.

As R is a programming language with a big geospatial user-base, multiple packages for remote sensing have been created¹⁹⁷. This is the case of the `sp` package (see <https://github.com/edzer/sp>), where objects like the `SpatialPixelsDataFrame` can be used for storing multispectral data, permitting the computation of spectral indices. More specific packages such as `raster` (see <https://github.com/rspatial/raster>) provides users with the `RasterLayer` and `RasterStack` objects for a fully remote sensing focus. As it was mentioned before, there are already multiple R packages for computing spectral indices for these kind of objects. However, the whole potential for Earth system research can be achieved by using the ASI catalogue in combination with the `stars` package (see <https://github.com/r-spatial/stars>), specialised for spatiotemporal arrays (i.e., data cubes). On the other hand, primitive objects such as numeric types as well as vectors, arrays, and data frames can still be used for a comprehensive use of spectral indices in different objects inside R.

Julia is a programming language for high performance scientific computing, created with the goals of composability and speed of execution¹⁹⁸. The latter aspect specifically makes the language suitable for large Earth system applications, and this is shown by the high number of related packages available such as `EarthDataLab.jl`², a package developed for ESDCs, and `GeoStats.jl`¹⁹⁹, among others. Thanks to the multiple dispatch nature of Julia, primitive objects, such as arrays, can be extended, serving as input for spatiotemporal array objects for spectral indices computation. Examples include the `Zarr.jl` package (see <https://github.com/JuliaIO/Zarr.jl>), which provides an implementation of chunked n -dimensional arrays, or the `YAXArrays.jl` package (see <https://github.com/JuliaDataCubes/YAXArrays.jl>), which can operate labeled data cubes from multiple data sources such as `netCDF` and `zarr` files. The thriving ML ecosystem in Julia has been growing and it now includes both production level software such as `Flux.jl`^{200,201} as well as state of the art libraries for scientific ML^{202–204}. This allows a seamless integration of these models with specific focus on Earth system research by using spectral indices from ASI, making the exploration of different modelling techniques quick and efficient.

Code availability

The ASI Catalogue code is open-source and can be found at <https://github.com/awesome-spectral-indices/awesome-spectral-indices> and Zenodo²⁰⁵. The `spyndex` Python package is open-source and can be found at <https://github.com/awesome-spectral-indices/spyndex>. It is also available through PyPI (<https://pypi.org/project/spyndex/>) and conda-forge (<https://anaconda.org/conda-forge/spyndex>). The catalogue in CSV format can also be downloaded from the `Espectro` Streamlit web app, which is available at <https://share.streamlit.io/davemlz/espectro/main/espectro.py>. The `Espectro` Streamlit web app code is open-source and can be found at <https://github.com/awesome-spectral-indices/espectro>.

Received: 13 December 2022; Accepted: 21 March 2023;

Published online: 08 April 2023

References

- Liang, S. & Wang, J. A systematic view of remote sensing. In *Advanced Remote Sensing*, chap. 1, 1–57, <https://doi.org/10.1016/b978-0-12-815826-5.00001-5>, second edn (Elsevier, 2020).
- Mahecha, M. D. *et al.* Earth system data cubes unravel global multivariate dynamics. *Earth System Dynamics* **11**, 201–234, <https://doi.org/10.5194/ESD-11-201-2020> (2020).
- Crowley, M. A. & Cardille, J. A. Remote Sensing's Recent and Future Contributions to Landscape Ecology. *Current Landscape Ecology Reports* **5**, 45–57, <https://doi.org/10.1007/s40823-020-00054-9> (2020).
- Emmanuel Johnson, J., Laparra, V., Piles, M. & Camps-Valls, G. Gaussianizing the Earth: Multidimensional information measures for Earth data analysis. *IEEE Geoscience and Remote Sensing Magazine* **9**, 191–208, <https://doi.org/10.1109/MGRS.2021.3066260> (2021).
- Zhao, Y. & Zhu, Z. ASI: An artificial surface Index for Landsat 8 imagery. *International Journal of Applied Earth Observation and Geoinformation* **107**, 102703, <https://doi.org/10.1016/j.jag.2022.102703> (2022).
- Xiao, J. *et al.* Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment* **233**, 111383, <https://doi.org/10.1016/j.rse.2019.111383> (2019).
- Wulder, M. A. *et al.* The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment* **185**, 271–283, <https://doi.org/10.1016/j.rse.2015.11.032> (2016).
- Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **202**, 18–27, <https://doi.org/10.1016/j.rse.2017.06.031> (2017).
- Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* **8**, 127–150, [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0) (1979).
- Zha, Y., Gao, J. & Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* **24**, 583–594, <https://doi.org/10.1080/01431160304987> (2003).
- Smith, A. M. *et al.* Testing the potential of multi-spectral remote sensing for retrospectively estimating fire severity in African Savannas. *Remote Sensing of Environment* **97**, 92–115, <https://doi.org/10.1016/j.rse.2005.04.014> (2005).
- Smith, A. M. *et al.* Production of Landsat ETM+ reference imagery of burned areas within Southern African savannas: comparison of methods and application to MODIS. *International Journal of Remote Sensing* **28**, 2753–2775, <https://doi.org/10.1080/01431160600954704> (2007).

13. Rouse, J. W., Haas, R. H., Schell, J. A. & Deering, D. W. Monitoring vegetation systems in the Great Plains with ERTS. Tech. Rep., NASA (1974).
14. Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **25**, 295–309, [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X) (1988).
15. Riggs, G. A., Hall, D. K. & Salomonson, V. V. Snow index for the Landsat Thematic Mapper and moderate resolution imaging spectroradiometer. *International Geoscience and Remote Sensing Symposium (IGARSS)* **4**, 1942–1944, <https://doi.org/10.1109/IGARSS.1994.399618> (1994).
16. McFeeters, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* **17**, 1425–1432, <https://doi.org/10.1080/01431169608948714> (1996).
17. Zeng, Y. *et al.* Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nature Reviews Earth and Environment* <https://doi.org/10.1038/s43017-022-00298-5> (2022).
18. Xue, J. & Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors* **2017**, <https://doi.org/10.1155/2017/1353691> (2017).
19. Yang, P., van der Tol, C., Campbell, P. K. & Middleton, E. M. Fluorescence Correction Vegetation Index (FCVI): A physically based reflectance index to separate physiological and non-physiological information in far-red sun-induced chlorophyll fluorescence. *Remote Sensing of Environment* **240**, 111676, <https://doi.org/10.1016/j.rse.2020.111676> (2020).
20. Zeng, Y. *et al.* Estimating near-infrared reflectance of vegetation from hyperspectral data. *Remote Sensing of Environment* **267**, <https://doi.org/10.1016/j.rse.2021.112723> (2021).
21. Badgley, G., Field, C. B. & Berry, J. A. Canopy near-infrared reflectance and terrestrial photosynthesis. *Science Advances* **3**, https://doi.org/10.1126/SCIADV.1602244/SUPPL_FILE/1602244_SM.PDF (2017).
22. Camps-Valls, G. *et al.* A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances* **7**, 7447–7473, https://doi.org/10.1126/SCIADV.ABC7447/SUPPL_FILE/ABC7447_SM.PDF (2021).
23. Dash, J. & Curran, P. J. The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing* **25**, 5403–5413, <https://doi.org/10.1080/0143116042000274015> (2004).
24. Frampton, W. J., Dash, J., Watmough, G. & Milton, E. J. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing* **82**, 83–92, <https://doi.org/10.1016/j.isprsjprs.2013.04.007> (2013).
25. Feyisa, G. L., Meilby, H., Fensholt, R. & Proud, S. R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment* **140**, 23–35, <https://doi.org/10.1016/j.rse.2013.08.029> (2014).
26. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing* **27**, 3025–3033, <https://doi.org/10.1080/01431160600589179> (2006).
27. Wang, X. *et al.* A robust Multi-Band Water Index (MBWI) for automated extraction of surface water from Landsat 8 OLI imagery. *International Journal of Applied Earth Observation and Geoinformation* **68**, 73–91, <https://doi.org/10.1016/j.jag.2018.01.018> (2018).
28. Jiang, W. *et al.* An Effective Water Body Extraction Method with New Water Index for Sentinel-2 Imagery. *Water* **13**, 1647, <https://doi.org/10.3390/W13121647> (2021).
29. Martín, Md. P. & Chuvieco, E. Cartografía de grandes incendios forestales en la península ibérica a partir de imágenes NOAA-AVHRR. *Serie Geográfica* **7**, 109–128 (1998).
30. Filipponi, F. BAIS2: Burned Area Index for Sentinel-2. *Proceedings* **2**, 364, <https://doi.org/10.3390/ECRS-2-05177> (2018).
31. Dixit, A., Goswami, A. & Jain, S. Development and Evaluation of a New “Snow Water Index (SWI)” for Accurate Snow Cover Delineation. *Remote Sensing* **11**, 2774, <https://doi.org/10.3390/RS11232774> (2019).
32. Kawamura, M., Jayamanna, S. & Tsujiko, Y. Relation Between Social and Environmental Conditions in Colombo, Sri Lanka and the Urban Index Estimated by Satellite Remote Sensing Data. In *XVIIIth ISPRS Congress*, 321–326 (1996).
33. Rasul, A. *et al.* Applying Built-Up and Bare-Soil Indices from Landsat 8 to Cities in Dry Climates. *Land* **7**, 81, <https://doi.org/10.3390/LAND7030081> (2018).
34. Lin, H., Wang, J., Liu, S., Qu, Y. & Wan, H. Studies on urban areas extraction from Landsat TM images. *International Geoscience and Remote Sensing Symposium (IGARSS)* **6**, 3826–3829, <https://doi.org/10.1109/IGARSS.2005.1525743> (2005).
35. Nguyen, C. T., Chidthaisong, A., Diem, P. K. & Huo, L. Z. A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. *Land* **10**, 231, <https://doi.org/10.3390/LAND10030231> (2021).
36. Henrich, V. *et al.* Development of an online indices database: Motivation, concept and implementation. In *6th EARSeL Imaging Spectroscopy SIG Workshop Innovative Tool for Scientific and Commercial Environment Applications* (Tel Aviv, 2009).
37. Henrich, V., Krauss, G., Götze, C. & Sandow, C. IDB-www.indexdatabase.de Entwicklung einer Datenbank für Fernerkundungsindizes Ziele und Eigenschaften der IDB. Tech. Rep., AK Fernerkundung, Bochum (2012).
38. Grizonnet, M. *et al.* Orfeo ToolBox: open source processing of remote sensing images. *Open Geospatial Data, Software and Standards* **2**, 1–8, <https://doi.org/10.1186/S40965-017-0031-6> (2017).
39. Conrad, O. *et al.* System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development* **8**, 1991–2007, <https://doi.org/10.5194/GMD-8-1991-2015> (2015).
40. Hoyer, S. & Hamman, J. J. xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software* **5**, <https://doi.org/10.5334/JORS.148> (2017).
41. Gitelson, A. A. Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology* **161**, 165–173, <https://doi.org/10.1078/0176-1617-01176> (2004).
42. Alcaras, E., Costantino, D., Guastaferro, F., Parente, C. & Pepe, M. Normalized Burn Ratio Plus (NBR+): A New Index for Sentinel-2 Imagery. *Remote Sensing* **14**, 1727, <https://doi.org/10.3390/RS14071727> (2022).
43. Alvarez-Mozos, J., Villanueva, J., Arias, M. & Gonzalez-Audicana, M. Correlation Between NDVI and Sentinel-1 Derived Features for Maize. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 6773–6776, <https://doi.org/10.1109/IGARSS47720.2021.9554099> (Institute of Electrical and Electronics Engineers (IEEE), 2021).
44. Apan, A., Held, A., Phinn, S. & Markley, J. Detecting sugarcane ‘orange rust’ disease using EO-1 Hyperion hyperspectral imagery. *International Journal of Remote Sensing* **25**, 489–498, <https://doi.org/10.1080/01431160310001618031> (2004).
45. Arreola-Esquivel, M. *et al.* Non-Binary Snow Index for Multi-Component Surfaces. *Remote Sensing* **13**, 2777, <https://doi.org/10.3390/RS13142777> (2021).
46. As-syakur, A. R., Adnyana, I. W. S., Arthana, I. W. & Nuarsa, I. W. Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area. *Remote Sensing* **4**, 2957–2970, <https://doi.org/10.3390/RS4102957> (2012).
47. Bannari, A., Asalhi, H. & Teillet, P. M. Transformed difference vegetation index (TDVI) for vegetation cover mapping. *International Geoscience and Remote Sensing Symposium (IGARSS)* **5**, 3053–3055, <https://doi.org/10.1109/IGARSS.2002.1026867> (2002).
48. Baret, F., Guyot, G. & Major, D. J. TSAVI: A vegetation index which minimizes soil brightness effects on LAI and APAR estimation. *Digest - International Geoscience and Remote Sensing Symposium (IGARSS)* **3**, 1355–1358, <https://doi.org/10.1109/IGARSS.1989.576128> (1989).
49. Baret, F. & Guyot, G. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment* **35**, 161–173, [https://doi.org/10.1016/0034-4257\(91\)90009-U](https://doi.org/10.1016/0034-4257(91)90009-U) (1991).

50. Bendig, J. *et al.* Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *International Journal of Applied Earth Observation and Geoinformation* **39**, 79–87, <https://doi.org/10.1016/J.JAG.2015.02.012> (2015).
51. Birth, G. S. & McVey, G. R. Measuring the Color of Growing Turf with a Reflectance Spectrophotometer. *Agronomy Journal* **60**, 640–643, <https://doi.org/10.2134/AGRONJ1968.00021962006000060016X> (1968).
52. Blanco, V. *et al.* Potential of UAS-Based Remote Sensing for Estimating Tree Water Status and Yield in Sweet Cherry Trees. *Remote Sensing* **12**, 2359, <https://doi.org/10.3390/RS12152359> (2020).
53. Bouhennache, R., Bouden, T., Taleb-Ahmed, A. & Cheddad, A. A new spectral index for the extraction of built-up land features from Landsat 8 satellite imagery. *Geocarto International* **34**, 1531–1551, <https://doi.org/10.1080/10106049.2018.1497094> (2018).
54. Broge, N. H. & Leblanc, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment* **76**, 156–172, [https://doi.org/10.1016/S0034-4257\(00\)00197-8](https://doi.org/10.1016/S0034-4257(00)00197-8) (2001).
55. Buschmann, C. & Nagel, E. *In vivo* spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing* **14**, 711–722, <https://doi.org/10.1080/01431169308904370> (1993).
56. Cao, Y.-G., Li-Juan, Y. & Zheng, Z.-Z. Extraction of Information on Geology Hazard from Multi-Polarization SAR Images. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1529–1532 (2008).
57. Ceccato, P., Flasse, S. & Grégoire, J. M. Designing a spectral index to estimate vegetation water content from remote sensing data: Part I: Theoretical approach. *Remote Sensing of Environment* **82**, 188–197, [https://doi.org/10.1016/S0034-4257\(02\)00037-8](https://doi.org/10.1016/S0034-4257(02)00037-8) (2002).
58. Chen, J. M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Canadian Journal of Remote Sensing* **22**, 229–242, <https://doi.org/10.1080/07038992.1996.10855178> (1996).
59. Clevers, J. G. Application of a weighted infrared-red vegetation index for estimating leaf Area Index by Correcting for Soil Moisture. *Remote Sensing of Environment* **29**, 25–37, [https://doi.org/10.1016/0034-4257\(89\)90076-X](https://doi.org/10.1016/0034-4257(89)90076-X) (1989).
60. Coffelt, J. L. & Livingston, R. K. Second U.S. Geological Survey Wildland Fire Workshop: Los Alamos, New Mexico, October 31–November 3, 2000. Tech. Rep., USGS, <https://doi.org/10.3133/OFR0211> (2002).
61. Crippen, R. E. Calculating the vegetation index faster. *Remote Sensing of Environment* **34**, 71–73, [https://doi.org/10.1016/0034-4257\(90\)90085-Z](https://doi.org/10.1016/0034-4257(90)90085-Z) (1990).
62. Datt, B. Remote Sensing of Chlorophyll a, Chlorophyll b, Chlorophyll a+b, and Total Carotenoid Content in Eucalyptus Leaves. *Remote Sensing of Environment* **66**, 111–121, [https://doi.org/10.1016/S0034-4257\(98\)00046-7](https://doi.org/10.1016/S0034-4257(98)00046-7) (1998).
63. Daughtry, C. S., Walthall, C. L., Kim, M. S., De Colstoun, E. B. & McMurtrey, J. E. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sensing of Environment* **74**, 229–239, [https://doi.org/10.1016/S0034-4257\(00\)00113-9](https://doi.org/10.1016/S0034-4257(00)00113-9) (2000).
64. Dechant, B. *et al.* NIRVP: A robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales. *Remote Sensing of Environment* **268**, 112763, <https://doi.org/10.1016/J.RSE.2021.112763> (2022).
65. Deng, Y., Wu, C., Li, M. & Chen, R. RNDISI: A ratio normalized difference soil index for remote sensing of urban/suburban environments. *International Journal of Applied Earth Observation and Geoinformation* **39**, 40–48, <https://doi.org/10.1016/J.JAG.2015.02.010> (2015).
66. Escadafal, R. & Huete, A. Etude des propriétés spectrales des sols arides appliquée à l'amélioration des indices de végétation obtenus par télédétection. *Comptes Rendus de l'Académie des Sciences* **132**, 1385–1391 (1991).
67. Estoque, R. C. & Murayama, Y. Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imagery: A comparative assessment of various spectral indices. *Ecological Indicators* **56**, 205–217, <https://doi.org/10.1016/J.ECOLIND.2015.03.037> (2015).
68. Feng, D. A New Method for Fast Information Extraction of Water Bodies Using Remotely Sensed Data. *Remote Sensing Technology and Application* **24**, 167–171, <https://doi.org/10.11873/J.ISSN.1004-0323.2009.2.167> (2009).
69. Fisher, A., Flood, N. & Danaher, T. Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sensing of Environment* **175**, 167–182, <https://doi.org/10.1016/J.RSE.2015.12.055> (2016).
70. Gerard, F. *et al.* Forest Fire Scar Detection in the Boreal Forest with Multitemporal SPOT-VEGETATION Data. *IEEE Transactions on Geoscience and Remote Sensing* **41**, 2575–2585, <https://doi.org/10.1109/TGRS.2003.819190> (2003).
71. Gillespie, A. R., Kahle, A. B. & Walker, R. E. Color enhancement of highly correlated images. II. Channel ratio and “chromaticity” transformation techniques. *Remote Sensing of Environment* **22**, 343–365, [https://doi.org/10.1016/0034-4257\(87\)90088-5](https://doi.org/10.1016/0034-4257(87)90088-5) (1987).
72. Gitelson, A. & Merzlyak, M. N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology* **22**, 247–252, [https://doi.org/10.1016/1011-1344\(93\)06963-4](https://doi.org/10.1016/1011-1344(93)06963-4) (1994).
73. Gitelson, A. & Merzlyak, M. N. Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology* **143**, 286–292, [https://doi.org/10.1016/S0176-1617\(11\)81633-0](https://doi.org/10.1016/S0176-1617(11)81633-0) (1994).
74. Gitelson, A. A., Kaufman, Y. J. & Merzlyak, M. N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment* **58**, 289–298, [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7) (1996).
75. Gitelson, A. A. & Merzlyak, M. N. Signature Analysis of Leaf Reflectance Spectra: Algorithm Development for Remote Sensing of Chlorophyll. *Journal of Plant Physiology* **148**, 494–500, [https://doi.org/10.1016/S0176-1617\(96\)80284-7](https://doi.org/10.1016/S0176-1617(96)80284-7) (1996).
76. Gitelson, A. A., Merzlyak, M. N., Chivkunova, O. & Olga, B. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochemistry and Photobiology* **74**, 38–45, <https://doi.org/10.1562/0031> (2001).
77. Gitelson, A. A., Kaufman, Y. J., Stark, R. & Rundquist, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment* **80**, 76–87, [https://doi.org/10.1016/S0034-4257\(01\)00289-9](https://doi.org/10.1016/S0034-4257(01)00289-9) (2002).
78. Gitelson, A. A., Gritz, Y. & Merzlyak, M. N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology* **160**, 271–282, <https://doi.org/10.1078/0176-1617-00887> (2003).
79. Goel, N. S. & Qin, W. Influences of canopy architecture on relationships between various vegetation indices and LAI and Fpar: A computer simulation. *Remote Sensing Reviews* **10**, 309–347, <https://doi.org/10.1080/02757259409532252> (1994).
80. Gong, P., Pu, R., Biging, G. S. & Larrieu, M. R. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* **41**, 1355–1362, <https://doi.org/10.1109/TGRS.2003.812910> (2003).
81. Gu, Y. *et al.* A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. *Geophysical Research Letters* **34**, <https://doi.org/10.1029/2006GL029127> (2007).
82. Guo, Y. *et al.* Modified Red Blue Vegetation Index for Chlorophyll Estimation and Yield Prediction of Maize from Visible Images Captured by UAV. *Sensors* **20**, 5055, <https://doi.org/10.3390/S20185055> (2020).
83. Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J. & Dextra, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment* **81**, 416–426, [https://doi.org/10.1016/S0034-4257\(02\)00018-4](https://doi.org/10.1016/S0034-4257(02)00018-4) (2002).
84. Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J. & Strachan, I. B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* **90**, 337–352, <https://doi.org/10.1016/J.RSE.2003.12.013> (2004).

85. Haboudane, D., Tremblay, N., Miller, J. R. & Vigneault, P. Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* **46**, 423–436, <https://doi.org/10.1109/TGRS.2007.904836> (2008).
86. Han, J., Zhang, Z. & Cao, J. Developing a New Method to Identify Flowering Dynamics of Rapeseed Using Landsat 8 and Sentinel-1/2. *Remote Sensing* **13**, 105, <https://doi.org/10.3390/RS13010105> (2020).
87. Hancock, D. W. & Dougherty, C. T. Relationships between Blue- and Red-based Vegetation Indices and Leaf Area and Yield of Alfalfa. *Crop Science* **47**, 2547–2556, <https://doi.org/10.2135/CROPSCI2007.01.0031> (2007).
88. Hardisky, M. A., Klemas, V. & Smart, R. M. The Influence of Soil Salinity, Growth Form, and Leaf Moisture on-the Spectral Radiance of *Spartina alterniflora* Canopies. *Photogrammetric Engineering and Remote Sensing* **49**, 77–83 (1983).
89. Holden, Z. A., Smith, A. M., Morgan, P., Rollins, M. G. & Gessler, P. E. Evaluation of novel thermally enhanced spectral indices for mapping fire perimeters and comparisons with fire atlas data. *International Journal of Remote Sensing* **26**, 4801–4808, <https://doi.org/10.1080/01431160500239008> (2005).
90. Huete, A. R., Liu, H. Q., Batchily, K. & Van Leeuwen, W. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment* **59**, 440–451, [https://doi.org/10.1016/S0034-4257\(96\)00112-5](https://doi.org/10.1016/S0034-4257(96)00112-5) (1997).
91. Hunt, E. R. & Rock, B. N. Detection of changes in leaf water content using Near- and Middle-Infrared reflectances. *Remote Sensing of Environment* **30**, 43–54, [https://doi.org/10.1016/0034-4257\(89\)90046-1](https://doi.org/10.1016/0034-4257(89)90046-1) (1989).
92. Hunt, E. R. *et al.* A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International Journal of Applied Earth Observation and Geoinformation* **21**, 103–112, <https://doi.org/10.1016/J.JAG.2012.07.020> (2013).
93. Jiang, Z., Huete, A. R., Didan, K. & Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment* **112**, 3833–3845, <https://doi.org/10.1016/J.RSE.2008.06.006> (2008).
94. Jiang, H. *et al.* A shadow-eliminated vegetation index (SEVI) for removal of self and cast shadow effects on vegetation in rugged terrains. *International Journal of Digital Earth* **12**, 1013–1029, <https://doi.org/10.1080/17538947.2018.1495770> (2018).
95. Jordan, C. F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **50**, 663–666, <https://doi.org/10.2307/1936256> (1969).
96. Jurgens, C. The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data. *International Journal of Remote Sensing* **18**, 3583–3594, <https://doi.org/10.1080/014311697216810> (1997).
97. Karnieli, A., Kaufman, Y. J., Remer, L. & Wald, A. AFRI—aerosol free vegetation index. *Remote Sensing of Environment* **77**, 10–21, [https://doi.org/10.1016/S0034-4257\(01\)00190-0](https://doi.org/10.1016/S0034-4257(01)00190-0) (2001).
98. Kaufman, Y. J. & Tanré, D. Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing* **30**, 261–270, <https://doi.org/10.1109/36.134076> (1992).
99. Kawashima, S. & Nakatani, M. An Algorithm for Estimating Chlorophyll Content in Leaves Using a Video Camera. *Annals of Botany* **81**, 49–54, <https://doi.org/10.1006/ANBO.1997.0544> (1998).
100. Keshri, A. K., Shukla, A. & Gupta, R. P. ASTER ratio indices for supraglacial terrain mapping. *International Journal of Remote Sensing* **30**, 519–524, <https://doi.org/10.1080/01431160802385459> (2008).
101. Kim, Y. & Van Zyl, J. Comparison of forest parameter estimation techniques using SAR data. *International Geoscience and Remote Sensing Symposium (IGARSS)* **3**, 1395–1397, <https://doi.org/10.1109/IGARSS.2001.976856> (2001).
102. Kwak, Y. *et al.* Prompt Proxy Mapping of Flood Damaged Rice Fields Using MODIS-Derived Indices. *Remote Sensing* **7**, 15969–15988, <https://doi.org/10.3390/RS71215805> (2015).
103. Lacaux, J. P., Tourre, Y. M., Vignolles, C., Ndione, J. A. & Lafaye, M. Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote Sensing of Environment* **106**, 66–74, <https://doi.org/10.1016/J.RSE.2006.07.012> (2007).
104. Li, H. *et al.* Mapping Urban Bare Land Automatically from Landsat Imagery with a Simple Index. *Remote Sensing* **9**, 249, <https://doi.org/10.3390/RS9030249> (2017).
105. Liu, S., Zheng, Y., Dalponte, M. & Tong, X. A novel fire index-based burned area change detection approach using Landsat-8 OLI data. *European Journal of Remote Sensing* **53**, 104–112, <https://doi.org/10.1080/22797254.2020.1738900> (2020).
106. Louhaichi, M., Borman, M. M. & Johnson, D. E. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. *Geocarto International* **16**, 65–70, <https://doi.org/10.1080/1010640108542184> (2001).
107. Major, D. J., Baret, F. & Guyot, G. A ratio vegetation index adjusted for soil brightness. *International Journal of Remote Sensing* **11**, 727–740, <https://doi.org/10.1080/01431169008955053> (1990).
108. Martín, M. P., Gómez, I. & Chuvieco, E. Burnt Area Index (BAIM) for burned area discrimination at regional scale using MODIS data. *Forest Ecology and Management* **234**, S221, <https://doi.org/10.1016/J.FORECO.2006.08.248> (2006).
109. Mathieu, R., Pouget, M., Cervelle, B. & Escadafal, R. Relationships between Satellite-Based Radiometric Indices Simulated Using Laboratory Reflectance Data and Typic Soil Color of an Arid Environment. *Remote Sensing of Environment* **66**, 17–28, [https://doi.org/10.1016/S0034-4257\(98\)00030-3](https://doi.org/10.1016/S0034-4257(98)00030-3) (1998).
110. Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B. & Rakitin, V. Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum* **106**, 135–141, <https://doi.org/10.1034/J.1399-3054.1999.106119.X> (1999).
111. Meyer, G. E., Hindman, T. W. & Laksmi, K. Machine vision detection parameters for plant species identification. In *Proc. SPIE 3543, Precision Agriculture and Biological Quality*, vol. 3543, 327–335, <https://doi.org/10.1117/12.336896> (SPIE, 1999).
112. Meyer, G. E. & Neto, J. C. Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture* **63**, 282–293, <https://doi.org/10.1016/J.COMPAG.2008.03.009> (2008).
113. Milczarek, M., Robak, A. & Gadawska, A. Sentinel Water Mask (SWM) - New Index for Water Detection On Sentinel-2 Images. Tech. Rep., Crisis Information Centre at Space Research Centre, Polish Academy of Sciences (2017).
114. Mishra, S. & Mishra, D. R. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sensing of Environment* **117**, 394–406, <https://doi.org/10.1016/J.RSE.2011.10.016> (2012).
115. Mitchard, E. T. *et al.* Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: Overcoming problems of high biomass and persistent cloud. *Biogeosciences* **9**, 179–191, <https://doi.org/10.5194/BG-9-179-2012> (2012).
116. Nasirzadehdizaji, R. *et al.* Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage. *Applied Sciences* **9**, 655, <https://doi.org/10.3390/APP9040655> (2019).
117. Pasqualotto, N., Delegido, J., Van Wittenberghe, S., Rinaldi, M. & Moreno, J. Multi-Crop Green LAI Estimation with a New Simple Sentinel-2 LAI Index (SeLI). *Sensors* **19**, 904, <https://doi.org/10.3390/S19040904> (2019).
118. Peñuelas, J., Baret, F. & Filella, I. Semi-empirical indices to assess carotenoids/chlorophyll alpha ratio from leaf spectral reflectance. *Photosynthetica* **31**, 221–230 (1995).
119. Periasamy, S. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1. *Remote Sensing of Environment* **217**, 537–549, <https://doi.org/10.1016/J.RSE.2018.09.003> (2018).
120. Pinder, J. E. III & McLeod, K. W. Indications of Relative Drought Stress in Longleaf Pine from Thematic Mapper Data. *Photogrammetric Engineering and Remote Sensing* **65**, 495–501 (1999).

121. Pinty, B. & Verstraete, M. M. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetatio* **101**, 15–20, <https://doi.org/10.1007/BF00031911> (1992).
122. Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H. & Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sensing of Environment* **48**, 119–126, [https://doi.org/10.1016/0034-4257\(94\)90134-1](https://doi.org/10.1016/0034-4257(94)90134-1) (1994).
123. Rad, A. M., Kreitler, J. & Sadeh, M. Augmented Normalized Difference Water Index for improved surface water monitoring. *Environmental Modelling and Software* **140**, 105030, <https://doi.org/10.1016/j.envsoft.2021.105030> (2021).
124. Rikimaru, A., Roy, P. S. & Miyatake, S. Tropical forest cover density mapping. *Tropical Ecology* **43**, 39–47 (2002).
125. Rondeaux, G., Steven, M. & Baret, F. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment* **55**, 95–107, [https://doi.org/10.1016/0034-4257\(95\)00186-7](https://doi.org/10.1016/0034-4257(95)00186-7) (1996).
126. Roujean, J. L. & Breon, F. M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment* **51**, 375–384, [https://doi.org/10.1016/0034-4257\(94\)00114-3](https://doi.org/10.1016/0034-4257(94)00114-3) (1995).
127. Saberioon, M. M. *et al.* Assessment of rice leaf chlorophyll content using visible bands at different growth stages at both the leaf and canopy scale. *International Journal of Applied Earth Observation and Geoinformation* **32**, 35–45, <https://doi.org/10.1016/j.jag.2014.03.018> (2014).
128. Saito, A. & Yamazaki, E. Characteristics of Spectral Reflectance for Vegetation Ground Surfaces with Snow-cover; Vegetation Indices and Snow Indices. *Journal of Japan Society of Hydrology and Water Resources* **12**, 28–38, <https://doi.org/10.3178/JJSHWR.12.28> (1999).
129. Shen, L. & Li, C. Water body extraction from Landsat ETM+ imagery using adaboost algorithm. In *18th International Conference on Geoinformatics, Geoinformatics 2010*, 1–4, <https://doi.org/10.1109/GEOINFORMATICS.2010.5567762> (2010).
130. Sims, D. A. & Gamon, J. A. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment* **81**, 337–354, [https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/10.1016/S0034-4257(02)00010-X) (2002).
131. Sinha, P., Verma, N. K. & Ayele, E. Urban Built-up Area Extraction and Change Detection of Adama Municipal Area using Time-Series Landsat Images. *International Journal of Advanced Remote Sensing and GIS* **5**, 1886–1895, <https://doi.org/10.23953/CLOUD.IJARSG.67> (2016).
132. Sripada, R. P., Heiniger, R. W., White, J. G. & Weisz, R. Aerial Color Infrared Photography for Determining Late-Season Nitrogen Requirements in Corn. *Agronomy Journal* **97**, 1443–1451, <https://doi.org/10.2134/AGRONJ2004.0314> (2005).
133. Stathakis, D., Perakis, K. & Savin, I. Efficient segmentation of urban areas by the VIBI. *International Journal of Remote Sensing* **33**, 6361–6377, <https://doi.org/10.1080/01431161.2012.687842> (2012).
134. Sulik, J. J. & Long, D. S. Spectral considerations for modeling yield of canola. *Remote Sensing of Environment* **184**, 161–174, <https://doi.org/10.1016/j.rse.2016.06.016> (2016).
135. Tian, Y., Chen, H., Song, Q. & Zheng, K. A Novel Index for Impervious Surface Area Mapping: Development and Validation. *Remote Sensing* **10**, 1521, <https://doi.org/10.3390/RS10101521> (2018).
136. Trigg, S. & Flasse, S. An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. *International Journal of Remote Sensing* **22**, 2641–2647, <https://doi.org/10.1080/01431160110053185> (2001).
137. Trudel, M., Charbonneau, F. & Leconte, R. Using RADARSAT-2 polarimetric and ENVISAT-ASAR dual-polarization data for estimating soil moisture over agricultural fields. *Canadian Journal of Remote Sensing* **38**, 514–527, <https://doi.org/10.5589/M12-043> (2012).
138. USGS. Landsat Normalized Burn Ratio 2 | U.S. Geological Survey (2022).
139. Veraverbeke, S., Harris, S. & Hook, S. Evaluating spectral indices for burned area discrimination using MODIS/ASTER (MASTER) airborne simulator data. *Remote Sensing of Environment* **115**, 2702–2709, <https://doi.org/10.1016/j.rse.2011.06.010> (2011).
140. Viaña-Borja, S. P. & Ortega-Sánchez, M. Automatic Methodology to Detect the Coastline from Landsat Images with a New Water Index Assessed on Three Different Spanish Mediterranean Deltas. *Remote Sensing* **11**, 2186, <https://doi.org/10.3390/RS11182186> (2019).
141. Vincini, M., Frazzi, E. & D'Alessio, P. A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precision Agriculture* **9**, 303–319, <https://doi.org/10.1007/S11119-008-9075-Z/> (2008).
142. Vincini, M. & Frazzi, E. Comparing narrow and broad-band vegetation indices to estimate leaf chlorophyll content in planophile crop canopies. *Precision Agriculture* **12**, 334–344, <https://doi.org/10.1007/S11119-010-9204-3/> (2011).
143. Wang, F.-m, Huang, J.-f, Tang, Y.-l & Wang, X.-z New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. *Rice Science* **14**, 195–203, [https://doi.org/10.1016/S1672-6308\(07\)60027-4](https://doi.org/10.1016/S1672-6308(07)60027-4) (2007).
144. Wang, L. & Qu, J. J. NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophysical Research Letters* **34**, <https://doi.org/10.1029/2007GL031021> (2007).
145. Wang, C. *et al.* A snow-free vegetation index for improved monitoring of vegetation spring green-up date in deciduous ecosystems. *Remote Sensing of Environment* **196**, 1–12, <https://doi.org/10.1016/j.rse.2017.04.031> (2017).
146. Wang, Z., Liu, J., Li, J. & Zhang, D. D. Multi-Spectral Water Index (MuWI): A Native 10-m Multi-Spectral Water Index for Accurate Water Mapping on Sentinel-2. *Remote Sensing* **10**, 1643, <https://doi.org/10.3390/RS10101643> (2018).
147. Waqar, M. M., Mirza, J. F., Mumtaz, R. & Hussain, E. Development of New Indices for Extraction of Built-Up Area and Bare Soil from Landsat. *Data. Open Access Scientific Reports* **1**, 1–4, <https://doi.org/10.4172/scientificreports.136> (2012).
148. Wilson, E. H. & Sader, S. A. Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment* **80**, 385–396, [https://doi.org/10.1016/S0034-4257\(01\)00318-2](https://doi.org/10.1016/S0034-4257(01)00318-2) (2002).
149. Woebbecke, D. M., Meyer, G. E., Von Bargen, K. & Mortensen, D. A. Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions. *Transactions of the ASAE* **38**, 259–269, <https://doi.org/10.13031/2013.27838> (1995).
150. Wolf, A. Using WorldView 2 Vis-NIR MSI Imagery to Support Land Mapping and Feature Extraction Using Normalized Difference Index Ratios. *Tech. Rep., Digital Globe* (2010).
151. Wu, C., Niu, Z., Tang, Q. & Huang, W. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology* **148**, 1230–1241, <https://doi.org/10.1016/j.agrformet.2008.03.005> (2008).
152. Wu, W. The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization. *Remote Sensing* **6**, 1211–1233, <https://doi.org/10.3390/RS6021211> (2014).
153. Xiao, X., Shen, Z. & Qin, X. Assessing the potential of VEGETATION sensor data for mapping snow and ice cover: A Normalized Difference Snow and Ice Index. *International Journal of Remote Sensing* **22**, 2479–2487, <https://doi.org/10.1080/01431160110019766> (2001).
154. Xiao, X. *et al.* Satellite-based modeling of gross primary production in an evergreen needleleaf forest. *Remote Sensing of Environment* **89**, 519–534, <https://doi.org/10.1016/j.rse.2003.11.008> (2004).
155. Xing, N. *et al.* A Transformed Triangular Vegetation Index for Estimating Winter Wheat Leaf Area Index. *Remote Sensing* **12**, 16, <https://doi.org/10.3390/RS12010016> (2019).
156. Xu, H. A new index for delineating built-up land features in satellite imagery. *International Journal of Remote Sensing* **29**, 4269–4276, <https://doi.org/10.1080/01431160802039957> (2008).
157. Xu, H. Analysis of impervious surface and its impact on Urban heat environment using the normalized difference impervious surface index (NDISI). *Photogrammetric Engineering and Remote Sensing* **76**, 557–565, <https://doi.org/10.14358/PERS.76.5.557> (2010).
158. Yan, D., Huang, C., Ma, N. & Zhang, Y. Improved Landsat-Based Water and Snow Indices for Extracting Lake and Snow Cover/Glacier in the Tibetan Plateau. *Water* **12**, 1339, <https://doi.org/10.3390/W12051339> (2020).

159. Yang, W. *et al.* A semi-analytical snow-free vegetation index for improving estimation of plant phenology in tundra and grassland ecosystems. *Remote Sensing of Environment* **228**, 31–44, <https://doi.org/10.1016/j.rse.2019.03.028> (2019).
160. Yue, J., Tian, J., Tian, Q., Xu, K. & Xu, N. Development of soil moisture indices from differences in water absorption between shortwave-infrared bands. *ISPRS Journal of Photogrammetry and Remote Sensing* **154**, 216–230, <https://doi.org/10.1016/j.isprsjprs.2019.06.012> (2019).
161. Zhang, R., Rao, N. & Liao, K. Approach for a Vegetation Index Resistant to Atmospheric Effect. *Acta Botanica Sinica* **38**, 53–62 (1996).
162. Zhao, H. & Chen, X. Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+. *International Geoscience and Remote Sensing Symposium (IGARSS)* **3**, 1666–1668, <https://doi.org/10.1109/IGARSS.2005.1526319> (2005).
163. Zheng, Q., Huang, W., Cui, X., Shi, Y. & Liu, L. New Spectral Index for Detecting Wheat Yellow Rust Using Sentinel-2 Multispectral Imagery. *Sensors* **18**, 868, <https://doi.org/10.3390/S18030868> (2018).
164. Zhou, W., Li, Z., Ji, S., Hua, C. & Fan, W. A New Index Model NDVI-MNDWI for Water Object Extraction in Hybrid Area. *Communications in Computer and Information Science* **482**, 513–519, https://doi.org/10.1007/978-3-662-45737-5_51 (2015).
165. Niu, L. *et al.* Triangle Water Index (TWI): An Advanced Approach for More Accurate Detection and Delineation of Water Surfaces in Sentinel-2 Data. *Remote Sensing* **14**, 5289, <https://doi.org/10.3390/RS14215289> (2022).
166. Gamon, J. A. *et al.* A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 13087–13092, https://doi.org/10.1073/PNAS.1606162113/SUPPL_FILE/PNAS.201606162SI.PDF (2016).
167. Griffith, C. Box: Python dictionaries with advanced dot notation access (2022).
168. Montero, D. *et al.* Awesome Spectral Indices. *Zenodo* <https://doi.org/10.5281/zenodo.7424467> (2022).
169. Colvin, S. pydantic: Data parsing and validation using Python type hints (2021).
170. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362, <https://doi.org/10.1038/s41586-020-2649-2> (2020).
171. Reback, J. *et al.* pandas-dev/pandas: Pandas 1.4.1. *Zenodo* <https://doi.org/10.5281/ZENODO.6053272> (2022).
172. Jordahl, K. *et al.* geopandas/geopandas: v0.10.2. *Zenodo* <https://doi.org/10.5281/ZENODO.5573592> (2021).
173. Montero, D. eemont: A Python package that extends Google Earth Engine. *Journal of Open Source Software* **6**, 3168, <https://doi.org/10.21105/joss.03168> (2021).
174. Montero, D., Aybar, C., Mahecha, M. D. & Wieneke, S. spectral: Awesome Spectral Indices deployed via the Google Earth Engine JavaScript API. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-4-W*, 301–306, <https://doi.org/10.5194/ISPRS-ARCHIVES-XLVIII-4-W1-2022-301-2022> (2022).
175. Knohl, A., Schulze, E. D., Kolle, O. & Buchmann, N. Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany. *Agricultural and Forest Meteorology* **118**, 151–167, [https://doi.org/10.1016/S0168-1923\(03\)00115-1](https://doi.org/10.1016/S0168-1923(03)00115-1) (2003).
176. Knohl, A. *et al.* FLUXNET2015 DE-Hai Hainich. *FluxNet; University of Goettingen, Bioclimatology* <https://doi.org/10.18140/FLX/1440148> (2012).
177. Roy, D. P. *et al.* A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment* **176**, 255–271, <https://doi.org/10.1016/j.rse.2016.01.023> (2016).
178. Roy, D. P. *et al.* 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, <https://doi.org/10.1016/j.rse.2017.06.019> (2017).
179. Roy, D. P., Li, Z. & Zhang, H. K. 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**, <https://doi.org/10.3390/rs9121325> (2017).
180. Domnich, M. *et al.* KappaMask: AI-Based Cloudmask Processor for Sentinel-2. *Remote Sensing* **13**, 4100, <https://doi.org/10.3390/RS13204100> (2021).
181. Zhu, Z. & Woodcock, C. E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment* **118**, 83–94, <https://doi.org/10.1016/j.rse.2011.10.028> (2012).
182. Zhu, Z., Wang, S. & Woodcock, C. E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. *Remote Sensing of Environment* **159**, 269–277, <https://doi.org/10.1016/j.rse.2014.12.014> (2015).
183. Pabon-Moreno, D. E., Migliavacca, M., Reichstein, M. & Mahecha, M. D. On the potential of Sentinel-2 for estimating Gross Primary Production. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 4409412, <https://doi.org/10.1109/TGRS.2022.3152272> (2022).
184. Vogelmann, J. E., Rock, B. N. & Moss, D. M. Red edge spectral measurements from sugar maple leaves. *International Journal of Remote Sensing* **14**, 1563–1575, <https://doi.org/10.1080/01431169308953986> (1993).
185. McMurtrey, J. E., Chappelle, E. W., Kim, M. S., Meisinger, J. J. & Corp, L. A. Distinguishing nitrogen fertilization levels in field corn (Zea mays L.) with actively induced fluorescence and passive reflectance measurements. *Remote Sensing of Environment* **47**, 36–44, [https://doi.org/10.1016/0034-4257\(94\)90125-2](https://doi.org/10.1016/0034-4257(94)90125-2) (1994).
186. Rowan, L. C. & Mars, J. C. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. *Remote Sensing of Environment* **84**, 350–366, [https://doi.org/10.1016/S0034-4257\(02\)00127-X](https://doi.org/10.1016/S0034-4257(02)00127-X) (2003).
187. Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E. & Papale, D. Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data. *Remote Sensing of Environment* **168**, 360–373, <https://doi.org/10.1016/j.rse.2015.07.015> (2015).
188. Tramontana, G. *et al.* Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* **13**, 4291–4313, <https://doi.org/10.5194/bg-13-4291-2016> (2016).
189. Jung, M. *et al.* Scaling carbon fluxes from eddy covariance sites to globe: Synthesis and evaluation of the FLUXCOM approach. *Biogeosciences* **17**, 1343–1365, <https://doi.org/10.5194/bg-17-1343-2020> (2020).
190. Emmanuel Johnson, J., Laparra, V., Pérez-Suay, A., Mahecha, M. D. & Camps-Valls, G. Kernel methods and their derivatives: Concept and perspectives for the earth system sciences. *PLoS ONE* **15**, <https://doi.org/10.1371/journal.pone.0235885> (2020).
191. Cortés-Andrés, J. *et al.* Physics-aware nonparametric regression models for Earth data analysis. *Environmental Research Letters* **17**, 054034, <https://doi.org/10.1088/1748-9326/AC6762> (2022).
192. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
193. Reichstein, M. *et al.* Modelling landsurface time-series with recurrent neural nets. *International Geoscience and Remote Sensing Symposium (IGARSS)* **2018**(July), 7640–7643, <https://doi.org/10.1109/IGARSS.2018.8518007> (2018).
194. Shi, X. *et al.* Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems* **2015**(Janua), 802–810 (2015).
195. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Tech. Rep., Google Research (2015).
196. Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems* **32**, <https://doi.org/10.48550/arxiv.1912.01703> (2019).
197. Lovelace, R., Nowosad, J. & Muenchow, J. *Geocomputation with R* (Chapman and Hall/CRC, 2019).

198. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: A Fresh Approach to Numerical Computing. *SIAM Review* **59**, 65–98, <https://doi.org/10.48550/arxiv.1411.1607> (2014).
199. Hoffmann, J. GeoStats.jl—High-performance geostatistics in Julia. *Journal of Open Source Software* **3**, 692, <https://doi.org/10.21105/JOSS.00692> (2018).
200. Innes, M. Flux: Elegant machine learning with Julia. *Journal of Open Source Software* **3**, 602, <https://doi.org/10.21105/JOSS.00602> (2018).
201. Innes, M. J. *et al.* Fashionable Modelling with Flux. *CoRR* <https://doi.org/10.48550/arxiv.1811.01457> (2018).
202. Rackauckas, C. *et al.* DiffEqFlux.jl - A Julia Library for Neural Differential Equations. *CoRR* <https://doi.org/10.48550/arxiv.1902.02376> (2019).
203. Zubov, K. *et al.* NeuralPDE: Automating Physics-Informed Neural Networks (PINNs) with Error Approximations. *CoRR* <https://doi.org/10.48550/arxiv.2107.09443> (2021).
204. Martinuzzi, F., Rackauckas, C., Abdelrehim, A., Mahecha, M. D. & Mora, K. ReservoirComputing.jl: An Efficient and Modular Library for Reservoir Computing Models. *Journal of Machine Learning Research* **23**, 1–8, <https://doi.org/10.48550/arxiv.2204.05117> (2022).
205. Montero, D. *et al.* Awesome Spectral Indices. *Zenodo* <https://doi.org/10.5281/zenodo.5508090> (2022).

Acknowledgements

This project was financially supported via the “Digital Forest” project, Ministry of Lower-Saxony for Science and Culture (MWK) via the program Niedersächsisches Vorab (ZN 3679) and the European Space Agency (ESA) via the Deep Earth System Data Lab (DeepESDL) project. M.S. and M.D.M. acknowledge funding by the German Research Foundation (DFG) via NFDI4Earth. We would like to thank the authors and creators of the multiple spectral indices cited in this paper and that are part of the ASI catalogue. We also thank the contributors of spectral indices to the catalogue and the developers of the software used for the Python library, which includes the Python libraries `numpy`, `xarray`, `pandas`, `geopandas`, `dask` and `earthengine-api`. We also thank the providers of public remote sensing data such as ESA, NASA, and USGS, as well as the Google Earth Engine and Planetary Computer platforms that were used for requesting and processing the data used. We acknowledge the support by the Open Access Publishing fund of Leipzig University.

Author contributions

D.M. conceptualized the Catalogue, curated the data, led the writing of the manuscript. D.M. and C.A. wrote the code for the Catalogue design. D.M. wrote the code for the Python library. M.D.M. and S.W. supervised the investigation. D.M. and M.S. created the static visualizations. M.S. created the interactive visualization. All authors contributed to the review and writing of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023