Evaluation framework for the generation of continental bare surface reflectance composites

Paul Karlshoefer [a],*, Pablo d'Angelo [a], Jonas Eberle [b], Uta Heiden [a]

[a] German Aerospace Center, The Remote Sensing Technology Institute, Münchener Str. 20, Weßling, 82234, Germany
[b] German Aerospace Center, Earth Observation Center, Münchener Str. 20, Weßling, 82234, Germany

ARTICLE INFO

ABSTRACT

Soils play a pivotal role in supporting ecosystems, human health, food security, and climate regulation. Since several years, temporal composites of bare soil reflectances derived from multispectral satellite data are used as input for soil property modeling. Due to the importance of these model inputs, the quality of the surface reflectance composites (SRC) is essential. The quality depends on the precise selection of pixels that are free of green and dry vegetation, cloud contamination and other atmospheric disturbances.

However, there is a lack of suitable concepts and tools to evaluate the impact of the diverse processing parameters for the generation of SRC, especially for large areas such as continents. This study presents a novel approach to evaluate the process of computing bare SRCs across large geographical areas. It can estimate the theoretical limit achievable with defined processing parameters (spectral indices, thresholds, specific filtering, etc.) and it is also suitable to compare the performance of different SRC concepts from the literature. The performance is derived from the angular spectral distance between reference spectra derived from the LUCAS survey and the SRC spectra. It is demonstrated that a linear combination of two spectral indices complemented with a regional threshold dataset keep the complexity of threshold data sets low while performing well across Europe. The results also show that regionalization is as crucial as the choice of the index itself. The additional outlier removal focusing on clouds and haze marginally improved the SRC at the continental scale but can be very effective for areas with more frequent clouds. The proposed method offers two main advantages. First, it allows for parameter customization tailored to the region of interest, or, at minimum, to areas well represented by the reference data. Second, it facilitates the systematic evaluation of successive adaptations in the SRC generation process, eliminating the labor-intensive and error-prone task of visually comparing images to assess improvements in the SRC final product. The final bare surface reflectance composite for Europe and adjacent regions provids a robust foundation for future large-scale soil and bare surface monitoring.

1. Introduction

The properties of bare soils play a pivotal role in supporting ecosystems, human health, food security, and climate regulation (Montanarella et al., 2015). Earth Observation can be used to directly model bare soil characteristics based on their spectral reflectance characteristics (Chabrillat et al., 2002; Ben-Dor et al., 2009; Gerighausen et al., 2012; Ward et al., 2020) or based on environmental covariates (Mulder et al., 2011; Minasny and McBratney, 2016; Poggio et al., 2021). Recently, temporal composites of bare soil reflectances derived from space-borne satellite data over multiple seasons are used as input for soil property mapping (Dvorakova et al., 2021; Zepp et al., 2021; Broeg et al., 2024; Tziolas et al., 2024). The advantage of this technique is to overcome the temporal coverage of soils with vegetation by selecting only bare soil reflectance pixels from the multispectral

time stack (Diek et al., 2017; Rogge et al., 2018; Demattê et al., 2018; Roberts et al., 2019; Heiden et al., 2022). This way, the area for which soil properties can be mapped is increased.

The quality of bare surface reflectance composites (SRC) depends on the selection of mostly undisturbed pixels representing bare soils and non-vegetated surfaces. Disturbances may arise from partial coverage of the soil by photosynthetically active and non-active vegetation (Rogge et al., 2018; Demattê et al., 2018) and by soil roughness and moisture (Dvorakova et al., 2023; Vaudour et al., 2021). The majority of studies use one or more spectral indices to disentangle such disturbances (Heiden et al., 2022) with the choice of indices and associated thresholds tailored to the local or regional conditions of the area of interest (Castaldi et al., 2023; Dvorakova et al., 2023; Broeg et al.,

---

2024). Studies aiming at larger extents, such as continental (Safanelli et al., 2020) to global scales (Rizzo et al., 2023) are rare due to the complexity of capturing the nuanced variations of soil properties across different regions (Gallo et al., 2018) and the computational complexity.

Regionalization of data models has been shown to improve results in various global environmental studies, and it is a key component of the framework presented in this work. Notable examples include the mapping of forest canopy height (Ku et al., 2021), biodiversity (Coops et al., 2018), bare ground gain (Ying et al., 2017) and streamflow of river basins (Odusanya et al., 2022). To the best of the authors' knowledge, this has not yet been tested for the European-wide or continent-scale SRC generation. The reason might be the lack of concepts that rigorously evaluate the performance of various spectral indices paired with a regionalized set of thresholds derived from local characteristics in the reflectance data.

Another significant factor affecting SRC quality is cloud cover. Most temporal compositing approaches use the cloud masks produced during conversion from radiance values (Level 1) to reflectance values (Level 2) or additional software such as FMask (Baetens et al., 2019). Cloud masks, while effective in detecting highly affected pixels, must be conservative by design to avoid excessive data exclusion. While thick clouds can be detected with low uncertainties, the flagging of thin and semi-transparent clouds remains difficult (Skakun et al., 2022). Studies also reported spectral confusion between clouds and high reflectance built-up areas (Corbane et al., 2020) as well as in areas with low vegetation and high albedo characteristics, typical for Mediterranean soils (Ben-Dor, 1994). Kempeneers and Soille (2017) reduce the cloud contamination by selecting images based on quicklook RGB information and a minimum index value for the Sentinel-2 blue band. Other methods typically address this by taking the median of reflectance values along the time axis, assuming that cloudy observations are outliers (Simonetti et al., 2021).

As an alternative to the median for the temporal averaging, the mean along the temporal axis of the reflectance values can be taken (Gasmi et al., 2021). A fundamental advantage is the greater asymptotic efficiency of the mean over the median in statistically consistent and normally distributed data. For instance, given 10 samples (samples being bare surface pixels), the relative efficiency of the median is only approximately 80% that of the mean (Serfling, 2011). The observed phenomenon is assumed to be relatively stable in time; however, the dynamics in the atmosphere (e.g. haze) and surface (e.g. surface moisture) lead to a high variance in the observed reflectance data and therefore a tail-heavy distribution. The mean is inherently more biased towards outliers and, therefore, it has to be combined with robust outlier detection to specifically remove pixels affected by residual clouds. In addition to its better convergence to the true mean of the distribution, the mean enables the derivation of statistical products centered around it, such as the standard deviation or the confidence interval, which allows to characterizes the spectral variability uncertainty of the SRC pixels.

All of the issues described above influence the quality of SRC and thus affect the subsequent derivation of soil properties. To develop and test new concepts to improve the SRC at large scales such as continents, a robust and efficient evaluation concept is necessary. In many studies, the quality of different SRC versions are indirectly evaluated using the prediction accuracy of the derived soil properties such as described in Vaudour et al. (2021), Žížala et al. (2019) and Tziolas et al. (2020a). However, this strategy introduces additional uncertainty related to the machine learning models. For local studies, the resulting SRC spectra have been directly compared with laboratory spectra using the soil line and spectral dispersion using principle component analysis (Demattê et al., 2018). Heiden et al. (2022) use spectral distance measures between the SRC and laboratory spectra of the Copernicus Land Use-Land Cover Area Frame Survey (LUCAS) for Bavaria in Germany .

The evaluation of SRC at large scale such as the European continent is still unsolved due to missing reference data that represent averaged spectral information across different seasons and years. However,

Nocita et al. (2015) proved the statistical relation between spectral reflectance measurements of soils and their chemical properties. As a consequence, many of the soil parameter mapping approaches have used the spectral and chemical analysis collected in local laboratory and field surveys (Castaldi et al., 2018; Tziolas et al., 2020b; Mzid et al., 2022) or in the LUCAS survey (Ward et al., 2020; Castaldi et al., 2019; Safanelli et al., 2020) as a reference. Thus, for an European-wide evaluation of SRCs, the LUCAS survey seems to be a suitable reference that contains chemical and spectral measurements of the soil following the same protocol (Orgiazzi et al., 2017). The challenge is the different measurement method between the SRC and LUCAS spectra. In particular, the spectral measurements of the LUCAS survey are derived from dried and sieved soil samples at high radiometric resolution, making direct comparisons with space-borne soil reflectance under natural field conditions challenging. To allow comparison with Sentinel-2-based surface reflectance, the high-resolution LUCAS soil spectra are often resampled to match the multi-spectral resolution of Sentinel-2 (Castaldi et al., 2019). This resampling process smooths out the distinctive narrow spectral absorption features, effectively filtering out high-frequency information. However, the remaining broad spectral shape — like convexity and overall curvature can be used to compare SRC and LUCAS spectra.

This study aims to introduce a systematic evaluation framework designed to quantify improvements in SRC generation methods, moving beyond traditional visual assessments. By optimizing key processing parameters (e.g. spectral indices and cloud-filtering) through an iterative process that maximizes concordance between SRC spectra and resampled LUCAS spectra at corresponding pixel locations, we aim to identify good processing settings with reasonable computational effort. We hypothesize that this framework can determine the choice of the spectral index and its corresponding thresholds (regional or global) and quantify the benefit of further subroutines in the SRC generation. This framework is applied to (1) evaluate SRC generation enhancements introduced in this work, while building upon developments from the Soil Composite Mapping Processor (SCMaP) as described in previous studies (Rogge et al., 2018; Heiden et al., 2022) and (2) it is used to compare other SRC generation concepts that have been published. The developed framework can provide quantitative and objective means to define the best SRC processing parameter for a specific area.

2. Material and methods

This section outlines the methodological foundation of the evaluation framework designed to quantify the influence of processing parameters on the quality of bare surface reflectance composites (SRCs).

A schematic overview of the evaluation framework is provided by Fig. 1. The main idea is to search for the optimal set of processing parameters (blue box in Fig. 1) identified by minimizing the angular distance between the derived SRC spectra and the reference spectra through an iterative refinement of those parameters. The processing parameters are generic and can be defined as proposed in this study or taken from the literature such as provided by Diek et al. (2017), Demattê et al. (2018) and Broeg et al. (2024). In this study, the proposed parameters include the selection of spectral indices and additional bare surface detection routines suitable to minimize the influence of green and dry vegetation and to filter out cloudy and hazy pixels.

A common temporal stack of Sentinel-2 L2A data (Section 2.1.1) is used ensuring consistency across all optimization runs and enabling a fair comparison of the performance of the spectral indices. The reference spectra are obtained from the European LUCAS 2015 survey (Section 2.1.3).

The SRC generation itself remains intentionally generic within this framework, with its key components determined during the optimization process (Section 2.2). Fig. 2 illustrates the generation of a SRC image, where the parameters found in the optimization loop are used
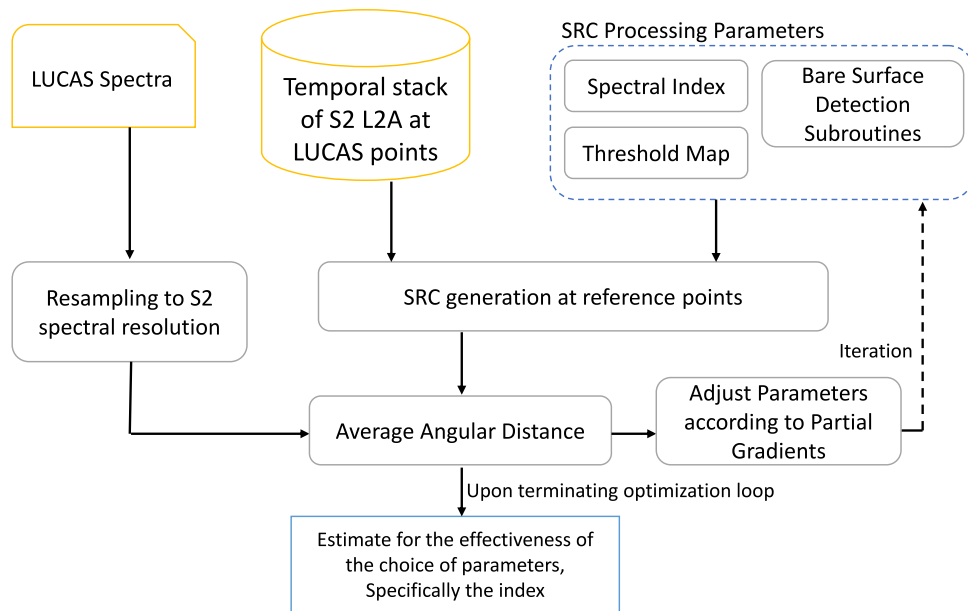
Fig. 1. The evaluation framework: The spectral reference data is taken from the LUCAS survey. The temporal stack of the Sentinel-2 L2A data is identical for any SRC, that is being generated in this study. Processing parameters are iteratively optimized via gradient descent until the spectral angular distance between reference and SRC spectra no longer decreases, providing an estimate of the performance of the chosen parameters.
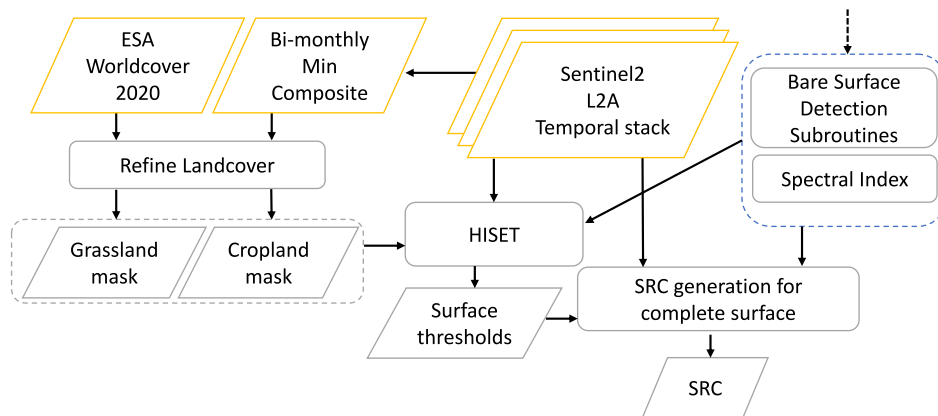

Fig. 2. SRC generation enhancements: Pixel-based surface thresholds are derived using HISET and a refined landcover data set, that is based on ESA Worldcover. The resulting surface thresholds are then used in combination with the entire S2 temporal stack and the processing parameters found prior (blue), to generate the SRC.

(blue). Instead of limiting the evaluation to LUCAS points, this approach enables the generation of a spatially continuous SRC image. A key component in this process is the HISET algorithm (Section 2.2.1), which is employed to construct surface thresholds essential for an area-wide SRC derivation. This procedure relies on cropland and grassland masks, initially based on the ESA WorldCover dataset (Section 2.1.2) and subsequently refined to align with the specific requirements of this study (Section 2.2.2).

This approach contrasts with existing studies, where the indices and thresholds are predefined. The proposed method offers two main advantages. First, it allows for parameter customization tailored to the region of interest, or, at minimum, to areas well represented by the reference data. Second, it facilitates the systematic evaluation of successive adaptations in the SRC generation process, eliminating the labor-intensive and error-prone task of visually comparing images to assess improvements in the final product (Section 2.3).

In this work, the term "bare surface" encompasses non-vegetated terrestrial surfaces including soils, barren rock, and sandy surfaces. Artificial and urban surfaces, snow and water bodies are excluded.

2.1. Data preparation

2.1.1. Sentinel-2

The foundation of the SRC is a temporal stack of Sentinel-2 multi-spectral imagery. This archive is well-suited for the analysis due to the Sentinel-2 constellation's ability to consistently capture data with a five-day revisit period (European Space Agency (ESA), 2015), providing a dense temporal resolution critical for capturing the rare observation opportunities of non-vegetated, bare surfaces.

The selection process for the input data used to generate all tested SRCs in this paper is filtered based on several criteria:

- **Metadata Filtering**: Images with a maximum cloud coverage of 80 percent are considered. This value is strategically chosen to maximize the density of the temporal stack, acknowledging that high cloud cover could still yield valuable data. Furthermore, the sun has to be elevated more than 20 degrees between the senor and the horizontal plane.
- **Temporal Range**: The study period spans from January 2018 to December 2022.

Table 1

Sentinel 2 bands used in this work. (Copernicus, 2020).

Band	Central wavelength (S2A)	Central wavelength (S2B)
B02 (blue)	492 nm	492 nm
B03 (green)	560 nm	559 nm
B04 (red)	664 nm	665 nm
B05 (red edge 1)	704 nm	704 nm
B06 (red edge 2)	741 nm	739 nm
B07 (red edge 3)	783 nm	780 nm
B08 (NIR)	833 nm	833 nm
B8A (red edge 4)	865 nm	864 nm
B11 (SWIR 1)	1614 nm	1610 nm
B12 (SWIR 2)	2202 nm	2186 nm

- **Spectral Band Selection**: Only bands with a spatial resolution at ground level of 20 m or better are included (see Table 1).

The spatial organization of the input images adheres to the Military Grid Reference System (MGRS) (Meyers, 2015), which facilitates systematic processing and analysis. The data were pre-processing to Level-2A (L2A) reflectances using the MAJA Atmospheric Correction processor (Hagolle et al., 2010). This step adjusts for atmospheric effects and cloud identification and ensures radiometric consistency across the image stack. The used scenes are corrected for adjacency and topographic effects (denoted by the suffix FRE) (Hagolle et al., 2018, 2021).

For this study, the analysis is confined to pixels classified as '0' by the geophysical mask (*_MG2_R2), effectively excluding areas represented as open water, clouds, snow, and those obscured by topographic shadows or unsuitable sun elevation angles. This selection criterion ensures that the focus remains on the surfaces that are most relevant to the SRC.

In total, a 1145 MGRS tiles and a total of 4 45 806 individual Sentinel-2 scenes (each comprised of 10 bands) contributed to the computation.

2.1.2. Landcover mask

The integration of land cover data, specifically the ESA World-Cover map (Zanaga et al., 2022), serves as an input for the estimation of surface thresholds and is therefore only part of HISET (Section 2.2.1). Therefore, it functions as an indirect input for the final composite rather than as a direct masking layer. Designed for the year 2020, ESA WorldCover is centrally positioned within the temporal stack of Sentinel-2 data. With a spatial resolution of 10 m and precise classification of critical land-cover types, such as grasslands and croplands, it is - with refinement - a central input to the threshold derivation, which is detailed in Section 2.2.1.

2.1.3. LUCAS spectral library

The LUCAS topsoil spectral database (2015) is used as reference for the general shape of soil spectra at their respective sample sites. Important for this work are the more than 22,000 topsoil spectra from European soil samples. To increase the likelihood of bare soil visibility for optical remote sensing, our analysis focuses specifically on the 8823 data points located within agricultural fields (labeled as "Cropland"). Another 149 points have been removed from the set due to low spectral variability in time (according to Eq. (6)) in their local vicinity, in an effort to exclude sites which have turned into inactive cropland since 2025. On such fields, their might be permanent vegetation with no option to observe the bare surface.

To allow for comparison between the LUCAS spectra, s_{lucas} and the Sentinel-2 spectra, the LUCAS spectra are projected into Sentinel-2 data space using the spectral response function sr_A and sr_B for Sentinel-2 A and B, respectively:

$$\vec{q} = \frac{sr_A + sr_B}{2} \vec{s}_{lucas} \tag{1}$$

Using Eq. (1), the high-resolution LUCAS spectra $\vec{s}_{lucas}$ are resampled to match the spectral resolution of Sentinel-2 by the average of the spectral response functions of Sentinel-2 A and B (Castaldi et al., 2019; Okujeni et al., 2024).

Due to the radiometric resampling to ten bands, only the general shape of the LUCAS spectra is preserved, while detailed absorption features are eliminated. This, combined with an angular distance metric that is invariant to the scaling of the spectral vector, ensures that the primary characteristic compared between composite spectra and $\vec{q}$ is their overall shape of the spectra.

2.2. Procedure to generate bare surface reflectance composites

Starting from the foundational element, a single observation, its multi-spectral cube is $O \in \mathbb{R}^{u \times v \times b}$ for an image of spatial dimension $u \times v$ and b spectral bands.

Given that observations are geometrically co-registered and mapped onto the Sentinel-2 tile grid, thus, each image sharing uniform dimensions, the tensor can be directly extended by a temporal component $S = (O_0, O_1, \dots, O_n)$ over n observations within a tile.

The synthesis of a SRC C from S is executed through two pivotal functions. Firstly, the classifier $f_{bare}(S) \to \{0|1\}^{u \times v \times b \times n}$, discriminates each element within S whether it is representative of undisturbed bare surface (1) or not (0). Using the mask will result in the set

$$S_{bare} := \{s \in S | f_{bare}(s) > 0\}. \tag{2}$$

In our case, f_{bare} is primarily based on a spectral index, z, thus:

$$f_{bare} := \begin{cases} 1, & \text{if } t_{0,(i,j)} < z(\vec{s}_{i,j,:,m}) < t_{1,(i,j)}, i \in \{1..u\}, j \in \{1..v\}, m \in \{1..n\} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Eq. (3) masks all pixels indexed by the coordinates i and j that are not in between their local set of thresholds $t_{0,(i,j)}$ and $t_{1,(i,j)}$ for index z. It is extended by an outlier detection and removal which is detailed in the next chapter.

Secondly, the aggregation $C(S) \to \mathbb{R}^{u \times v \times b}$ condenses the temporal dimension to a single multi-spectral image using the definition of S_{bare} (Eq. (2))

$$C(S_{bare}) = \frac{1}{|S_{bare}|} \sum_{r \in S_{bare}} r. \tag{4}$$

In this work, we deliberately chose the mean over the median for its greater asymptotic relative efficiency; thus, its tendency to converge faster to the true mean of the distribution (Serfling, 2011).

The delineation of bare surfaces from vegetated areas in our analysis hinges on the precise determination of thresholds applied to a carefully chosen spectral index, crucial for identifying natural bare surfaces. These thresholds, t_0 and t_1, are defined within the bounds of a spectral vegetation index z calculated only from a reflectance vector s, which is composed from the spectral band values $s = (s_{B2}, s_{B3}, s_{B4}, s_{B5}, s_{B6}, s_{B7}, s_{B8}, s_{B8A}, s_{B11}, s_{B12})^T$:

$$t_0 < z(s) < t_1 \tag{5}$$

The spectral index $z : \mathbb{R}^{10} \to \mathbb{R}$ serves as the basis for this classification, with higher index values indicating an increased presence of vegetation. The selection of t_1 is particularly critical as it delimits the transition from bare to vegetated surfaces. We hypothesize that these thresholds require careful calibration with respect to the local characteristics of soil and crops.

Within the scope of this work, spectral indices commonly found in the literature are investigated (Table 2). Single and multi-staged indices are tested, like NDVI/NBR2, NDVI/NBR, BCC/NDVI/NBR2, VNSIR/NDVI/NBR2, where the indices are applied to a spectrum consecutively, each with its own threshold. VNSIR and BCC are only tested as proposed by the authors staged with filtering by NDVI and NBR2, first. The BSI has been inverted so lower values correspond to bare

Table 2

Spectral indices tested in this work for their suitability to generate bare surface composites. Abbreviations of the indices and their original source (column 1), their respective formulas in terms of Sentinel-2 bands (column 2) and their full names and usage (column3) are shown.

Index	Formula	Usage
NDVI (Rouse et al., 1974)	$\left(\frac{B8-B4}{B8+B4}\right)$	Normalized Differences Vegetation Index used by Urbina-Salazar et al. (2021)
NBR2 (Deventer et al., 1997)	$\left(\frac{B11-B12}{B11+B12}\right)$	Normalized Burn Ratio 2 used by Vaudour et al. (2021)
NBR (García and Caselles, 1991)	$\left(\frac{B8-B12}{B8+B12}\right)$	Normalized Burn Ratio
BSI (Rikimaru et al., 2002)	$-\left(\frac{B12+B4-B8A-B2}{B12+B4+B8A+B2}\right)$	Bare soil index (inverted) used by Diek et al. (2017)
NDVI+NBR	$\left(\frac{B8-B4}{B8+B4}+\frac{B8-B12}{B8+B12}\right)$	Linear combination of NDVI and NBR used by Heiden et al. (2022)
NDVI/NBR2	$\left(\left(\frac{B8-B4}{B8+B4}\right),\left(\frac{B11-B12}{B11+B12}\right)\right)$	Staged NDVI and NBR2 with 2 separate sets of thresholds used in Dvorakova et al. (2021)
NDVI/NBR	$\left(\left(\frac{B8-B4}{B8+B4}\right),\left(\frac{B8-B12}{B8+B12}\right)\right)$	Staged NDVI and NBR with 2 separate sets of thresholds
VNSIR (Demattê et al., 2020)	$1-((2*B4-B3-B2)+3(B12-B8))/10\,000$	Visible to Shortwave Infrared Tendency Index staged with NDVI/NBR2 3 sets of thresholds, used by Demattê et al. (2020)
BCC (Gillespie et al., 1987)	$\left(\frac{B2}{B4+B3+B2}\right)$	Blue Chromatic Coordinate staged with NDVI/NBR2 3 sets of thresholds, used by Broeg et al. (2024)

surfaces (similar to NDVI, NBR2 etc.). Linear combinations are also possible, where two indices are combined and hence can be treated with a single threshold (like NDVI+NBR).

2.2.1. Surface thresholds (HISET)

This study employs two distinct strategies for deriving thresholds. The first strategy involves an optimization process, as detailed in Section 2.3, which computes a set of thresholds at each reference point (LUCAS sample site).

The second strategy derives thresholds for SRC generation using an extended version of the Histogram Separation Threshold (HISET) algorithm, introduced by Heiden et al. (2022). These thresholds, termed surface thresholds, are applicable to entire areas rather than singular points. Ultimately, the surface thresholds are utilized for the SRC generation, as the thresholds computed at the LUCAS points are specific to the sparse grid of reference points and are only used to determine the best spectral index.

The general principles of HISET, along with the modifications required to ensure its applicability across diverse conditions, are elaborated in this section.

The surface thresholds adhere to the Sentinel-2 tile-grid. It is chosen to be the spatial skeleton, where thresholds are determined for each tile in the grid. The granularity of roughly 100 km^2 ensures a balance in scale that avoids both the over-generalization of large regions and the over-fitting risks of excessively small ones. To refine spatial continuity and minimize edge effects within the resulting data, these thresholds were smoothed employing a radial basis function centered on the centroid of each tile, see result in Fig. 6. This allows for a seamless integration of tiles into the final composite. Heiden et al. (2022) details the HISET algorithm in the general case. It is basically designed to account for the spectral similarity for the non-photosynthetically active vegetation (NPV) and bare soil (Daughtry and Hunt, 2008; Dennison et al., 2019, 2023). Since Sentinel-2 only has two spectral bands in the SWIR, the usage of spectral absorption features to detect NPV is not an option. . Therefore, HISET measures the spectral overlap between bare soils and NPV and defines a threshold that minimizes the spectral overlap. HISET encompasses four primary steps:

1. **Spectral Index Calculation**: Compute the spectral vegetation index for each image and pixel in the temporal Sentinel-2 series
2. **Minimum Index Selection**: Selecting the minimum spectral index value across the temporal sequence per pixel ensures representation in its least vegetated state, typically indicating bare surfaces and non-photosynthetic vegetation (NPV) for croplands and grasslands, respectively.
3. **Histogram Aggregation**: All pixels that belong to cropland or grassland according to the land-cover map (Zanaga et al., 2022) are compiled into two separate histograms. Each histogram is normalized to an area of 1, effectively transforming it to a discrete probably density function.
4. **Threshold Identification**: The threshold is chosen to be the center of the histogram bin that minimizes the larger of both areas, that is intersected, at either side (see Fig. 4).

In addition to the value for the threshold, the area that has been minimized in step 4, gives an indication to the degree of separation of both histograms. Due to normalization, this value is between 0 for two disjointed histograms and 0.5 when one is entirely contained within the other. Multiplied by 100, this value is the quality score of the threshold. It serves as an indicator for the effectiveness of the surface threshold.

2.2.2. Enhancements to the surface threshold derivation

The initial methodology (HISET) for threshold derivation, as outlined in the previous Section 2.2.1, was further refined to address the challenges posed by the complexity of bare surfaces in Europe.

A notable concern was the presence of mixed pixels, that embody areas where agricultural land intersects with non-agricultural elements, such as roads or cart tracks, within the same pixel. These elements interfere with the spectrum and consequently the spectral index, introducing variance to the histograms. This effect leads to blurred histograms and consequently elevated scores of the threshold, that indicate a degraded or inaccurate thresholds. To reduce the effect of mixed pixels, we introduce a measure of spectral variability:

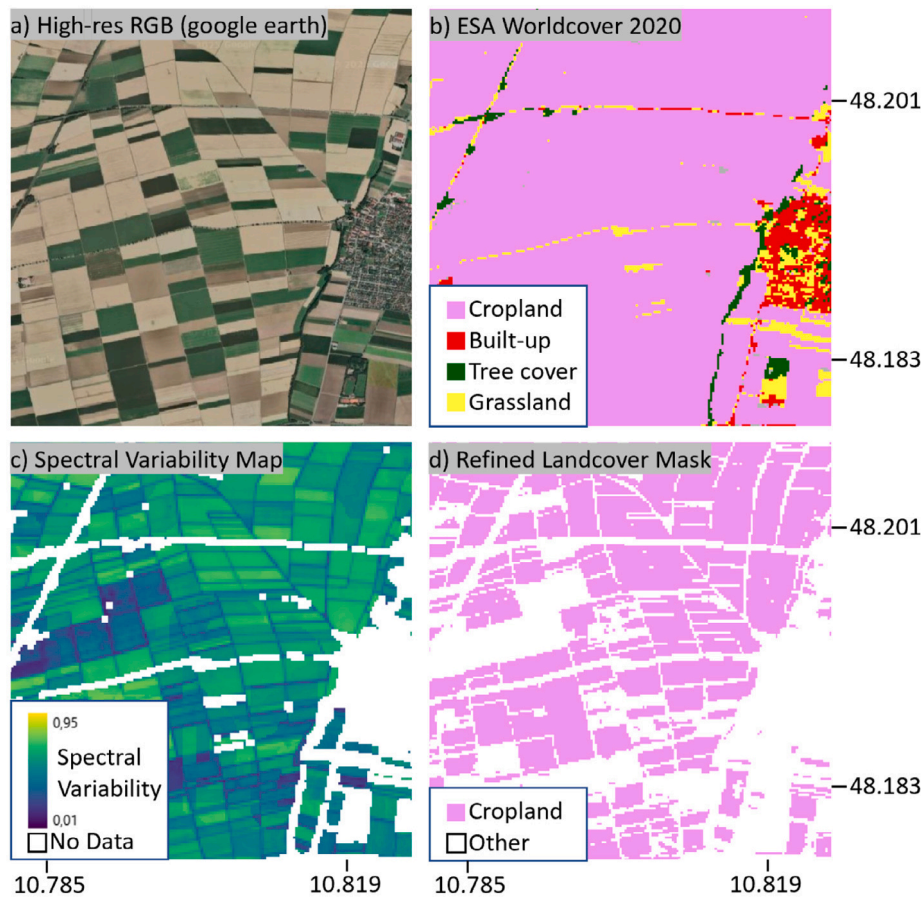$$v := \frac{1}{n-1}\sum_{i=0}^{n-1}\left|\frac{z(H_{i+1})-z(H_i)}{d_{i+1}-d_i}\right| \tag{6}$$

Fig. 3. The high-resolution google earth image of an agricultural area in southern Germany (a) shows the overestimation of cropland pixels by the ESA Worldcover landcover map (b). By excluding pixels with low spectral variability in the vegetation index (darker areas in the (c)), the land-cover map is refined (d). Some fields without active management are excluded by this process as well as many small roads.

Here, $H = (\bar{s}_0, \bar{s}_1, \ldots, \bar{s}_n)$ is a sequence of averaged spectra over two months in time with d indicating the respective central dates. The initial temporal averaging of two months of the reflectance data eliminates signals at a higher frequency, that are not related to the crop lifetime cycle.

The spectral variability v estimates the frequency and intensity of changes to agricultural surfaces per pixel and therefore the likelihood that a bare state is exposed at some point in time. It is comparable only at local scale, as climatic and seasonal effect or land management policies do impact it. Greater spectral variability translates to actively manged croplands, that are less likely to intersect a stable surface. Thus, the land-cover map for cropland is refined by this additional criterion. Depending on the overall abundance of cropland pixels, a percentile (up to the 20th-percentile) of less active pixels is removed or the pixel's contribution to the histogram is weighted according to its variability.

Beyond considering spectral variability, pixels that contribute to expansive, contiguous areas are prioritized. In contrast, isolated pixels and small patches — regardless of classification — are excluded from contributing to the histograms. They are found through a parallel flood-fill algorithm (van der Walt et al., 2014).

The resulting land-cover map can be seen in Fig. 3. It gives an example of the identification of both mixed pixels at field edges and fields with comparatively low variability in the vegetation index (red circle). Both are excluded, so the surface threshold estimation is based on more pixels, which actually represent the bare surface.

While the HISET methodology shows robust performance across the majority of Europe, challenges arise in tiles with limited cropland or grassland coverage, leading to imbalanced or erratic density functions

and ultimately, poor surface thresholds. Such cases are notably prevalent in coastal regions and islands where a large portion of the tile is water, as well as northern regions with inherently sparse cropland.

Various tests indicated, that both land-cover classes should be represented by at least two percent of all pixels in the tile, (at least $2e5$), thereby a tile is classified as fit. In instances where this criterion is not met, alternative approaches are employed to estimate thresholds:

- **Proximity Interpolation**: For tiles adjacent to one or more 'fit' tiles, thresholds are linearly interpolated, utilizing the centroid of each tile as a basis for distance calculation. This method is particularly relevant for coastal regions.
- **Histogram Aggregation**: In the absence of nearby 'fit' tiles, as often found with islands, pixel data from multiple adjacent tiles are combined into a singular histogram. The resulting threshold is used for all participating tiles.
- **Bio-geographic Aggregation**: In the case of northern Scandinavia, the system of multiple climatic zones led to an increased variance in the combined histograms. Here, pixels are aggregated by bio-geographic zone (EEA, 2016). Thresholds are then calculated for each zone and applied proportionally across the tiles based on the area of each bio-geographic zone.

These tailored approaches ensure the methodology's effectiveness, even in regions where traditional threshold derivation faced significant challenges.

2.2.3. Detection of residual clouds and haze

Eq. (3) can be extended by a set of cloud filters, integrated to specifically address the simpler challenge of separating cloudy and hazy pixels from those representing a bare surface. This excludes pixels that are affected by haze or clouds. Given the high brightness of clouds in contrast to most soils, this is important.

The first filter exploits the distinct spectral characteristics between soil and cloud pixels particularly in the SWIR. While soil spectra have generally higher values in the short-waved infrared (SWIR) compared to the visible to near-infrared (VNIR) (true for more than 99% of LUCAS soil spectra), the opposite is true for clouds. Therefore the condition can be formulated as

$$s_{B12} - s_{B8} \overset{!}{>} 0, s \in S_{bare}, \tag{7}$$

where the subscript denotes the selected Sentinel-2 band.

The second filter within f_{bare} focuses on the removal of pixels affected by haze or thin clouds. The influence is very prominent in the blue band (Band B02), which is consequently used for detection (B02 is also extensively used in ESA (2021)). Given the challenge that the reflectance in the B02 band of the underlying soil is unknown, this filter detects outliers along the temporal axis in relation to the median of the dataset in the blue band. The normalized mean absolute deviation (NMAD) for in the blue band $S_{soil,B2} = \{s_{B2}, \forall s \in S_{bare}\}$ is given by:

$$\tilde{S}_{soil,B2} = median_t(S_{soil,B2})$$

$$\text{Absolute Deviations} = \{|x - \tilde{S}_{soil,B2}| : x \in S_{soil,B2}\}$$

$$\text{NMAD} = k \cdot median(\text{Absolute Deviations}) \tag{8}$$

where $k = 1.4826$, a constant scaling factor for data adhering to a normal distribution and $median_t$ denotes the median along the temporal axis. Eventually, reflectance values are considered only if they fulfill the condition:

$$s_{B2} - \tilde{S}_{soil,B2} \overset{!}{<} \sigma \cdot \text{NMAD}, s \in S_{bare} \tag{9}$$

Typically, σ is set to 3.

The conditions (7) and (9) serve to adapt the scope of cloud exclusion of the MAJA processor to a stricter set of rules. Consequently, only pixels that satisfy these criteria are selected for inclusion in the SRC, contributing to a more accurate reflection of the spectral signature of the actual surface.

2.3. SRC evaluation

Determining the optimal spectral vegetation index, $z(s)$, is critical for the f_{bare} classifier, ensuring it accurately distinguishes bare soil and surface pixels from other pixels using only the Sentinel-2 spectral bands' reflectance vector, s. The index z is parameterized by a set of thresholds, so it is able to adapt to the diverse climatic and surface conditions. By finding the optimal set of thresholds for various choices of z at each evaluation pixel, one can draw conclusions about the suitability of the chosen spectral index. Ultimately, this performance estimation can serve as a baseline for testing the whole SRC generation and therefore can quantify improvements made to the processor.

To evaluate z, we measure the congruence between the SRCs generated using a particular z and its associated thresholds, and data measured on the surface. These reference spectra $\vec{q}$ are derived from the LUCAS topsoil database (2015) (Orgiazzi et al., 2018), which was resampled according to Section 2.1.3.

To quantify the difference between a single composite spectrum $\vec{c}$ and $\vec{q}$, literature proposes the spectral angle, which is derived from the cosine similarity:

$$\alpha(\vec{c}, \vec{s_2}) = cos^{-1}\left(\frac{\vec{c} \cdot \vec{q}}{\|\vec{c}\| \|\vec{q}\|}\right). \tag{10}$$

The measure is invariant to global changes in brightness, e.g. the length of the vector in its vector-space (van der Meer, 2006), which

is important since the reference spectra are scaled differently to the remote sensing spectra.

The establishment of a reference dataset, alongside a mechanism for its comparison against Sentinel-2 data sets the stage for an optimization problem. Our objective is to find, for a choice of z at each LUCAS point an ideal set of thresholds, that ensure maximum concordance with the shape of the LUCAS reference.

For the jth LUCAS point, a set of thresholds that are used to compute the composite $C(z, g)$ are required.

The dimension of this set depend on the form of z (for example z can be a set of multiple spectral indices). The set of (t_0, t_1) pairs is called τ to avoid confusing subscripts.

The task then becomes identifying the optimal $\tau_j = (\tau_{(j,(1))}, \tau_{(j,(2))}..)^T$ for each point that minimizes the angular distance between the SRC and the ground truth:

$$\phi_j(z) = \min_j \left(\alpha\left(C(z, \tau_j), \vec{q}_j\right)\right). \tag{11}$$

$C(z, \tau)$ and in result Eq. (11) are not smooth nor differentiable. However, for the optimization, this is necessary. Therefore, the set notation of Eq. (2) is replaced by a soft exclusion using the sigmoid $\sigma(x) = \frac{1}{1+e^{-\beta x}}$. The soft exclusion is based on the distance between the current guess of the thresholds of τ to the actual value of the spectral index for all $s \in S$.

$$\hat{C}(z, \tau) = \frac{\sum^i \sigma(z(X_i) - \tau)s}{\sum^i \sigma(z(X_i) - \tau)}, \forall s \in S \tag{12}$$

The parameter β is the sharpness of the soft exclusion. It is set to 15 in this case.

The performance p of z using its optimal set of thresholds τ^{opt} is the average angular distance over all m LUCAS points

$$p(z) = \frac{1}{m} \sum_{j=1}^{m} \phi_j(z). \tag{13}$$

This is the central measure to evaluate the SRC.

The τ_j^{opt} that minimizes Eq. (11) is found iteratively at each LUCAS point, using gradient descent:

$$\tau_j^{n+1} = \tau_j^n - \gamma \nabla \alpha\left(C(z, \tau_j^n), \vec{q}_j\right), \tag{14}$$

with $\nabla = \left(\frac{\partial}{\partial \tau_{j,(1)}}, \frac{\partial}{\partial \tau_{j,(2)}}, \dots\right)$ and a learning rate $\gamma > 0$. The partial derivatives are approximated numerically using finite differences. In addition to Eq. (11), there is an additional condition that the number of spectra composing the final composite must be at least five for each pixel. This requirement reduces the likelihood of overfitting the thresholds τ^{opt}.

For normalized spectral indices that are used in this work, γ is set to

$$\gamma = \frac{t_{max} - t_{min}}{200}, \quad z \rightarrow [t_{min}, t_{max}]. \tag{15}$$

Assuming that Eq. (11) is locally convex, the search described in Eq. (14) is terminated upon reaching a local minimum. By restarting the algorithm with different initial guesses that converge to the similar τ^{opt}, we assume that Eq. (11) is indeed convex in the vicinity of its minimum.

2.4. Software and tools

The computations related to the optimization were performed on the high-performance data analytics platform terrabyte hosted at the Leibniz Supercomputing Center. The source code is mostly written in Python and leverages the power of vectorized computations implemented in numpy and scipy for calculations, as well as gdal to handle IO. A spectral library of all Sentinel-2 observations in time at all LUCAS points allows to generate single pixel SRCs quickly, crucial for quick convergence during the gradient descent method.

The final SRC at continental scale was computed using the Soil Composite Mapping Processor (SCMaP) from DLR.

Table 3

Performance p according to Eq. (13), computed from a bare surface discriminator f_{bare} that is based on the spectral index z. Lower values of p indicate greater spectral similarity. For readability, p was normalized to the performance of NDVI+NBR.

Spectral index z	$[t_{min}, t_{max}]$	Initial guess τ^0	$p(z)$ normal.
NDVI $\left(\frac{B8-B4}{B8+B4}\right)$	$[-1, +1]$	$(-0.25, 0.25)^T$	1.55
NBR2 $\left(\frac{B11-B12}{B11+B12}\right)$	$[-1, +1]$	$(-0.3, 0.1)^T$	1.26
NDVI/NBR2	$[-1, +1]$	$((-0.25, 0.25), (-0.3, 0.1))^T$	1.11
BSI $-\left(\frac{B12+B4-B8A-B2}{B12+B4+B8A+B2}\right)$	$[-1, +1]$	$(-0.1, 0.1)^T$	1.09
NBR $\left(\frac{B8-B12}{B8+B12}\right)$	$[-1, +1]$	$(-0.3, 0.1)^T$	1.05
NDVI/NBR2/VNSIR	$[-1, +1]^2, \mathbb{R}$	$((-0.25, 0.25), (-0.3, 0.1), (0, 1))^T$	1.04
NDVI+NBR $\left(\frac{B8-B4}{B8+B4} + \frac{B8-B12}{B8+B12}\right)$	$[-2, +2]$	$(-0.6, 0.25)^T$	1
NDVI/NBR $\left(\frac{B8-B4}{B8+B4}, \frac{B8-B12}{B8+B12}\right)$	$[-1, +1]^2$	$((-0.6, 0.25), (-0.5, 0.1))^T$	0.98
NDVI/NBR2/BCC	$[-1, +1]^2, \mathbb{R}$	$((-0.25, 0.25), (-0.3, 0.1), (0, 0.3))^T$	0.96

3. Results

3.1. Selection of the spectral index

First, common choices for a spectral vegetation index z from publications are investigated (Heiden et al., 2022; Broeg et al., 2024; Demattê et al., 2020; Diek et al., 2017). Using the method outlined in Section 2.3 to obtain the best set of thresholds for each index at each LUCAS point, the Table 3 shows the performance $p(z)$ of these indices in calculating an accurate SRC spectrum at these points. The performance is measured according to Eq. (13), thus, lower values signify greater spectral similarity. For readability, the last column is normalized to the value of NDVI+NBR.

Table 3 illustrates that NDVI (1.55 times worse than NDVI+NBR) and NBR2 (1.26 times worse than NDVI+NBR) on their own are poor choices for the classification of bare surfaces. Combining them into a two-stage spectral index with separate, independent thresholds improves their performance to 1.11 times performance of NDVI+NBR.

Demattê et al. (2020) proposed to add a third stage to the classifier called VNSIR: $1 - ((2 * B4 - B3 - B2) + 3(B12 - B8))/10000$ (see Table 2). This improves the combined index substantially to just 1.04 time less accurate compared to the NDVI+NBR index.

The BSI (Rikimaru et al., 2002; Diek et al., 2017) manages to produce SRC spectra about 1.09 times less accurate than the NDVI+NBR index and the NBR index performs better, at just 1.05 times worse compared to NDVI+NBR. Eventually, by adding the NBR and the NDVI (NDVI+NBR), such as proposed by Heiden et al. (2022), we obtain the most effective, single-stage index combination of this list in selecting bare surface pixels at the LUCAS sites.

The two-staged NDVI/NBR index with distinct thresholds for both components or the three staged NDVI/NBR2/BCC (Broeg et al., 2024) improve performance (factors 0.98 and 0.96, respectively). However, further analysis will be based on the NDVI+NBR index. It has the highest performance of all indices that have only a one-dimensional set of thresholds (see column 2), which substantially reduces complexity and robustness in the actual SRC generation implementation.

3.2. Validating enhancements to the soil detection and threshold derivation

The thresholds τ^{opt} computed in the previous section reflect for the NDVI+NBR index the optimal threshold set at each LUCAS site. These thresholds are tuned to individual pixels and potentially biased to the selection of the LUCAS points; therefore, they cannot be applied at broader scales, such as that of Sentinel-2 tiles. Nevertheless, we can use the τ^{opt} to compare them to the performance p of the SRC generation. Incorporating successive refinements into f_{soil}, including additional cloud filter and land cover adjustments for threshold derivation, has quantifiably enhanced the accuracy of SRCs. Table 4 shows the

Table 4

Gradual improvement of the performance of the SRCs built by surface thresholds. Values are normalized to the limit of the index. Lower values correspond to greater spectral similarity (via lower angular distance) and in result, a better SRC.

Optimizations for SRC creation	Performance p of surface thresholds
Basic threshold database (HISET) Section 2.2.1	1.16
+ additional cloud filters for SRC creation Section 2.2.3	1.13
+ land-cover optimization Section 2.2.2	1.07
Theoretical limit of the index NDVI+NBR	1

reduction in the angular distance from 116% to 107% of the theoretical limit achievable with the NDVI+NBR index at pixel level.

Fig. 4 illustrates the effect of the optimization of the land cover map to estimate the thresholds. Using the vegetation variability defined in Eq. (6), areas with little change are excluded from HISET. The variability mask reveals that most often the removed areas are mixed pixels covering unmanaged fields, which consequently have less chance in exposing the bare surface. Removing these pixels (magenta region) from the threshold estimation allows for a cleaner separation of the two histograms (smaller score) and in result, stricter thresholds (solid line compared to dashed line).

Table 4 confirms that throughout Europe the addition of cloud filters and a sharper determination of thresholds result in SRCs closer to the reference data.

3.3. Performance of regional thresholds

The difference for the computed SRC at the LUCAS points between a set of regional thresholds versus a common, global threshold is illustrated in Fig. 5. This figure plots the performance of all existing composite spectra using the specified threshold versus the percentage of valid pixels in the bare surface composites. An invalid pixel means that less than five pixels in the temporal stack are labeled as bare surface, which is deemed insufficient for the composition (Dvorakova et al., 2023).

Ideally, a high performance (low angular distance between the LUCAS and the SRC spectra), close to the LUCAS reference data, and a high percentage of valid pixels, are desired (upper left corner). As shown in Fig. 5, the HISET surface thresholds (+) manage to compute composite spectra for over >82% of LUCAS points, while maintaining a high performance (an average angular distance of 0.058).

In contrast, all choices for a global threshold (x) either result in sparse SRCs (thresholds are excessively strict and exclude too many
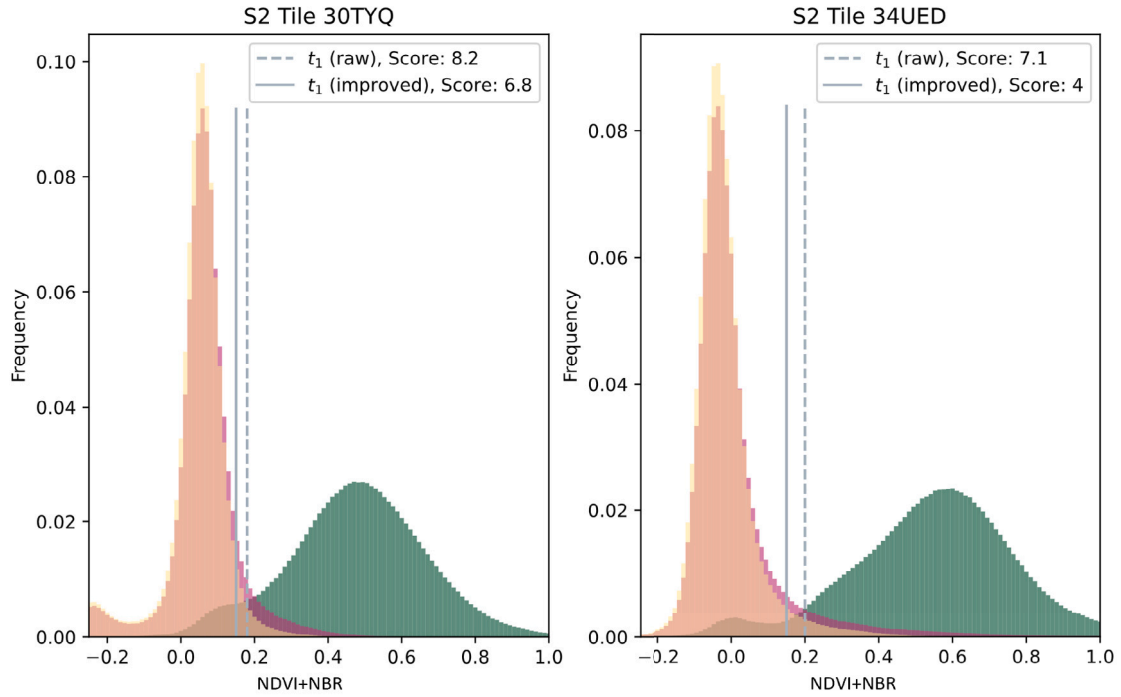
Fig. 4. Histograms of the spectral vegetation index NDVI+NBR in the Sentinel-2 tiles 30TYQ and 34UED for grassland (green) and croplands (orange) pixels (both normalized to an area of 1). The magenta region marks the pixels that have been removed by a refined landcover mask using the vegetation variability v, resulting in a sharper separation.
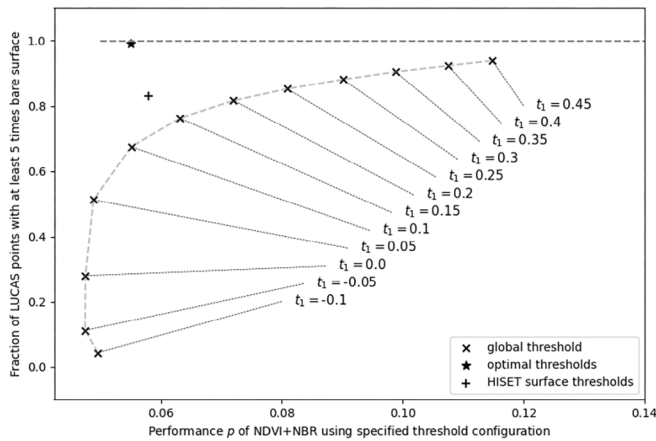

Fig. 5. Performance p of composite spectra using the regionalized surface HISET (marked with a plus (+)) versus a static global threshold (marked with an (x)) for the NDVI+NBR index. The star (*) indicates the theoretical upper limit, constrained by the condition to maintain at least 5 bare surface observations per SRC pixel. The fraction on the y-axis cannot exceed 1 (indicated by the horizontal dashed line). The dashed line connecting the (x) marker indicates that any choice for a global t_1 is further away from (*) than (+).

observations) or exhibit a low performance (larger values) to the reference, thus being less accurate.

The SRC using the pixel-based (optimal) thresholds, that are computed at each LUCAS point, are located at the star, representing the theoretical performance limit of the NDVI+NBR index. The star is closer to the HISET surface thresholds than any choice for a global threshold, indicating that HISET surface thresholds is always a better compromise than any globally fixed threshold.

In practice, comparing the performance of the regional HISET surface thresholds for the NDVI+NBR index with global thresholds found in the literature reveals a more significant improvement in quality

than initially suggested by Fig. 5. Table 5 illustrates, that soil spectra produced by our surface thresholds are substantially closer to the reference data, compared to the method proposed by Demattê et al. (2020) in row three.

This demonstrates that even with a well-chosen combination of indices, such as the case of the approach used in row three (the index performs only slightly worse than the NDVI+NBR, see Table 3), a global threshold can significantly impair results.

Broeg et al. (2024) (fourth line of Table 5) use a fixed, global threshold initially that is refined on a pixel-level in a second stage. While this work is designed for Germany, it was applied in this study at the European scale. Here, the performance is by a factor of 1.14 times lower for the European continent.

3.4. Threshold distribution for NDVI+NBR

Fig. 7 presents the resulting SRC from Sentinel-2 data between 2018 to 2022 displayed in true color (blue: B02, green: B03, red: B04). Pixels shown in white indicate areas with no data, where soil was not sufficiently often exposed during the observation period.

The interpolation process to display this large dataset may give the impression of a dense map at larger scales in some places like central Europe. However, upon zooming in, as highlighted in the window on the right, it becomes evident that the map is fragmented. This fragmentation is expected, as many surfaces, such as permanently vegetated areas or urban developments, do not expose bare, natural surfaces. Fig. 6 shows the distribution of HISET surface thresholds for the NDVI+NBR index in Europe. The map reveals a diverse and structured distribution of thresholds, spanning from −0.12 to 0.6.

4. Discussion

4.1. SRC evaluation method

The framework enables an automatic estimation of the performance of various processing parameters involved in the SRC generation

Table 5

Comparison of our bare surface reflectance composite generated from the HISET surface thresholds and NDVI+NBR index (second line) to two other choices found in literature: A global threshold (third line, Demattê et al., 2020) and a adapted threshold at pixel level (forth line, Broeg et al., 2024). The performance between the SRC created by the respective approach and the LUCAS reference is given in absolute figures (where lower distance is better) and relative to the SRC created by optimized NDVI+NBR thresholds.

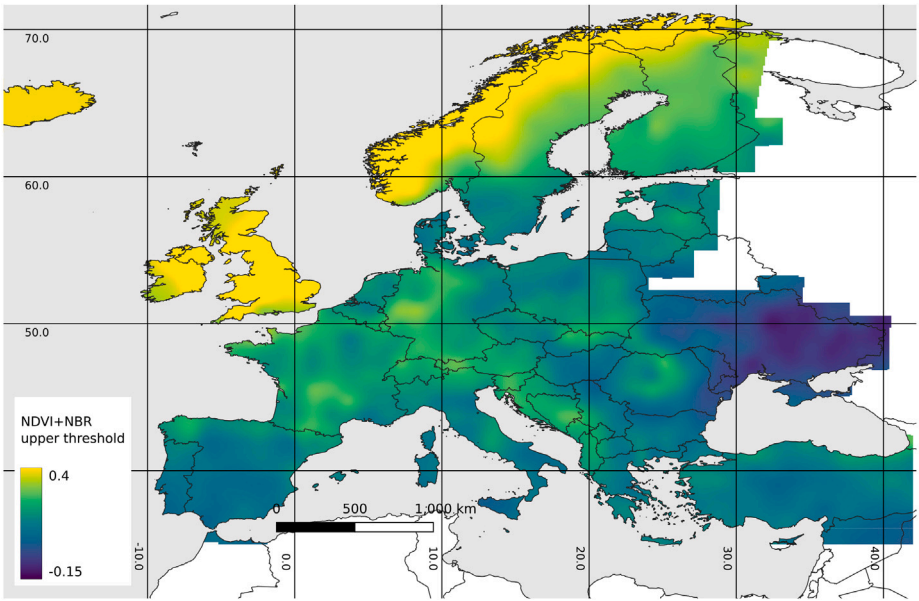| Approach	t scope	Performance p	
		Absolute	Relative
NDVI+NBR, optimal thresholds at LUCAS points	Pixel-level	0.055	1
NDVI+NBR, regional surface thresholds using HISET	Regional	0.058	1.07
$-0.25 <$ NDVI < 0.25, $-0.3 <$ NBR2 < 0.1, VNSIR < 0.9	Global	0.191	3.58
NDVI < 0.45, NBR2 < 0.16, BCC < 0.3; 0.4 percentile NBR2, median	Pixel-level	0.063	1.14

Fig. 6. HISET surface thresholds t_1 for the spectral vegetation index NDVI+NBR that separate bare surfaces and soils from surfaces covered by vegetation in Europe. Values range between +0.6 in England to −0.12 in Ukraine.
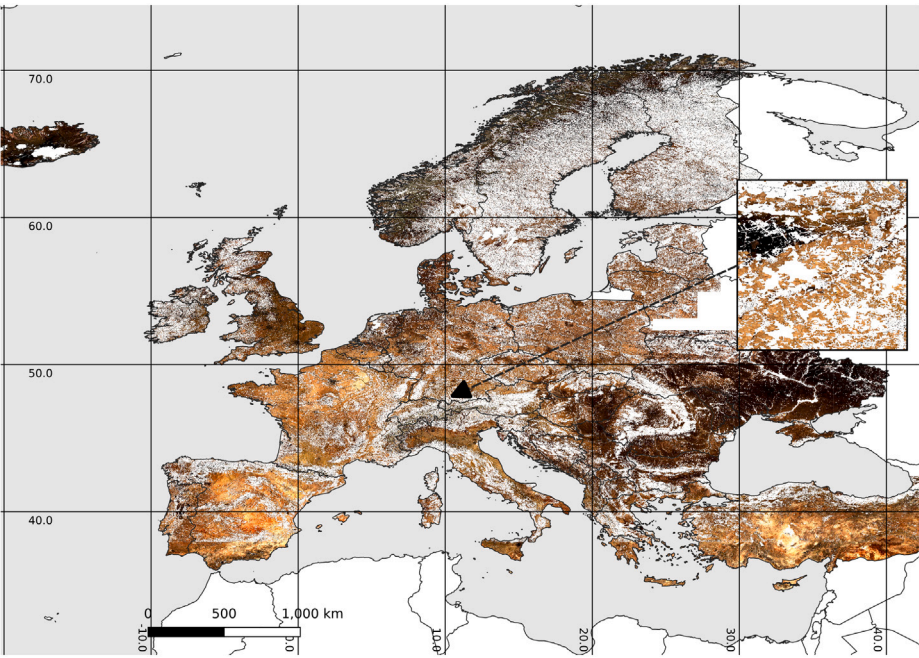

Fig. 7. SRC in true color (RGB: B04, B03, B02) with equal scaling (500–2000). The zoomed-in cutout highlights the local structure. It visually shows the diversity in spectral reflectance in Europe, ranging from dark carbon rich soils in Ukraine to soils rich in calcite in central France.

at reference points, such as the choice of the spectral index, additional subroutines like cloud detection and filtering, and concepts of thresholds (regional vs. global).

To the knowledge of the authors, such a systematic framework is novel. It is objective in nature and allows evaluating the choice of processing parameters at reference points before the actual large-scale computation that generates the entire SRC map. The framework provides a theoretical upper limit to the achievable performance based on the chosen spectral index, thus offering an estimate of the expected product quality and facilitating a comparison between successive releases or SRCs developed using different methods.

The optimization approach employed in this study is based on a guided descent along partial gradients, chosen for its simplicity and decent performance for low-dimensional problems. In theory, gradient-based methods yield robust solutions when the objective function is smooth (differentiable) and convex (Schmidt and Roux, 2013). However, the SRC generation problem presents challenges in this regard: it is inherently non-smooth due to the binary decision of including or excluding spectral observations. To mitigate this, a soft exclusion strategy was implemented, weighing the contribution of each spectrum (Eq. (12)). However, this complicates an analytic solution of the partial derivatives, which, in result, have to be estimated numerically by finite differences.

While global convexity cannot be guaranteed, empirical tests suggest that the problem is locally convex in the vicinity of optimal solutions. This was confirmed by initializing the algorithm with multiple perturbed initial guesses that converge to similar solutions (Table 3).

Here, an alternative implementation using adaptively refined sparse grids was considered, that would not require a smooth, differentiable problem. But, this would come at the cost of slower convergence (and more evaluations of the object function) and; thus, was not used in the final version (Bungartz and Griebel, 2004).

Nevertheless, the method is sensitive to tuning of meta-parameters of the object function to archive convergence such as step size, termination criteria and the finite difference distances (Eq. (15)). For instance, the smoothing factor β has to maintain smooth transition with a monotonous derivative (convexity) while preserving the general shape of the object function.

Multiple restarts of the optimization loop and the large volume of Sentinel-2 L2A data at the LUCAS points, make the problem computationally expensive. To address this, data was structured in a relational database with parallel I/O and deployed on a high-performance parallel file system. Furthermore, the computation can be easily run in parallel, as individual pixels are spatially independent, enabling the distribution of tasks on multiple processing nodes.

A fundamental requirement for applying the method is the availability of a reference dataset that adequately represents the spectral variability of bare surfaces. Ideally, the reference data should be temporally aligned with the remote sensing observations to minimize discrepancies caused by changes in land management or surface evolution. The used LUCAS samples were collected in 2015, while the Sentinel-2 time series spans 2018–2022. Although this is not ideal, the study by Fernández-Ugalde et al. (2020) has indicated that observable soil property changes typically require periods of at least six years. Moreover, no alternative harmonized topsoil dataset of comparable scale is currently available. Despite the temporal offset, the LUCAS dataset remains a robust choice due to its demonstrated ability to represent spatial variability in European surfaces. For instance, Ballabio et al. (2016) and de Brogniez et al. (2015) successfully used LUCAS data to model topsoil physical properties and organic carbon maps, respectively.

A second key requirement is the availability of a sufficiently long temporal archive of remote sensing observations to increase the likelihood of capturing bare surface conditions and to stabilize SRC spectra and to have meaningful gradients. This is relevant for the optimization

technique, as the gradual changes towards the reference spectra requires a large pool of observations. In practice, the five-year S2 archive used in this study is sufficient, where LUCAS sites have at least 50, oftentimes more than 100 spectra.

The method presumes that cosine similarity is a suitable, monotonous measure for spectral similarity, particularly for largely similar spectra. Despite the widespread use of the spectral angle or cosine similarity for clustering spectra in the literature, it might not be the optimal choice to find nuanced differences in largely similar spectra. An alternative approach worth exploring might involve measuring the distance from the entire distribution S_{bare} to the LUCAS ground reference. This method would consider information otherwise lost during the stage of the temporal averaging, potentially offering a more comprehensive measure.

4.2. Quality of the bare surface reflectance composites

The impact of a defined processing parameter set on the quality of the generated composite spectra was evaluated based on the average angular spectral distance (performance). It could be shown that, in general, four main aspects impact the quality of an SRC, (1) the chosen spectral index, (2) the usage of a well-calibrated regional set of threshold as opposed to a single, globally used threshold, (3) the trade-off between spatial coverage to spectral fidelity, and (4) a cloud filter, specific to bare surface.

Table 3 summarizes the optimal performance of each spectral index at all LUCAS points. It highlights the important role of the selection of spectral index. The NDVI that separates photosynthetically active from photosynthetically non-active surfaces is very important but shows limited performance when taken alone. Incorporating short-wave infrared information (e.g., NBR, NBR2) significantly enhances performance. Therefore, other authors have used these indices in sequence. The best results are achieved using either multiple indices (e.g., NDVI/NBR2/VNSIR, NDVI/NBR2/BCC, NDVI/NBR) or combined indices (e.g. NDVI+NBR). Among these, NDVI+NBR stands out as both effective and practical, while offering strong performance with a simple one-dimensional threshold set. This is especially an advantage for operational processing because it reduces the effort to define additional thresholds.

Secondly, complementing the index with a regional threshold dataset significantly improved the accuracy of the SRC, as demonstrated in Fig. 5. This enhancement is as critical as the choice of the index itself. This is evident from the comparison in Table 5, where a well-suited index (NDVI/NBR2/VNSIR) combined with a globally static threshold was outperformed by NDVI+NBR in conjunction with the HISET surface thresholds.

Third, for the effective application of SRCs in spectral soil analysis, a trade-off must be maintained between threshold stringency and spatial coverage. As shown in Table 4 and Fig. 4, lower threshold values refine pixel selection but reduce coverage, affecting both the overall SRC extent (Fig. 5) and the number of valid bare surface observations per SRC pixel. Using HISET surface thresholds, the SRC bare surface pixels over Europe are composed by on average 10.2 distinct S2 observations. For spectral reliability, at least 7–10 contributing pixels are recommended (Dvorakova et al., 2023; Heiden et al., 2022). This limitation is particularly relevant in Mediterranean regions, where the pixels are often characterized by a spectral mixture between green and dry vegetation and soils (e.g. tree crops). When thresholds are too strict, this reduces the bare surface coverage substantially.

Fourth, although the overall impact on SRC quality was minor, the treatment of cloud contamination warrants discussion. The SRC generation process includes several steps to ensure cloud-free spectral composites, beginning with input selection (scenes with less than 80% cloud cover) and application of the L2A MAJA cloud mask (MG2). Baetens et al. (2019) has shown that MAJA detects clouds with slightly higher
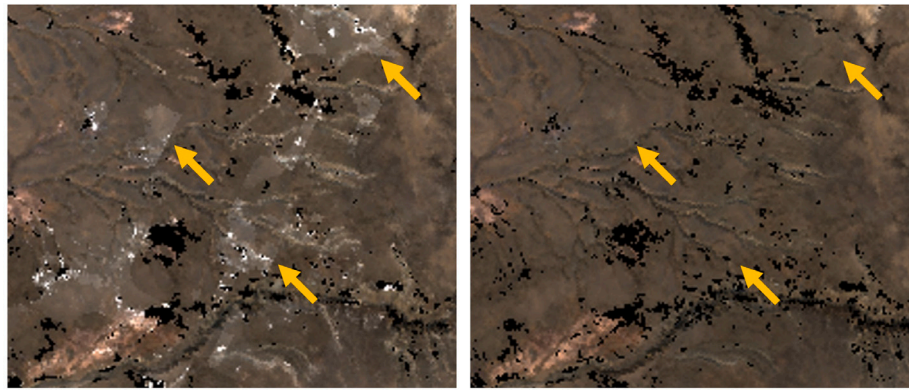
Fig. 8. Illustration of the benefit of an outlier-based cloud filter (here in central Iceland). Left: SRC without the filter, showing sharp edges from MAJA cloud masks in the composite. Right: With the filter applied, these artifacts are removed.

accuracy compared to Sen2Cor and Fmask (90%). While the quantitative impact of the additional cloud filters specific to bare surface at the European scale was limited (Table 4) — likely due to the localized nature of cloud artifacts — it proved effective in reducing visually disruptive contamination (Fig. 8) without being overly restrictive.

Finally, the HISET method for deriving surface thresholds could be compared to other threshold derivation concepts. The HISET surface thresholds demonstrated performance at least equal to the other method highlighted in Table 5. The HISET surface thresholds have an advantage over methods that fix thresholds to a defined percentile (Broeg et al., 2024) because it performs well regardless of the actual fraction of bare surface observations in the temporal stack of a pixel. For instance, in a stack composed solely of bare pixels, choosing a fixed percentile would result in an excessively strict threshold, ultimately excluding many valid observations. On the other hand, HISET relies on an external land cover (ESA WorldCover) data set that may introduce additional uncertainties and dependencies. In this study it was necessary to refine it to yield good performance of spectra generated using the HISET surface thresholds (Table 4).

Based on the pixel-specific surface NDVI+NBR thresholds (Fig. 6) and the soil-specific cloud handling, a European SRC has been generated based on the mean of all valid bare surface pixels (Fig. 7). It shows a diversity of bare surface reflectance spectra across the continent and illuminates the European continent from a novel perspective, offering valuable insights into soil characteristics. It is already used as the foundation for future research within the domain (van Wesemael et al., 2024).

5. Conclusion and outlook

The work developed and tested a novel strategy to evaluate the process of generating bare surface reflectance composites at large scale — in this case for the European continent. The evaluation method was specifically designed to overcome the common yet problematic practice of assessing the quality of large-scale maps based on small, selectively chosen, and potentially biased highlight areas. By replacing subjective human judgment with a systematic and reproducible evaluation framework, the approach ensures a more objective and robust assessment of the SRC generation process and, thereby significantly increasing confidence in the processor's performance.

By using guided optimization, it was demonstrated that the spectral index NDVI+NBR, complemented with a regional threshold dataset, performs well in Europe. Additionally, they showed that it is beneficial to build upon available cloud masks by using additional soil-specific filters to reduce the effects of residual clouds and haze. HISET to derive surface thresholds is a suitable and effective method, when the input data is cleaned and cases of too few grassland or cropland pixels are dealt with.

Although the NDVI+NBR index has been validated for its current utility, it may not represent the ideal choice for other regions than Europe. Expanding the optimization problem to include a search for other spectral indices could potentially uncover more adept indices for specific regions that may increase the performance.

Future SRC generating methods will leverage hyperspectral data to enable a more refined selection of bare surface pixels. While current hyperspectral archives lack the temporal density required to generate a SRC directly for large-scale applications (Chabrillat et al., 2024), other options are feasible. In particular, current research explores training machine learning models on paired Sentinel-2 multi-spectral data and fractional vegetation cover vectors (photosynthetic, non-photosynthetic vegetation, bare surfaces and soils) derived from near-coincident EnMAP acquisitions. Initial results, as presented by Schwind et al. (2024) are promising.

This framework may serve as important instruments to determine the quality of bare surface composites as input to soil property mapping tasks. Therefore, they could contribute to the monitoring of the soil health using satellite data to support the ambitious goal of achieving healthy soils in Europe by 2050 (Pieper et al., 2023).

CRediT authorship contribution statement

Paul Karlshoefer: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Pablo d'Angelo:** Writing – review & editing, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jonas Eberle:** Software, Data curation. **Uta Heiden:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Data availability

The final SRC and other derived products (e.g. standard deviation of bare surface pixels per pixel, count of valid and bare surface observations) named `SoilSuite` have been made publicly available under the license CC-BY-4.0 (SoilSuite, 2024).

References

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of copernicus sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. Remote. Sens. (ISSN: 2072-4292) 11 (4), http://dx.doi.org/10.3390/rs11040433, URL https://www.mdpi.com/2072-4292/11/4/433.

Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma (ISSN: 0016-7061) 261, 110–123. http://dx.doi.org/10.1016/j.geoderma.2015.07.006, URL https://www.sciencedirect.com/science/article/pii/S0016706115300173.

Ben-Dor, E., 1994. A precaution regarding cirrus cloud detection from airborne imaging spectrometer data using the 1.38 ym water vapor band. Remote Sens. Environ. 50, 346–350.

Ben-Dor, E., Chabrillat, S., Demattê, J., Taylor, G., Hill, J., Whiting, M., Sommer, S., 2009. Using imaging spectroscopy to study soil properties. Remote Sens. Environ. (ISSN: 0034-4257) 113, S38–S55. http://dx.doi.org/10.1016/j.rse.2008.09.019, URL https://www.sciencedirect.com/science/article/pii/S0034425709000753. Imaging Spectroscopy Special Issue.

Broeg, T., Don, A., Gocht, A., Scholten, T., Taghizadeh-Mehrjardi, R., Erasmi, S., 2024. Using local ensemble models and landsat bare soil composites for large-scale soil organic carbon maps in cropland. Geoderma (ISSN: 0016-7061) 444, 116850. http://dx.doi.org/10.1016/j.geoderma.2024.116850, URL https://www.sciencedirect.com/science/article/pii/S001670612400079X.

Bungartz, H.-J., Griebel, M., 2004. Sparse grids. Acta Numer. 13, 147–269. http://dx.doi.org/10.1017/S0962492904000182.

Castaldi, F., Chabrillat, S., Chartin, C., Genot, V., Jones, A.R., van Wesemael, B., 2018. Estimation of soil organic carbon in arable soil in Belgium and Luxembourg with the LUCAS topsoil database. Eur. J. Soil Sci. (ISSN: 1365-2389) 69 (4), 592–603. http://dx.doi.org/10.1111/ejss.12553.

Castaldi, F., Chabrillat, S., Don, A., van Wesemael, B., 2019. Soil organic carbon mapping using LUCAS topsoil database and sentinel-2 data: An approach to reduce soil moisture and crop residue effects. Remote. Sens. (ISSN: 2072-4292) 11 (18), 2121. http://dx.doi.org/10.3390/rs11182121.

Castaldi, F., Halil Koparan, M., Wetterlind, J., Žydelis, R., Vinci, I., Özge Savaş, A., Kıvrak, C., Tunçay, T., Volungevičius, J., Obber, S., Ragazzi, F., Malo, D., Vaudour, E., 2023. Assessing the capability of sentinel-2 time-series to estimate soil organic carbon and clay content at local scale in croplands. ISPRS J. Photogramm. Remote Sens. (ISSN: 0924-2716) 199, 40–60. http://dx.doi.org/10.1016/j.isprsjprs.2023.03.016, URL https://www.sciencedirect.com/science/article/pii/S0924271623000771.

Chabrillat, S., Foerster, S., Segl, K., Beamish, A., Brell, M., Asadzadeh, S., Milewski, R., Ward, K.J., Brosinsky, A., Koch, K., Scheffler, D., Guillaso, S., Kokhanovsky, A., Roessner, S., Guanter, L., Kaufmann, H., Pinnel, N., Carmona, E., Storch, T., Hank, T., Berger, K., Wocher, M., Hostert, P., van der Linden, S., Okujeni, A., Janz, A., Jakimow, B., Bracher, A., Soppa, M.A., Alvarado, L.M., Buddenbaum, H., Heim, B., Heiden, U., Moreno, J., Ong, C., Bohn, N., Green, R.O., Bachmann, M., Kokaly, R., Schodlok, M., Painter, T.H., Gascon, F., Buongiorno, F., Mottus, M., Brando, V.E., Feilhauer, H., Betz, M., Baur, S., Feckl, R., Schickling, A., Krieger, V., Bock, M., La Porta, L., Fischer, S., 2024. The EnMAP spaceborne imaging spectroscopy mission: Initial scientific results two years after launch. Remote Sens. Environ. (ISSN: 0034-4257) 315, 114379. http://dx.doi.org/10.1016/j.rse.2024.114379, URL https://www.sciencedirect.com/science/article/pii/S003442572400405X.

Chabrillat, S., Goetz, A.F., Krosley, L., Olsen, H.W., 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. Remote Sens. Environ. (ISSN: 0034-4257) 82 (2), 431–445. http://dx.doi.org/10.1016/S0034-4257(02)00060-3, URL https://www.sciencedirect.com/science/article/pii/S0034425702000603.

Coops, N.C., Kearney, S.P., Bolton, D.K., Radeloff, V.C., 2018. Remotely-sensed productivity clusters capture global biodiversity patterns. Sci. Rep. (ISSN: 2045-2322) 8 (1), http://dx.doi.org/10.1038/s41598-018-34162-8.

Copernicus, E., 2020. MultiSpectral instrument (MSI) overview. https://web.archive.org/web/20201017053209/https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument. (Accessed 15 April 2025).

Corbane, C., Politis, P., Kempeneers, P., Simonetti, D., Soille, P., Burger, A., Pesaresi, M., Sabo, F., Syrris, V., Kemper, T., 2020. A global cloud free pixel- based image composite from sentinel-2 data. Data Brief (ISSN: 2352-3409) 31, 105737. http://dx.doi.org/10.1016/j.dib.2020.105737, URL https://www.sciencedirect.com/science/article/pii/S2352340920306314.

Daughtry, C., Hunt, E., 2008. Mitigating the effects of soil and residue water contents on remotely sensed estimates of crop residue cover. Remote Sens. Environ. (ISSN: 0034-4257) 112 (4), 1647–1657. http://dx.doi.org/10.1016/j.rse.2007.08.006.

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.A., Montanarella, L., van Wesemael, B., 2015. A map of the topsoil organic carbon content of europe generated by a generalized additive model. Eur. J. Soil Sci. 66 (1), 121–134. http://dx.doi.org/10.1111/ejss.12193, URL https://bsssjournals.onlinelibrary.wiley.com/doi/abs/10.1111/ejss.12193. arXiv:https://bsssjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12193.

Demattê, J.A.M., Fongaro, C.T., Rizzo, R., Safanelli, J.L., 2018. Geospatial soil sensing system (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sens. Environ. (ISSN: 0034-4257) 212, 161–175. http://dx.doi.org/10.1016/j.rse.2018.04.047, URL https://www.sciencedirect.com/science/article/pii/S0034425718302049.

Demattê, J.A., Safanelli, J.L., Poppiel, R.R., Rizzo, R., Silvero, N.E.Q., Mendes, W.d., Bonfatti, B.R., Dotto, A.C., Salazar, D.F.U., Mello, F.A.d., da Silveira Paiva, A.F., Souza, A.B., dos Santos, N.V., Nascimento, C.M., de Mello, D.C., Bellinaso, H., Neto, L.G., Amorim, M.T.A., de Resende, M.E.B., da Souza Vieira, J., de Queiroz, L.G., Gallo, B.C., Sayão, V.M., da Silva Lisboa, C.J., 2020. Bare earth's surface spectra as a proxy for soil resource monitoring. Sci. Rep. (ISSN: 2045-2322) 10, http://dx.doi.org/10.1038/s41598-020-61408-1.

Dennison, P.E., Lamb, B.T., Campbell, M.J., Kokaly, R.F., Hively, W.D., Vermote, E., Dabney, P., Serbin, G., Quemada, M., Daughtry, C.S., Masek, J., Wu, Z., 2023. Modeling global indices for estimating non-photosynthetic vegetation cover. Remote Sens. Environ. (ISSN: 0034-4257) 295, 113715. http://dx.doi.org/10.1016/j.rse.2023.113715.

Dennison, P.E., Qi, Y., Meerdink, S.K., Kokaly, R.F., Thompson, D.R., Daughtry, C.S.T., Quemada, M., Roberts, D.A., Gader, P.D., Wetherley, E.B., Numata, I., Roth, K.L., 2019. Comparison of methods for modeling fractional cover using simulated satellite hyperspectral imager spectra. Remote. Sens. (ISSN: 2072-4292) 11 (18), 2072. http://dx.doi.org/10.3390/rs11182072.

Deventer, v., Ward, A., Gowda, P., Lyon, J., 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. Photogramm. Eng. Remote Sens. 63, 87–93.

Diek, S., Fornallaz, F., Schaepman, M.E., De Jong, R., 2017. Barest pixel composite for agricultural areas using landsat time series. Remote. Sens. (ISSN: 2072-4292) 9 (12), http://dx.doi.org/10.3390/rs9121245, URL https://www.mdpi.com/2072-4292/9/12/1245.

Dvorakova, K., Heiden, U., Pepers, K., Staats, G., van Os, G., van Wesemael, B., 2023. Improving soil organic carbon predictions from a sentinel-2 soil composite by assessing surface conditions and uncertainties. Geoderma (ISSN: 0016-7061) 429, 116128. http://dx.doi.org/10.1016/j.geoderma.2022.116128, URL https://www.sciencedirect.com/science/article/pii/S0016706122004359.

Dvorakova, K., Heiden, U., van Wesemael, B., 2021. Sentinel-2 exposed soil composite for soil organic carbon prediction. Remote. Sens. (ISSN: 2072-4292) 13 (9), http://dx.doi.org/10.3390/rs13091791, URL https://www.mdpi.com/2072-4292/13/9/1791.

EEA, 2016. Biogeographical regions Europe. URL https://www.eea.europa.eu/en/datahub/datahubitem-view/11db8d14-f167-4cd5-9205-95638dfd9618.

ESA, 2021. Level-2A algorithm theoretical basis document. X.

European Space Agency (ESA), 2015. Sentinel-2 user handbook. URL https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook.

Fernández-Ugalde, O., Ballabio, C., Lugato, E., Scarpa, S., Jones, A., 2020. Assessment of changes in topsoil properties in LUCAS samples between 2009/2012 and 2015 surveys. Publ. Off. Eur. Union http://dx.doi.org/10.2760/5503.

Gallo, B.C., Demattê, J.A.M., Rizzo, R., Safanelli, J.L., Mendes, W.D.S., Lepsch, I.F., Sato, M.V., Romero, D.J., Lacerda, M.P.C., 2018. Multi-temporal satellite images on topsoil attribute quantification and the relationship with soil classes and geology. Remote. Sens. (ISSN: 2072-4292) 10 (10), http://dx.doi.org/10.3390/rs10101571, URL https://www.mdpi.com/2072-4292/10/1/571.

García, M.L., Caselles, V., 1991. Mapping burns and natural reforestation using thematic mapper data. Geocarto Int. 6 (1), 31–37. http://dx.doi.org/10.1080/10106049109354290.

Gasmi, A., Gomez, C., Lagacherie, P., Zouari, H., Laamrani, A., Chehbouni, A., 2021. Mean spectral reflectance from bare soil pixels along a landsat-TM time series to increase both the prediction accuracy of soil clay content and mapping coverage. Geoderma (ISSN: 0016-7061) 388, 114864. http://dx.doi.org/10.1016/j.geoderma.2020.114864, URL https://www.sciencedirect.com/science/article/pii/S0016706120326197.

Gerighausen, H., Menz, G., Kaufmann, H., 2012. Spatially explicit estimation of clay and organic carbon content in agricultural soils using multi-annual imaging spectroscopy data. Appl. Environ. Soil Sci. 2012, 868090. http://dx.doi.org/10.1155/2012/868090, URL https://www.hindawi.com/journals/aess/2012/868090/.

Gillespie, A.R., Kahle, A.B., Walker, R.E., 1987. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. Remote Sens. Environ. (ISSN: 0034-4257) 22 (3), 343–365. http://dx.doi.org/10.1016/0034-4257(87)90088-5, URL https://www.sciencedirect.com/science/article/pii/0034425787900885.

Hagolle, O., Colin, J., Coustance, S., Kettig, P., D'Angelo, P., Auer, S., Doxani, G., Desjardins, C., 2021. Sentinel-2 surface reflectance products generated by CNES and DLR: Methods, validation and applications. ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci. V-1-2021, 9–15. http://dx.doi.org/10.5194/isprs-annals-V-1-2021-9-2021, URL https://isprs-annals.copernicus.org/articles/V-1-2021/9/2021/.

Hagolle, O., Huc, M., Desjardins, C., Auer, S., Richter, R., 2018. MAJA algorithm theoretical basis document. http://dx.doi.org/10.5281/zenodo.1209633.

Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, venus, LANDSAT and SENTINEL-2 images. Remote Sens. Environ. (ISSN: 0034-4257) 114 (8), 1747–1755. http://dx.doi.org/10.1016/j.rse.2010.03.002.

Heiden, U., d'Angelo, P., Schwind, P., Karlshöfer, P., Müller, R., Zepp, S., Wiesmeier, M., Reinartz, P., 2022. Soil reflectance composites—Improved thresholding and performance evaluation. Remote. Sens. (ISSN: 2072-4292) 14, http://dx.doi.org/10.3390/rs14184526.

Kempeneers, P., Soille, P., 2017. Optimizing sentinel-2 image selection in a big data context. Big Earth Data 1 (1–2), 145–158. http://dx.doi.org/10.1080/20964471.2017.1407489, arXiv:https://doi.org/10.1080/20964471.2017.1407489.

Ku, N.-W., Popescu, S., Eriksson, M., 2021. Regionalization of an existing global forest canopy height model for forests of the southern United States. Remote. Sens. (ISSN: 2072-4292) 13 (9), http://dx.doi.org/10.3390/rs13091722, URL https://www.mdpi.com/2072-4292/13/9/1722.

Meyers, J., 2015. Sentinel-2 UTM tiling grid (ESA). https://github.com/justinelliotmeyers/Sentinel-2-Shapefile-Index.

Minasny, B., McBratney, A., 2016. Digital soil mapping: A brief history and some lessons. Geoderma (ISSN: 0016-7061) 264, 301–311. http://dx.doi.org/10.1016/j.geoderma.2015.07.017, URL https://www.sciencedirect.com/science/article/pii/S0016706115300276. Soil mapping, classification, and modelling: history and future directions.

Montanarella, L., Badraoui, M., Chude, V., Baptista, I., Mamo, T., Yemefack, M., Aulakh, M., Yagi, K., Hong, S.-Y., Vijarnsorn, P., Zhang, G.-L., Arrouays, D., Black, H., Krasilnikov, P., Sobocká, J., Alegre, J., Henríquez, C., Mendonça Santos, M., Taboada, M., McKenzie, N., 2015. The Status of the World's Soil Resources (Main Report). UN, URL http://www.fao.org/3/a-i5199e.pdf.

Mulder, V., de Bruin, S., Schaepman, M., Mayr, T., 2011. The use of remote sensing in soil and terrain mapping - A review. Geoderma (ISSN: 0016-7061) 162 (1), 1–19. http://dx.doi.org/10.1016/j.geoderma.2010.12.018, URL https://www.sciencedirect.com/science/article/pii/S0016706110003976.

Mzid, N., Castaldi, F., Tolomio, M., Pascucci, S., Casa, R., Pignatti, S., 2022. Evaluation of agricultural bare soil properties retrieval from landsat 8, sentinel-2 and PRISMA satellite data. Remote. Sens. (ISSN: 2072-4292) 14 (3), 714. http://dx.doi.org/10.3390/rs14030714.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. In: Advances in Agronomy. Elsevier, pp. 139–159. http://dx.doi.org/10.1016/bs.agron.2015.02.002.

Odusanya, A.E., Schulz, K., Mehdi-Schulz, B., 2022. Using a regionalisation approach to evaluate streamflow simulated by an ecohydrological model calibrated with global land surface evaporation from remote sensing. J. Hydrol.: Reg. Stud. (ISSN: 2214-5818) 40, 101042. http://dx.doi.org/10.1016/j.ejrh.2022.101042, URL https://www.sciencedirect.com/science/article/pii/S2214581822000556.

Okujeni, A., Kowalski, K., Lewińska, K.E., Schneidereit, S., Hostert, P., 2024. Multidecadal grassland fractional cover time series retrieval for Germany from the landsat and sentinel-2 archives. Remote. Sens. Environ. (ISSN: 0034-4257) 302, 113980. http://dx.doi.org/10.1016/j.rse.2023.113980, URL https://www.sciencedirect.com/science/article/pii/S0034425723005321.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2017. LUCAS soil, the largest expandable soil dataset for Europe: a review. Eur. J. Soil Sci. (ISSN: 1365-2389) 69 (1), 140–153. http://dx.doi.org/10.1111/ejss.12499.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernandez-Ugalde, O., 2018. LUCAS soil, the largest expandable soil dataset for Europe: a review. Eur. J. Soil Sci. 69 (1), 140–153. http://dx.doi.org/10.1111/ejss.12499, URL https://bsssjournals.onlinelibrary.wiley.com/doi/abs/10.1111/ejss.12499, arXiv:https://bsssjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12499.

Pieper, S., Frauenstein, J., Ginsky, H., Glante, F., Grimski, D., Kotschik, P., Marx, K., 2023. The upcoming European soil health law -chances and challenges for an effective soil protection. Umweltbundesamt.

Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7 (1), 217–240. http://dx.doi.org/10.5194/soil-7-217-2021, URL https://soil.copernicus.org/articles/7/217/2021/.

Rikimaru, A., Roy, P., Miyatake, S., et al., 2002. Tropical forest cover density mapping. Trop. Ecol. 43 (1), 39–47.

Rizzo, R., Wadoux, A.M.-C., Demattê, J.A., Minasny, B., Barrãn, V., Ben-Dor, E., Francos, N., Savin, I., Poppiel, R., Silvero, N.E., da Silva Terra, F., Rosin, N.A., Rosas, J.T.F., Greschuk, L.T., Ballester, M.V., Gãmez, A.M.R., Belllinaso, H., Safanelli, J.L., Chabrillat, S., Fiorio, P.R., Das, B.S., Malone, B.P., Zalidis, G.,

Tziolas, N., Tsakiridis, N., Karyotis, K., Samarinas, N., Kalopesa, E., Gholizadeh, A., Shepherd, K.D., Milewski, R., Vaudour, E., Wang, C., Salama, E.S.M., 2023. Remote sensing of the earth's soil color in space and time. Remote. Sens. Environ. (ISSN: 0034-4257) 299, 113845. http://dx.doi.org/10.1016/j.rse.2023.113845, URL https://www.sciencedirect.com/science/article/pii/S0034425723003966.

Roberts, D., Wilford, J., Ghattas, O., 2019. Exposed soil and mineral map of the Australian continent revealing the land at its barest. Nat. Commun. (ISSN: 2041-1723) 10 (1), http://dx.doi.org/10.1038/s41467-019-13276-1.

Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., Heiden, U., 2018. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with landsat imagery (1984–2014). Remote. Sens. Environ. (ISSN: 0034-4257) 205, 1–17. http://dx.doi.org/10.1016/j.rse.2017.11.004, URL https://www.sciencedirect.com/science/article/pii/S003442571730514X.

Rouse, J., Haas, R., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the great plains with ERTS. In: Freden, S.C., Mercanti, E.P., Becker, M. (Eds.), Third Earth Resources Technology Satellite–1 Syposium. Volume I: Technical Presentations. NASASP-351, NASA, Washington, D.C., pp. 309–317.

Safanelli, J.L., Chabrillat, S., Ben-Dor, E., Demattê, J.A.M., 2020. Multispectral models from bare soil composites for mapping topsoil properties over europe. Remote. Sens. (ISSN: 2072-4292) 12 (9), 102277. http://dx.doi.org/10.3390/rs12091369, URL https://www.mdpi.com/2072-4292/12/9/1369.

Schmidt, M., Roux, N.L., 2013. Fast convergence of stochastic gradient descent under a strong growth condition. arXiv:1308.6370. URL https://arxiv.org/abs/1308.6370.

Schwind, P., Kühl, K., Marshall Ingram, D., Bachmann, M., Heiden, U., 2024. Using deep learning to generate fractional vegetation cover from multispectral data. In: 13th EARSeL Workshop on Imaging Spectroscopy.

Serfling, R., 2011. Asymptotic relative efficiency in estimation. In: Encyclopedia of Statistical Sciences. Springer, p. 1. http://dx.doi.org/10.1007/978-3-642-04898-2_126, URL https://www.researchgate.net/publication/260387595_Asymptotic_Relative_Efficiency_in_Estimation.

Simonetti, D., Pimple, U., Langner, A., Marelli, A., 2021. Pan-tropical sentinel-2 cloud-free annual composite datasets. Data Brief (ISSN: 2352-3409) 39, 107488. http://dx.doi.org/10.1016/j.dib.2021.107488, URL https://www.sciencedirect.com/science/article/pii/S2352340921007691.

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J.-C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022. Cloud mask intercomparison exercise (CMIX): An evaluation of cloud masking algorithms for landsat 8 and sentinel-2. Remote. Sens. Environ. (ISSN: 0034-4257) 274, 112990. http://dx.doi.org/10.1016/j.rse.2022.112990, URL https://www.sciencedirect.com/science/article/pii/S0034425722001043.

SoilSuite, 2024. Entinel-2 5-year (2018–2022) composites at European scale. http://dx.doi.org/10.15489/qkud8cudg596, URL https://geoservice.dlr.de/web/datasets/soilsuite_eur_5y.

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020a. Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. Remote. Sens. (ISSN: 2072-4292) 12 (9), http://dx.doi.org/10.3390/rs12091389, URL https://www.mdpi.com/2072-4292/12/9/1389.

Tziolas, N., Tsakiridis, N., Heiden, U., van Wesemael, B., 2024. Soil organic carbon mapping utilizing convolutional neural networks and earth observation data, a case study in Bavaria state Germany. Geoderma (ISSN: 0016-7061) 444, 116867. http://dx.doi.org/10.1016/j.geoderma.2024.116867, URL https://www.sciencedirect.com/science/article/pii/S001670612400096X.

Tziolas, N., Tsakiridis, N., Ogen, Y., Kalopesa, E., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020b. An integrated methodology using open soil spectral libraries and earth observation data for soil organic carbon estimations in support of soil-related SDGs. Remote. Sens. Environ. (ISSN: 0034-4257) 244, 111793. http://dx.doi.org/10.1016/j.rse.2020.111793.

Urbina-Salazar, D., Vaudour, E., Baghdadi, N., Ceschia, E., Richer-de-Forges, A.C., Lehmann, S., Arrouays, D., 2021. Using sentinel-2 images for soil organic carbon content mapping in croplands of southwestern France. The usefulness of sentinel-1/2 derived moisture maps and mismatches between sentinel images and sampling dates. Remote. Sens. (ISSN: 2072-4292) 13 (24), http://dx.doi.org/10.3390/rs13245115, URL https://www.mdpi.com/2072-4292/13/24/5115.

van der Meer, F., 2006. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. Int. J. Appl. Earth Obs. Geoinf. (ISSN: 1872826X) 8, 3–17. http://dx.doi.org/10.1016/j.jag.2005.06.001.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., Yu, T., the scikit-image contributors, 2014. Scikit-image: image processing in python. PeerJ (ISSN: 2167-8359) 2, e453. http://dx.doi.org/10.7717/peerj.453.

van Wesemael, B., Abdelbaki, A., Ben-Dor, E., Chabrillat, S., d'Angelo, P., Demattê, J.A., Genova, G., Gholizadeh, A., Heiden, U., Karlshoefer, P., Milewski, R., Poggio, L., Sabetizade, M., Sanz, A., Schwind, P., Tsakiridis, N., Tziolas, N., Yagüe, J., Žížala, D., 2024. A European soil organic carbon monitoring system leveraging sentinel 2 imagery and the LUCAS soil data base. Geoderma (ISSN: 0016-7061) 452, 117113. http://dx.doi.org/10.1016/j.geoderma.2024.117113, URL https://www.sciencedirect.com/science/article/pii/S0016706124003422.

Vaudour, E., Gomez, C., Lagacherie, P., Loiseau, T., Baghdadi, N., Urbina-Salazar, D., Loubet, B., Arrouays, D., 2021. Temporal mosaicking approaches of sentinel-2 images for extending topsoil organic carbon content mapping in croplands. Int. J. Appl. Earth Obs. Geoinf. (ISSN: 1569-8432) 96, 102277. http://dx.doi.org/10.1016/j.jag.2020.102277, URL https://www.sciencedirect.com/science/article/pii/S030324342030920X.

Ward, K.J., Chabrillat, S., Brell, M., Castaldi, F., Spengler, D., Foerster, S., 2020. Mapping soil organic carbon for airborne and simulated EnMAP imagery using the LUCAS soil database and a local PLSR. Remote. Sens. (ISSN: 2072-4292) 12 (20), http://dx.doi.org/10.3390/rs12203451, URL https://www.mdpi.com/2072-4292/12/20/3451.

Ying, Q., Hansen, M.C., Potapov, P.V., Tyukavina, A., Wang, L., Stehman, S.V., Moore, R., Hancher, M., 2017. Global bare ground gain from 2000 to 2012 using landsat imagery. Remote Sens. Environ. (ISSN: 0034-4257) 194, 161–176. http://dx.doi.org/10.1016/j.rse.2017.03.022, URL https://www.sciencedirect.com/science/article/pii/S0034425717301244.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., Arino, O., 2022. ESA WorldCover 10 m 2021 v200. http://dx.doi.org/10.5281/zenodo.7254221.

Zepp, S., Heiden, U., Bachmann, M., Wiesmeier, M., Steininger, M., van Wesemael, B., 2021. Estimation of soil organic carbon contents in croplands of bavaria from SCMaP soil reflectance composites. Remote. Sens. (ISSN: 2072-4292) 13 (16), http://dx.doi.org/10.3390/rs13163141, URL https://www.mdpi.com/2072-4292/13/16/3141.

Žížala, D., Minařík, R., Zádorová, T., 2019. Soil organic carbon mapping using multispectral remote sensing data: Prediction ability of data with different spatial and spectral resolutions. Remote. Sens. (ISSN: 2072-4292) 11 (24), http://dx.doi.org/10.3390/rs11242947, URL https://www.mdpi.com/2072-4292/11/24/2947.