



## Article

# Soil Reflectance Composites—Improved Thresholding and Performance Evaluation

Uta Heiden <sup>1,\*</sup>, Pablo d'Angelo <sup>1</sup>, Peter Schwind <sup>1</sup>, Paul Karlshöfer <sup>1</sup>, Rupert Müller <sup>1</sup>, Simone Zepp <sup>2</sup>, Martin Wiesmeier <sup>3</sup> and Peter Reinartz <sup>1</sup>

<sup>1</sup> German Aerospace Center (DLR), The Remote Sensing Technology Institute, Muenchener Str. 20, 82234 Wessling, Germany

<sup>2</sup> German Aerospace Center (DLR), German Remote Sensing Data Center, Muenchener Str. 20, 82234 Wessling, Germany

<sup>3</sup> Bavarian State Research Center for Agriculture, Institute for Organic Farming, Soil and Resource Management, Lange Point 6, 85354 Freising, Germany

\* Correspondence: uta.heiden@dlr.de

**Abstract:** Reflectance composites that capture bare soil pixels from multispectral image data are increasingly being analysed to model soil constituents such as soil organic carbon. These temporal composites are used instead of single-date multispectral images to account for the frequent vegetation cover of soils and, thus, to get broader spatial coverage of bare soil pixels. Most soil compositing techniques require thresholds derived from spectral indices such as the Normalised Difference Vegetation Index (NDVI) and the Normalised Burn Ratio 2 (NBR2) to separate bare soils from all other land cover types. However, the threshold derivation is handled based on expert knowledge of a specific area, statistical percentile definitions or in situ data. For operational processors, such site-specific and partly manual strategies are not applicable. There is a need for a more generic solution to derive thresholds for large-scale processing without manual intervention. This study presents a novel Histogram SEparation Threshold (HISSET) methodology deriving spectral index thresholds and testing them for a Sentinel-2 temporal data stack. The technique is spectral index-independent, data-driven and can be evaluated based on a quality score. We tested HISSET for building six soil reflectance composites (SRC) using NDVI, NBR2 and a new index combining the NDVI and a short-wave infrared (SWIR) band (PV+IR2). A comprehensive analysis of the spectral and spatial performance and accuracy of the resulting SRCs proves the flexibility and validity of HISSET. Disturbance effects such as spectral confusion of bare soils with non-photosynthetic-active vegetation (NPV) could be reduced by choosing grassland and crops as input LC for HISSET. The NBR2-based SRC spectra showed the highest similarity with LUCAS spectra, the broadest spatial coverage of bare soil pixels and the least number of valid observations per pixel. The spatial coverage of bare soil pixels is validated against the database of the Integrated Administration and Control System (IACS) of the European Commission. Validation results show that PV+IR2-based SRCs outperform the other two indices, especially in spectrally mixed areas of bare soil, photosynthetic-active vegetation and NPV. The NDVI-based SRCs showed the lowest confidence values (95%) in all bands. In the future, HISSET shall be tested in other areas with different environmental conditions and LC characteristics to evaluate if the findings of this study are also valid.

**Keywords:** temporal composites; soil reflectance; LUCAS survey; CORINE survey; spectral; SCMaP; Sentinel-2; performance measures; uncertainty



**Citation:** Heiden, U.; d'Angelo, P.; Schwind, P.; Karlshöfer, P.; Müller, R.; Zepp, S.; Wiesmeier, M.; Reinartz, P. Soil Reflectance Composites—Improved Thresholding and Performance Evaluation. *Remote Sens.* **2022**, *14*, 4526. <https://doi.org/10.3390/rs14184526>

Academic Editor: Jeroen Meersmans

Received: 31 July 2022

Accepted: 7 September 2022

Published: 10 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In the last few years, large-scale and systematic information about the status and development of soils has moved into focus [1,2] because there is an increased demand for agricultural production [3]. At the same time, soil erosion [4] and soil-health degradation can be observed [5]. Environmental covariates derived from Earth observation (EO) have

already been used successfully for modelling soil information on a global [6] and continental scale [7]. Since the opening of the Landsat archives, the temporal compositing technique has become increasingly popular for many applications [8–12]. Optical multispectral EO archives have been explored for detecting bare soils from temporal data stacks on a per-pixel basis. Then, the bare soil pixels are averaged to create soil reflectance composites (SRC). Thus, the temporary covering of mostly agricultural soils with vegetation in single data takes can be partially circumvented, resulting in enlarged bare soil areas in the reflectance composites. Example developments are presented by [13] that have used the Landsat archive to derive a global map of bare ground gain by using 147 multitemporal metrics as input for tree-based regression and classification algorithms. The first operational SRC processors were developed by [14,15]. Ref. [14] uses functionalities and data from the Google Earth Engine (Geospatial Soil Sensing System (GEOS3)). Ref. [15] takes advantage of a processing environment and algorithms from the German Aerospace Center (DLR) for L2A processing (Soil Composite Mapping Processor—SCMAP). Both systems have been proven successful for agricultural areas at different spatial scales. Ref. [16] presented a global study to explore the spectra of the bare Earth surface as input for soil resource monitoring. Ref. [17] presented an exposed soil composite for the entire continent of Australia and tackled the challenge of applying a processor to very different biogeographical and climatic environments.

Several studies have created soil reflectance composites (SRC) and derived soil parameters such as soil organic matter (SOM) and clay content [18–20], soil organic carbon content [21–24], calcium carbonates [25] and other top-soil properties [26]. The combination of sensors such as SAR and Sentinel-3 improves the bare SRC and retrievals [19,27,28]. Refs. [28–30] investigate expert-knowledge-based systems to select pixels with a minimal influence of disturbing factors such as crop residues, surface roughness and soil moisture.

The principles of almost all of the compositing techniques mentioned earlier are similar: (1) spectral reflectance indices are used to select bare soil pixels from the temporal data stack, (2) index thresholds are used to optimise the selection of undisturbed bare soil pixels and (3) the temporal length of the multitemporal data stack is defined. Based on previous scientific studies, temporal length definition depends on the used EO sensor and the application and the region of interest. It ranges between 1 year [28], 2–4 years for Sentinel-2 [20,23,27,31] and 5 to more than 30 years for Landsat [14,15,18,25,30]. In general, the temporal length of Landsat-based studies is larger due to its lower revisiting time of up to 16 days. In regions without measurable changes in top-soil organic matter contents, such as in Bavaria, Germany, larger time series such as 30 years of available Landsat datasets (1984–2014) can be used [30,32] with the advantage of covering the maximum amount of bare soils in the composite. However, by using such a long time series, the averaging of disturbing factors such as different soil moisture conditions, varying structural soil occurrences and remaining effects from non-photosynthetic active vegetation (NPV) is expected [28]. These factors influence the quality of bare SRC [14,15] as well as the accuracy of the soil parameter retrievals [28,29]. Refs. [33,34] improved the underlying S2 Level 2A database by selecting dates after specific index thresholds, and [23,28] selected specific seasons to reduce disturbing effects, such as soil moisture as well as crop residues. Specifically, the “greening-up” period is promising since crop residues play a minor role. Ref. [29] defined the “greening-up” time as the last date of acquisition where the soil is exposed before the crop develops and is mostly smoothed. However, this selection reduces the number of pixels available for composition, and an increase in temporal length across years might be necessary.

A wide range of strategies can be found in selecting the spectral index, and the threshold derivation technique. In a wide range of studies, the well-known normalised difference vegetation index (NDVI) is often used [35] to separate the photosynthetically active areas (PV) from the non-active areas (NPV). Refs. [13,20] used an NDVI of  $<0.27$ , and [14] used an NDVI of  $<0.25$ , followed by other spectral indices to account for the above-described disturbance factors. Ref. [14] tested different empirically derived normalised

burn ratio II (NBR2) thresholds (0.025–0.350) for a test area in Brazil using laboratory soil spectra. Ref. [25] expanded this technique by using LUCAS top-soil spectral measurements to derive NDVI (−0.05 to 0.30) and NBR2 (−0.15 and 0.15) thresholds. Ref. [31] uses empirically derived thresholds from NDVI (<0.20) and the bare soil index (BSI) of >0.08 for test areas in Italy. Ref. [18] applied the BSI threshold (<0.021) that has been derived using airborne imaging spectrometer data from the Airborne Prism Experiment (APEX). Ref. [19] combined NDVI and NBR2 and tested different index ranges by selecting the optimum threshold based on the best correlation between resampled LUCAS spectra and SRC spectra. Ref. [15] developed a new index specifically for Landsat that combines the NDVI with a ratio of the near-infrared (NIR) and blue band to account for cloud artefacts that remain after cloud removal. The index thresholds have been derived using manually defined percentile measures. They have been proven stable across different years but not across different biogeographic regions in Germany. Thus, it has been found that percentiles are not transferable and strongly depend on the test site characteristics [30]. This manual step should be avoided for operational processors to reduce human intervention during processing to a minimum.

Using thresholds for continental or even global scales requires robust and transferable rules for their definition. However, the rules defined in most of the studies mentioned above are not applicable at large (e.g., global) scales due to site-specific conditions. Only a few authors have used data-driven techniques such as regression trees [13] and high-dimensional statistical methods [17] for spectral compositing. Machine learning and artificial intelligence techniques sound promising but lack an independent training database that contains the variability of bare soil surfaces across the globe.

A more fundamental problem lies in the limited spectral information content of multispectral EO data (e.g., Sentinel-2) to discriminate between soil and spectrally similar surfaces such as urban areas and NPV. Such confusion (disturbance) can influence the modelling performance of soil parameter retrievals and, thus, are an essential research topic [33,36]. In this sense, index thresholds are just a compromise implying that a certain percentage of false positives needs to be minimised. On the other hand, it is a practical, easy-to-use and computationally efficient technique for large areas and massive data processing. By reviewing the literature mentioned above, we can conclude that there is currently no agreement in the community about the “best spectral index”. The fixed thresholds vary widely depending on the biogeographic zone and the climatic condition of the test area. The exclusion of urban areas is not critical since profound databases such as the Global Urban Footprint [37], World Settlement Footprint [11] and the Global Human Settlement Layer [38] exist and are updated regularly to mask urban areas out. However, more significant problems are caused by NPV. These surfaces can only be distinguished from soils by using data resolving the narrow spectral absorption features at, e.g., 1.7 and 2.1  $\mu\text{m}$  that characterise the cellulose and lignin content. Good experiences have been had with the NBR2 that is often used to account for the soil-NPV confusion, although it can only minimise the influence of spectrally similar surfaces. However, the definition of a transferable threshold remains a difficult task and has been solved differently by authors.

More insights are needed about the diversity of index thresholds across large areas and their impact on the resulting SRCs. For this purpose, there is a need for a more generic solution to derive thresholds for large-scale processing. The solution shall (1) be generally applicable, (2) allow for regionalised threshold derivation, (3) account especially for spectral similarities between bare soil and NPV and (4) be independent of specific spectral index selections.

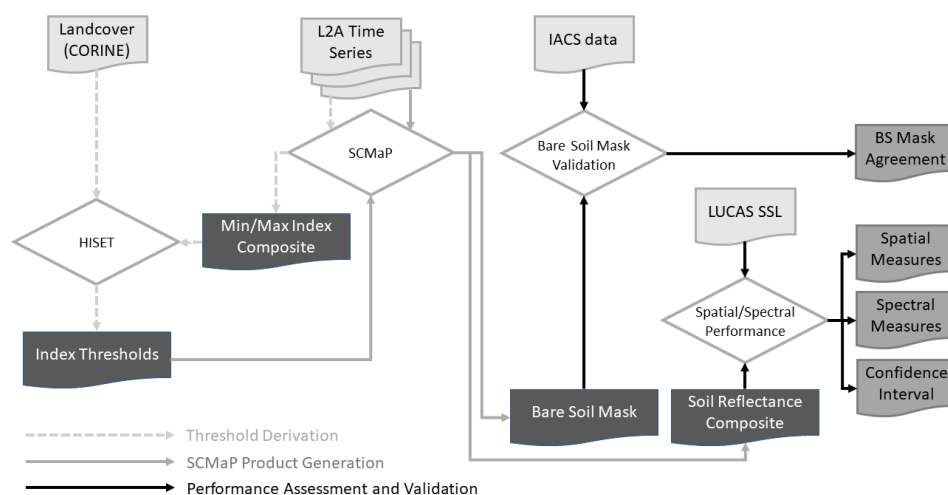
This study aims to develop a generic methodology for deriving spectral index thresholds used to derive SRCs. For this purpose, the two most commonly used indices represented in the soil-related literature are selected: NDVI and NBR2. Further, a new index called PV+IR2 is developed that combines the information from the visible to near-infrared (VNIR) and the short-wave infrared (SWIR) wavelength region. Several technical objectives have to be attained: (1) a fully automatised threshold derivation technique to process large

and diverse areas by objective means, (2) the evaluation of the three above-mentioned spectral indices to detect bare soils and to exclude pixels covering NPV and (3) the development of statistical measures to evaluate and compare the performance of SRC, including novel data products to assess the spectral uncertainty of the SRC.

This study gives more insights into the impact of selected indices and thresholds on resulting SRCs, which could influence the subsequent modelling of soil parameters. Moreover, it introduces measures and novel image products to evaluate the performance of SRCs as a database for the modelling part.

## 2. Materials and Methods

The analyses in this study are focused on part of Germany (Section 2.1). For the Sentinel-2 tiles covering this area (Section 2.2) and specific landcover types Section 2.3), the spectral index thresholds are derived using the SCMaP (Section 2.6) and the novel Histogram SEparation Threshold (HISET) methods (Section 2.7). Based on the set of thresholds, SCMaP is used again to generate six different SRCs that are compared within a performance assessment and validation step (Section 2.8). For this purpose, the LUCAS Soil Spectral Library (SSL) (Section 2.4) and the agricultural information from the *Integrated Administration and Control System* (IACS) (Section 2.5) are used to derive several spatial and spectral information products describing the performance and accuracy of the derived SRCs. An overview of the main analyses carried out in this study is given in Figure 1.

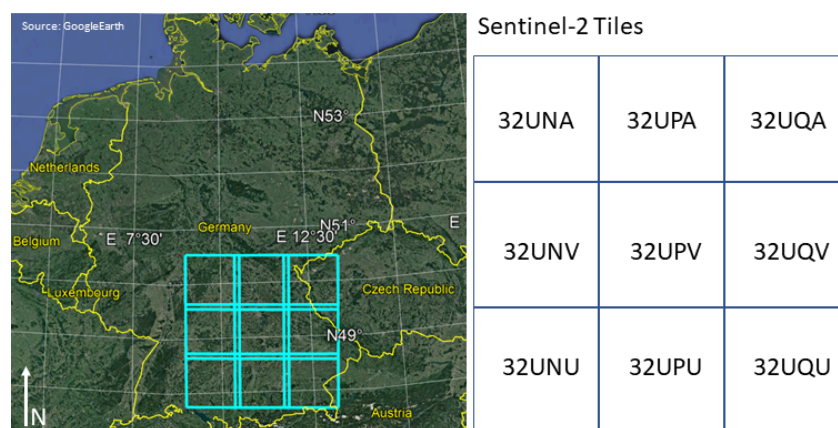


**Figure 1.** Summary of the main analyses and processing steps carried out in this study; BS = Bare Soil.

### 2.1. Test Area

All processing and analysis are performed for a mid-European test site covering large parts of the German federal state of Bavaria, Western parts of the Czech Republic and parts of Austria (Figure 2).

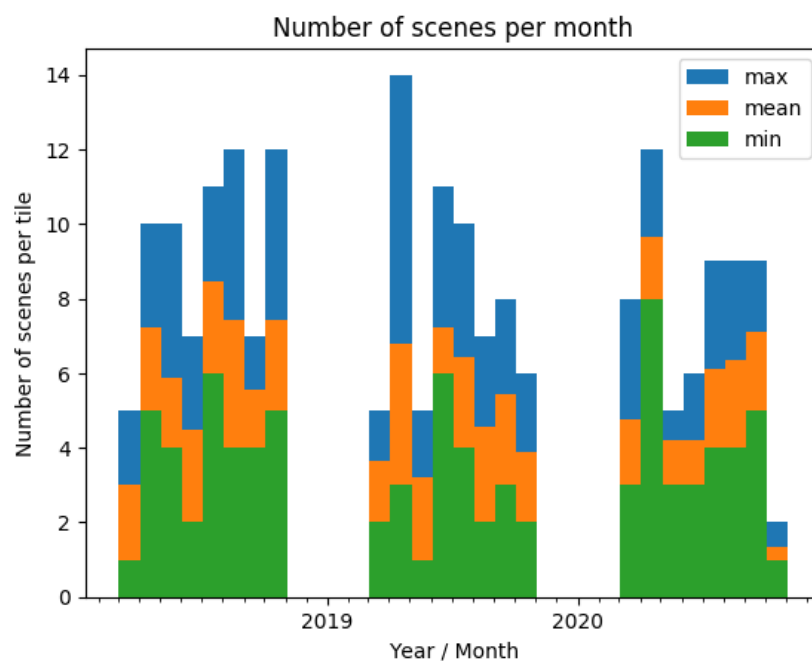
The area of 108.504 km<sup>2</sup> is framed by larger forested areas such as the hilly Spessart in the Northwest, the Thuringian forest in the North and the Bavarian Forest and Sumava National Park in the East. The Southern half of the test site is dominated by several river systems such as the Danube, Lech, Iller, Isar and Inn. Especially in the South, several old peat bog areas, such as the Donaumoos, are mainly used for agricultural purposes. Based on the World Reference Base (IUSS Working Group WRB, 2015) classification, the dominating soil types of the test site are Cambisols and Luvisols [6,39], with SOC values ranging from 0.26% to nearly 20.00% in the (former) peat bog areas. The test site has been chosen because (1) it covers a high percentage of cropland interspersed with forest and grassland patches, and (2) it contains a wide range of SOC values and soil types.



**Figure 2.** Themid-European test site highlighted in cyan covers large parts of the German federal state of Bavaria, Western parts of the Czech Republic and small parts of Austria.

## 2.2. Sentinel-2 Data

For the development of the SRCs, all available S2 images acquired by both sensors (A and B) between 2018 and 2020 are considered for the area covered by nine tiles from the UTM zone 32 North (Figure 2). The available images are then further reduced by accepting images only with a cloud coverage of  $< 80\%$  and excluding the winter months (November to February). Winter months are excluded due to the solar zenith angle of higher than  $70\%$ , for which a reliable scene-based estimation of atmospheric parameters (water vapour, aerosols) cannot be guaranteed [40]. For the nine tiles, statistics of the number of scenes per tile are given in Figure 3. In 2018, the test site was characterised by arid and cloudless weather, which resulted in a constant high number of scenes across the observation window from March to October. The distribution of Sentinel-2 scenes per month and year shown in Figure 3 in relation to the climatic conditions of the respective year gives the first view of potential soil conditions and, thus, SRC variability that is integrated into the SRCs. In total, 1202 single images are used from 2018 to 2020. The three Sentinel-2 bands with a pixel resolution of 60 m (B1, B9, B10) are excluded, while the four 10-m bands (B2, B3, B4 and B8) are resampled to 20 m pixel resolution using the nearest neighbour.



**Figure 3.** Distribution of available Sentinel-2 images per month for the area of nine tiles constrained by a cloud coverage  $< 80\%$  and excluding winter months; only March to October scenes are considered.

For generating SRCs, bottom-of-atmosphere reflectance data (L2A) and per-pixel cloud, haze and snow information are required. All orthorectified S2 data (L1C) available for the test site are downloaded, filtered by the constraint mentioned above and ingested into DLR's internal processing environment. In the next step, L1C data are processed to L2A reflectances using MACCS-ATCOR Joint Algorithm (MAJA) [41,42].

### 2.3. Corine Land Cover (CLC) Data Processing

For the compositing approach (see Section 2.6), SCMaP can assimilate two thresholds to separate bare soils from spectrally similar land cover (LC) types without using ancillary data [15]. It requires understanding the temporal behaviour of certain LC types (see Section 2.7). The Corine Land Cover (CLC) mapping is used to select the LC types of interest. It is a European-wide harmonised and long-term available LC database coordinated and integrated by EEA and available as a Copernicus service [43]. It currently contains LC layers and change layers from 1990 to 2018 and will be regularly updated in the future. The long time for which CLC is available is used to calibrate the threshold model on stable areas across time. Thus, areas are excluded that had a recent change and might not be representative of the LC type.

In order to analyse the temporal behaviour of certain LC types, the CLC need to be filtered and aggregated according to the procedure described in [30]. Bare soils most likely occur in agricultural areas after harvesting and before sowing. Therefore, all CLC types linked to intense agriculture have been combined and renamed crops (LC crops: combination from CLC classes 211, 212 and 213). NPV is known as spectrally similar to bare soils in multispectral EO data. This cover type mostly occurs in grasslands in a dry state and in deciduous forests when they are in a leaf-off condition. In order to analyse spectral-temporal NPV characteristics, grassland-type LC classes CLC 231, 321 and 322 are chosen and combined with LC grassland. Further, deciduous forests show NPV mostly in spring and autumn under leaf-off conditions. Therefore, CLC class 311 has been selected for LC deciduous. Spectral similarity can also be found in urban areas such as impervious surfaces, and thus CLC classes 111 and 121 are combined to build LC urban for this study.

For all available CLC surveys from 1990 to 2018, the described CLC classes are aggregated while keeping the vector format. Only areas that remain unchanged across all CLC surveys are used. Additionally, the CLC change layers (CHA) are used to filter our small-scale changes that are not covered by the general CLC surveys. Borders of 100 m are excluded from the stable LC pixel areas [30]. In the final step, the vectors are rasterised in the same spatial resolution as the multitemporal S2 database (20 m). For the resulting stable LC pixels, the spectral-temporal characteristics are measured by spectral indices as the primary information for defining the thresholds (see Table 1).

### 2.4. Lucas Soil Spectral Library (SSL)

The LUCAS Soil Spectral Library (SSL) is part of the Land Use/Cover Area frame Survey (LUCAS). It is the only existing European-wide soil, and landcover survey repeated regularly and measured in a standardised way [44]. Besides chemical analyses, all soil samples are spectrally measured by a single laboratory. The LUCAS soil survey is updating our knowledge about the status and development of European soils. Further, it is an elementary part of the European soil mappings published regularly by the Joint Research Center (JRC). In this study, the SRC spectra are compared to the LUCAS SSL spectra of 2015 to evaluate the spectral performance of the SRCs and their potential for further European-wide soil modelling and prediction. However, the SRC spectra are a collection of natural soil conditions containing different roughness and moisture conditions. In contrast, LUCAS SSL contains spectra measured in oven-dried and sieved conditions. Since it can be expected that this treatment mainly impacts the albedo and that spectral absorption features of mineral components remain, LUCAS and SRC spectra are compared using the spectral angle (see Section 2.8).

The LUCAS spectra were resampled to S2A/(B) L2A reflectances using the following formula (see also [45]):

$$\rho_{band} = \frac{\int \rho(\lambda) RSR_{band}(\lambda) d\lambda}{\int RSR_{band}(\lambda) d\lambda} \quad (1)$$

where  $\rho(\lambda)$  corresponds to the LUCAS spectra,  $RSR_{band}(\lambda)$  to the response function for each of the S2 bands and  $\rho_{band}$  is the band integrated reflectance value. The spectral response functions from S2 bands are taken from ESA's web resource [46].

### 2.5. European IACS Data

The Integrated Administration and Control System (IACS) of the European Commission (EC) is the main pillar of controlling European payments to farmers. One of the elements is a spatial database for all agricultural plots (land parcels) in EU countries for which payments have been made. The IACS mapping and database are controlled by independent facilities on behalf of the EC. For each year, land parcel outlines are stored with information about the broad crop category (arable land, grassland, permanent crops, permanent grassland, others) and the specific crop type harvested in the respective year. In this study, a subset of the IACS database for 2018, 2019 and 2020 covering the test area of Bavaria, Germany, is used to validate the SRC bare soil mask. The mask contains the coverage of the SRC pixels. The validation is based on the assumption that the SRC shall cover crop types with a harvesting event in the respective year since they will appear in bare soil conditions. Therefore, all areas covering a crop type with a harvest event within the period of interest are collected in the bare soil mask representing the ground truth for the spatial representation of the SRC pixels. In this study, three years from 2018–2020 are integrated to build the IACS bare soil mask for comparison with the SRC mask of bare soils.

### 2.6. Soil Composite Mapping Processor (SCMaP) Update

The generation of the SRCs follows the principles published in [15] and is developed to analyse multitemporal Landsat databases. It computes several temporal, spectral and index composites, frequency products and statistical products from a multispectral image time series that can support digital soil mapping and spectral soil modelling. The main output is the SRC that collects predominantly pure bare soil pixels from a multispectral time stack. The development of SCMaP started with Landsat data inputs and is now extended to using other multispectral multitemporal data cubes, such as Sentinel-2.

SCMaP has been primarily developed for temperate and continental climates, where bare soils occur mainly in agricultural areas. These areas are characterised by a change from the vegetated condition during the green-up and crop growing stages to bareness after harvesting and when fields are in the seedbed condition. The frequency and time of bareness depend on the crop cycle. After filtering out cloud pixels and snow pixels using the normalised difference snow index (NDSI), SCMaP computes a spectral index (see Table 1) for each scene and calculates the minimum and maximum index composites. These index composites are used for the index threshold derivation described in Section 2.7. SCMaP uses two spectral index thresholds ( $t_{min}$ ,  $t_{max}$ ) to separate non-vegetated areas from vegetated areas in every single scene of the temporal data stack. Only pixels showing a change between vegetated and non-vegetated states are recognised as bare soil pixels. These pixels are different from permanently sealed surfaces. The number of required changes can be defined as ( $min_{soil}$ ). It mitigates the effect of spurious outliers and allows the computation of meaningful quality layers, such as the standard deviation ( $s$ ) of the SRC spectra. It should be noted that the SCMaP principle differs from other techniques that use ancillary data to exclude other areas such as urban or permanent vegetation (e.g., [18]).

The processor can handle different spectral indices. Table 1 provides an overview of the spectral indices that are used in this study. NDVI and NBR2 are selected because they are the most commonly used indices described in the literature. PV+IR2 combines information from the VNIR and the SWIR wavelength region. Both regions are relevant to separating green vegetation and NPV from bare soils. The SCMaP has been used for

the generation of all compositing image products with varying parameter settings that are described in Section 3.1.

**Table 1.** Spectral indices used by the SCMaP in this study.

Index Name	Index Short	Formula	Reference
Normalised Difference Snow Index	NDSI	$(B3 - B11)/(B3 + B11)$	[47]
Normalised Difference Vegetation Index	NDVI	$(B4 - B8)/(B4 + B8)$	[48]
Combined NDVI and SWIR2 index	PV + IR2	$(B8 - B4)/(B8 + B4) + (B8 - B12)/(B8 + B12)$	experimental
Normalised Burned Ratio 2	NBR2	$(B11 - B12)/(B11 + B12)$	[49]

### 2.7. Histogram SEparation Threshold (HISET) Method

Two spectral index thresholds are used to separate bare soils from all other surfaces. In the case of SCMaP, the histograms of the minimum and maximum index composites (Section 2.6) for all stable LC pixels (Section 2.3) are used to analyse their spectral-temporal characteristics. To separate bare soils from vegetation, especially when NPV dominates vegetation, histograms of the minimum index composite for crops and grassland/deciduous are used. The separation of bare soils and urban areas is based on the histograms of the maximum index composite for crops and urban areas. In both cases, the histograms of the respective LC types overlap and prevent full separation. In order to remain flexible in the choice of the used index and regional specifications, a new strategy has been developed that can separate two LCs with minimum overlap between them. The threshold is based on the optimal separation of two histograms and, therefore, is called the HIStogram SEparation Threshold (HISET). The following calculations are performed iteratively over the range of the two input histograms or probability density functions (PDF) to calculate the HISET and the corresponding HISET score (see also Appendix A.1):

- For each of the histograms/PDFs, the proportion of the bins left and right of a potential threshold is calculated.
- For the two calculated proportions on each side of the threshold, the minimum is determined.
- The maximum of the two minima is determined for the potential threshold (preliminary HISET score).
- The final HISET score is defined by the minimum of all the calculated scores in the previous step. This minimum also determines the final histogram separation threshold.

The resulting HISET score is a quality measure describing how well the two input histograms could be separated. It can be described as the smallest possible proportion of any distribution of A lying on the same side of the threshold as distribution B while making sure that distribution B does not have a higher proportion lying on the other side of the threshold than distribution A. For a threshold separating the two histograms completely, the HISET score would be 0%. For the worst realistic separation, for example, for two identical histograms that cannot be isolated, the HISET score would be approximately 50%. For example, a worse score is possible in theory for two identical histograms containing only one bin. For the presented use case, however, realistically, this should never occur.

It is worth mentioning that it makes no difference for the HISET algorithm if the provided inputs are histograms or PDFs since it works with proportions. For a graphical representation of the threshold together with the two distributions, however, it makes sense to use the PDFs, especially if the relative size of the two distributions varies significantly.



## 2.8. Spatial and Spectral Performance Assessment

The SRCs have been proven as a suitable database for large-scale soil modelling [18,25], and their quality has been mostly assessed indirectly by metrics describing the resulting soil parameter prediction. Commonly, they are calculated as the coefficient of determination ( $R^2$ ), the root mean square error (RMSE), the performance of deviation (RPD) and the ratio of performance to interquartile distance (RPIQCV), indicating the model and the prediction performance [22,34,50,51]. The quality of the underlying SRCs is not explicitly evaluated. The best way to express and compare the quality of an SRC product is to calculate its uncertainty. Although it is a standard for single-pixel measurements [52,53] that include uncertainties due to atmospheric parameters, such as aerosol optical thickness or water vapour, the uncertainty is even more complex for a pixel-by-pixel combination of spectra collected over time. The observed surface has its own spectral, spatial and temporal variability and, thus, sources of uncertainty [54]. Even more, temporal compositing can depend on varying cloud and snow cover. Thus, temporal data integration is not equal across space, leading to sampling uncertainty [55].

To the authors' knowledge, there is currently no standardised procedure or set of measures to evaluate the accuracy or uncertainty of temporal bare SRCs in terms of spectral agreement with a reference soil measurement. Due to the lack of a statistically reliable number of matching in situ and S2 spectral measurements, the focus of this study is on spectral and spatial metrics (Sections 2.8.1 and 2.8.3), enabling a comparison of different SRCs and proving a pixel-wise estimate of the variability and confidence of the resulting SRC spectra (Section 2.8.4). Further, the spatial coverage of the bare soil pixels (bare soil mask) is validated against the IACS database (Section 2.8.2).

### 2.8.1. Spatial Performance

The percentage of bare soil pixels is a general measure to assess the spatial coverage for which soil parameters can be directly derived from EO data using spectral soil modelling approaches. Especially for large-scale soil modelling (e.g., Europe) that is dependent on sparsely distributed calibration points derived from standardised soil core measurements, a higher SRC coverage can be essential for higher model performance. Thus, the **percentage of bare soil pixels** and the **number of LUCAS points** counted that are covered by the SRC.

### 2.8.2. Spatial Validation of the Bare Soil Mask

However, the percentage of bare soil pixels is not suitable to assess the spatial validity and completeness of the bare soil area since spectrally similar surfaces, such as deciduous forests in leaf-off conditions or NPV-containing grasslands, might be detected as bare soils due to the spectral similarities [56]. Few authors have presented spatial metrics such as the percentage of detected bare soil area and the spatial agreement of the bare soil area with the crop area [18,28,30]. The latter is especially applicable in the temperate climate zone, where bare soil occurs mainly during harvesting and ploughing of agricultural sites and barely on natural sites.

The external IACS database containing land parcel-specific crop types has been used for the quantitative validation of the bare soil area. It is a European database containing all agriculturally used areas and is updated yearly (see Section 2.5). The derived mask of bare soils is compared to the bare soil mask of the SRCs to perform a simple pixel-wise accuracy assessment. It comprises true and false positives and true and false negatives for the entire test area available as an image product and as a confusion matrix (Figure 4).

		Predicted class		
Actual class	True positives (TP)	False negative (FN)	$\frac{TP}{TP+FN} * 100$ $\frac{FN}{TP+FN} * 100$	
	False positive (FP)	True negative (TN)	$\frac{TN}{FP+TN} * 100$ $\frac{FP}{FP+TN} * 100$	
	$\frac{TP}{TP+FP} * 100$ $\frac{FP}{TP+FP} * 100$	$\frac{TN}{TN+FN} * 100$ $\frac{FN}{TN+FN} * 100$	$\frac{TP + TN}{TP+TN+FP+FN}$	

Figure 4. Quantitative measures for the spatial validation of the SRC bare soil mask.

### 2.8.3. Spectral Performance

The quality of the spectral representation of the SRCs is essential for calibrating the soil spectral models. However, the assessment of the quality is not trivial. A composite spectrum integrates different natural conditions over the selected time range. In the case of soils, the spectrum changes with different soil moisture contents, roughness conditions and other disturbance factors such as NPV coverage that can only be minimised by using multispectral data for generating spectral composites. For optimal soil modelling, the soil condition would be dry, and the surface is smoothed because calibration points available on a large scale, such as LUCAS, are measured under laboratory conditions (dried and sieved). Such soil conditions rarely occur in natural environments. Therefore, the spectral comparison will be affected by albedo differences. Ref. [14] compared laboratory and composite spectra using derivatives such as the soil line and dispersion between selective bands using principal components for both laboratory and composite spectra. In a preliminary study, Ref. [57] used the spectral angle [58] between LUCAS SSL spectra and the respective SRC spectra for estimating the similarity of the spectral pairs by neglecting albedo differences. Although these spectral similarity measures are not suitable for estimating absolute spectral agreements, they can be used to evaluate the improvement or degradation of SRCs generated by varying processing parameters. In order to compare the spectral performance of the different SRCs, the spectral similarity between the spectral lab measurement of the LUCAS points and the SRC spectra is calculated using the **spectral angle**. The spectral angle has been widely used in geological hyperspectral applications and is known to be insensitive to illumination and albedo effects [58]. The band-wise **mean reflectance difference** between both measurements is additionally calculated for completeness.

### 2.8.4. Spectral Variance of SRCs

Due to the averaging of several bare soil measurements in the SRC spectrum, different soil conditions are covered, although there are concepts to reduce disturbing effects by choice of index, selection of specific seasons [29] or the use of SAR data [28] to account for soil moisture variability. Therefore, instead of getting a perfect Sentinel-2 spectrum,

information about the spectral variance of the averaged soil conditions ( $n - 1$ ) and the standard deviation ( $s$ ) for each SRC spectrum ( $\bar{x}$ ) is calculated as:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (2)$$

The SRC pixel  $s$  is calculated using different composite parameters that can be considered a quality parameter for the resulting composites. Thus, the  $s$  is a parameter for evaluating the spectral variance within an SRC image, and it can reveal if natural patterns of variance exist, such as for field edges, where more spectral variability due to disturbances from, e.g., NPV, can be expected.

However, the spatial representation of  $s$  can be biased by the number of **valid pixels** used for the SRC averaging. For each SRC, per-pixel information about the number of valid bare soil observations is stored in an image product. The number mostly depends on cloud, haze, snow coverage and the geographical position on Earth. On the one hand, there is a higher overlap of S2 orbits in higher latitudes resulting in denser time series, but on the other hand, the large-scale probability of cloud-free images is decreased. The general assumption is that the more pixels are available, the more soil conditions are integrated into the mean SRC pixel spectrum. A higher number of valid pixels can be positive because it increases the confidence and reliability of the averaged SRC spectrum. However, this increases the likelihood of integrating more outliers in the pixel spectrum, such as more moisture levels. Ref. [29] found that instead of integrating all available bare soil conditions, carefully selected observations, predominantly before the green-up phase where soils are smoothed and dry, can result in higher SOC prediction performances. In this study, the minimum number of valid pixels is set to three observations to reduce the number of extremes and outliers. For comparing SRCs generated by the different processing parameters, the minimum (*min*), maximum (*max*), mean (*mean*) and standard deviation ( $s$ ) are calculated for the complete area of the SRCs, additionally to the pixel-based image product.

Finally, the 95% confidence interval (CI) of the SRC spectrum ( $\bar{x}$ ) is a measure for estimating the uncertainty of this pixel spectrum [59]. The measure can be taken as quality criteria for a potential calibration point selection. The CI for each spectral band is defined as:

$$CI = \bar{x} \pm \left( t_{1-\frac{\alpha}{2}, df} \frac{s}{\sqrt{n}} \right) \quad (3)$$

where  $\alpha$  is 0.05,  $df$  is the degrees of freedom, defined as  $n - 1$ ,  $n$  is the number of reflectance values and  $t$  is the  $t$ -value, which is 2.093 for the 95% CI.

### 3. Results

#### 3.1. Experimental Settings

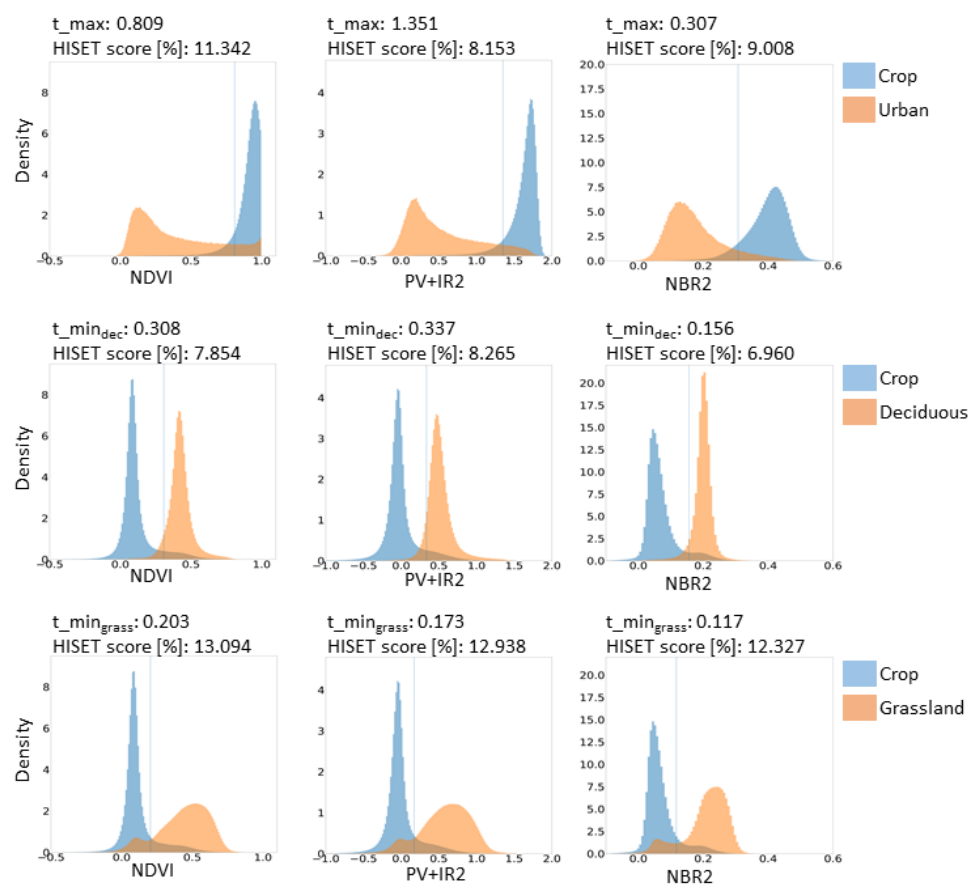
Six SRCs have been created using three indices (see Table 2) and two  $t_{min}$  threshold versions. The thresholds vary according to the results of the HSET analysis described in Section 3.2. The processing parameters for each SRC are listed in Table 2.

**Table 2.** Soil reflectance composite (SRC) processing parameter settings.

Processing Parameter	Value
S2 Level 2A reflectance processor	MAJA
S2 Level 2A cloud mask	MG2 (MAJA output)
Time range (year)	2018–2020
Month	03–10
Cloud cover	<80%
NDSI	0.00
Minimum count	3
Index (see Table 1)	NDVI, PV+IR2, NBR2

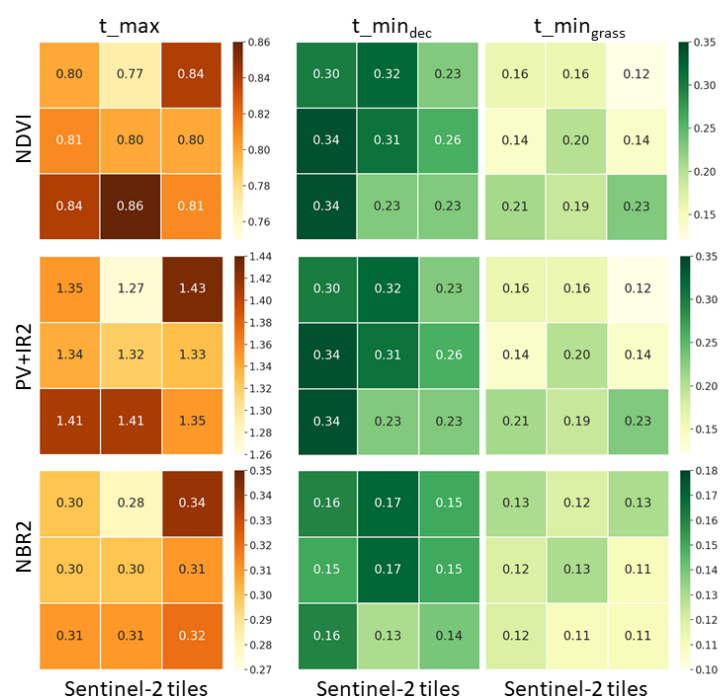
### 3.2. Thresholds

The *min* and *max* index composites have been processed for each spectral index (NDVI, PV+IR2 and NBR2) and per S2 tile (see also Table 2). HISET thresholds and scores are derived for each of the S2 tiles and for a spatial mosaic of all tiles. Figure 5 shows the histogram plots for all pixels covering the selected LC types in the complete test area. For  $t_{max}$ , the crop pixels are compared with the urban pixels.  $T_{min}$  has been created in two versions, (1) by comparing crops with deciduous trees and (2) by comparing crops with grassland. Figure 5 shows histogram overlaps for all indices and thresholds. For calculating  $t_{max}$ , the *max* index composite is used, where crops shall be in a photosynthetic-active state and, therefore, well separable from urban areas (upper row). The overlap happens when urban LC pixels are also covering urban green. The HISET score for PV+IR2 is 8.2% and, thus, has the lowest score value. The highest score can be found for NDVI (11.3%). Independent from the index, deciduous (middle row) and grassland (lower row)  $t_{min}$  index values show clear differences. Compared to grasslands, histograms of deciduous show minor overlaps, resulting in thresholds with HISET scores between 7.0% and 8.3%. For grassland, a bimodal and wider distribution of index values can be observed for all indices. It highlights the above-described spectral similarity of NPV and bare soil that cannot be fully separated using multispectral data, especially when the NPV-containing LC is in a dry condition (first local maximum). Although the HISET score is higher than for deciduous trees, each index's threshold is defined as lower, resulting in fewer pixels that are selected for building the SRC. Thus, fewer disturbance effects are expected. Comparing the three used spectral indices, the HISET score for NBR2 is the lowest for both  $t_{min}$  versions.



**Figure 5.** Density plots showing the spectral-temporal variability and separability of certain LC types in the complete test site measured by three spectral indices (NDVI, PV+IR2 and NBR2). LC types are agricultural fields (crop), deciduous forests/trees (deciduous) and grasslands (grassland), which build the basis for the threshold definition.

In the next step, the derived HISET thresholds and scores have been analysed to reveal potential dependencies on specific LC characteristics of each S2 tile. The  $t_{max}$  distribution pattern (left column of Figure 6) is similar for NDVI and PV+IR2 since the index calculation is similar (see Table 1). At the same time, values for S2 tile 32UPA are the lowest for each of the indices. Threshold differences between  $t_{min_{dec}}$  and  $t_{min_{grass}}$  are obvious, and similarities between NDVI and PV+IR2 are visible. NBR2 shows different spatial patterns for both  $t_{min}$  versions. An S2 tile-based variability of HISET scores can also be observed (see Figure A1). To clarify, if the derived HISET values are related to the percentage of occurring LC type pixels, a preliminary multiple regression analysis (MRA) has been performed by defining the HISET score or threshold as a predictor and the two respective LC types as coefficients. The test has been performed by considering all nine S2 tiles. The coefficients of determination,  $R^2$ , for all tests are below 0.52, with a mean  $R^2$  for all thresholds and indices of 0.30. The results show no measurable relationship between the percentages of LC types used to derive the HISET parameter.



**Figure 6.** Spatial variability of  $t_{max}$  and  $t_{min}$  thresholds across S2 tiles.

### 3.3. Composite Generation

Based on the threshold analyses described above, the  $t_{max}$  and  $t_{min}$  index values have been derived for each spectral index for the complete test site (9 tiles = 95.964 km<sup>2</sup>). The SRCs are processed using the  $t_{max}$  and  $t_{min}$  listed in Figure 5. The performance of each resulting SRC is quantified using the methods described in Section 2.8, and the image-based results are summarised in Table 3. Figure 7 shows an SRC subset around Noerdlingen, Germany, generated from PV+IR2 and using  $t_{min_{grass}}$ . Additionally, two example subsets, A and B, for all processed versions (see Table 3) are displayed for various image data products. The colour stretching is equal for each row of the example subsets to enable the visual comparison of the image results.

Compared with the WRB soil types, the area East of the city of Noerdlingen, Germany (example A), is covered by luvisols and appears reddish in the images. West of the river Woeritz (example B), the images have a greenish colour. Cambisols and stagnisols cover these areas.

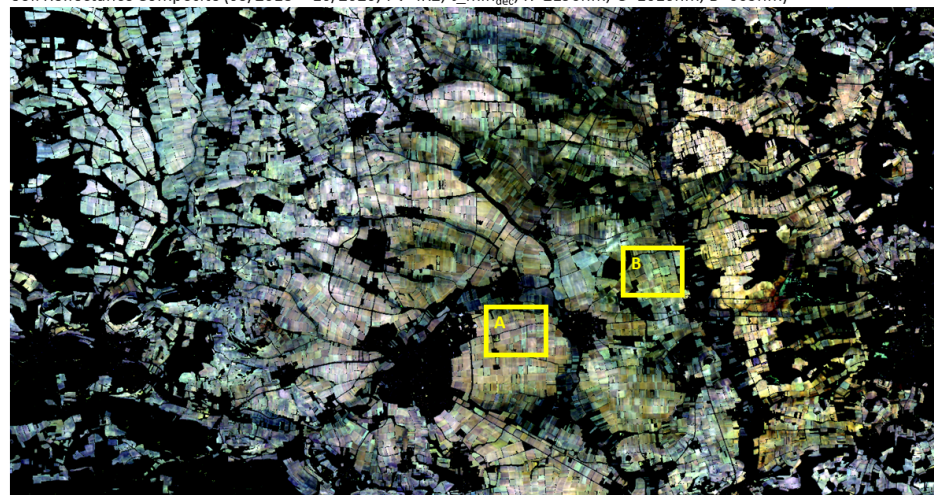
A closer look at examples A and B shows that the resulting composites differ in spatial coverage and spectral variability, driven by the selection of the number and type of valid pixels. In principle, the number of valid pixels is higher for all  $t_{min_{dec}}$  versions

independent from the index. This is in line with the generally lower threshold values for the  $t_{min_{grass}}$  listed in Table 3. Since the valid pixel images show patterns related to the fields, it can be assumed that the occurrence of bare soil pixels is related to the crop type. This effect is visible for each of the indices and  $t_{min}$  versions. In the examples in Figure 7, the NBR2/ $t_{min_{grass}}$  parameter combination produces the least coverage of SRC pixels. However, for the entire test side, the PV+IR2/ $t_{min_{grass}}$  combination shows the lowest coverage of bare soil pixels (Table 3), while the mean number of valid pixels is slightly higher compared to NBR2/ $t_{min_{grass}}$  (16.19% vs. 15.76%). From all  $t_{min}$  versions, NBR2 resulted in the highest coverage of bare soil pixels (29.64%). The number of LUCAS points that are covered by the SRCs is similar for  $t_{min_{dec}}$  and  $t_{min_{grass}}$  versions, respectively.

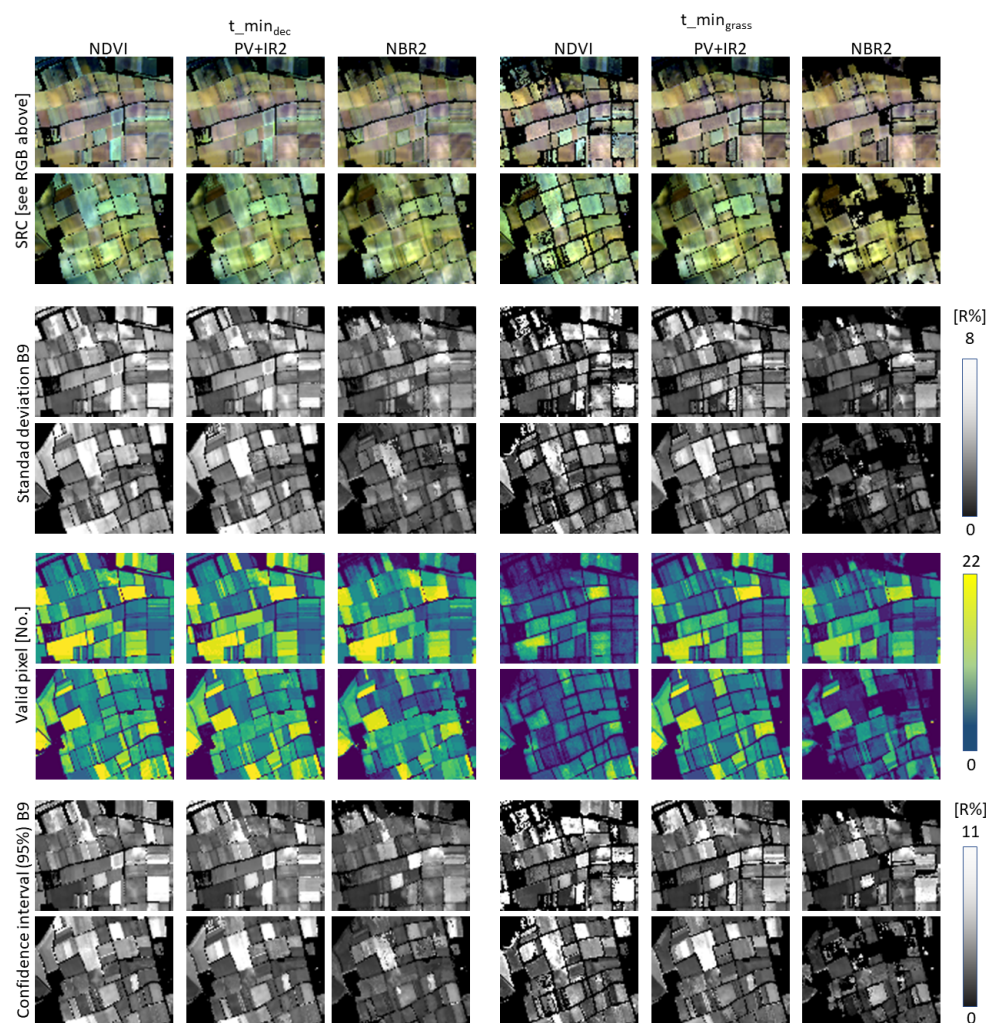
**Table 3.** Performance assessment of soil reflectance composites (9 tiles = 95,964 km<sup>2</sup>) generated with varying  $t_{min}$ ;  $s$  = standard deviation.

Soil Reflectance Composite (SRC) Versions							
Index		NDVI		PV+IR2		NBR2	
$t_{max}$		0.809		1.351		0.307	
$t_{min}$		0.308	0.203	0.337	0.173	0.156	0.117
Performance indicators							
Bare soil pixels (%)		34.7	28.5	32.8	28.1	35.1	29.6
No. of valid observations per pixel ( $min = 3$ )	<i>max</i>	97	97	94	93	99	99
	<i>mean</i>	23.6	18.1	20.6	16.2	20.9	15.8
	<i>s</i>	12.5	10.1	11.1	9.2	11.4	9.3
No. of LUCAS points covered by SRC		231	209	233	209	233	211
Spectral angle (rad) for all covered LUCAS points	<i>mean</i>	0.085	0.066	0.081	0.063	0.071	0.058
	<i>median</i>	0.083	0.063	0.074	0.057	0.069	0.055
	<i>min</i>	0.032	0.027	0.028	0.026	0.030	0.026
LUCAS points	<i>max</i>	0.178	0.192	0.252	0.240	0.147	0.123
	<i>s</i>	0.026	0.023	0.032	0.026	0.020	0.017
	<i>mean</i>	11.9	11.7	11.8	11.6	11.9	11.8
Mean refl. diff. (%) for all covered LUCAS points	<i>median</i>	12.1	11.6	12.1	12.0	12.0	12.1
	<i>min</i>	3.52	3.22	3.23	2.99	3.53	3.60
	<i>max</i>	20.2	20.3	20.1	19.9	21.1	20.3
	<i>s</i>	3.07	3.03	3.00	2.96	3.01	3.03

Soil Reflectance Composite (03/2018 – 10/2020; PV+IR2;  $t_{min_{dec}}$ ; R=2190nm, G=1610nm, B=665nm)



**Figure 7.** Cont.



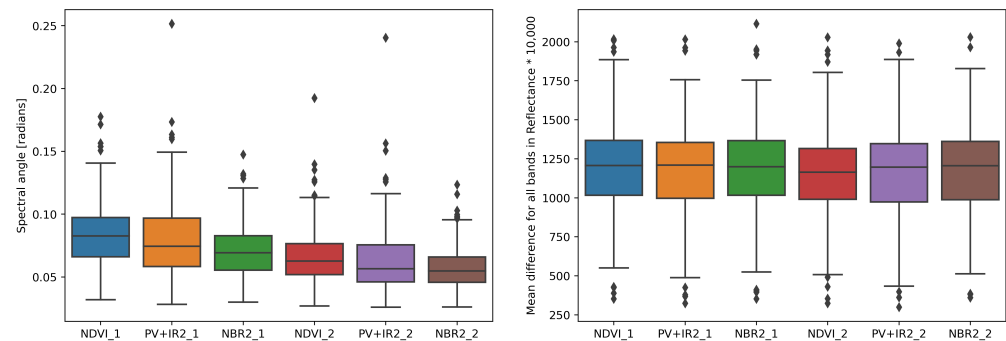
**Figure 7.** SRC subset around Noerdlingen, Germany. Subsets A and B for all processed versions (see Table 3) show details about the spatial and spectral variability of the SRCs. The RGB of examples A and B are the same as used for the large subset; R = Reflectance.

### 3.4. Performance Assessment

As expected from Section 3.2, SRCs generated using  $t_{min_{grass}}$  have a lower percentage of exposed soils than those processed with  $t_{min_{dec}}$  due to the more strict threshold value. The SRCs generated using the NBR2 cover the largest percentage of exposed soils. However, values between the index versions differ less than between  $t_{min}$  versions. Interestingly, the mean number of valid observations is highest for the NDVI thresholds, indicating that the NDVI threshold allows more pixels to be integrated with the SRC pixels. The most restricted selection of valid pixels is performed for the NBR2/ $t_{min_{grass}}$  version. The spectral comparison between S2 resampled LUCAS and SRC spectra using the spectral angle and the mean reflectance difference (Section 2.8.3) was made for all LUCAS points that cover an SRC pixel. Further, only those LUCAS points are considered that are covered in every SRC to ensure the comparability of the spectral angle and mean reflectance difference values. Figure 8 presents a comparative view of the statistic.

The summarised spectral angle statistics show a gradient from NDVI, PV+IR2 to NBR2 for all  $t_{min}$  versions. The best match can be found for the NBR2/ $t_{min_{grass}}$  version with a mean spectral angle of 0.058, which marks a drop of more than 30% compared to the highest spectral angle value calculated for the NDVI/ $t_{min_{dec}}$  version. Additionally, the interquartile range of SRC/ $t_{min_{grass}}$  versions is smaller compared to the SRC/ $t_{min_{dec}}$  versions, with the smallest interquartile range of the NBR2/ $t_{min_{grass}}$  SRC. In contrast,

the calculated mean reflectance differences do not show significant changes across the SRC versions.



**Figure 8.** Spectral statistics per SRC version. Index\_1 represents SRCs based on  $t_{min_{dec}}$ , and Index\_2 stands for SRC versions calculated with  $t_{min_{grass}}$ .

### 3.5. Validating the Bare Soil Mask

The bare soil mask built during the SCMaP processing collects pixel positions showing bare soils between March 2018 and October 2020 (excluding winter months). This bare soil mask has been validated using the preprocessed IACS data (see Section 2.5) that contains the same information as the SCMaP bare soil mask. Therefore, the validation is a pixel-based comparison of both masks to quantify their agreement. Both binary masks cover the same spatial extent, determined by the common overlap between the IACS and the SRC data. This analysis has been performed for the six SRC versions resulting in six confusion matrices, where grey numbers in the coloured boxes are given in the percentage of matched/unmatched pixels (Figure 9). The most crucial target, class 1, represents bare soils, whereas class 0 collects crops with permanent vegetation coverage. The accuracy of class 1 is measured by true negatives (TN), according to Figure 4.

		0		1							
$t_{min_{dec}}$	IACS	0	61.66	6.66	90.25 9.75	61.90	6.42	90.60 9.40	60.60	7.72	88.70 11.30
		1	1.66	29.44	94.66 5.34	1.33	29.77	95.71 4.29	1.54	29.57	95.06 4.94
			97.38 2.62	81.56 18.44	0.916	97.89 2.11	82.26 17.74	0.922	97.53 2.47	79.30 20.70	0.907
	IACS	0	65.94	2.38	96.52 3.48	65.85	2.46	96.39 3.61	64.47	3.85	94.37 5.63
		1	3.02	28.08	90.29 9.71	2.71	28.40	91.30 8.70	2.96	28.14	90.48 9.52
			95.62 4.38	92.20 7.80	0.946	95.44 4.56	92.89 7.11	0.948	95.61 4.39	87.97 12.03	0.932
		SRC / NDVI		SRC / PV+IR2		SRC / NBR2					

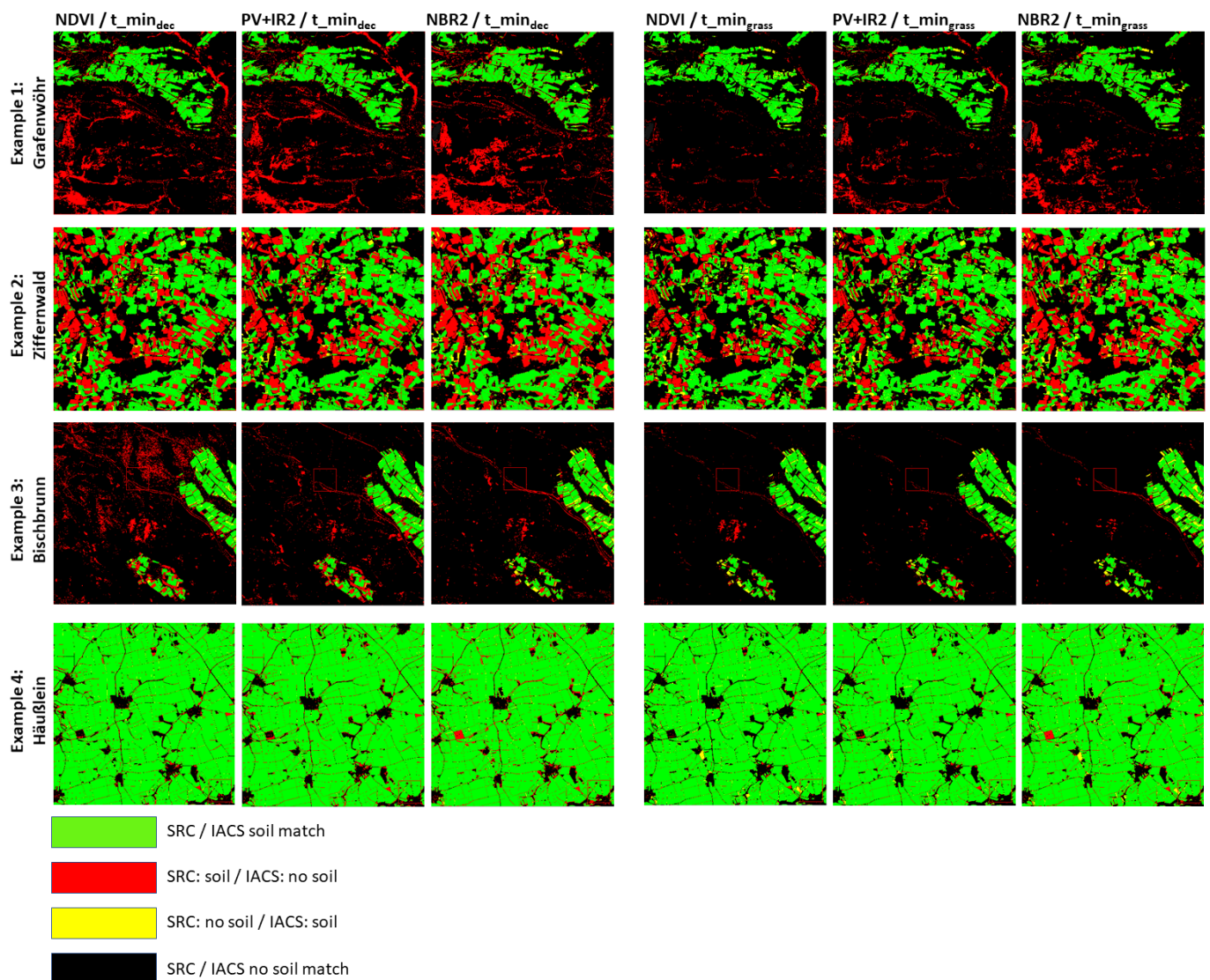
**Figure 9.** Validation of the SRC bare soil mask using the European IACS data. Results are presented as a confusion matrix; numbers are given in (%) except for the overall accuracy (dark grey field).

The overall accuracy for all index/ $t_{min}$  versions is  $>0.9$  for all six SRCs. The  $t_{min_{grass}}$  versions have higher accuracies than  $t_{min_{dec}}$  for all versions. Most important are the matrix values of the middle quarter at the bottom that summarizes the percentage of SRC pixels also covered by the IACS mask (TN as the green number). In the same box, the percentage of SRC pixels not covered by the IACS mask is listed (FN as the red number). While the



mean number of valid pixels is always higher for all  $t_{min_{dec}}$  versions in comparison with the  $t_{min_{grass}}$  version, the TN percentages are lower and more misclassification occurs. In return, the  $t_{min_{grass}}$  version shows less misclassification for all indices. Comparing the three indices, PV+IR2 seems to perform best, and NBR2 has the lowest accuracy values.

However, since the overall accuracy numbers are similar for all index/ $t_{min}$  versions, Figure 10 selects some image examples that mainly visualize problematic areas. Example one shows an area around the military area, Grafenwoehr, Germany, with crops limited by grassland along the river Creussen in the North and a mixture of natural grasslands, open soil surfaces and mixed forests in the Southern part of the image. This example highlights that stricter thresholds such as  $t_{min_{grass}}$  improve the exclusion of NPV, such as grassland and peatland in the military area. However, it also contains a few real bare soil surfaces correctly identified. In general, it can be observed that NBR2 is strong for detecting and excluding pure NPV surfaces (such as along river Creussen in the North) but fail for peat bog areas (e.g., Southern end of the example image). If mixtures of NPV, bare soils and photosynthetic-active vegetation are present, NDVI-based indices perform better.



**Figure 10.** Selective example cases for the spatial agreement between the SRC and IACS bare soil masks.

In the second example, the area is dominated by crops. None of the indices and  $t_{min}$  versions seems to work correctly. The red fields are not in the IACS mask because they

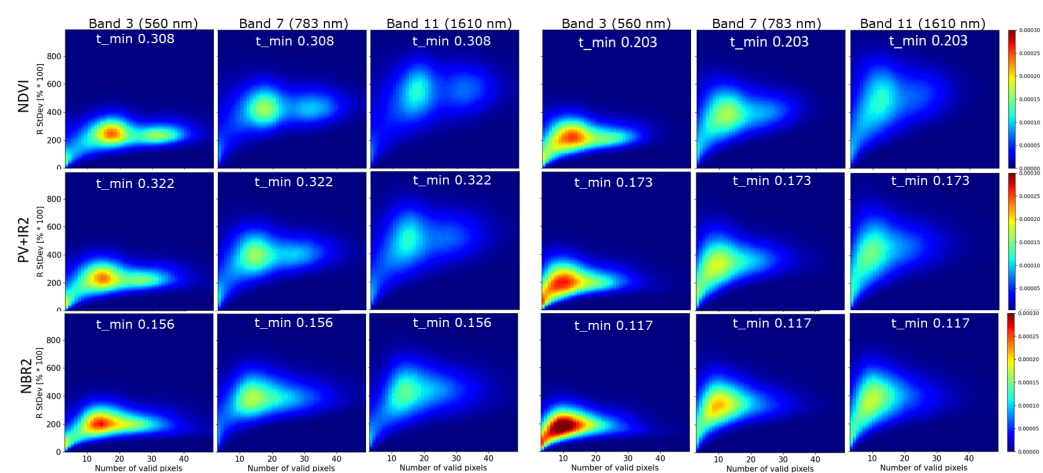
are classified as permanent crops. These crop types are excluded from the IACS mask. However, the area is dominated by hope cultivation characterised by bare soil exposure in the spring month before the green-up phase. Those areas are correctly detected as bare soils within the SRCs.

Example three shows the area Northwest of the small town of Bischbrunn at the Spessart, rich in old deciduous forests (a mixture of oak, beech, larch and other tree types). Specifically, the  $NDVI/t_{min_{dec}}$  version seems to fail here. In all  $t_{min_{grass}}$  versions, deciduous trees could be correctly excluded from the SRCs.

Opposite to the above-described three examples, example four shows purely crops, where SRC and IACS mask mainly match. The  $t_{min_{grass}}$  versions show a few yellow pixels highlighting areas where SRC did not contain bare soils, while IACS expects it.

### 3.6. Spectral Variability of SRCs

In the logic of a perfect spectrum as input for spectral modelling, the soil measured by an EO instrument shall be dry and smooth to be comparable with laboratory measurements of a soil parameter, such as soil organic carbon. In this case, the spectral variability of the SRC would be very low. However, this is very unlikely to occur in natural environments. In Figure 11, the pixel-based spectral variability is shown for each SRC version as a function of the number of valid pixels for three selected S2 bands.

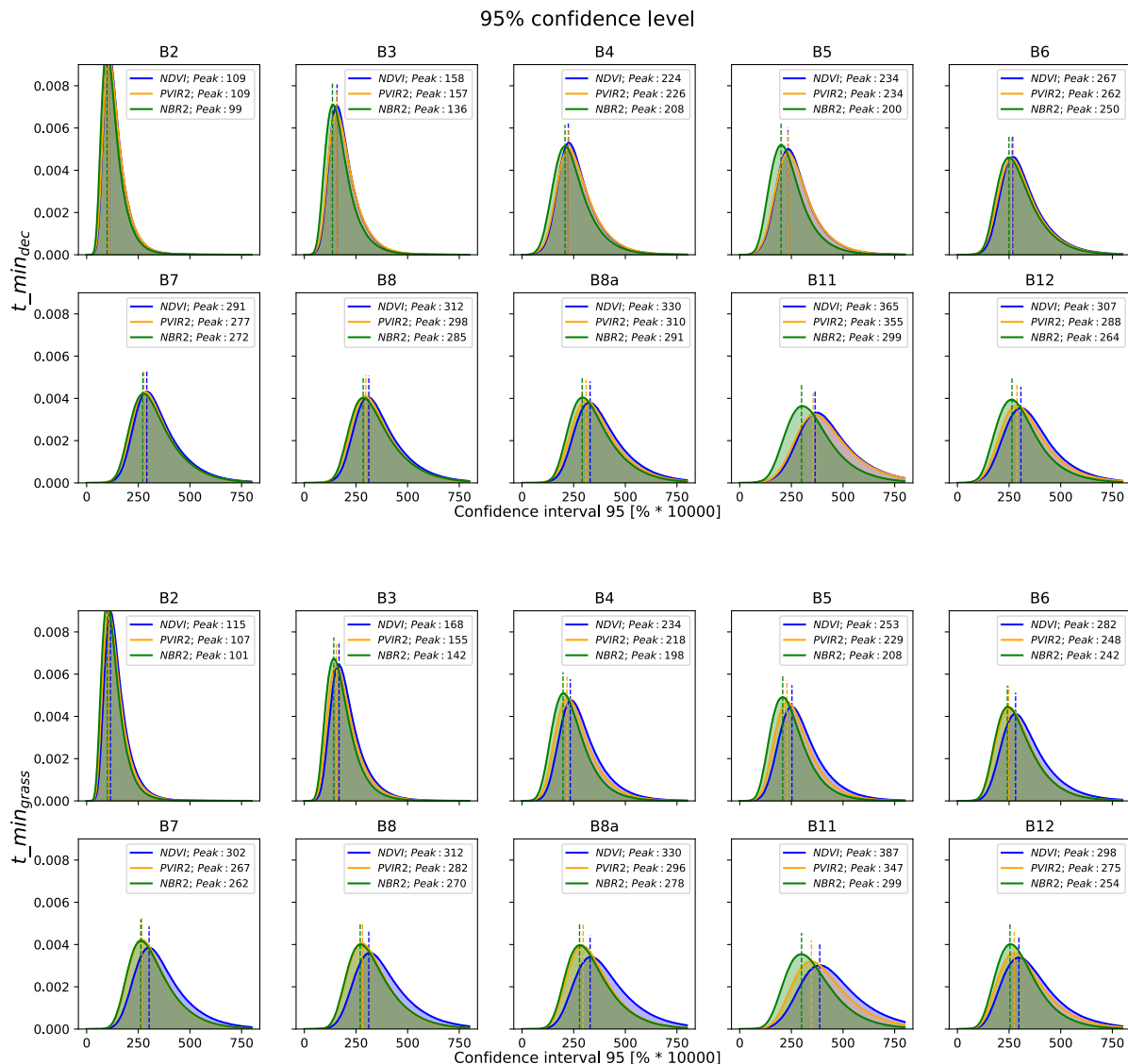


**Figure 11.** Histogram plots for the generated SRC version showing the distribution of the SRC standard deviation  $s$  to the number of valid pixels for selective S2 bands (centre wavelength of the S2 bands are given in brackets according to [46]).

Since  $s$  depends on the mean SRC reflectance, it is low in bands where the reflectance is also low for soils. Band 11 shows higher  $s$  values than band 3. At the same time, the spread of values is higher for higher band numbers. The majority of pixels have a  $s$  around 1–3% for band 3, around 4% for band 7 and 4–6% for band 11. Concerning the number of valid pixels,  $s$  values seem to achieve an equilibrium, where the mean  $s$  values are not increasing anymore with the number of valid pixels. Moreover, the variability is getting smaller, indicating that all soil conditions are integrated into the SRC reflectance. Comparing the three  $t_{min_{dec}}$  index versions, the state of equilibrium gets higher with the bands and shows a slight decrease from NDVI to NBR2. The same can be observed at the  $t_{min_{grass}}$  versions. However, the  $s$  hotspot and the number of valid pixels are lower for all bands.

For a more generic view of the dependency of the number of valid pixels and  $s$ , the 95% CI per used S2 bands for all SRCs is estimated (Figure 12). Image histograms were calculated with a bin size of 75. For comparison, histograms are normalised such that the total area of the histogram equals 1. The CIs differ per spectral index and  $t_{min}$  SRC version, but they simultaneously vary per band. The largest CI differences can be seen at band 11, followed by bands 5, 8a and 12. The CI of NBR2 is the lowest in almost all bands and all  $t_{min}$  SRC versions. The  $t_{min_{grass}}$  SRC versions show a gradient from NDVI with

the highest to NBR2 with the lowest CI. For bands 2, 6, 7, 8 and 8a, the PV+IR2 and NBR2 histograms match. In Figure 7, the CIs are displayed for two example regions. It shows field-to-field differences as well as higher values at the field edges. Further, NBR2 has the lowest CI for both  $t_{min}$  SRC versions.



**Figure 12.** The band-wise histogram plots compare the distribution of the CI for the generated SRC spectral index versions. A kernel density estimate is overlaid to highlight the differences between the spectral indices.

## 4. Discussion

### 4.1. HISET Methodology

HISET is the core novel methodology in this study that enables the comparison of all generated SRC versions. HISET defines a threshold and a quality score that is used for building the SRCs. It is based on the minimum overlap between two LC histograms. In this study, histograms are generated from a spectral index calculated for two LC types that are critical for generating reliable SRCs. HISET compensates for the manual step of percentile definition described in [15,30]. It is fully automated and data-driven instead of expert knowledge-based (e.g., [18]). Expert knowledge about the threshold can be advantageous if a specific area with known temporal-spectral characteristics of the LC types exists, such as in [14]. However, HISET was developed for areas in which this knowledge is missing or

very diverse, such as for continents (e.g., Europe). In this study, it is used for the complete test area covering nine S2 tiles.

The quality of the derived thresholds to separate two LC types is measured by the HISET score. According to Figure A1, the HISET scores are in the same order of magnitude for all tiles and independent of the number of LC input pixels. This positive observation is essential since the size and distribution of the LC pixels can vary a lot per spatial unit (e.g., S2 tiles). In the test site, the percentage of crop pixels varies from 10% to 20% and grassland from 2% to 8%. These numbers can also be much more diverse depending on the region of interest. Huge differences between histograms are compensated by normalising them and using PDFs of a histogram. In preliminary tests, no dependency of the HISET results from the number of LC pixels could be measured. This observation does not exclude the dependency of spectral LC type characteristics of the respective area. However, it is unlikely that this is the case in Bavaria and adjacent areas, where climate and environmental conditions are similar. Future tests should be expanded to other regions with a broader coverage or a wider spread of spectral LC characteristics.

This study uses the European-wide available CLC to have the same LC mapping source for the three countries covered by the test area. Due to the coarse spatial scale of a minimum mapping unit of 25 ha, LC pixels can be a mixture of different cover types. In this context, the LC mapping itself can introduce a level of uncertainty. Moreover, the quality of the CLC survey per country can differ because each EU member state takes responsibility for their own survey. HISET works independently from a specific LC dataset. In future studies, other, more suitable mappings can be tested, such as ESA's new 10-m resolution WorldCover map [60]. In summary, HISET supports various tests since it is independent of the selected LC type, the area size, the spectral index or the multispectral sensor. Thus, it can be implemented in automated processors.

#### 4.2. Selected LC Types and Resulting $T_{min}$ SRC Versions

This study tests deciduous trees and grasslands as input LC for the threshold derivation to minimise disturbing effects such as remaining NPV cover on soils. In general, the derived threshold for deciduous trees ( $t_{min_{dec}}$ ) is less restrictive (relaxed condition) than the threshold derived based on grasslands ( $t_{min_{grass}}$ ) independent from the spectral index. On the other hand, the HISET score for  $t_{min_{dec}}$  is lower than for  $t_{min_{grass}}$ , which does not mean that the resulting SRCs are higher in quality. The spectral evaluation confirms that the spectral similarity between LUCAS and SRC spectra is higher for  $t_{min_{grass}}$  versions, while the spectral mean reflectance difference is almost the same. In fact, due to the larger spectral overlap of dry grasslands with bare soils, the threshold is shifted toward lower index values, and thus, fewer pixels are included in the SRC (see also Table 3). Figure 9 additionally shows higher spatial accuracies for the  $t_{min_{grass}}$  versions, although, or even because, fewer pixels are selected. A lower percentage of permanent crops are included in the  $t_{min_{grass}}$  SRC version (FN ranges between 2% and 4% instead of 6% and 8% for the  $t_{min_{dec}}$  version).

Although the study presents a thoroughly spatial validation of the bare soil masks of the SRCs, spectral mixtures of NPV and bare soils known for crop residues are not considered. Although such an evaluation would be valuable, it is difficult to achieve on a large scale due to missing field data. Ref. [61] has analysed the suitability of the NBR2 index calculated from S2 to exclude NPV disturbances and found that it is poorly correlated with residue cover due to the sensitivity of the NBR2 with soil moisture. It indicates that the complex soil system is influenced by various overlapping effects such as NPV residuals, soil moisture and surface conditions [62], altering the spectral signals. Further analyses are necessary to disentangle such effects to improve the bare soil selection for SRCs or to calibrate the resulting bare soil spectra. The diversity of soil types and texture in such analyses should be considered. Promising techniques are the development of radiative transfer models for soils [63] that are recently expanded to account for smooth and rough soil surfaces [64,65].

#### 4.3. Spectral Index and Resulting $T_{min}$ SRC Versions

Three spectral indices (NDVI, PV+IR2 and NBR2) have been tested to evaluate their performance in generating SRCs. Generally, a quality gradient can be observed from NDVI with the lowest and NBR2 with the highest performance values. The only exception is the validation of the bare soil mask, where the PV+IR2 SRC versions performed best for both  $t_{min}$ , which is confirmed by examples shown in Figure 10. Specifically, if mixtures of green vegetation, NPV and bare soil occur, PV+IR2 seems to perform best. Example 2 (hops cultivation) in Figure 10 produces the same error for all indices and, thus, should not influence this assessment.

By analysing the distribution of the SRC  $s$  concerning the number of valid pixels (Figure 11), spectral variability decreases with the number of valid pixels. The previously described equilibrium state of  $s$  values is the lowest for NBR2 with a gradient from NDVI, PV+IR2 to NBR2. On the one hand, the NBR2 threshold is very restrictive in selecting bare soil pixels resulting in lower  $s$  values. On the other hand, the spread of values is also lower, which is especially visible in the higher band numbers in Figure 11. Further, it can be assumed that using NBR2 for building SRCs needs fewer observations to cover the variability of bare soil conditions.

The comparison of the 95% CI level per band and index clearly shows the higher reliability of the NBR2 in almost every band, which is even more visible for the  $t_{min_{grass}}$  version. The performance of PV+IR2 was higher for the  $t_{min_{grass}}$  version for several bands, such as B3 to B12. Specifically, in B6, B7, B8 and B8a, the CI showed almost complete agreement with the NBR2. This finding is supported by the highest spatial agreement of the PV+IR2/ $t_{min_{grass}}$  SRC with the IACS bare soil mask. It reveals that a spectral index combining VNIR and SWIR bands of multispectral data can lead to reliable SRC generation. A more detailed analysis should be performed for areas with mixtures of NPV, green vegetation and bare soils, where PV+IR2 showed the best performances based on visual inspections, which are shown for the example regions in Figure 11.

This regional analysis showed that a single index for selecting bare soil pixels could be sufficient. It should be tested if the findings described above are transferable to other regions with different environmental conditions, such as in the Mediterranean region with other crop types such as vineyards [35]. Several authors used two or more indices for the bare soil pixel selection, such as [34] for an area in the Northeast of Germany and [19] for Northern Greece. However, it requires an additional index threshold definition for each spectral index. The spectral index thresholds can vary according to the environmental conditions of each biogeographic region. Future studies should test if selecting an index combining the most used wavelength regions of VNIR and SWIR (such as for the PV+IR2) can be an advantage, especially for regions with stronger spectral mixtures between PV, NPV and bare soils.

#### 4.4. Performance Measures

Different spatial and spectral performance measures are developed and used for comparing the generated SRCs versions. The evaluation of the best SRC version might depend on the purpose for which the SRC will be used. Spectral measures such as spectral similarity with reference spectra and information about the variability of the SRC spectra should be especially considered when the SRC is used for spectral soil modelling. However, spatial performance measures and validation can be most meaningful if spatial coverage is essential, such as for statistical assessments of intensively used crop coverage or erosion monitoring.

The histogram analyses of the  $s$  values compared to the number of valid pixels can be used to derive the minimum number of pixels needed to create a robust and reliable SRC spectrum. The number can define the optimal temporal length of an SRC product and, thus, define the conditions for reliable monitoring of a soil parameter. For this study area, 7–15 observations are recommended depending on the used spectral index. Ref. [66] found

a decrease in SOC uncertainty of predictions when the number of scenes per pixel increases by keeping a minimum of more than six observations per pixel.

The CI is provided per test area and pixel and can be used to filter out specific bands or even pixels with very high CIs for improved calibration of the spectral soil models. It could be observed that pixels on field edges have higher CI values. Thus, pixels with high CI values could be excluded from the SRC generation to improve the SRC spectral accuracy. Finally, the band-specific CI can also be an additional quality layer for machine learning or artificial intelligence algorithms considering the importance of specific bands for the respective output.

## 5. Conclusions

In this study, the novel HISET methodology has been tested to calculate spectral index thresholds and quality scores used to build SRCs. Six different versions are generated using three different spectral indices and two threshold versions ( $t_{min}$ ). Several spectral and spatial performance measures are introduced to evaluate the resulting SRCs. The coverage of occurring base soil pixels in the SRCs is validated against the reference from the IACS data. The main findings of the study are:

- HISET is a very flexible, objective and data-driven method for deriving image-based thresholds by reducing the spectral overlap between spectrally similar LC types. Due to its fully automated character, HISET can also be used for large-scale analyses and the regionalisation of thresholds.
- Using grassland as LC input for HISET results in stricter (lower) thresholds and, thus, less bare soil coverage. At the same time, the resulting SRCs are of higher spectral and spatial quality.
- Thresholds derived from the NBR2 index produce SRCs with the best general performance except for the differentiation of urban and crops. If an external urban mask is used to exclude urban areas from the bare soil mask, NBR2 shall be the preferred spectral index.
- The PV+IR2/ $t_{min,grass}$  SRC version performs well for the differentiation between NPV and bare soils and for the separation of urban areas and bare soils and, thus, can be used if no external urban mask is available.
- NDVI always shows the lowest performance and should not be used for further soil parameter retrievals.

Based on the results of this study, two directions for the future emerge. First, it should be analysed if selective pixel-based performance measures such as the CI can help improve soil model calibration and prediction. Second, HISET shall be used to investigate if the findings of this study regarding the spectral index and the  $t_{min}$  versions are also valid for other areas with different environmental conditions and LC characteristics.

**Author Contributions:** Conceptualisation, U.H., P.S. and P.d.; methodology, U.H., P.S., P.d., P.K. and S.Z.; software, P.S., P.d., P.K. and U.H.; validation, U.H., P.K., P.d., S.Z. and M.W.; formal analysis, U.H.; investigation, U.H., P.S. and P.d.; data curation, U.H., P.d., S.Z. and M.W.; writing—original draft preparation, U.H.; visualisation, U.H., P.S., P.d. and P.K.; supervision, U.H., R.M. and P.R.; project administration, U.H. and R.M.; funding acquisition, U.H., R.M. and P.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research was carried out in the frame of the Worldsoils project financed by the European Space Agency and the support is gratefully acknowledged.

**Acknowledgments:** We thank the Bavarian agencies, the Bavarian Environment Agency (LfU) and the Bavarian State Research Center for Agriculture (LfL) for providing the soil databases.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Appendix A. Supplementary Information

### Appendix A.1. HISET Pseudocode

**procedure** HISET(*histogram*<sub>1</sub>, *histogram*<sub>2</sub>)

*HISET*<sub>score</sub>  $\leftarrow \infty$

*threshold*<sub>min</sub>  $\leftarrow \text{None}$

**for** *threshold*  $\leftarrow$  *histogram\_values*<sub>min</sub> to *histogram\_values*<sub>max</sub> **do**

*left*<sub>prop</sub><sub>1</sub>  $\leftarrow$  proportion of entries smaller than *threshold* for *histogram*<sub>1</sub>

*left*<sub>prop</sub><sub>2</sub>  $\leftarrow$  proportion of entries smaller than *threshold* for *histogram*<sub>2</sub>

*left*<sub>min</sub>  $\leftarrow$  MIN(*left*<sub>prop</sub><sub>1</sub>, *left*<sub>prop</sub><sub>2</sub>)

*right*<sub>prop</sub><sub>1</sub>  $\leftarrow$  proportion of entries bigger than *threshold* for *histogram*<sub>1</sub>

*right*<sub>prop</sub><sub>2</sub>  $\leftarrow$  proportion of entries bigger than *threshold* for *histogram*<sub>2</sub>

*right*<sub>min</sub>  $\leftarrow$  MIN(*right*<sub>prop</sub><sub>1</sub>, *right*<sub>prop</sub><sub>2</sub>)

**if** MAX(*left*<sub>min</sub>, *right*<sub>min</sub>) < *HISET*<sub>score</sub> **then**

*HISET*<sub>score</sub>  $\leftarrow$  MAX(*left*<sub>min</sub>, *right*<sub>min</sub>)

*threshold*<sub>min</sub>  $\leftarrow$  *threshold*

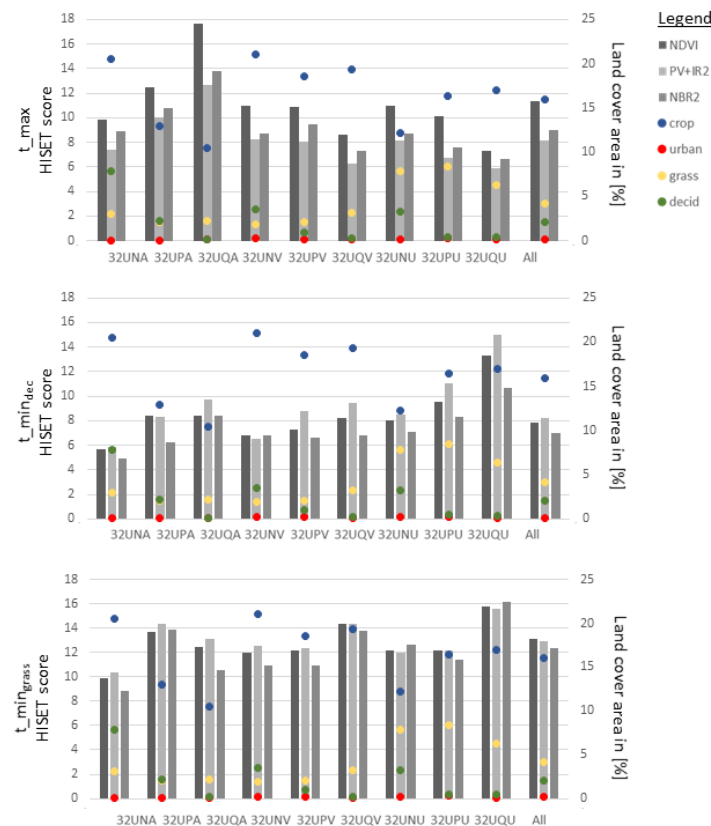
**end if**

**end for**

**return** *HISET*<sub>score</sub>, *threshold*<sub>min</sub>

**end procedure**

### Appendix A.2. HISET Scores Per S2 Tile



**Figure A1.** Performance of the thresholds  $t_{max}$ ,  $t_{min_{dec}}$  and  $t_{min_{grass}}$  derived from three spectral indices (NDVI, PV+IR2, NBR2) to separate between spectrally similar LC types measured by the HISET score for each S2 tile separately and for all tiles together. The lower the HISET score, the better the separability between the respective LC types.

## References

- Vogel, H.J.; Bartke, S.; Daedlow, K.; Helming, K.; Kögel-Knabner, I.; Lang, B.; Rabot, E.; Russell, D.; Stöfel, B.; Weller, U.; et al. A systemic approach for modeling soil functions. *SOIL* **2018**, *4*, 83–92. [[CrossRef](#)]
- Bispo, A.; Andersen, L.; Angers, D.A.; Bernoux, M.; Brossard, M.; Cécillon, L.; Comans, R.N.J.; Harmsen, J.; Jonassen, K.; Lamé, F.; et al. Accounting for Carbon Stocks in Soils and Measuring GHGs Emission Fluxes from Soils: Do We Have the Necessary Standards? *Front. Environ. Sci.* **2017**, *5*. [[CrossRef](#)]
- Löbmann, M.T.; Maring, L.; Prokop, G.; Brils, J.; Bender, J.; Bispo, A.; Helming, K. Systems knowledge for sustainable soil and land management. *Sci. Total Environ.* **2022**, *822*. [[CrossRef](#)] [[PubMed](#)]
- Borrelli, P.; Alewell, C.; Alvarez, P.; Anache, J.A.A.; Baartman, J.; Ballabio, C.; Bezak, N.; Biddoccu, M.; Cerdà, A.; Chalise, D.; et al. Soil erosion modelling: A global review and statistical analysis. *Sci. Total Environ.* **2021**, *780*. [[CrossRef](#)] [[PubMed](#)]
- IPCC. *Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*; Technical Report; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2019. [[CrossRef](#)]
- Poggio, L.; de Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL* **2021**, *7*, 217–240. [[CrossRef](#)]
- Hengl, T.; Miller, M.A.; Križan, J.; Shepherd, K.D.; Sila, A.; Kilibarda, M.; Antonijević, O.; Glušica, L.; Dobermann, A.; Haefele, S.M.; et al. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. *Sci. Rep.* **2021**, *11*, 6130. [[CrossRef](#)] [[PubMed](#)]
- White, J.C.; Wulder, M.A.; Hobart, G.W.; Luther, J.E.; Hermosilla, T.; Griffiths, P.; Coops, N.C.; Hall, R.J.; Hostert, P.; Dyk, A.; et al. Pixel-Based Image Compositing for Large-Area Dense Time Series Applications and Science. *Can. J. Remote Sens.* **2014**, *40*, 192–212. [[CrossRef](#)]
- Griffiths, P.; Nendel, C.; Hostert, P. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sens. Environ.* **2019**, *220*, 135–151. [[CrossRef](#)]
- Hansen, M.C.; Egorov, A.; Roy, D.P.; Potapov, P.; Ju, J.; Turubanova, S.; Kommareddy, I.; Loveland, T.R. Continuous fields of land cover for the conterminous United States using Landsat data: First results from the Web-Enabled Landsat Data (WELD) project. *Remote Sens. Lett.* **2011**, *2*, 279–288. [[CrossRef](#)]
- Marconcini, M.; Metz-Marconcini, A.; Üreyen, S.; Palacios-Lopez, D.; Hanke, W.; Bachofer, F.; Zeidler, J.; Esch, T.; Gorelick, N.; Kakarla, A.; et al. Outlining where humans live—The World Settlement Footprint 2015. *Sci. Data* **2020**, *7*. [[CrossRef](#)]
- Maynard, J.J.; Levi, M.R. Hyper-temporal remote sensing for digital soil mapping: Characterizing soil-vegetation response to climatic variability. *Geoderma* **2017**, *285*, 94–109. [[CrossRef](#)]
- Ying, Q.; Hansen, M.C.; Potapov, P.V.; Tyukavina, A.; Wang, L.; Stehman, S.V.; Moore, R.; Hancher, M. Global bare ground gain from 2000 to 2012 using Landsat imagery. *Remote Sens. Environ.* **2017**, *194*, 161–176. [[CrossRef](#)]
- Demattè, J.A.M.; Fongaro, C.T.; Rizzo, R.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens. Environ.* **2018**, *212*, 161–175. [[CrossRef](#)]
- Rogge, D.; Bauer, A.; Zeidler, J.; Mueller, A.; Esch, T.; Heiden, U. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). *Remote Sens. Environ.* **2018**, *205*, 1–17. [[CrossRef](#)]
- Demattè, J.A.M.; Safanelli, J.L.; Poppiel, R.R.; Rizzo, R.; Silvero, N.E.Q.; de Sousa Mendes, W.; Bonfatti, B.R.; Dotto, A.C.; Salazar, D.F.U.; de Oliveira Mello, F.A.; et al. Bare Earth's Surface Spectra as a Proxy for Soil Resource Monitoring. *Sci. Rep.* **2020**, *10*, 4461. [[CrossRef](#)]
- Roberts, D.; Wilford, J.; Ghattas, O. Exposed soil and mineral map of the Australian continent revealing the land at its barests. *Nat. Commun.* **2019**, *10*, 5297. [[CrossRef](#)]
- Diek, S.; Fornallaz, F.; Schaepman, M.E.; De Jong, R. Barest Pixel Composite for Agricultural Areas Using Landsat Time Series. *Remote Sens.* **2017**, *9*, 1245. [[CrossRef](#)]
- Tziolas, N.; Tsakiridis, N.; Ben-Dor, E.; Theocharis, J.; Zalidis, G. 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.* **2020**, *12*, 1389. [[CrossRef](#)]
- Loiseau, T.; Chen, S.; Mulder, V.L.; Román Dobarco, M.; Richer-de-Forges, A.C.; Lehmann, S.; Bourennane, H.; Saby, N.P.A.; Martin, M.P.; Vaudour, E.; et al. Satellite data integration for soil clay content modelling at a national scale. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101905. [[CrossRef](#)]
- Zepp, S.; Heiden, U.; Bachmann, M.; Wiesmeier, M.; Steining, M.; van Wesemael, B. Estimation of Soil Organic Carbon Contents in Croplands of Bavaria from SCMaP Soil Reflectance Composites. *Remote Sens.* **2021**, *13*, 3141. [[CrossRef](#)]
- Urbina-Salazar, D.; Vaudour, E.; Baghdadi, N.; Ceschia, E.; Richer-de-Forges, A.C.; Lehmann, S.; Arrouays, D. 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.* **2021**, *13*, 5115. [[CrossRef](#)]
- Žižala, D.; Minařík, R.; Zádorová, T. Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. *Remote Sens.* **2019**, *11*, 2947. [[CrossRef](#)]
- Blasch, G.; Spengler, D.; Itzerott, S.; Wessolek, G. Organic Matter Modeling at the Landscape Scale Based on Multitemporal Soil Pattern Analysis Using RapidEye Data. *Remote Sens.* **2015**, *7*, 11125–11150. [[CrossRef](#)]



25. Safanelli, J.L.; Chabrillat, S.; Ben-Dor, E.; Demattê, J.A.M. Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe. *Remote Sens.* **2020**, *12*, 1369. [CrossRef]
26. Silvero, N.E.Q.; Demattê, J.A.M.; Amorim, M.T.A.; dos Santos, N.V.; Rizzo, R.; Safanelli, J.L.; Poppiel, R.R.; de Sousa Mendes, W.; Bonfatti, B.R. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. *Remote Sens. Environ.* **2021**, *252*, 112–117. [CrossRef]
27. Lin, C.; Zhu, A.X.; Wang, Z.; Wang, X.; Ma, R. The refined spatiotemporal representation of soil organic matter based on remote images fusion of Sentinel-2 and Sentinel-3. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *89*, 102–094. [CrossRef]
28. Vaudour, E.; Gomez, C.; Lagacherie, P.; Loiseau, T.; Baghdadi, N.; Urbina-Salazar, D.; Loubet, B.; Arrouays, D. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *96*, 102–277. [CrossRef]
29. Dvorakova, K.; Heiden, U.; van Wesemael, B. Sentinel-2 Exposed Soil Composite for Soil Organic Carbon Prediction. *Remote Sens.* **2021**, *13*, 1791. [CrossRef]
30. Zepp, S.; Jilge, M.; Metz-Marconcini, A.; Heiden, U. The influence of vegetation index thresholding on EO-based assessments of exposed soil masks in Germany between 1984 and 2019. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 366–381. [CrossRef]
31. Mzid, N.; Pignatti, S.; Huang, W.; Casa, R. An Analysis of Bare Soil Occurrence in Arable Croplands for Remote Sensing Topsoil Applications. *Remote Sens.* **2021**, *13*, 474. [CrossRef]
32. Möller, M.; Zepp, S.; Wiesmeier, M.; Gerighausen, H.; Heiden, U. Scale-Specific Prediction of Topsoil Organic Carbon Contents Using Terrain Attributes and SCMaP Soil Reflectance Composites. *Remote Sens.* **2022**, *14*, 2295. [CrossRef]
33. Vaudour, E.; Gholizadeh, A.; Castaldi, F.; Saberioon, M.; Borůvka, L.; Urbina-Salazar, D.; Fouad, Y.; Arrouays, D.; Richer-de Forges, A.C.; et al. Satellite Imagery to Map Topsoil Organic Carbon Content over Cultivated Areas: An Overview. *Remote Sens.* **2022**, *14*, 2917. [CrossRef]
34. Castaldi, F.; Chabrillat, S.; Don, A.; van Wesemael, B. Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects. *Remote Sens.* **2019**, *11*, 2121. [CrossRef]
35. Vaudour, E.; Gomez, C.; Fouad, Y.; Lagacherie, P. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sens. Environ.* **2019**, *223*, 21–33. [CrossRef]
36. Tziolas, N.; Tsakiridis, N.; Chabrillat, S.; Demattê, J.A.M.; Ben-Dor, E.; Gholizadeh, A.; Zalidis, G.; van Wesemael, B. Earth Observation Data-Driven Cropland Soil Monitoring: A Review. *Remote Sens.* **2021**, *13*, 4439. [CrossRef]
37. Esch, T.; Heldens, W.; Hirner, A.; Keil, M.; Marconcini, M.; Roth, A.; Zeidler, J.; Dech, S.; Strano, E. Breaking new ground in mapping human settlements from space—The Global Urban Footprint. *ISPRS J. Photogramm. Remote Sens.* **2017**, *134*, 30–42. [CrossRef]
38. Pesaresi, M.; Ehrlich, D.; Ferri, S.; Florczyk, A.; Carneiro Freire, S.; Halkia, S.; Julea, A.; Kemper, T.; Soille, P.; Syrris, V. *Operating Procedure for the Production of the Global Human Settlement Layer from Landsat Data of the Epochs 1975, 1990, 2000, and 2014*; Technical Report LB-NA-27741-EN-C (print), LB-NA-27741-EN-N (online); JRC: Luxembourg, 2016. [CrossRef]
39. SOILGRIDS Information. Available online: <https://soilgrids.org/> (accessed on 4 July 2022).
40. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [CrossRef]
41. Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* **2010**, *114*, 1747–1755. [CrossRef]
42. Hagolle, O.; Huc, M.; Desjardins, C.; Auer, S.; Richter, R. MAJA Algorithm Theoretical Basis Document, 2017. Available online: <https://zenodo.org/record/1209633#.YxwrRrRBxPY> (accessed on 4 April 2022).
43. Copernicus Land Monitoring Service—CORINE Land Cover (CLC). Available online: <https://land.copernicus.eu/pan-european/corine-land-cover> (accessed on 4 April 2022).
44. Orgiazzi, A.; Ballabio, C.; Panagos, P.; Jones, A.; Fernández-Ugalde, O. LUCAS Soil, the largest expandable soil dataset for Europe: A review. *Eur. J. Soil Sci.* **2017**, *69*, 140–153. [CrossRef]
45. Alonso, K.; Bachmann, M.; Burch, K.; Carmona, E.; Cerra, D.; de los Reyes, R.; Dietrich, D.; Heiden, U.; Hölderlin, A.; Ickes, J.; et al. Data Products, Quality and Validation of the DLR Earth Sensing Imaging Spectrometer (DESI). *Sensors* **2019**, *19*, 4471. [CrossRef]
46. Sentinel-2 Spectral Response Functions (S2-SRF). Available online: [https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset\\_publisher/Wk0TKajilSaR/content/sentinel-2a-spectral-responses](https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/document-library/-/asset_publisher/Wk0TKajilSaR/content/sentinel-2a-spectral-responses) (accessed on 4 July 2022).
47. Hall, D.K.; Riggs, G.A.; Salomonson, V.V. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sens. Environ.* **1995**, *54*, 127–140. [CrossRef]
48. Rouse, J.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS. In *Third Earth Resources Technology Satellite—1 Symposium. Volume I: Technical Presentations*; Freden, S.C., Mercanti, E.P., Becker, M., Eds.; NASASP-351; NASA: Washington, DC, USA, 1974; pp. 309–317.
49. Pereira, J.M.C.; Sá, A.C.L.; Sousa, A.M.O.; Silva, J.M.N.; Santos, T.N.; Carreiras, J.M.B. Spectral characterisation and discrimination of burnt areas. In *Remote Sensing of Large Wildfires: In the European Mediterranean Basin*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 123–138. [CrossRef]
50. Sorenson, P.; Shirliffe, S.; Bedard-Haughn, A. Predictive soil mapping using historic bare soil composite imagery and legacy soil survey data. *Geoderma* **2021**, *401*, 1–11. [CrossRef]

51. Castaldi, F.; Chabrillat, S.; van Wesemael, B. Sampling Strategies for Soil Property Mapping Using Multispectral Sentinel-2 and Hyperspectral EnMAP Satellite Data. *Remote Sens.* **2019**, *11*, 309. [[CrossRef](#)]
52. Vermote, E.F.; Kotchenova, S. Atmospheric correction for the monitoring of land surfaces. *J. Geophys. Res. Atmos.* **2008**, *113*. [[CrossRef](#)]
53. de los Reyes, R.; Langheinrich, M.; Schwind, P.; Richter, R.; Pflug, B.; Bachmann, M.; Müller, R.; Carmona, E.; Zekoll, V.; Reinartz, P. PACO: Python-Based Atmospheric Correction. *Sensors* **2020**, *20*, 1428. [[CrossRef](#)]
54. Melin, F.; Boss, E.; Bulgarelli, B.; Doerffer, R.; Franz, B.A.; Hieronymi, M.; Hu, C.; Kwiatkowska, E.; Neukermans, G.; et al. *Uncertainty in Ocean Colour Remote Sensing*; Technical Report; International Ocean Colour Coordinating Group: Dartmouth, NS, Canada, 2019. [[CrossRef](#)]
55. Reuter, M.; Thomas, W.; Mieruch, S.; Hollmann, R. A Method for Estimating the Sampling Error Applied to CM-SAF Monthly Mean Cloud Fractional Cover Data Retrieved From MSG SEVIRI. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2469–2481. [[CrossRef](#)]
56. Elvidge, C.D. Visible and near infrared reflectance characteristics of dry plant materials. *Int. J. Remote Sens.* **1990**, *11*, 1775–1795. [[CrossRef](#)]
57. Heiden, U.; D'Angelo, P.; Schwind, P.; de los Reyes, R.; Mueller, R. Evaluating Soil Reflectance Composites generated by SCMaP using different Sentinel-2 reflectance data inputs. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 495–498.
58. Kruse, F.; Lefkoff, A.; Boardman, J.; Heidebrecht, K.; Shapiro, A.; Barloon, P.; Goetz, A. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [[CrossRef](#)]
59. Hayslett, H.; Murphy, P. *Statistics; Made Simple Books*; Elsevier Science: Amsterdam, The Netherlands, 2014.
60. Zanaga, D.; Van De Kerchove, R.; De Keersmaecker, W.; Souverijns, N.; Brockmann, C.; Quast, R.; Wevers, J.; Grosu, A.; Paccini, A.; Vergnaud, S.; et al. ESA WorldCover 10 m 2020 v100, 2021. Available online: <https://zenodo.org/record/5571936/export/xm#.YxwtWrRBxPY> (accessed on 4 July 2022).
61. Dvorakova, K.; Shi, P.; Limbourg, Q.; van Wesemael, B. Soil Organic Carbon Mapping from Remote Sensing: The Effect of Crop Residues. *Remote Sens.* **2020**, *12*, 1913. [[CrossRef](#)]
62. Prudnikova, E.; Savin, I. Some Peculiarities of Arable Soil Organic Matter Detection Using Optical Remote Sensing Data. *Remote Sens.* **2021**, *13*. [[CrossRef](#)]
63. Bablet, A.; Vu, P.; Jacquemoud, S.; Viallefont-Robinet, F.; Fabre, S.; Briottet, X.; Sadeghi, M.; Whiting, M.; Baret, F.; Tian, J. MARMIT: A multilayer radiative transfer model of soil reflectance to estimate surface soil moisture content in the solar domain (400–2500 nm). *Remote Sens. Environ.* **2018**, *217*, 1–17. [[CrossRef](#)]
64. Jacquemoud, S.; Dupiau, A.; Briottet, X.; Fabre, S.; Viallefont-Robinet, F. Monitoring soil water content from space in the solar domain: The power of radiative transfer models. In Proceedings of the ESA Living Planet Symposium 2022, Bonn, Germany, 23–27 May 2022.
65. Dupiau, A.; Jacquemoud, S.; Briottet, X.; Viallefont-Robinet, F.; Fabre, S.; Philpot, W.; Biagio, C.D.; Formenti, P. Modeling soil reflectance spectra in the solar domain (400–2500 nm) as a function of moisture content: Improvement of the MARMIT model. In *OSA Optical Sensors and Sensing Congress 2021 (AIS, FTS, HISE, SENSORS, ES)*; Optica Publishing Group: Washington, DC, USA, 2021. [[CrossRef](#)]
66. Dvorakova, K.; Heiden, U.; Pepers, K.; Staats, G.; van Os, G.; van Wesemael, B. Improving soil organic carbon predictions from Sentinel 2 soil composites by assessing surface conditions and uncertainties. *Geoderma* **2022**, *accepted*. [[CrossRef](#)]