SoilSuite for Rwanda, Kenya and Ethiopia

Data and Method Description

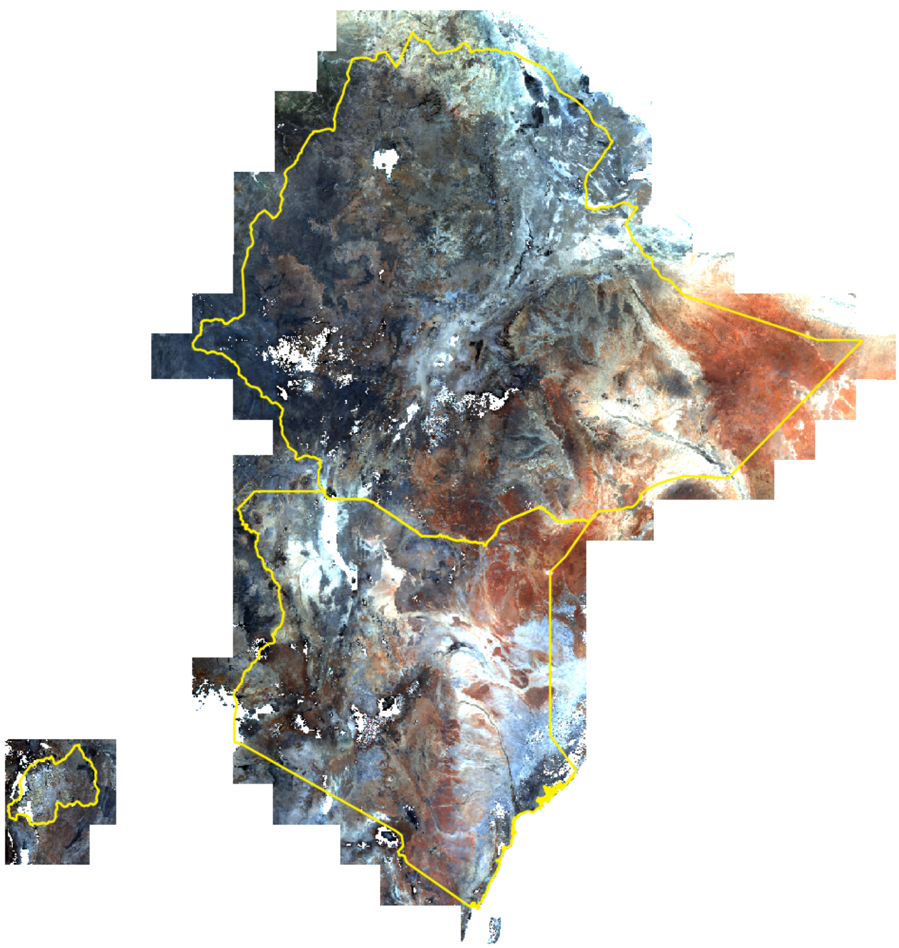
Uta Heiden[a], Pablo d'Angelo[a], Paul Karlshöfer[a], Kevin Kuehl[a]

[a]–German Aerospace Center (DLR), The Remote Sensing Technology Institute (IMF)" Muenchener Str.20, Wessling, 82234, Bavaria, Germany

---

Keywords: bare surface, soil, spectral reflectance, Sentinel-2, soil variability, temporal statistics, composites

---



Email addresses: `uta.heiden@dlr.de` (Uta Heiden), `pablo.angelo@dlr.de` (Pablo d'Angelo), `paul.karlshoefer@dlr.de` (Paul Karlshöfer), `kevin.kuehl@dlr.de` (Kevin Kuehl)

How to cite this document: Heiden, U., d'Angelo, P., Karlshoefer, P., Kuehl, K., 2025. SoilSuite for Rwanda, Kenya and Ethiopia. Technical report, January 2025

1. Purpose of the document

The document provides a preliminary description of the SoilSuite data package published at the EOC Geoservice platform. It contains a short description of the data generation, data access, metadata, potential application and limitations of the SoilSuite data package. It shall enable the reader to evaluate the content and purpose of the data products and shall help to employ the data into further models at the best possible way.

The authors of this data package welcome any feedback from the recipient of the data in order to improve the processors and thus the quality of the data. It is expected that the exchange will take place in a lively and informal manner. Please send any feedback to:

- uta.heiden@dlr.de
- pablo.angelo@dlr.de
- paul.karlshoefer@dlr.de

A detailed publication of the SoilSuite data package is in preparation and will be published soon. In the meantime, please use the following references:

- Karlshoefer P., d'Angelo P., Eberle J., Heiden U., 2025, Evaluation framework for the generation of continental bare surface reflectance composites, Geoderma, Volume 459, `https://doi.org/10.1016/j.geoderma.2025.117340`.
- Heiden U., d'Angelo P., Karlshoefer P., Kuehl K., 2025. SoilSuite for East-Africa. Technical report, January 2025, URL: ?
- SoilSuite (2024): Sentinel-2 4-year (2018-2021) composites at European Scale, German Aerospace Center (DLR), 10.15489/qkud8cudg596.
- Heiden, U., d'Angelo, P., Schwind, P., Karlshöfer, P., Müller, R., Zepp, S., Wiesmeier, M., Reinartz, P., 2022. Soil Reflectance Composites—Improved Thresholding and Performance Evaluation. Remote Sens. 2022, 14, 4526. `https://doi.org/10.3390/rs14184526`. (Transfer of the basic bare soil reflectance composite concept to Sentinel-2 data and evaluation)
- Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., Heiden, U., 2018. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014), Remote Sensing of Environment, 205, 1-17, ISSN 0034-4257, `https://doi.org/10.1016/j.rse.2017.11.004`. (Description of the principle idea of the SoilSuite products based on Landsat data)

2. SoilSuite data package overview

The SoilSuite contains a collection of different image data products that provide information about the spectral and statistical properties of soils from three East-African countries and other bare surfaces such as rocks and sand. It is created using DLR's Soil Composite Mapping Processor (ScMAP), which utilises the Sentinel-2 data archive. SCMaP is a specialised processing chain for detecting and analysing bare soils/surfaces on a large (continental) scale. Bare surface and soil pixels are selected using a combined NDVI and NBR index (PVIR2) that optimises the exclusion of photosynthetically active and non-active vegetation. The index is calculated and applied for each individual pixel. All SoilSuite products are calculated based on the available Sentinel-2 scenes recorded between January 2018 and December 2021. The data package excludes all scenes with a cloud cover of > 80 %. The spectral composite products are calculated from the mean value after extensive removal of clouds, haze and snow effects at both scene and pixel level. The spectral data products are available at a pixel size of 20 m and contain 10 Sentinel-2 bands (B02, B03, B04, B05, B06, B07, B08, B08A, B11, B12).

In the following, all image data products of the SoilSuite are introduced. The quicklook of each data product shows a subset that spans the southern part of Kenya. The full data set covers the area shown in Figure 1.

Bare Surface Reflectance Composite - Mean	
Product abbreviation Data type NoData value Number of bands Pixel size File Format Compression	[*]_SRC.tif Int16 -10.000 10 20 m Cloud optimized GeoTiff (COG) LZW
Formula	$$\overline{R_{\text{bare}}} = \frac{1}{\lvert R_{\text{bare}} \rvert} \sum_{r \in R_{\text{bare}}} r \quad (1)$$ where R_{bare} is the set of valid bare surface reflectances.
Band 1	B2 bare surface reflectance average
Band 2	B3 bare surface reflectance average
Band 3	B4 bare surface reflectance average
Band 4	B5 bare surface reflectance average
Band 5	B6 bare surface reflectance average
Band 6	B7 bare surface reflectance average
Band 7	B8 bare surface reflectance average
Band 8	B8A bare surface reflectance average
Band 9	B11 bare surface reflectance average
Band 10	B12 bare surface reflectance average
General use of the product	The Bare Surface Reflectance Composite (Mean) contains reflectance values scaled between 0 and 10.000. It represents the mean reflectance of all Sentinel-2 bare surface occurrences between 2018 - 2021 excluding urban areas, water bodies, clouds, haze, cloud shadows and snow. This product is one of the main inputs for spectral as well as digital soil mapping approaches.

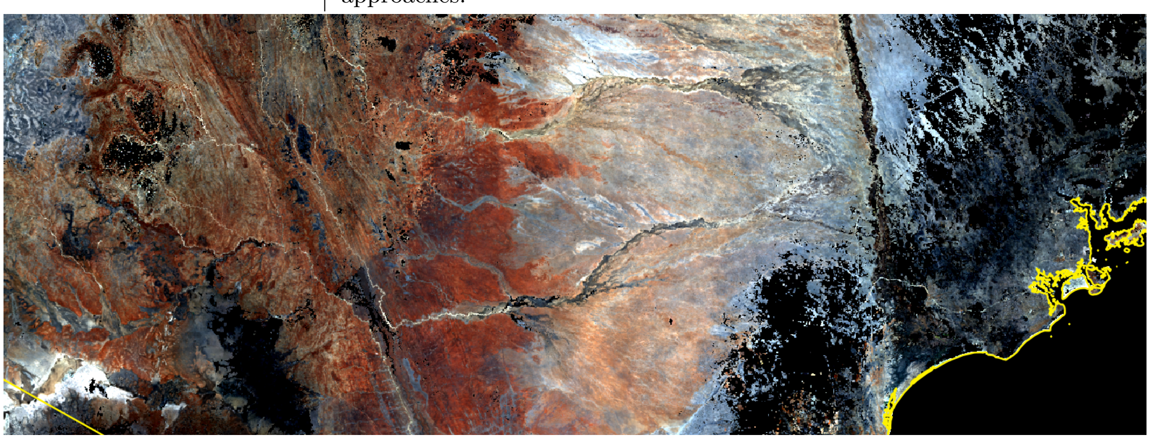


Table 1: The characteristics of the Bare Surface Reflectance Composite - Mean (SRC) product (RGB - B4/B3/B2)

Bare Surface Reflectance Composite - Standard Deviation	
Product abbreviation Data type NoData value Number of bands Pixel size File Format Compression	[*]_SRC-STD.tif Int16 -10 10 20 m Cloud optimized GeoTiff (COG) LZW
Formula	$$s_{R_{\text{bare}}} = \sqrt{\frac{1}{\lvert R_{\text{bare}} \rvert} \sum_{r \in R_{\text{bare}}} (r - \overline{R_{\text{bare}}})^2} \quad (2)$$
Band 1	B2 bare surface reflectance standard deviation
Band 2	B3 bare surface reflectance standard deviation
Band 3	B4 bare surface reflectance standard deviation
Band 4	B5 bare surface reflectance standard deviation
Band 5	B6 bare surface reflectance standard deviation
Band 6	B7 bare surface reflectance standard deviation
Band 7	B8 bare surface reflectance standard deviation
Band 8	B8A bare surface reflectance standard deviation
Band 9	B11 bare surface reflectance standard deviation
Band 10	B12 bare surface reflectance standard deviation
General use of the product	The Bare Surface Reflectance Composite (Std Dev) contains the standard deviation of all Sentinel-2 bare surface reflectance occurrences between 2018 - 2021. It is scaled between 0 and 10.000 for each band and excludes urban areas, water bodies, clouds, haze, cloud shadows and snow. This product quantifies e.g. the spectral dynamic of soils, which might differ according to their physical and chemical properties and thus, can be taken as an additional characteristic of the soil.

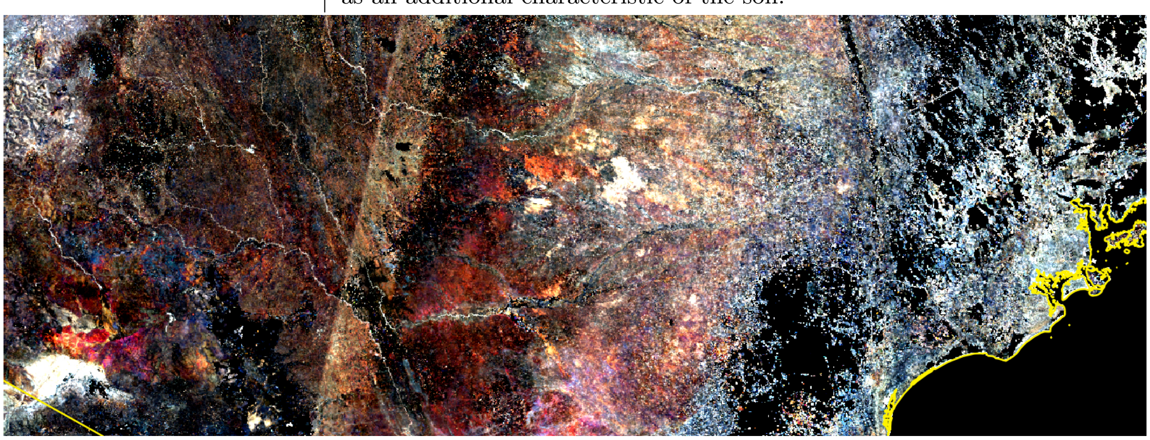
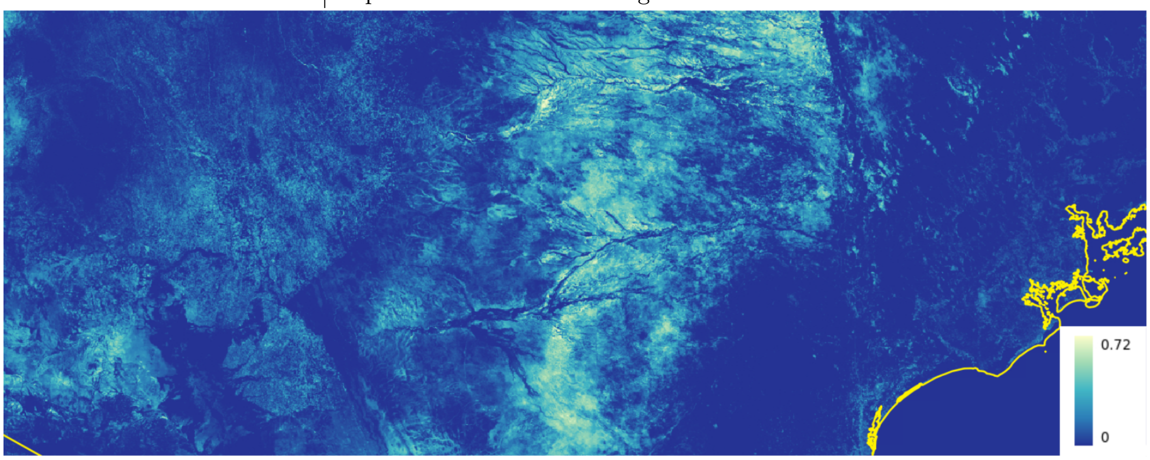
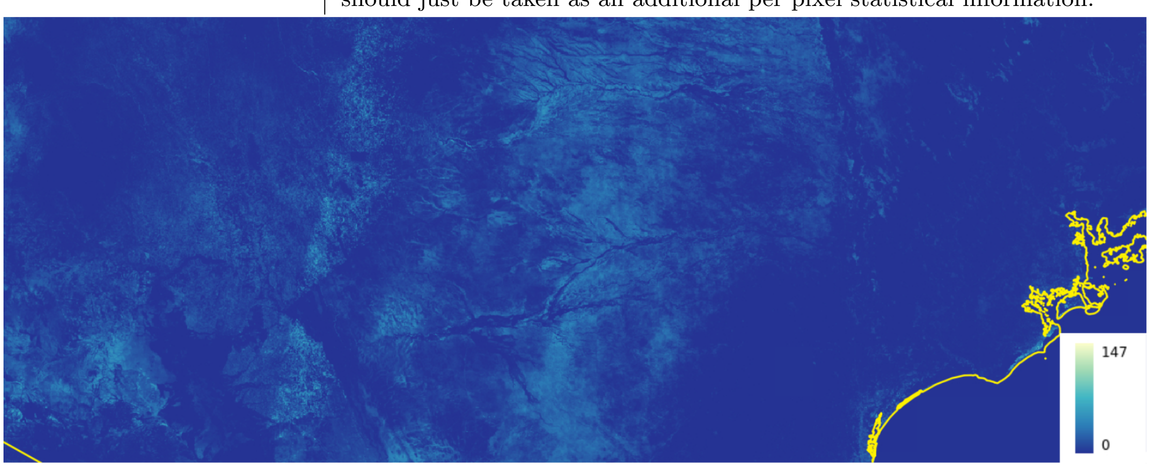


Table 2: The characteristics of the Bare Surface Reflectance Composite - Standard Deviation (SRC-STD) product (RGB - B4/B3/B2)

Bare Surface Statistics	
Product abbreviation	[*]_SFREQ.tif
Data type	Float32
NoData value	-10
Number of bands	3
Pixel size	20 m
File Format	Cloud optimized GeoTiff (COG)
Compression	LZW
Band 1	Bare surface frequency (BSF)
General use of the product	The Bare Surface Frequency (BSF) is scaled between 0 and 1 and quantifies the ratio of bare surface occurrences over the total number of valid Sentinel-2 observations $\lvert R_{\text{bare}}\rvert/\lvert R\rvert$ in the observed time period between 2018 - 2021. The BSF is comparable across the region of interest. In agricultural areas, the BSF gives an indication about the intensity of use of agricultural soils. It also can show areas, where bare surfaces are not expected such as in eroded grassland areas or clear-cuts in forests.



Band 2	Bare surface count $\lvert R_{\text{bare}}\rvert$ (BSC)
General use of the product	The BSC is defined as the number of occurrences detected as bare surface between 2018 - 2021. It is not normalized to the total number of valid Sentinel-2 observations. The striping visible in the BSC product reveals the dependence of the pixel position on the Sentinel-2 ground track that can lead to almost double observations in overlapping orbits. Therefore, the resulting BSC is not comparable across the region of interest and should just be taken as an additional per pixel statistical information.



Band 3	Valid pixel count $|R|$ (VPC)
General use of the product	This VPC contains the total number of valid Sentinel-2 observations in the time period between 2018 - 2021 excluding cloud, haze, cloud shadow, snow, etc..

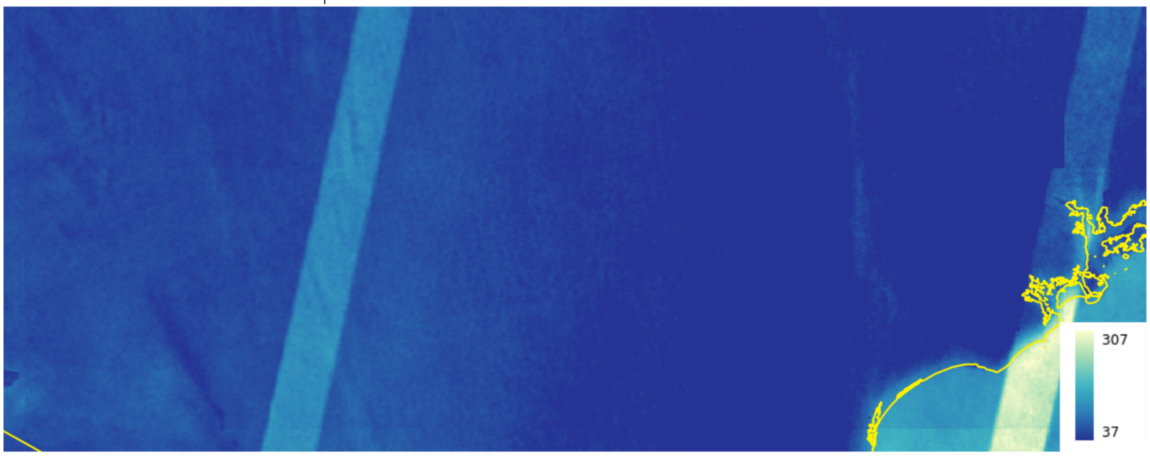


Table 3: The characteristics of the Bare Surface statistics collected in the SFREQ product

Mask	
Product abbreviation Data type NoData value Number of bands Pixel size File Format Compression	[*]_MASK.tif Byte 0 1 20 m Cloud optimized GeoTiff (COG) LZW
Band 1	Three-class mask: (1) Bare surface occurrence, (2) Permanent vegetation, (3) Water bodies, urban areas, roads etc.
General use of the product	The MASK product aggregates simple landcover classes. The value 1 stands for the bare surface occurrence class with pixels that are characterized by an alternation of vegetation and bare soil/surface. Especially areas used for agriculture fall into this mask. It should be noted that a pixel only get 1, if bare surface was detected at least 3 times in the time period to reduce noise in the data. The value 2 describes pixels with permanent vegetation such as coniferous forests and permanent grasslands and excludes areas with bare surface occurrences. As an example, a clear-cut in a forest could appear as value 1 and not value 2, if bare surface is detected at least 3 times. All other pixels receive the value 3, which collects pixels with permanent non-vegetation surface such as water bodies, urban areas, roads, etc.).

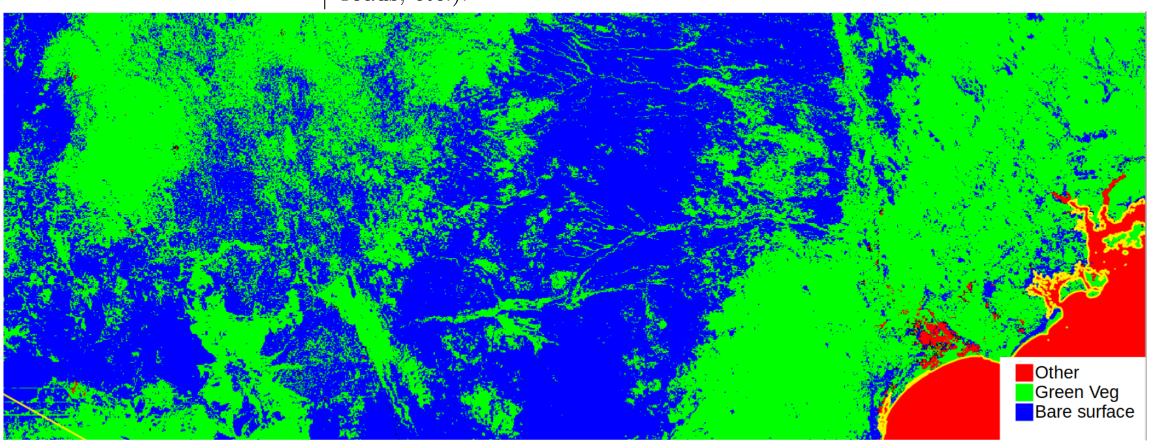


Table 4: The characteristics of the MASK product.

Reflectance Composite - Mean	
Product abbreviation	[*]_MREF.tif
Data type	Int16
NoData value	-10.000
Number of bands	10
Pixel size	20 m
File Format	Cloud optimized GeoTiff (COG)
Compression	LZW
Formula	$$\overline{R} = \frac{1}{\lvert R \rvert} \sum_{r \in R} r \tag{3}$$ where R is the set all valid surface reflectances.
Band 1	B2 reflectance average
Band 2	B3 reflectance average
Band 3	B4 reflectance average
Band 4	B5 reflectance average
Band 5	B6 reflectance average
Band 6	B7 reflectance average
Band 7	B8 reflectance average
Band 8	B8A reflectance average
Band 9	B11 reflectance average
Band 10	B12 reflectance average
General use of the product	The Reflectance Composite (Mean) contains reflectance values per band, scaled between 0 and 10.000. It represents the mean reflectance of all valid Sentinel-2 observations between 2018 - 2021. Valid observations include vegetation, bare and other surfaces and exclude clouds, haze, cloud shadows and snow.

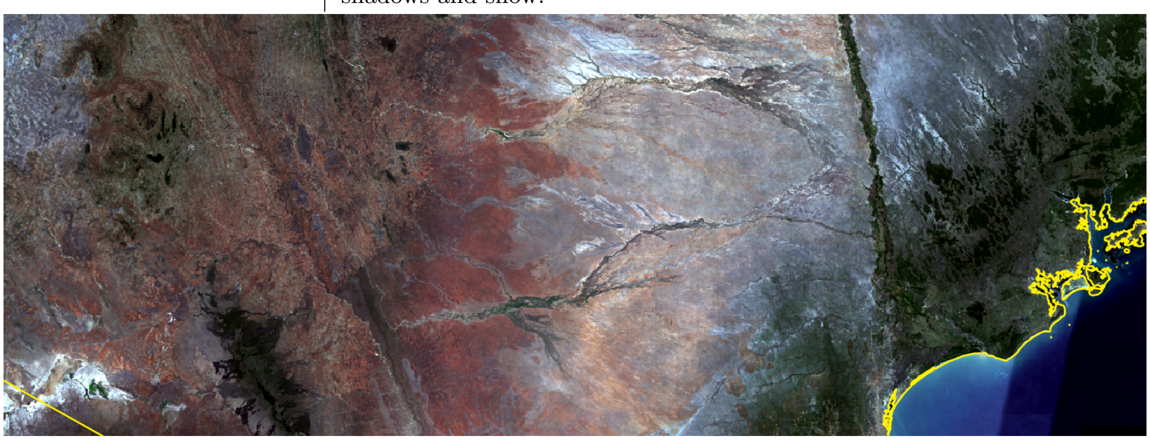


Table 5: The characteristics of the Reflectance Composite - Mean (MREF) product (RGB - B4/B3/B2)

Reflectance Composite - Standard Deviation	
Product abbreviation Data type NoData value Number of bands Pixel size File Format Compression	[*]_MREF-STD.tif Int16 -10.000 10 20 m Cloud optimized GeoTiff (COG) LZW
Formula	$$s_R = \sqrt{\frac{1}{\lvert R \rvert} \sum_{r \in R} (r - \overline{R})^2} \quad (4)$$
Band 1	B2 reflectance standard deviation
Band 2	B3 reflectance standard deviation
Band 3	B4 reflectance standard deviation
Band 4	B5 reflectance standard deviation
Band 5	B6 reflectance standard deviation
Band 6	B7 reflectance standard deviation
Band 7	B8 reflectance standard deviation
Band 8	B8A reflectance standard deviation
Band 9	B11 reflectance standard deviation
Band 10	B12 reflectance standard deviation
General use of the product	The Reflectance Composite (Std Dev) contains the standard deviation per band for all valid Sentinel-2 observations between 2018 - 2021, scaled from 0 to 10.000. The valid observations include vegetation, bare surfaces and other surfaces and it excludes clouds, haze, cloud shadows and snow. The product quantifies the spectral dynamic of all land cover types occurring in the images. As an example, it shows a higher standard deviation in agricultural areas compared to natural grasslands.

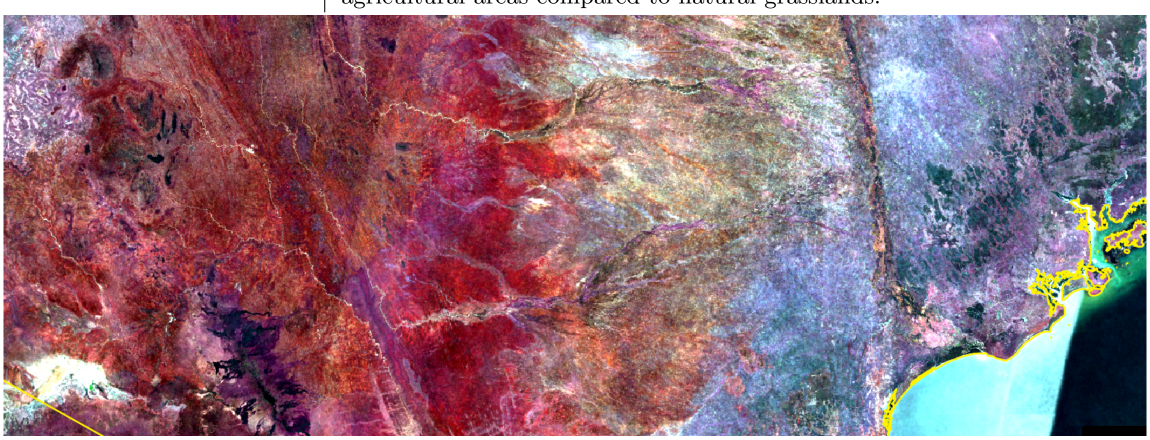


Table 6: The characteristics of the Reflectance Composite - Standard Deviation (MREF-STD) product (RGB - B4/B3/B2)

3. Data access

The SoilSuite is a data set has been published in July 2025 under the license CC-BY-4.0. The data is available from two platforms listed below. The platforms provide several web services and the download capability. Further, the data is described by STAC metadata.

- EOC Geoservice of the DLR: `https://geoservice.dlr.de/web/datasets/soilsuite_afr_4y` or DOI: 10.15489/qkud8cudg596

The processed area comprise the complete coverage of the three African countires Kenya, Ethiopia and Rwanda using tiles of size 200x200 km (see also Figure 1). The SoilSuite Africa product is reprojected from native Sentinel-2 UTM-tiling grid to EPSG:3395 - WGS 84 coordinates.

Figure 1: SoilSuite Africa tiling grid.

All products follow a strict naming convention that shall explain the character of the product, the underlying data base and the main processing parameter. Further, urban areas, water bodies (for SRC* products only) and permanently open areas (for SRC* products only) are masked based on the 10m WorldCover product of ESA (`https://esa-worldcover.org/en`). The naming convention for all files is as follows:

SCOMP_[sensor]_[tileID]_[timeperiodStart][timeperiodEnd]_[L2Ainput]-[spectralindex]
_[SCMaP_version]_[productabbreviation].tif

The tile ID contains the two leading digits of the easting and northing coordinates of the lower left corner. The files of each tile are stored in the <easting>/<northing> subdirectory.

4. Methodology

The text marked in *italic* in section 4 is partly copied from van Wesemael et al. (2024). We take the liberty of doing it because this is an unpublished document intended as data description until the official publication is available.

The generation of the SoilSuite is based on the SCMaP processor that mainly follows the principles published in Rogge et al. (2018). SCMaP has been further enhanced to analyse multitemporal Landsat and Sentinel-2 archives (Heiden et al., 2022) and to develop additional data products informing about the data quality and reliability (Karlshoefer et al., 2025). SCMaP has been primarily developed for temperate and continental climates, where bare soils occur mainly in agricultural areas. It has been extended for climate regimes occurring in East-Africa. SCMaP computes several temporal, spectral and index composites, frequency products and statistical products from a multispectral data stack that can support digital as well as spectral soil mapping and the retrieval of other soil-related information.

The advantage of temporal compositing is to (1) generate results that reflect not just one single point in time but a temporal average that balances out strong seasonal dynamics and preserves permanent differences between regions; (2) in case of bare SRC, it enlarges the area for which direct soil modelling can be applied. The larger this area, the better the prediction results since direct spectral models show mostly better cross-validation results. The challenge is to develop a rigorous pixel selection process to collect undisturbed bare soil pixels by covering an area as large as possible at the same time.

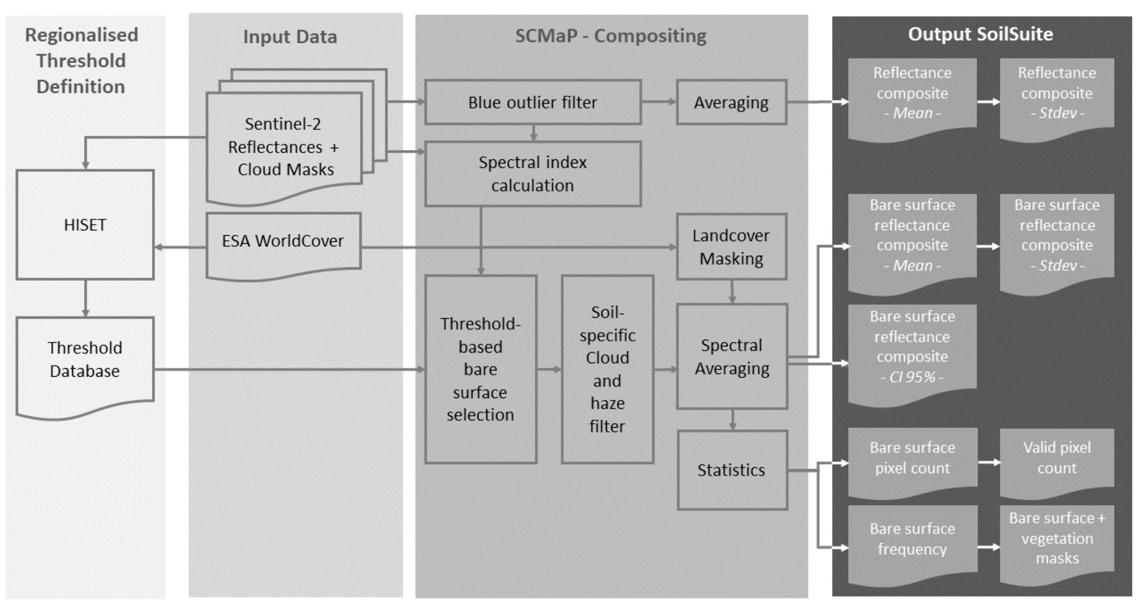


Figure 2: SCMaP flowchart modified after van Wesemael et al. (2024)

The processing is divided in 3 parts: (1) spectral index threshold derivation, which is an automated process to derive pixel-based threshold values to define whether a pixel is bare or not, (2) input data preparation that includes preprocessing of the data archive to surface reflectance and filtering the archive after defined rules and (3) the SCMaP compositing process including a couple of cloud and haze filtering and land cover masking steps.

4.1. Input Sentinel-2 Data

The Table 7 translates the Sentinel-2 bands to band numbers of the SCMaP products. The center wavelengths per band are taken from the ESA webpage `https://sentiwiki.copernicus.eu/web/s2-mission(access:13.06.2024).`

Sentinel-2 band number	**SCMaP band number**	**Center wavelength**	**Status**
B1	–	443 nm	not taken
B2	1	493 nm	resampled to 20 m pixel size
B3	2	560 nm	resampled to 20 m pixel size
B4	3	665 nm	resampled to 20 m pixel size
B5	4	704 nm	
B6	5	741 nm	
B7	6	783 nm	
B8	7	832 nm	resampled to 20 m pixel size
B8A	8	865 nm	
B9	–	945 nm	not taken
B10	–	1374 nm	not taken
B11	9	1614 nm	
B12	10	2202 nm	

Table 7: Used Sentinel-2 bands

The input data stack, which comprises all Sentinel-2 scenes in the target area (see Fig. 1) acquired from 2018 - 2021, was filtered by cloud coverage and sun elevation angle. All scenes with a cloud coverage of > 80 % and a sun elevation angle of $< 20°$ are excluded. The remaining scenes are then converted from the Level 1C format to L2A bottom of atmosphere surface reflectance using the MACCS-ATCOR Joint Algorithm (MAJA) as described in (Hagolle et al., 2018). For the further SCMaP processing, we used the L2A result as well as the MG2 mask informing about the cloud, snow, water and shadow coverage. L2A scenes where the overall mean of the blue band of the whole scene is larger than three times the standard deviation + blue mean of all scenes are ignored. This typically removes the first scene of the MAJA time series.

4.2. Index threshold derivation

Initially, thresholds need to be defined, which are calculated analysing the time period between 2019–2021. The regionalisation of the thresholds is necessary to reflect the differences in the environmental conditions of the region that emerged from a different climate, terrain properties and thus, affects the characteristics of the land cover.

The targeted 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 (based on the MAJA cloud mask MG2) and snow pixels using the normalised difference snow index (NDSI), SCMaP computes a spectral index (see Table 8) for each scene and calculates the minimum index composite. The index composite is used for the index threshold derivation using the Histogram SEparation Threshold (HISET) Method described in (Heiden et al., 2022). *The frequency distribution of spectral index* *values for the agricultural fields are then compared with the frequency of grasslands by overlaying the land cover classes grassland and cropland from the WorldCover data set for 2020 and 2021. These two landcover classes are used to optimize the separation between PV and NPV. In general, the clear separation of both land covers is not possible due to the limited spectral information of Sentinel-2. The histogram separation threshold method Heiden et al. (2022) is then used to derive a tile-based threshold that is further the baseline to select bare surfaces.*

SCMaP uses one spectral index threshold to separate non-vegetated areas from vegetated areas (photosynthetically active and non-active vegetation) per pixel of the temporal data stack. For this purpose, the PV+IR2 index is used combining the Normalized Difference Vegetation Index (NDVI, Rouse et al. (1974)) and the Normalized Burned Ration (NBR, Garcia and Caselles (1991)). The index performance and evaluation is detailed in Heiden et al. (2022). Further, a comparison with the single (non-combined) NBR2 and NDVI is given.

$$\mathrm{PV{+}IR2} = \left(\frac{\mathrm{B8} - \mathrm{B4}}{\mathrm{B8} + \mathrm{B4}}\right) + \left(\frac{\mathrm{B8} - \mathrm{B12}}{\mathrm{B8} + \mathrm{B12}}\right) \tag{5}$$

where Bx defines the Sentinel-2 band number as listed in Table 7.

4.3. Pixel-based cloud and haze filtering

The input L2A time series often contains residual haze and cloud pixels not detected by the L2A processor (and stored in the MG2 mask). These pixel affect the mean and standard deviations of both the overall mean reflectance composite as well as the bare surface reflectance composite. Two filters are used to detect and mask these outliers to exclude them from the further compositing step.

- Blue outlier filter: Pixels with $r_{\mathrm{B2}} > \sigma * NMAD(R_{\mathrm{B2}}) + median(R_{\mathrm{B2}})$ are ignored. R_{B2} are the blue reflectances and σ is a factor for the robust standard deviation filter. As the data itself contains outliers, median and NMAD (normalized median absolute deviation) are used in place of the more common mean and standard deviation.

- NIR/SWIR filter: For almost all soil types the reflectance values of B8 are lower than for B11, whereas cloudy or hazy pixels typically show the opposite behaviour. The NIR/SWIR filter only accepts pixels as bare surface if $(B11-B8)/(B11+B8) \geq 0.02$ is true.

The blue outlier filter is applied to all input pixels with $\sigma = 4$, and again with a stricter threshold of $\sigma = 3$ on the detected bare ground pixels.

4.4. Summary of the processing parameters

In Table 8, the processing parameter for the generation of the described out are listed and explained. In the final process, mean is calculated for the composite mean products (SRC, MREF) and the standard deviation as defined in table 2 and 6 for SRC-STD and MREF-STD respectively. For the spectral confidence interval product of the SRC, the formula in table **??** is used. All other products are generated as described in the tables of the section 2.

Processing parameter	**Value**	**Explanation**
S2 Level 2A reflectance processor	MAJA (Hagolle et al., 2021)	Atmospheric correction optimized for long time series.
S2 Level 2A cloud mask	MG2 (MAJA output)	
Time range (year)	2018 - 2021	
Time range (month)	1 - 12	
Cloud cover filtering Sun elevation filtering	$< 80\%$ $> 20°$	Only scenes with less than 80 percent cloud cover were used during compositing Only scenes with more than 20 degree sun elevation were used during compositing
Bad scene filtering	applied	Detect scenes where the L2A processor produced an invalid cloud mask.
NDSI value		No filtering based on NDSI was applied.
Minimum soil count	3	A pixel falls into the SRC mask, if it is detected three times as bare soil. This mitigates the effect of spurious outliers.
Spectral Index	PV+IR2	(Heiden et al., 2022)
Blue outlier filter	$\sigma = 4$ (mean reflectance), $\sigma = 3$ (soil reflectance)	Cloud and haze pixels not detected by the L2A cloud mask are removed.
NIR/SWIR filter	applied	
SCMaP revision	8925e96c	
Product version	version 1-13	
Output coordinate system	EPSG:3395 - WGS 84	

Table 8: Processing parameter settings

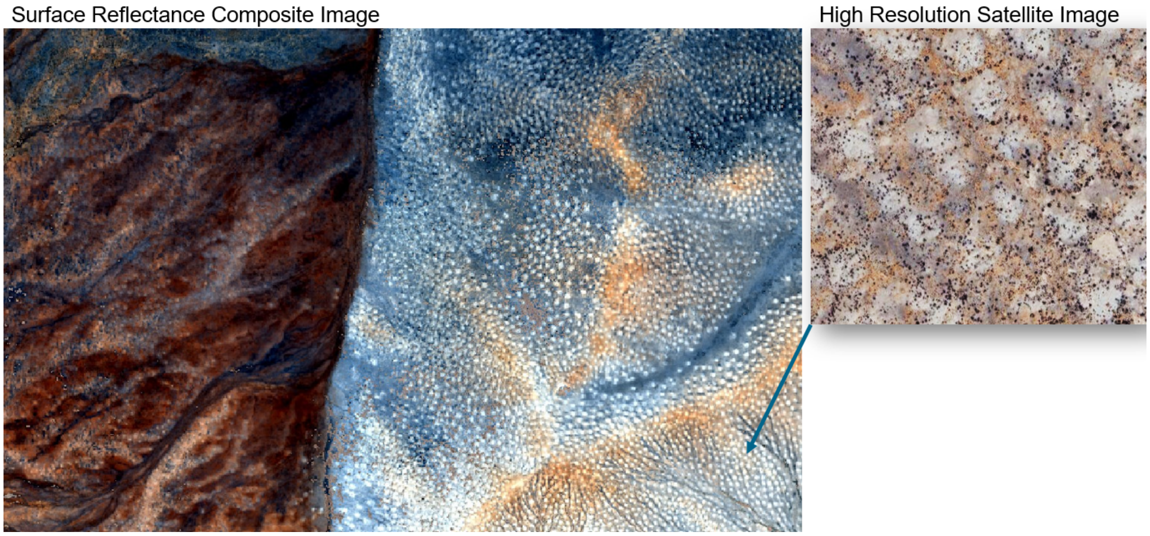


Figure 3: Dense SRC of arid areas in Central Kenya

5. Application and limitation of the data

The SoilSuite data is primarily developed as input for spectral and digital soil parameter models. The products comprise spectral reflectance characteristics of the predominant land cover (*SRC*, *MREF*), their spectral dynamics (*SRC-STD*, *MREF-STD*) and also statistical information (*SFREQ*) about the changes of these land covers. In the following, the meaningfulness, value and limitations are shortly and exemplarily discussed.

5.1. MASK product

The MASK product contains three classes and is basically a by-product of SCMaP to generate the SRC and related products. It allows to detect pixels that show bare surfaces in the observed time period (soil-mask) and it also highlights area, where bare surfaces are never been visible from space in the observed time period (vegetation-mask). In the ESA WorldSoils project, these mask are the basis for assigning pixels to one of the two implemented soil models (digital soil model for vegetated soils and spectral soil model for bare soils). In temperate climates, the soil mask can be taken as approximation for areas that are intensively used for agriculture. However, care should be taken in arid areas, where agricultural activity is completely different due to the different climate conditions and crop types such as orchards or agroforestry. Although these areas are intensively used for food production, they do not show a very distinct change from vegetation to bareness.

5.2. Bare surface frequency

The bare surface frequency is a proxy for the temporal length a surface is bare and thus is very important product for soil erosion, soil degradation and soil carbon sequestration processes. However, this exclude the "activity" of these surfaces, since it does not quantify how often the surface cover changes from bareness to vegetation. It is valid for all areas and can be compared among different regions since it is normalized by the valid pixel count and thus, considers the areas of the overlapping orbits. Preliminary analyses have shown that this product is very significant in digital soil mapping approaches. See 1) Figure 4.

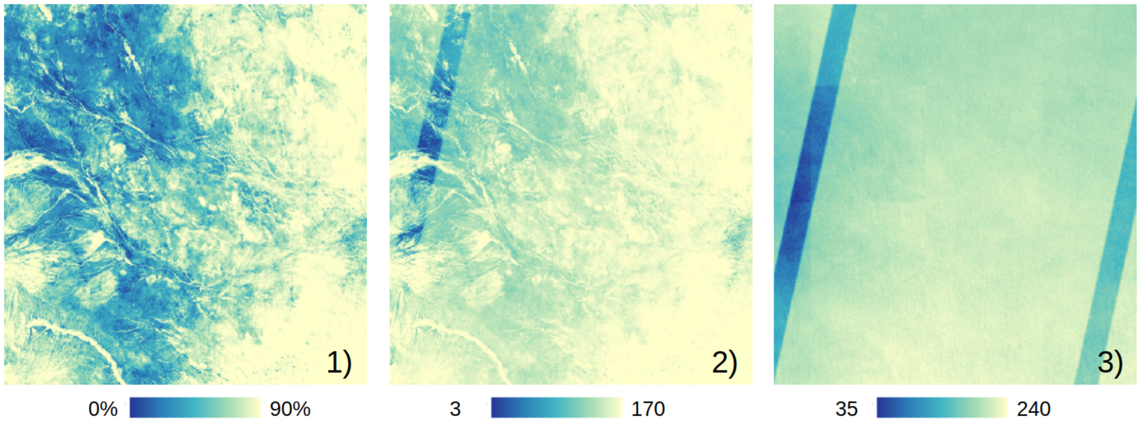


Figure 4: Bare surface frequency (1), bare surface count (2) and valid pixel count (3), central Kenia

5.3. Bare surface count

The bare surface count is in combination with the valid pixel count (see below) the basis for the calculation of the bare surface frequency. It should not be used as covariate in digital soil mapping approaches because it shows the different amount of pixels due to the overlapping orbits in form of stripes (see Figure 4). For this purpose, the bare surface frequency should be used. See 2) Figure 4.

5.4. Bare surface count

The valid pixel count shows the general amount of pixels that are available for the SCMaP processing and excludes clouds, cloud shadows, haze, snow and other data artefacts. This information is the basis for all subsequent products (e.g. SRC and MREF products). It clearly highlights the overlapping orbits in which it happens that the valid pixel count is almost doubled. It is an important information for the data distribution across the complete region and thus, the comparability and reliability of the products. It also highlights areas of permanent higher cloud cover (humid climates) compared to lower permanent cloud cover (arid climates). See 3) Figure 4.

5.5. MREF products

MREF comprise the mean of all valid pixels and MREF-STD the standard deviation of the mean product. Both products contain valuable information for digital soil mapping approaches. An example is given in Figure 5. The MREF-STD reveals information about the spectral dynamic of the landscape. The reddish and whitish areas are characterised by a very high dynamic compared to the greenish grassland areas along the rivers. Due to the different soil of the peatbog area (histosol), the crop type and the dynamic of the agricultural cycle is different from the whitish areas. There is a high potential of this product for soil mapping approaches and other land cover classification approaches.

It should be noted that the MREF-STD product is affected by the overlapping orbits and thus show minor striping effects as shown for the SFREQ valid pixel cound (band 3). The standard deviation values are correct of the complete area and stripes are not errors. However, additional preprocessing is necessary in order to reduce these BRDF effects and produce a comparable product also for the orbit overlaps. This is work in progress and not solved yet. For AI or DL models it is expected that this effect can be learned by introducing both, the valid pixel count as well as the standard deviation product. The striping effects are very minor and barely visible.

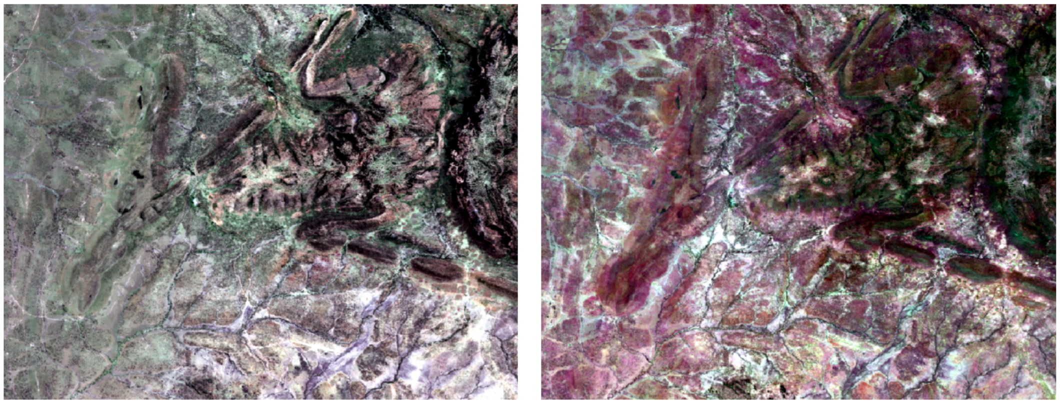


Figure 5: MREF products for a subset in Kenia; especially the MREF-STD product uncovers structures otherwise hidden; RGB: B4/B3/B2

5.6. SRC products

One of the main products of the SCMaP processor is the bare surface reflectance composite (**SRC**) that collects all bare surface pixels in a given time period using the index listed in Table 8. The product is a very important for soil parameter prediction models that directly correlate the spectral reflectance of the soil/surface with the soil parameter from ground data. It is therefore essential to collect just purely bare surfaces and sorting out pixels with disturbances from green and dry vegetation as well as from remaining clouds and haze. The spectral index threshold definition technique HISET is especially adapted to those disturbances by being very strict in the bare surface selection (low index threshold).

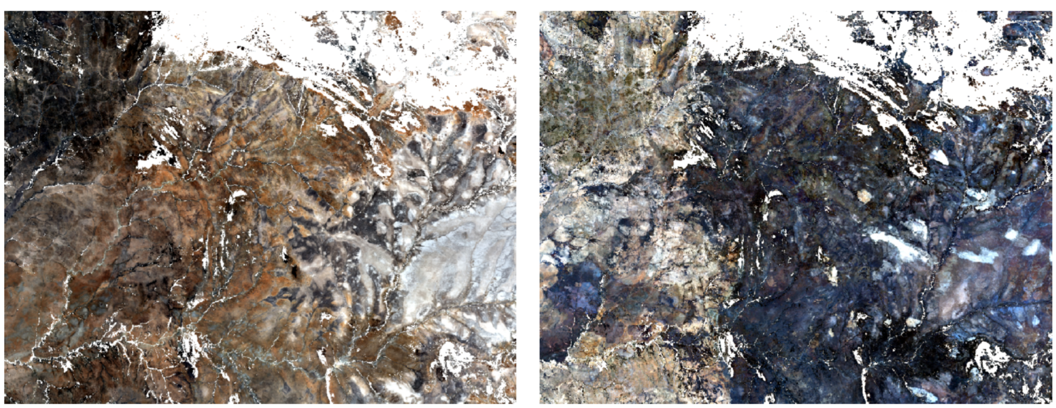


Figure 6: The SRC products contain pixels that covering the bare appearance of soils, for all white areas (no data), the soils are always covered with vegetation (such as for forest, grasslands, etc.); the left image shows the mean of all valid surface/soil pixels, the right image (standard deviation of the mean) reveal the temporal behavior of the different soil types, visualized in the RGB bands.

Care should be taken for areas that are characterised by small agricultural fields with a mixture of trees and low vegetation smaller than the 3 times the spatial resolution (3 x 3 pixel window). Orchards and agroforestry areas are also affected by spectral mixtures and thus, are mainly not included in the SRC products. They are correctly excluded due to spectral mixtures that would distort the pure soil spectral reflectance signal. This might happen in several areas of the Mediterranean (e.g. Southern Spain and Greece), where fields are often small or there are olive groves. In these areas, only few pixels are available for the SRC product (Figure 6). Care should be taken when checking the completeness of the SRC with the usual Google Maps because Google may show just one single snapshot in time and can deviate from the real bareness in the analysed time period.

The pixel number for the SRC could be increased by increasing the spectral index threshold. However, this would integrate spectral information from green and dry vegetation and can have a negative impact on the soil modeling and mapping. Our suggestion is to test and use the MREF and the MREF-STD as alternative products in these areas for any soil modeling task.

The idea for generating a standard deviation product from SRC (**SRC-STD**) is to quantify the different dynamics of surfaces / soils due to their varying mineral composition. Very clay-rich soil holds moisture much longer than a sandy soil and therefore has a different spectral dynamic that is captured in the SRC-STD product. It is an experimental product that can be tested for spectral as well as digital soil mapping approaches. The coverage is the same as for the SRC product.

We also provide the 95% confidence interval of the SRC product (**SRC-CI95**). It contains the half-width of the 95% confidence interval (CI) of the SRC that contains Sentinel-2 reflectance values of all bare surface occurrences collected between 2018 - 2021. A large confidence interval conveys large uncertainty in the SRC value and a small interval indicates a high degree of certainty. Thus, the visible stripes coincides with the lower number of valid pixels (see product SFREQ-VPC), which makes the SRC measurement less reliable. Moreover, higher values (high uncertainty) appear at the edges of agricultural fields, where pixels are more affected by spectral mixtures and lower number of bare surface occurrences. As an application example, it could be explored, if the filtering of pixels with a very low uncertainty (high confidence interval) from soil parameter model generation leads to more robust models.

References

M. L. Garcia and V. Caselles. Mapping burns and natural reforestation using thematic mapper data. *Geocarto International*, 6(1):31–37, 1991. doi: 10.1080/10106049109354290. URL `https://doi.org/10.1080/10106049109354290`.

O. Hagolle, M. Huc, C. Desjardins, S. Auer, and R. Richter. Maja algorithm theoretical basis document, Mar. 2018. URL `https://doi.org/10.5281/zenodo.1209633`.

O. Hagolle, J. Colin, S. Coustance, P. Kettig, P. D'Angelo, S. Auer, G. Doxani, and C. Desjardins. Sentinel-2 surface reflectance products generated by cnes and dlr: Methods, validation and applications. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2021:9–15, 2021. doi: 10.5194/isprs-annals-V-1-2021-9-2021. URL `https://isprs-annals.copernicus.org/articles/V-1-2021/9/2021/`.

U. Heiden, P. d'Angelo, P. Schwind, P. Karlshöfer, R. Müller, S. Zepp, M. Wiesmeier, and P. Reinartz. Soil reflectance composites—improved thresholding and performance evaluation. *Remote Sensing*, 14, 9 2022. ISSN 2072-4292. doi: 10.3390/rs14184526.

P. Karlshoefer, P. d'Angelo, J. Eberle, and U. Heiden. Evaluation framework for the generation of continental bare surface reflectance composites. *Geoderma*, 459:117340, 2025. ISSN 0016-7061. doi: https://doi.org/10.1016/j.geoderma.2025.117340. URL `https://www.sciencedirect.com/science/article/pii/S0016706125001788`.

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

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

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