Development of a HAMSTER: Hyperspectral Albedo Maps dataset with high Spatial and TEmporal Resolution

Giulia Roccetti[1,2], Luca Bugliaro[3], Felix Gödde[1], Claudia Emde[1], Ulrich Hamann[4], Mihail Manev[1], Michael Fritz Sterzik[2], and Cedric Wehrum[1]

[1]Ludwig-Maximilian University of Munich, Geschwister-Scholl-Platz 1, 80539, Munich, Germany
[2]European Southern Observatory, Karl-Schwarzschild-Straße 2, 85748, Garching bei München, Germany
[3]Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[4]MeteoSwiss, Locarno Monti, Switzerland

Correspondence: Giulia Roccetti (giulia.roccetti@eso.org)

Abstract. Surface albedo is an important parameter in radiative transfer simulations of the Earth's system, as it is funda-mental to correctly calculate the energy budget of the planet. The Moderate Resolution Imaging Spectroradiometer (MODIS) instruments on NASA's Terra and Aqua satellites continuously monitor daily and yearly changes in reflection at the planetary surface. The MODIS Surface Reflectance dataset (MCD43C3, Version 6.1) gives detailed albedo maps in seven different spec-tral bands in the visible and near-infrared range. These albedo maps allow us to classify different Lambertian surface types and their seasonal and yearly variability and change, albeit only in seven spectral bands. However, a complete set of albedo maps covering the entire wavelength range is required to simulate radiance spectra, and to correctly retrieve atmospheric and cloud properties from Earth's remote sensing. We use a Principal Component Analysis (PCA) regression algorithm to generate hyperspectral albedo maps of Earth. Combining different datasets of hyperspectral reflectance laboratory measurements for various dry soils, vegetation surfaces, and mixtures of both, we reconstruct the albedo maps in the entire wavelength range from 400 to 2500 nm. The PCA method is trained with a 10-years average of MODIS data for different times of the year. We obtain hyperspectral albedo maps with a spatial resolution of 0.05° in latitude and longitude, a spectral resolution of 10 nm, and a temporal resolution of 8 days. Using the hyperspectral albedo maps, we estimate the spectral profiles of different land surfaces, such as forests, deserts, cities and icy surfaces, and study their seasonal variability. These albedo maps shall enable to refine calculations of Earth's energy budget, its seasonal variability, and improve climate simulations.

1 Introduction

The surface albedo of the planet plays a crucial role within the climate system, governing the proportion of reflected solar light over the incoming solar radiation at the surface. This holds significant importance as it effectively regulates Earth's surface energy budget (Liang et al., 2010; He et al., 2014). The role of albedo extends to climate regulation, with snow and ice albedo feedback exerting a significant influence on climate change dynamics. Snow and ice possess much higher reflectivity compared to the surfaces they overlay. As temperatures rise, the diminishing extent of snow and ice cover leads to a decline in the planet's albedo. Consequently, this intensifies surface warming through a positive feedback mechanism.

Land surface albedo displays remarkable variability, both spatially and temporally. Notable fluctuations in surface albedo coincide with changes in land cover and surface conditions, including factors like vegetation (Loarie et al., 2011; Lyons et al.,
25 2008), snow (He et al., 2013), soil moisture (Govaerts and Lattanzio, 2008; Zhu et al., 2011), and urban development (Offerle et al., 2005). In addition, soil and vegetation surfaces show different reflectance behaviours as a function of wavelength, which are usually not incorporated in Earth system models (ESMs).

In the last decades, the advancement of satellite remote sensing techniques has enabled more accurate monitoring of Earth's surface, enhancing radiative transfer and climate models. This progress allows for continuous acquisition of extensive land sur-
30 face observation data. However, climate models still struggle to capture albedo's temporal and spatial variations. In particular, global and regional climate models often necessitate albedo products with absolute accuracy of 0.02–0.03 (Sellers et al., 1995; He et al., 2014). Zhang et al. (2010) compared Moderate Resolution Imaging Spectroradiometer (MODIS) albedo products with the Coupled Model Intercomparison Project Phase 3 (CMIP3) model results spanning 2000 to 2008, revealing discrepancies in globally averaged albedo of up to 0.06. In addition, validation of different satellites land surface products shows global
35 absolute differences up to 0.02–0.06, with the largest variation occasionally exceeding 0.1 (Shao et al., 2021).

The divergence of different albedo products is not the only source of uncertainty in ESM radiative transfer calculations. Most ESMs use a two-stream approach for the land component, where soil albedo has fixed values in two spectral broadband regions: the photosynthetically active radiation band (PAR, 400–700 nm) and the near infrared band (NIR, 700–2500 nm). However, broadband radiative transfer schemes show a strong spectral discontinuity at 700 nm (Braghiere et al., 2023). This divergence
40 in surface reflectance propagates into other radiative partitioning terms, such as absorptance and transmittance at the top of atmosphere (TOA).

More in general, in cloud-free simulations over land, the dominant factor impacting the TOA VIS/NIR radiance is surface reflection (Vidot and Borbás, 2014). Varied surface optical properties exhibit distinct spectral signatures contingent on the type of surface. Furthermore, within the VIS/NIR range, surface optical properties showcase a robust geometrical reliance that changes
45 according to solar and satellite directions. To elucidate the spectral reliance of the surface, an assumption of Lambertian behaviour can be made, implying isotropic luminance regardless of the viewer's angle. The albedo quantifies the proportion of reflected light under the assumption of isotropic radiation reflection.

Polar-orbiting satellites, like NASA's Terra and Aqua, provide global albedo maps, which are vital for spectral, temporal, and spatial global albedo assessment. The MODIS instrument, on NASA's Terra and Aqua satellites, offers coverage of Earth's sur-
50 face every 1 to 2 days, enhancing our understanding of terrestrial, oceanic, and atmospheric processes. In the VIS/NIR range, MODIS features seven spectral bands delivering data on land surface characteristics. However, radiative transfer simulations demand precise radiance calculations across all wavelengths, which necessitates hyperspectral albedo maps. Such comprehensive data are lacking due to the impracticality of obtaining albedo maps from satellites for every wavelength. As a result, various assumptions are incorporated into radiative transfer codes to overcome this lack of information.
55 The MODIS albedo measurements are derived simultaneously from the Bidirectional Reflectance Distribution Function (BRDF), depicting radiation discrepancies resulting from the scattering (anisotropy) of individual pixels. This methodology relies on multi-date, atmospherically corrected, and cloud-cleared input data obtained over 16 day intervals. The spatial resolution is set

at 0.05° in latitude and longitude (equivalent to 5.6 km at the equator) using the Climate Modeling Grid (CMG). To derive climatological averages, the MODIS albedo datasets are averaged over a 10-year span in steps of 8 days, and albedo maps are

60 then extrapolated in time using these metrics.

In this work, we introduce a novel methodology for creating hyperspectral albedo maps based on the seven representative bands of the MODIS instrument. Using a Principal Component Analysis (PCA) regression approach, we combine different soils, rocks and vegetation datasets representative of different parts of the world, along with the Lambertian surface albedo maps from the MCD43C3 product (Schaaf and Wang, 2021) (version 6.1) derived from the Terra and Aqua satellites. These

65 maps cover the seven bandpasses relevant for land surface albedos. Employing a PCA algorithm, as previously done by Vidot and Borbás (2014) and Jiang and Fang (2019), enables us to reduce the problem's high dimensionality and to generate new albedo maps by interpolating between the measured bandpasses.

These hyperspectral albedo maps of Lambertian surfaces hold significance in various climate and radiative transfer models for Earth's system. Using an ESM with coupled atmosphere-land simulations, Braghiere et al. (2023) demonstrated the impact of

70 making simplistic assumptions on albedo maps, using only two broadband values, compared to hyperspectral albedo maps. They estimated a divergence of the radiative forcing of 3.55 W m^{-2}, which subsequently impacts the net solar flux at TOA (> 3.3 W m^{-2}), cloudiness, rainfall, surface temperature and latent heat fluxes. Braghiere et al. (2023) also highlights the impact of implementing hyperspectral albedo maps in regional models, where differences in latent heat can be higher than 5 W m^{-2}, showing the implications for regional climate variability and prediction of extreme events.

75 In the near future, the launch of new satellite missions like NASA's Earth Surface Mineral Dust Source Investigation (EMIT), will allow obtaining hyperspectral soils and vegetations data and to benchmark the accuracy of the model generated hyperspectral maps.

2 Data and Methods

2.1 MODIS surface albedo climatology

80 NASA's MODIS instruments (Salomonson et al., 1989) aboard the Terra and Aqua satellites, launched in 1999 and 2002, observe the Earth in 36 spectral bands. Two channels, centred at 645 and 858 nm (see Tab. 1), have a spatial resolution of 250 m and five channels (centred at 469, 555, 1240, 1640, 2130 nm), including three in the shortwave infrared, have a spatial resolution of 500 m. All other channels have a resolution of 1 km.

The science dataset MCD43C3 (Schaaf and Wang, 2021), Version 6.1, is the combined Aqua+Terra L3 MODIS Surface Re-

85 flectance product and provides daily global estimates of directional hemispherical surface reflectance (black-sky albedo) and bihemispherical surface reflectance (white-sky albedo) for the seven (2+5) MODIS bands mentioned above, and for three broadband spectral intervals (visible 300–700 nm, near-infrared 700-5000 nm and shortwave 300–5000 nm) with a spatial resolution of 0.05° in latitude and longitude (corresponding to 5600 m at the equator). MODIS cloud-free observations are collected over 16 days and corrected for atmospheric gases and aerosols to derive surface albedo for land pixels (water bodies are not consid-

90 ered). Data are temporally weighted to the ninth day of the retrieval period, and this day appears in the file name. Each surface

Table 1. Spectral bands of MODIS in the VIS and NIR providing information about the land surface. For each band, we specify the central wavelength and the bandwidth.

band	central λ [nm]	bandwidth [nm]
1	645	620–670
2	858	841–876
3	469	459–479
4	555	545–565
5	1240	1230–1250
6	1640	1628–1652
7	2130	2105–2155

reflectance pixel contains the best possible measurement of the period, selected on the basis of high observation coverage, low view angle, the absence of clouds or cloud shadow, and aerosol loading. Usually, due to the sun-synchronous orbit of the Terra and Aqua satellites (equatorial crossing times at 10:30 AM and 1:30 PM respectively), only pixels with local solar noon zenith angle up to approx. 80° are provided with an albedo value.

95 The MODIS land surface products have been validated against in situ measurements and other satellite-based land surface albedo. Globally, the MODIS product is less accurate for high solar zenith angles (Sánchez-Zapero et al., 2023). We compile a black-sky and a white-sky albedo climatology for the seven MODIS spectral bands and the three broadband intervals starting from the MCD43C3 product from 2013 to 2022 in steps of 8 days, starting on January 1, i.e., from day of the year (DOY) 1 to DOY 361. This results in 46 albedo maps for each albedo type and spectral band with a spatial resolution of 0.05° in latitude

100 and longitude and a temporal resolution of 8 days. The aim is a complete surface albedo climatology map for all grid boxes that are illuminated by the Sun, i.e., up to a local solar noon zenith angle of 90°. Pixels that are in the dark (i.e., the Sun is always below the horizon) during the entire DOY are left unfilled. For the computation of the climatology, we proceed in the following way:

1. First, we select MCD43C3 albedo retrievals with albedo quality between 0 and 4 (see Tab. 2) and compute the mean
105 value of the surface albedo for every grid box over the 10 years for a given DOY. After this averaging procedure, some pixels remain unfilled due, e.g., to cloudiness and because of the local solar noon zenith angle constraints mentioned above.

2. Thus, for every DOY we fill the missing values with the mean of the albedo at DOY-$n \cdot 8$ and DOY+$n \cdot 8$ (temporal averages obtained in 1), with $n \in [1, 5]$. The mean value with the smallest n is the one that is used, i.e., the value that is
110 closest in time.

Table 2. Meaning of the MCD43C3 albedo quality flag.

Flag value	Description
0	Best quality, 100% with full inversions
1	Good quality, 75% or more with best full inversions and 90% or more with full inversions
2	Relative good quality, 75% or more with full inversions
3	Mixed, 75% or less full inversions and 25% or less fill values
4	All magnitude inversions or 50% or less fill values
5	50% or more fill values

3. In a third step, remaining missing values for a given DOY are replaced with the spatial average for the same DOY over an area of $m \times m$ grid boxes around each missing value, with $m \in 5, 7, 9$. The mean value with the smallest m is the one that is used, i.e., the value corresponding to the smallest surrounding area.

4. Further remaining missing values are replaced with the mean over longitude of surface albedo in $2°$ latitude bands for the same DOY.

5. If missing values still exist at this stage for given grid boxes and a given DOY, the mean value over all DOY during the 10 years under consideration is used to replace them.

6. The very last missing values for a given DOY are filled using the mean surface albedo value over all grid boxes available from the temporal average created in 1) for a given DOY.

7. Finally, since the MCD43C3 product only retrieves land properties, we compute an albedo value for the ocean pixels in each of the seven MODIS bands using the "deep ocean" spectrum from the ECOSTRESS library of the US Geological Survey (USGS) database (Baldridge et al., 2009; Meerdink et al., 2019). To this end, incoming solar spectral irradiance (Kurucz, 1992) is first convolved with the spectral response function of the given MODIS channels. Then, under the assumption of no atmosphere, reflected spectral irradiance at the surface is computed upon multiplication with the spectral ocean albedo and integrated over wavelength. This value is finally divided by the integral of the incoming spectral irradiance computed above to obtain the band albedo values for the open ocean. These values are used everywhere for the global water bodies and at every time. Of course, we are aware that water surfaces are better characterised using a bidirectional reflection distribution function (BRDF) in order to take care of specular reflection (Cox and Munk, 1954a, b; Nakajima, 1983).

This climatology is the starting point to build the hyperspectral albedo maps, where the average sea ice and snow cover is automatically included.

2.2 Soils and Vegetations spectra

To create the hyperspectral albedo maps for each DOY, we use laboratory and in-situ hyperspectral measurements of different soils, rocks and vegetation surfaces. Jiang and Fang (2019) developed hyperspectral soil reflectance eigenvectors to improve canopy radiative transfer. Studying the impact of different regional datasets, they found an increase of accuracy and robustness with a global sample coverage of different soils and vegetations spectra, compared to the performances of regional datasets. Following this prescription, we select three dry soils and vegetations datasets, covering different countries and different surface materials:

1. the ECOSTRESS library (Baldridge et al., 2009; Meerdink et al., 2019) includes 1023 surface spectra from the United States, among which 487 are vegetations spectra, 62 non-photosynthetic vegetations, 381 rocks, 40 soils, 45 man-made materials, and 8 are water ice and snow spectra;

2. the ICRAF–ISRIC dataset (ICRAF-ISRIC, 2021), which is a global dataset with 4440 spectra of different soils from 58 different countries;

3. the LUCAS dataset (Orgiazzi et al., 2018), which contains 21782 different soils spectra from 28 European Union countries (including the United Kingdom), where we selected the 30° viewing angle. As shown by Shepherd et al. (2003), LUCAS spectra are problematic between 400 and 500 nm, where they reach negative values. Following Jiang and Fang (2019), we use the multiple linear regression algorithm from scikit learn (`sklearn.linear_model.LinearRegression`) (Pedregosa et al., 2011), trained on the ISAC–ISRIC dataset to reconstruct the LUCAS spectra in the 400 to 500 nm spectral range.

All the datasets cover the 400–2500 nm spectral range, with different spectral resolutions. LUCAS dataset has a spectral resolution of 0.5 nm, while ICRAF and ECOSTRESS of 10 nm. We interpolated the least resolved datasets to obtain all spectra with a resolution of 1 nm.

In total, we use 26635 dry soils and vegetations spectra from 82 different countries as input to extract the principal components. In Fig. 1, we show some representative soil and vegetations spectra from the ECOSTRESS library. Jiang and Fang (2019) also study the influence of humid soils on the PCA regression algorithm. They find that the effect of soil moisture is non-linear, causing a general reduction of reflectance due to a total internal reflection effect of the water surface. This effect is more prominent in the near infrared (1100–2500 nm). They conclude that to treat the dry and humid soils separately leads to a more applicable soil reflectance model (Jiang and Fang, 2019). A comprehensive and global database of humid soils is currently not available in the literature, and the inclusion of humid soils is outside the scope of our work.

2.3 Principal Component Analysis

The vector of the MODIS albedo data (Sect. 2.1) at the seven wavelengths ($\boldsymbol{R}$) can, in general, be decomposed as

$$\boldsymbol{R} = \boldsymbol{c}\mathrm{U}, \tag{1}$$
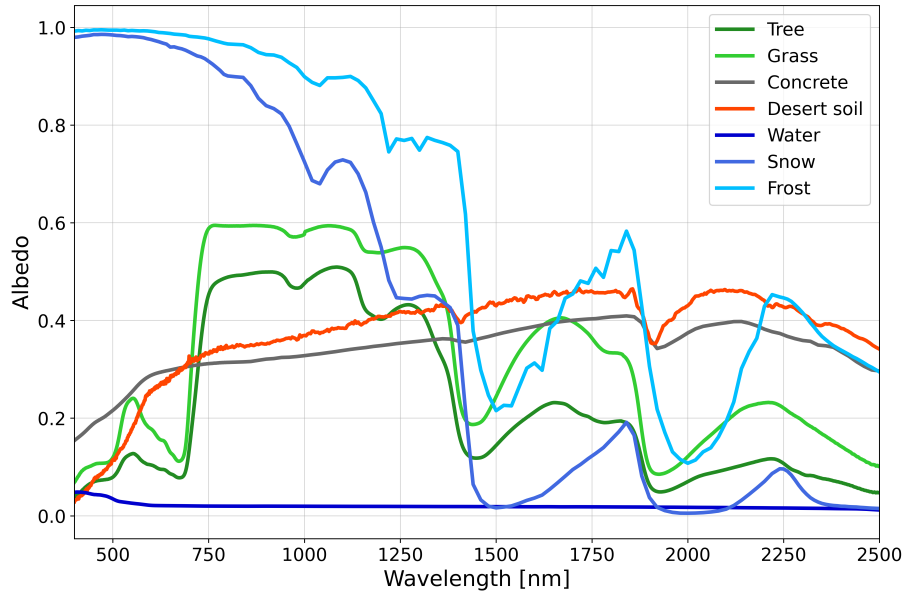
Figure 1. Albedo spectral signatures of some typical soils, vegetation and water bodies from the ECOSTRESS library.

where $\boldsymbol{R} = (r_1, ..., r_n)$ is the albedo vector, with n the number of wavelengths, $\boldsymbol{c} = (c_1, ..., c_m)$ is the coefficient vector, with m representing the number of surface spectra and $\mathbf{U}$ is an $m \times n$ matrix with the laboratory spectra of different soils and
165   vegetation types. In order to calculate the hyperspectral albedo maps, we first need to compute the coefficient vector $\boldsymbol{c}$ at every pixel, by inverting Eq. 1. Since the $\mathbf{U}$ matrix is not square, the correct inverse equation is:

$$\boldsymbol{c} = \boldsymbol{R}\mathbf{U}^T(\mathbf{U}\mathbf{U}^T)^{-1}. \tag{2}$$

From the MODIS dataset, $\boldsymbol{R}$ is available only for seven spectral bands (see Tab. 1), while the goal of this work is to fill the spectral gaps between the bands and reconstruct a full spectrum from VIS to NIR, with a fine spectral resolution. Computing
170   Eq. 2, with a dimensionality of $m = 26635$ is too computationally expensive. In order to reduce the dimensionality of this problem, as in Vidot and Borbás (2014), we apply a Principal Component Analysis (PCA) algorithm, which is an unsupervised machine learning algorithm, and extract the principal components from the matrix $\mathbf{U}$.

We need seven principal components (or eigenvectors) to solve our problem. As done by Vidot and Borbás (2014), we generate six principal components and we use a constant value for the seventh one, because this has been tested and shown to improve
175   the performances. The other six principal components are generated from the three dry datasets described in the previous section. Since these datasets account for different surface types (vegetations spectra are only given in ECOSTRESS), and come in different numbers, we cannot directly merge the spectra of the three datasets altogether. Thus, we balance the number of spectra from the different datasets clustering them using a k-means algorithm `sklearn.cluster.KMeans` (Pedregosa et al., 2011), as done in Liu et al. (2023). In this way, we obtain 100 representative soils spectra for the ICRAF–ISRIC and
180   the LUCAS datasets each, and 128 representative spectra for the ECOSTRESS datasets extracted as follows: 40 vegetations

spectra, 10 non-photosynthetic vegetations spectra, 40 soils spectra, 20 rocks spectra, 10 man-made materials spectra and 8 water bodies spectra. The water bodies spectra, which include snow of different granular sizes, frost, deep ocean, costal ocean and tap water, were not reduced in their dimensionality. Without accounting for this number difference, the vegetation and water surfaces present in the ECOSTRESS dataset would be outweighed by the number of soils spectra from the other datasets, resulting in a sensibly lower performance of the algorithm.

We use the `scikitlearn.decomposition.PCA` implementation of PCA, which follows a Singular Value Decomposition (SVD) of the data as in Halko et al. (2009). From this process, we end up with a matrix $\tilde{U}_\lambda$ with the same spectral resolution as the laboratory spectra, where λ represents the hyperspectral nature of this matrix. To combine it with the albedo data vector $\boldsymbol{R}$, which is only given at the seven MODIS bands, we need to convolve the full matrix $\tilde{\mathbf{U}}_\lambda$ with the average satellite response function of Terra and Aqua satellites of each band. This convolution is necessary to correctly estimate the measured albedo for the central wavelength of each band, crucial to generate the hyperspectral albedo maps with the PCA.

The result of the convolution is a square matrix $\tilde{\mathbf{U}}$ at the seven MODIS wavelengths available from satellites data. Since $\tilde{\mathbf{U}}$ is a square matrix, we can simply calculate

$$\boldsymbol{c} = \boldsymbol{R}\tilde{\mathbf{U}}^{-1}. \tag{3}$$

In this way, we have seven equations for seven coefficients, which allow us to estimate the coefficient vector $\boldsymbol{c}$. Once $\boldsymbol{c}$ is known, it is possible to calculate the albedo maps at all selected wavelengths as

$$\boldsymbol{R}_\lambda = \boldsymbol{c}\tilde{\mathbf{U}}_\lambda, \tag{4}$$

where the λ subscript indicates the hyperspectral nature of the elements. The same process is applied to all the pixels in the map to generate a final albedo map with a spatial resolution of $0.05°$ in latitude and longitude, and to all the different days of the year, taking into account also Earth's seasonal variability.

Vidot and Borbás (2014) created BRDFs maps using a PCA algorithm for their radiative transfer code. They use the ASTER library (today called ECOSTRESS library), which at the time contained much fewer soils and vegetations spectra, to create average maps to include hyperspectral reflectivity of the soils in their radiative transfer simulations. Jiang and Fang (2019) demonstrated that increasing the sample of different soils from various countries of the world helps to validate several dataset among each other. Without accounting for satellite data to create Earth's albedo maps, Jiang and Fang (2019) calculated eigenvectors, using a SVD algorithm, to study the hyperspectral properties of canopy trees in radiative transfer simulations, also including small humid soils local datasets. For the scope of this work, it was not possible to directly use the three eigenvectors generated by Jiang and Fang (2019), as we regress the hyperspectral albedo maps from the seven MODIS bands, thus seven eigenvectors are needed.

As a result of the method explained above, we obtain an hyperspectral climatology of black surface albedo over the entire globe from a wavelength of 400 to 2500 nm in steps of 10 nm. While the interpolation is done with a 1 nm resolution of the hyperspectral albedo maps, the final HAMSTER dataset has a spectral resolution of 10 nm to reduce the size of the single maps. This hyperspectral climatology has a spatial resolution of $0.05°$ in latitude and longitude and a temporal resolution of 8 days. It incorporates the information contained in the MODIS climatology and extends it to wavelengths that were not available before.
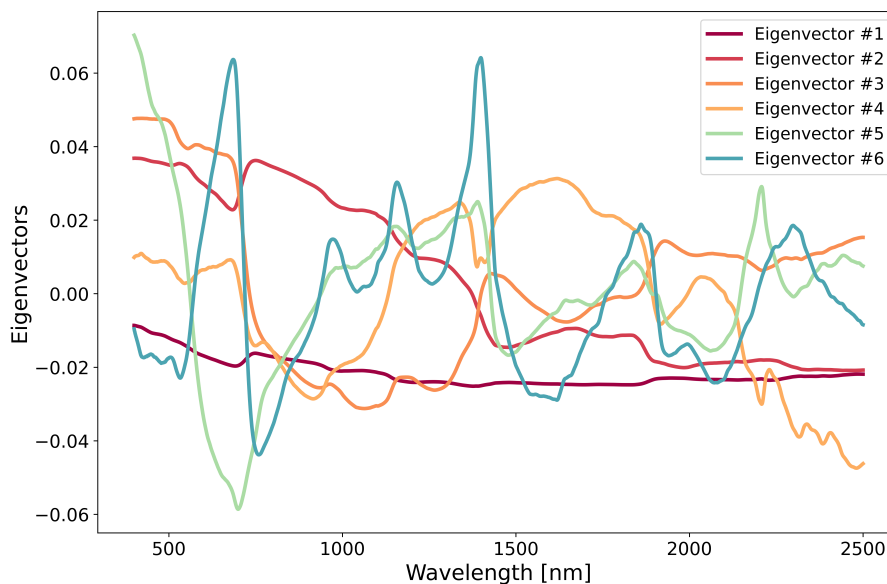
Figure 2. Eigenvectors generated by the PCA starting from the LUCAS, ICRAF–ISRIC and ECOSTRESS datasets. These eigenvectors are used to build the hyperspectral albedo maps. Eigenvectors are plotted in order of importance as calculated from the PCA.

2.4 Validation

As a first test, we use the hyperspectral albedo maps to reconstruct the MODIS channels black sky albedo of the climatology. We multiply the hyperspectral maps by the satellite spectral response function and we estimate the Root Mean Square Error (RMSE) for all the seven channels. For all MODIS channels (see Tab. 1), the RMSE is less than 0.0003. This confirms that the computed hyperspectral albedo maps are able to reconstruct the original MODIS climatology with great accuracy.

To validate the PCA retrieved maps (HAMSTER dataset), we compare them with the land surface albedo product of the SEVIRI instrument aboard the geostationary Meteosat Second Generation (MSG) satellite (Schmetz et al., 2002). SEVIRI has three channels in the VIS/NIR range, which are reported in Tab. 3. As MSG is a geostationary satellite, we cannot compare the entire world map, but only the Earth's "disk" that includes Africa and parts of Europe, South America and the Middle East. SEVIRI channels have spectral response functions that are broader than the analogous MODIS bands and are centred at slightly different wavelengths, and thus we convolved the hyperspectral maps to account for that. In particular, the SEVIRI channel centred at 810 nm touches the vegetation "ramp" starting from 700 nm and is expected to show higher albedo values than the first SEVIRI channel.

The SEVIRI land surface albedo product MDAL (Geiger et al., 2008; Juncu et al., 2022, product identifier LSA-101) is offered daily by the Land Surface Analysis (LSA) Satellite Application Facility (SAF) on the native SEVIRI grid with a spatial resolution of 3 km at the sub-satellite point, and is similar to the MODIS-based MCD43C3 product, against which it has been evaluated (Carrer et al., 2010). As for MCD43C3, both the bi-hemispherical (white-sky) and directional-hemispherical (black-

Table 3. Spectral bands of SEVIRI in the VIS and NIR providing information about the land surface. For each band, we specify the central wavelength and the bandwidth.

band	central λ [nm]	bandwidth [nm]
1	635	600–680
2	810	775–850
3	1640	1550–1750

sky) albedo are available. In order to compare with the HAMSTER hyperspectral albedo maps constructed from MODIS, we reprojected the SEVIRI data to the MCD43C3 grid (0.05° resolution in latitude and longitude), to allow for a consistent comparison. We selected two different DOY, one in the late boreal winter (March 5th, DOY 65) and one in the middle boreal

235 summer (July 30th, DOY 209) both in 2016 to compare the surface reflectivity during two different vegetation stages with possible snow cover in winter and no snow in summer over Northern Europe. The results are shown in Figures 3 and 4.

We compare the three solar satellite channels offered by SEVIRI with the reconstructed channels from the HAMSTER clima-
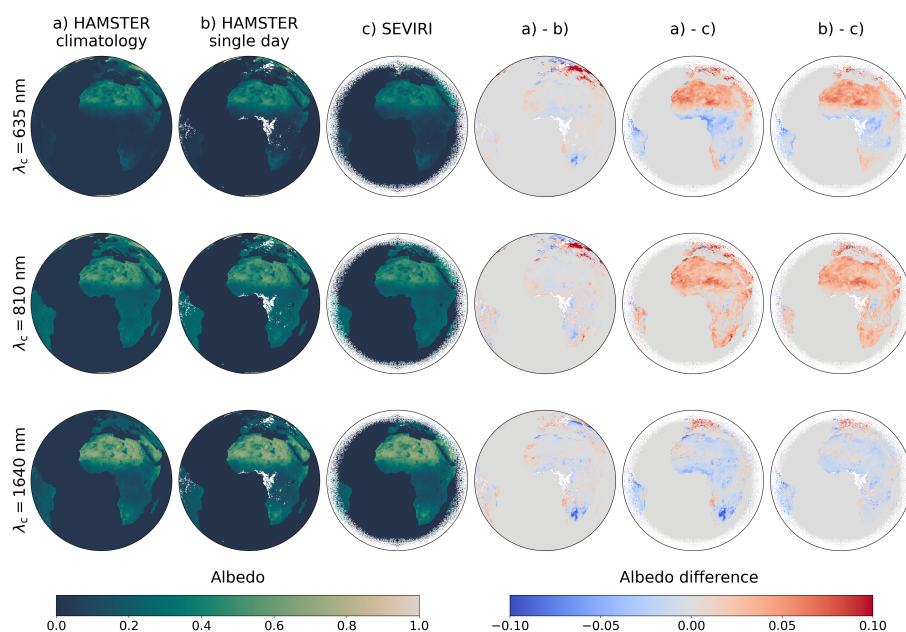
Figure 3. Comparison between HAMSTER climatology, HAMSTER single day and SEVIRI in the late boreal winter (March 5th 2016, DOY 65) for the three SEVIRI VIS/NIR channels. The first three columns show the albedo value for (a) the HAMSTER climatology and (b) the HAMSTER single day integrated over each SEVIRI channel, and (c) the SEVIRI albedo product. In the last three columns, we display the albedo difference between the three different albedo products or reconstructions between -0.10 to 0.10.
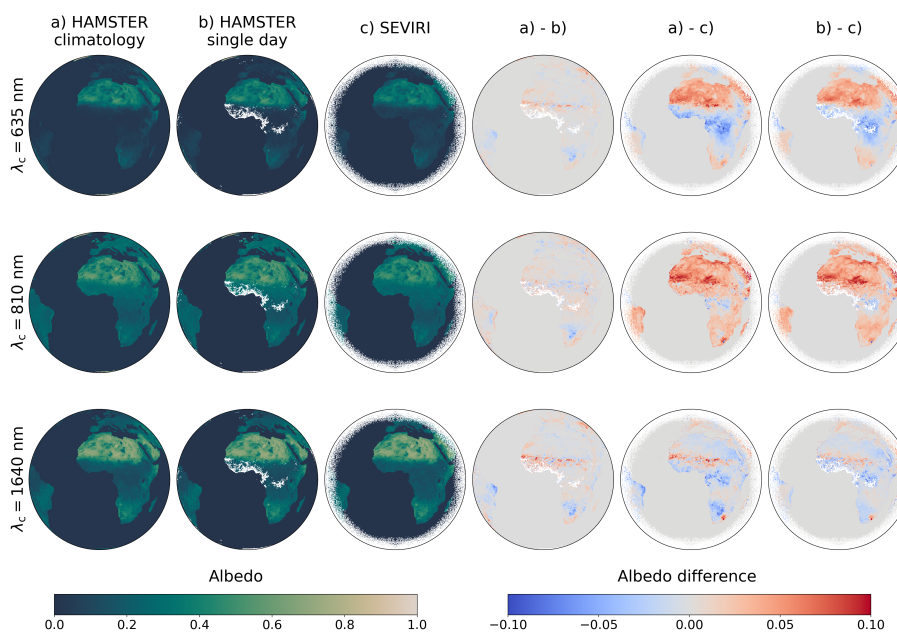
Figure 4. Comparison between HAMSTER climatology, HAMSTER single day and SEVIRI in the boreal summer (July 30th 2016, DOY 209) for the three SEVIRI VIS/NIR channels. The first three columns show the albedo value for (a) the HAMSTER climatology and (b) the HAMSTER single day integrated over each SEVIRI channel, and (c) the SEVIRI albedo product. In the last three columns, we display the albedo difference between the three different albedo products or reconstructions between -0.10 to 0.10.

tology and HAMSTER single day (first three columns in Figs. 3 and 4). SEVIRI's channel 3 has the same central wavelength (λ_c = 1640 nm) as MODIS's band 6, which offers an almost direct comparison between MODIS and SEVIRI land surface

240 products. However, the hyperspectral nature of the retrieved HAMSTER maps is still used to convolve around the 1640 nm MODIS band. The same happens for SEVIRI's channel 1 and MODIS's band 1, for which there is only a 10 nm difference in the central wavelength. On the other hand, SEVIRI's channel 2 (λ_c = 810 nm) is outside any MODIS band. This last case allows us to make a comparison between the reconstructed albedo maps and the SEVIRI measurements, rather than between the land surface products of the two instruments.

245 In addition, in Figures 3 and 4, we also assess the difference between the HAMSTER climatological average (first column) and a single day HAMSTER reconstruction (second column), without accounting for the 10-year average of the climatology. White pixels in the HAMSTER single day correspond to pixels without any albedo value from the MODIS MCD43C3 product. The climatological average shows less features, in particular over Europe, which might be due to the fluctuations of a single day, while in the HAMSTER single day we notice a larger dependence on the seasonality. The effect of the climatology is

250 shown in the forth column, where we plot the albedo difference between HAMSTER climatology and HAMSTER single day. In Fig. 3 we clearly see discrepancies of the order of 0.10 in the first two channels, while in SEVIRI's channel three we notice a lower albedo over Southern Africa for the HAMSTER climatology. Less differences are found for DOY 209, in the boreal
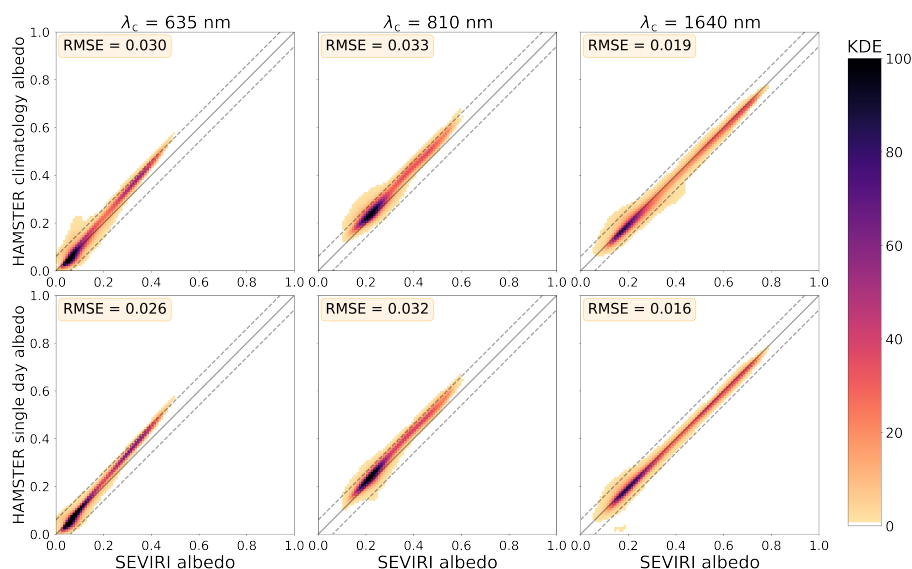
Figure 5. Kernel Density estimation (KDE) between HAMSTER climatology, HAMSTER single day and SEVIRI albedo data March 5th 2016 (DOY 65) for the three central wavelengths of SEVIRI channels (different columns). The first row accounts for HAMSTER climatology hyperspectral albedo maps, while the second row for the single day reconstruction. The soild line represent a perfect linear fit, while the dashed lines show linear fit with an offset of 0.06.

summer (Fig. 4). To conclude, the last two columns of Figs. 3 and 4 display the difference between HAMSTER (climatology and single day) integrated over the SEVIRI channels minus the SEVIRI land surface product. We notice an overestimation of

255   the order of 0.05 for the reconstructed HAMSTER hysperspectral albedo maps in the first two channels over the Sahara desert, while vegetated areas over Africa and part of Europe and South America show a negative (SEVIRI channel 1) and positive (SEVIRI channel 2) discepancy compared to SEVIRI of around the same order. On the other hand, SEVIRI channel 3 (λ_c = 1640 nm) is mostly underestimated by HAMSTER, with a smaller albedo difference compared to the other two channels. Since HAMSTER is built from the MODIS land surface product, our results are in accordance to the discrepancies found by

260   Shao et al. (2021), which points towards difference between various land surface products of up to 0.06. While we describe the different offset arising from this comparison, we can conlude that the reconstructed maps are consistent, in their validation, to the discrepancies arising from different satellite data products.

In Figs. 5 and 6, we show the probability density function (pdf), calculated as the Kernel Density Estimation (KDE) (i.e., a Gaussian-kernel-based probability density, Scott, 1992) between HAMSTER (climatology and single day) and the SEVIRI

265   land surface products for the two DOY selected. For each comparison, we estimate the RMSE and we represent the discrepancies between the different albedo products with the KDE.

We notice that the RMSE is always very small, compatible with intrinsical differences between different retrieval of the albedo products. The RMSE is larger for SEVIRI channel 2 (centred at λ_c = 810 nm) for both DOY, which is also the SEVIRI channel furthest from any MODIS channel. We also notice that the comparison with the hyperspectral maps built on the single day
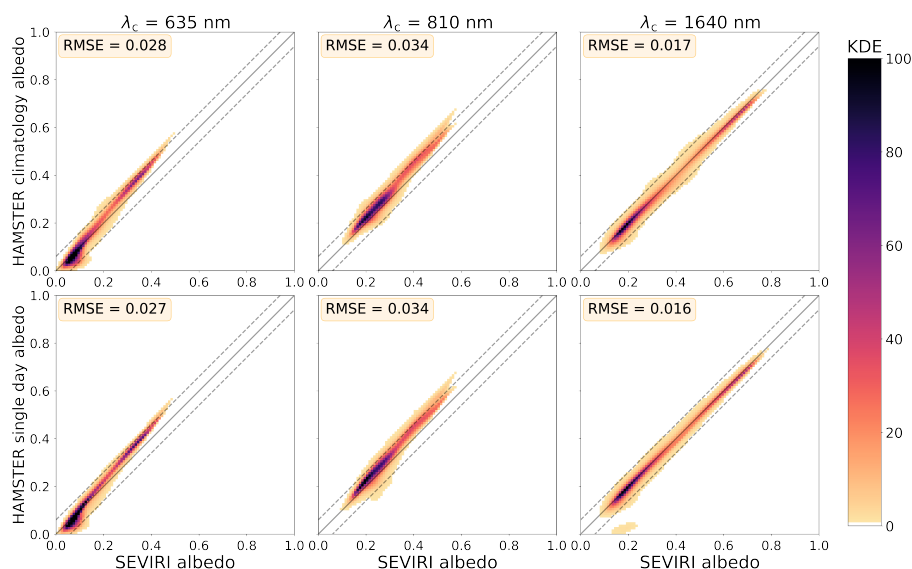
Figure 6. Kernel Density estimation (KDE) between HAMSTER climatology, HAMSTER single day and SEVIRI albedo data on July 30th 2016 (DOY 209) for the three central wavelengths of SEVIRI channels (different columns). The first row accounts for HAMSTER climatology hyperspectral albedo maps, while the second row for the single day reconstruction. The soild line represent a perfect linear fit, while the dashed lines show linear fit with an offset of 0.06.

270 albedos have always a slightly smaller RMSE, since the climatology can only reproduce a climatological vegetation state and snow coverage pattern of a certain DOY.

As a last test, we compare the hyperspectral albedo maps with the TROPOMI Lambertian Equivalent Reflectivity (LER) product from https://www.temis.nl/surface/albedo/tropomi_ler.php (Tilstra et al., 2021, 2023). TROPOMI LER product (sub-satellite pixel size of 0.125° x 0.125°) is remarkably different from MODIS MCD43C3 product, as it provides surface albedo

275 for snow/ice-free and snow/ice conditions separately. The snow/ice conditions are also averages over a month, which do not allow for a direct comparison with MODIS. On the other hand, TROPOMI bands are very narrow, of just 1 nm, and they provide many channels in the Vegetation Red Edge (VRE) rump. For this reason, we validate our hyperspectral albedo maps with the TROPOMI product only for the African continent, to avoid comparisons with snow/ice products which are not fully consistent. Due to the narrow satellite bands of TROPOMI, it was not needed to convolve for its satellite response function, and we esti-

280 mated the RMSE between TROPOMI LER and our HAMSTER hyperspectral albedo product. The results are shown in Tab. 4. The RMSE is comparable to what we find for SEVIRI and with known discrepancies among different surface albedo products, and it remains relatively small in the TROPOMI bands between 670 and 772 nm, inside the VRE and far from MODIS bands. This confirms good performances of the hyperspectral albedo maps also far from the MODIS bands on which they have been retrieved from. In Fig. 7, we select three TROPOMI bands and we compare the albedo value over Africa between HAMSTER

285 climatology (first column) and the TROPOMI albedo product (second column). We select the TROPOMI monthly product for the month of March (average from 2018 to 2023), and we compare it with the combination of the HAMSTER DOY 065,

073, 081 and 089 (corresponding to all DOY in March). In the thrid column, we again plot the albedo difference between the two products. For λ_c = 463 nm, we notice a very good agreement, with discrepancies of less than 0.025 over Africa. For λ_c = 747 nm, inside the VRE, the discrepancies are larger, with a general albedo overestimation of HAMSTER compared to TROPOMI, reaching differences of up to 0.10. We also compared the two products with a band in the far NIR (λ_c = 2314 nm) and we found an overestimation of dry and desert areas and an underestimation of vegetated regions due to HAMSTER. Also in this last band, albedo products reach differences of up to 0.10, in particular over deserts.
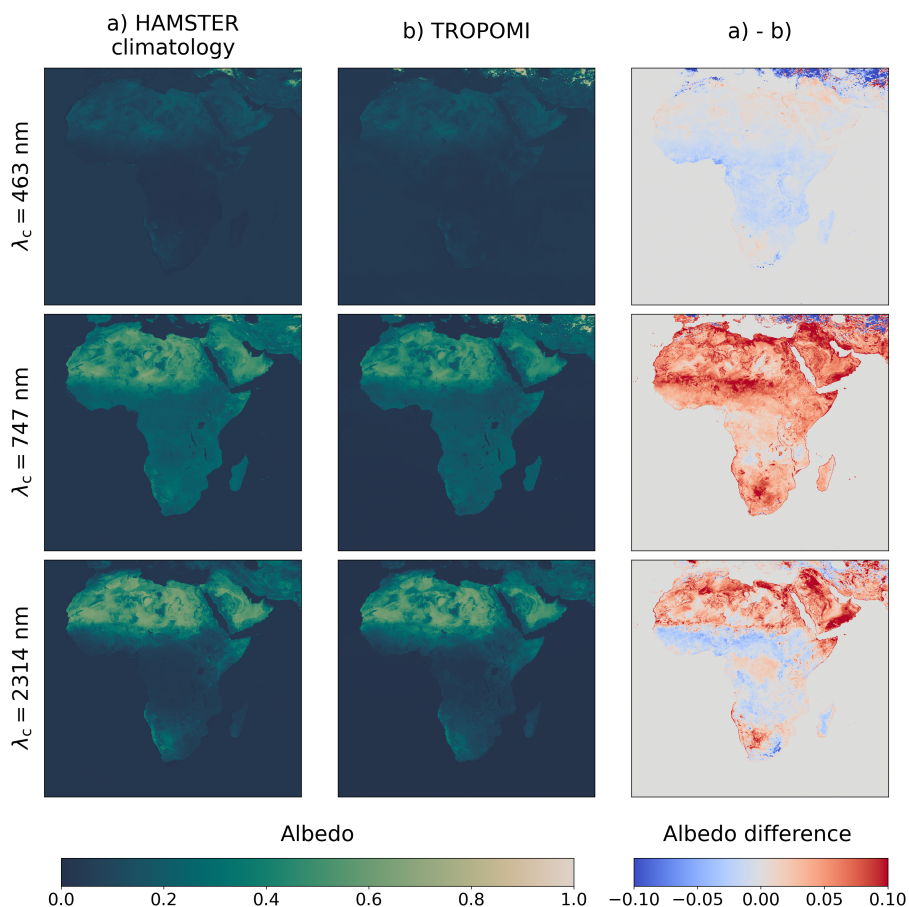
Figure 7. Comparison between HAMSTER climatology (first column) and TROPOMI in the late boreal winter (month of March) for three selected wavelengths among the TROPOMI VIS/NIR channels. The third column shows the albedo difference between HAMSTER climatology and TROPOMI LER albedo product.

Table 4. Spectral bands of TROPOMI LER product in the VIS and NIR, and RMSE of the comparison with the HAMSTER hyperspectral albedo maps over Africa.

λ [nm]	402	416	425	440	463	494	670	685	697	712	747	758	772	2314
RMSE	0.029	0.028	0.029	0.028	0.028	0.029	0.039	0.039	0.043	0.045	0.059	0.056	0.053	0.030

3   Results

In this section we present the two main results of this paper: the MODIS surface albedo climatology for the seven bands and,
295   building on that, the extended HAMSTER hyperspectral surface albedo dataset.

3.1   MODIS climatology dataset

As described in Sect. 2.1, we have derived a 10-year climatology of surface albedo for different DOY as a starting point to
generate the hyperspectral albedo maps. This climatological average, with a temporal resolution of 8 days, allows studying the
temporal variability of the albedo of the planet, as shown in Fig. 8. Since albedo values are not available for every pixel of
300   Earth's surface during the year due to missing solar illumination during winter, we study the temporal evolution of the mean
global albedo between 67° N and 67° S. Among these latitudes, we always have an estimate of the albedo of every single pixel
for all DOY. As a consequence, we are excluding from the mean albedo estimation both the Arctic and Antarctica regions,
as well as other high latitude land surface in the Northern Hemisphere. For this reason, the mean albedo value should not be
intended as a global estimate for Earth, but more as an indicator of its temporal variation.
305   In Fig. 8, we notice that the mean albedo is higher in the NIR bands, after the vegetation red edge (VRE) peaks. At 858 nm,
which peaks right after the VRE, we notice the largest albedo value for the planet, followed by 1240 nm. Continuing in the
NIR, with 1640 and 2130 nm, the albedo values decrease. On the contrary, in the VIS range there is a very small variation in the
albedo among the three bands. The VIS bands show a clear seasonal trend due to the melting of ice and snow in the Northern
Hemisphere, followed by a subsequent blossoming of vegetation. Thus, Earth's albedo is peaking in the late boreal winter in
310   the VIS and then decreasing in the boreal summer. This large variability trend can be interpreted with seasonal differences in
snow coverage, and it mainly follows the variability of the Northern Hemisphere since it host almost 80% of all the land in the
globe. However, in the NIR bands, we notice other features around boreal late spring and autumn which are due to blossoming
of flowers and reddening of leaves, which decrease the general reflectivity of green leaves.

In Fig. 9, we study the spatial variability of the albedo throughout the year for a particular wavelength for the entire 10 years
315   climatological average. Here we selected MODIS band 2 centred at the 858 nm. In particular, we plot the difference between
the maximum and minimum albedo values during the entire year, independently of when the maximum and minimum are
reached. For instance, the maximum of reflectivity over high latitudes in the Northern Hemisphere is reached during the boreal
summer, while over the coast around Antarctica it happens during the austral summer, due to sea ice melting. Moreover, since
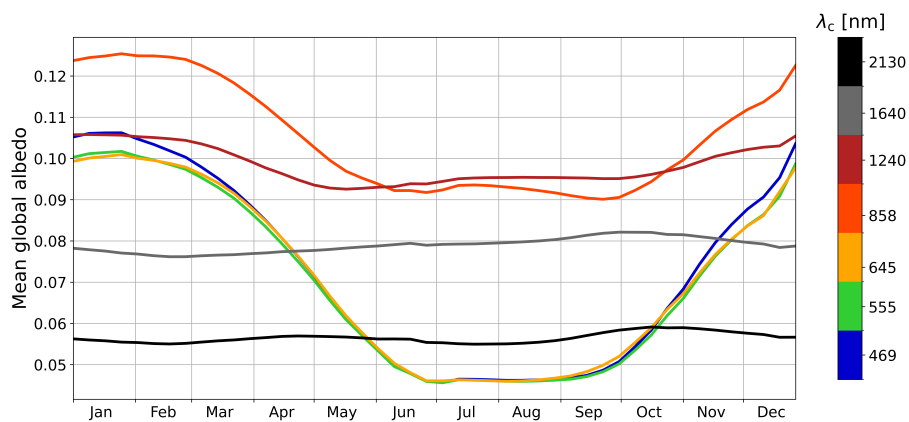
Figure 8. Yearly cycle of the MODIS climatology data black-sky albedo between 67°N and 67°S. The different curves represent the different MODIS channels indicated by their central wavelength.
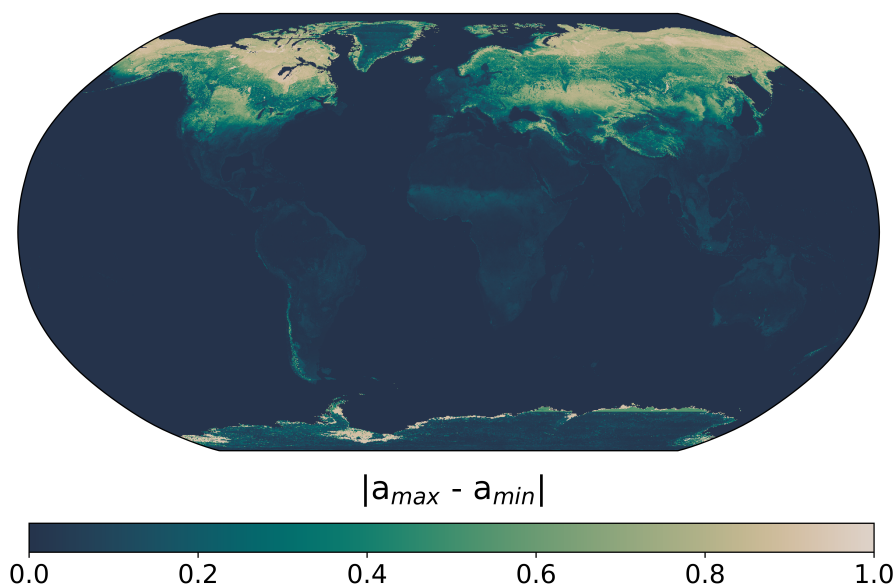
Figure 9. Spatial variation of the MODIS climatology, showing the difference between the maximum and minimum albedo value for each pixel during the year.

albedo data are not available during boreal winter (summer) for the Northern (Southern) Hemisphere, for high latitude regions
320   (north and south of 67°) the difference between the maximum and minimum albedo is calculated over a shorter time period, corresponding to the data coverage of the region.

Calculating this reflectivity variation for every pixel, the map in Fig. 9 highlights those regions that carry the largest variations. In particular, Arctic and Antarctic regions exhibit high reflectivity variations due to sea ice melting, clearly visible in the map.

Mainland Greenland also shows more variability than mainland Antarctica, possibly pointing towards melting of Greenland's
glaciers during boreal summer. Deserts all over the world, like the Sahara and Australian deserts, show the least variability,
remaining almost constant throughout the year. Also, tropical rainforests, like the Amazon rainforest, do not exhibit a significant
seasonal variability. On the contrary, temperate and boreal forests show a pronounced variation due to the difference between
the snow cover during winter and the summer months.

3.2 Hyperspectral albedo maps

From the MODIS climatology data, we build the hyperspectral albedo maps with a PCA regression algorithm, as described
in Sect. 2.3. The hyperspectral albedo maps allow us to combine the spectral features of different soils, vegetations and water
surfaces with the high spatial and temporal resolution of the MODIS climatology data. This could potentially have many pos-
sible applications, from the possible implementation in climate models (where Braghiere et al. (2023) already demonstrated its
feasibility) to the improvement of remote sensing retrieval frameworks. The new hyperspectral albedo maps have been imple-
mented in the radiative transfer software package libRadtran (http://www.libradtran.org/doku.php, (Mayer and Kylling, 2005;
Emde et al., 2016).

As a first application, we use the hyperspectral maps to calculate the mean global albedo value close to the equinoxes. In this
way, we have almost all pixels filled with an albedo value and it is possible to assess a mean albedo value for the entire globe
as a function of wavelength (see Fig. 10). The main difference between the spring and autumn equinoxes relies on the snow
coverage over the Northern Hemisphere, which increases the reflectivity during the spring boreal equinox. This affects mostly
the VIS wavelengths, following the typical albedo profile of snow and frost (see Fig. 1). From these hyperspectral albedo maps,
we could recover that the mean global albedo is around 0.21 in the VIS during March and around 0.17 over autumn, while
it decreases below 0.10 in the NIR. The dots in Fig. 10 represent the average over the MODIS channels, without taking into
account the hyperspectral albedo maps.

In addition, we apply the hyperspectral maps to study the VRE, which shows a steep increase in the reflectivity of vegetation
due to chlorophyll. In Fig. 11, we show the progression from 700 nm to 850 nm (with steps of 50 nm) of vegetation reflectivity
for DOY 65 (March 5th). We notice a substantial increase of the albedo for all kinds of forests, from tropical to boreal ones,
with the largest increase between 700 and 750 nm, as expected for the VRE. This comparison is only possible having albedo
maps which account for their hyperspectral dimension. Using only the MODIS wavelengths, we would have missed the entire
VRE transition because the closest bands are only at 645 and 858 nm.

As a last application, we study the spectral profile of different regions of the world, accounting also for their seasonal variabil-
ity. We select different examples of rainforests, boreal forests, deserts, urban areas, and ice-covered regions as shown in Fig. 12.
Inside the boundaries of the selected areas highlighted in Fig. 12, we average the spectra of all pixels in the region in order
to obtain an average spectrum representative of the entire region. The averages are made for the four seasons separately. The
first comparison pertains to forest spectra (dark green regions in Fig. 12). We selected three different rainforests, the Amazon,
Borneo and Congo rainforests, two different boreal forests, over Canada and Russia, and a Savanna region over Kenya and
Tanzania. The selection of the different areas is made by maximizing the possible land area hosting the same properties, but
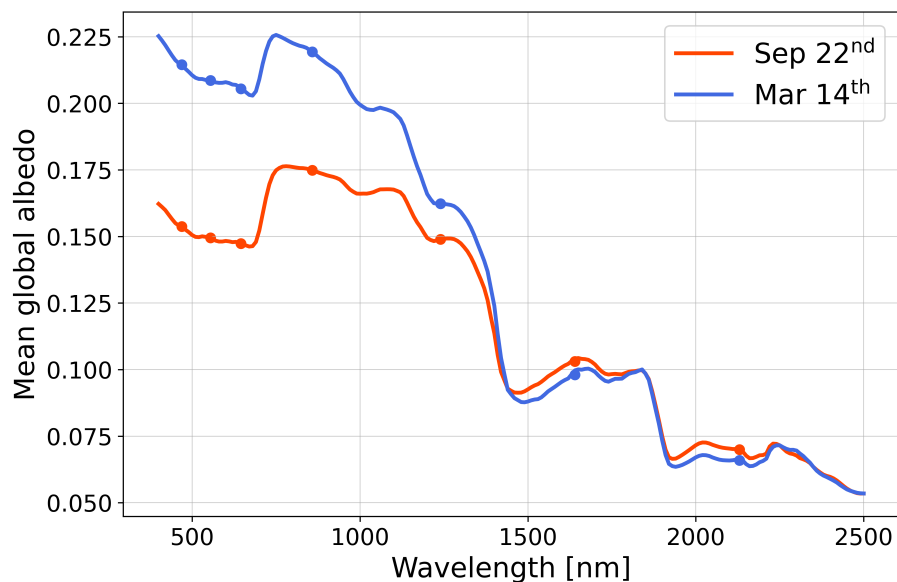
Figure 10. Mean global albedo as a function of wavelength over the entire globe. We select the two DOY closest to the equinoxes, where we almost have all pixels filled with an albedo value. The seven dots show the albedo value of the seven MODIS bands, while the curves are built from the average of all pixels of the HAMSTER hyperspectral albedo maps for a given wavelength.
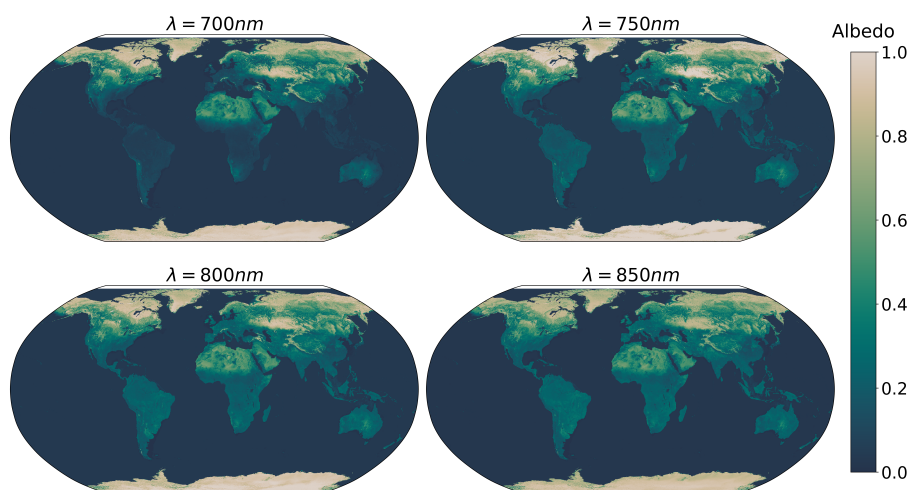
Figure 11. Spectral evolution of the surface albedo for March 5th (DOY 65). From $\lambda = 700$ nm to $\lambda = 850$ nm, there is a steep increase of the albedo over forests, due to the VRE.

avoiding mixing the region with urbanized soils and different lands.

Fig. 13 shows the comparison between the spectrum of different forests. We notice a similar trend among all kind of forests, with similar spectral features. In particular, all forest shows three jumps in reflectivity of decreasing amplitude. The main dif-
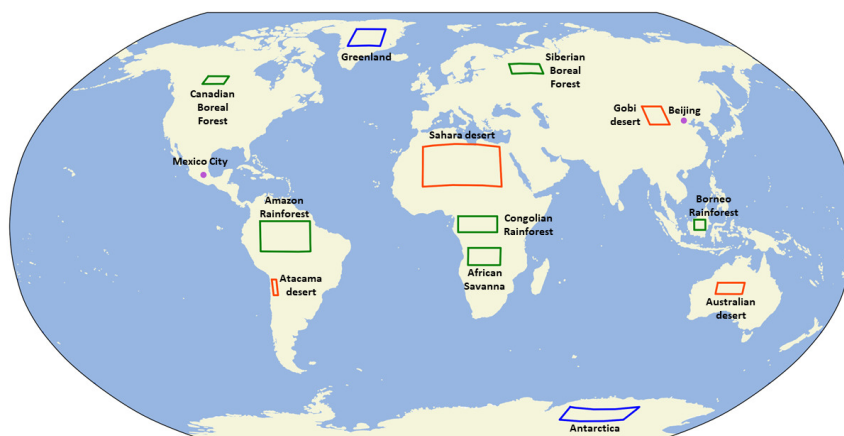
Figure 12. Regions of the world investigated. The green boxes represent the forests, the orange boxes the deserts, the blue boxes the ice sheets, and the purple circles the cities.

ference between tropical rainforests and boreal forests resides, as expected, in their seasonal variability. Tropical rainforests do not exhibit almost any seasonal change, being very similar among each other. On the other hand, boreal forests experience an important decrease of reflectivity from boreal winter to boreal summer. This is due to the melting of snow over boreal forests, which also happens with different timescales. There are also some small differences within tropical rainforests. Borneo shows
365   the least seasonal variation, while Congo shows the smallest reflectivity.

The final spectra are always combinations of different soils and vegetations, and the small differences we find are due to the different trees soils grounds and tree coverage of the different forests. If we compare the obtained spectra with Fig.1, we find an overall agreement with their main spectral features, but our final spectra are modulated by the combination of many different soils and averaged over seasons and different pixels. We extend the comparison to desert areas (orange regions in Fig. 12). We
370   select the Sahara desert, the Australian desert, the Gobi desert and the Atacama desert to extract the spectral properties from the hyperspectral albedo maps. Fig. 14 shows the comparison among different arid regions. We find that, among deserts, the reflectivity profile can greatly vary, depending on the mineralogy and composition of different soils and sands. In addition, as already discussed in Fig. 9, Sahara and Australian deserts do not display any significant seasonal change. This is not the case for the Gobi desert, which shows an enhanced reflectivity in the winter months, due to partial snow coverage.
375   In general, deserts exhibit common spectral shape, with a steep increase of reflectivity up to 750 nm, similar spectral features until the NIR, and a more or less steep decrease of reflectivity around 2150 nm. Different desert areas show larger discrepancies among themselves than forests.

The same methodology is applied to study the Greenland and Antarctica ice sheets (blue areas in Fig. 12). We select two regions which are always snow-covered to study their spectral features and seasonal patterns (see Fig. 15). As expected for fully
380   snow covered surfaces, their reflectivity is very high, reaching almost 1 in the VIS, and it decreases in the NIR range. During
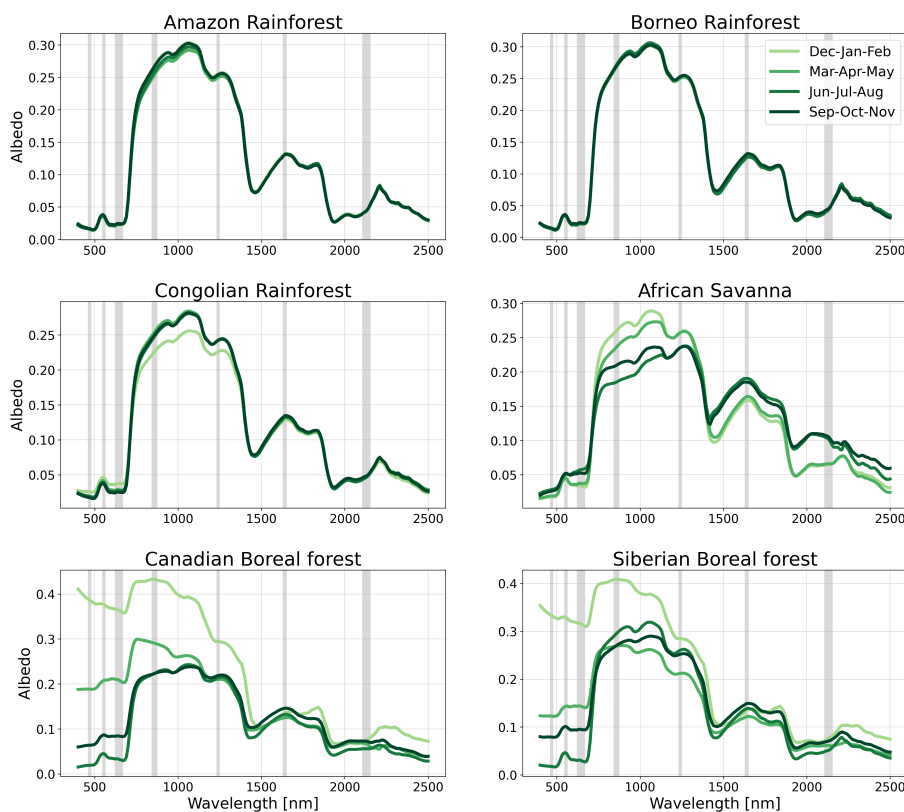
Figure 13. Spectra of different forests of the world obtained by averaging over all pixels in the corresponding region using the hyperspectral albedo maps. Seasonal variability is shown averaging the spectra over three-month periods, indicated with different colours. Gray bands represent the MODIS bandwidths.

Greenland and Antarctica's winters not all the pixels where always available, thus we averaged on less DOY and less pixels to estimate their winter seasonal spectra. In Fig. 1, we see that snow and frost show different patterns in reflectivity, in particular in the NIR. This can explain the spread in the NIR spectra of both Antarctica and Greenland. This also needs to be combined with the formation of clear liquid water lakes on the surface of the glaciers during the melting season, which lowers the total
385 reflectivity of the surface. For Greenland and Antarctica, we find similar behaviours in the NIR, with winter seasons exhibiting a higher reflectivity than summer seasons. We also notice that in the VIS there is almost no seasonal spectral variability over Antarctica, while Greenland shows two distinct trends between boreal autumn and winter and boreal spring and summer. To conclude, we also extracted spectral profiles of two different cities: the urban areas of Beijing and Mexico City. Fig. 16 shows Beijing having a larger seasonal variability than Mexico City. In general, the spectra of the two cities look different, but with
390 some common spectral features. The steep increase in the VIS might be due to vegetation, while other features in the NIR come from man-made materials and different soils present in the training dataset. The peak of the reflectivity for urbanized areas is low, as expected. In general, extracting the spectra of different surface types, we found a good agreement among the typical
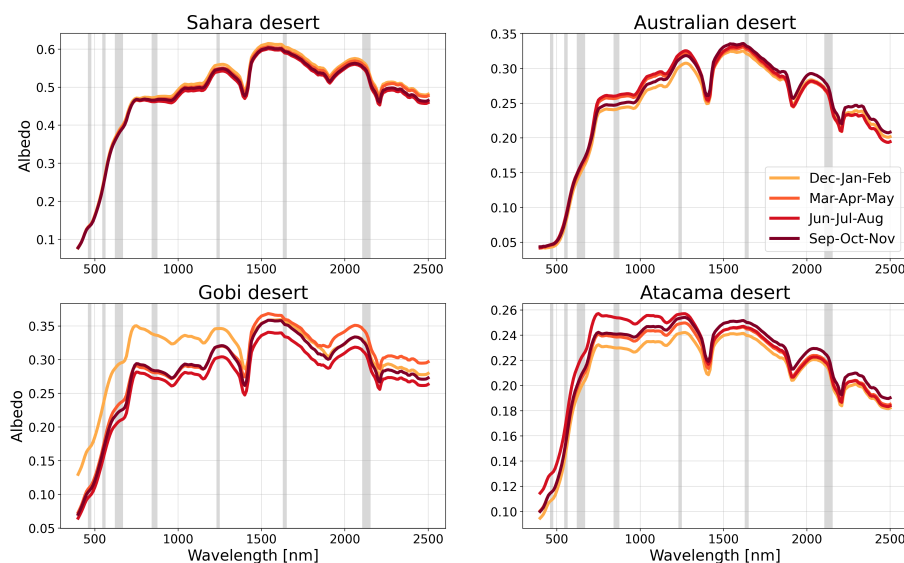
Figure 14. Spectra of different deserts of the world obtained from the average over different pixels from the hyperspectral albedo maps. Seasonal variability is shown averaging spectra over three-month periods, indicated with different colours. Gray bands represent the MODIS bandwiths.
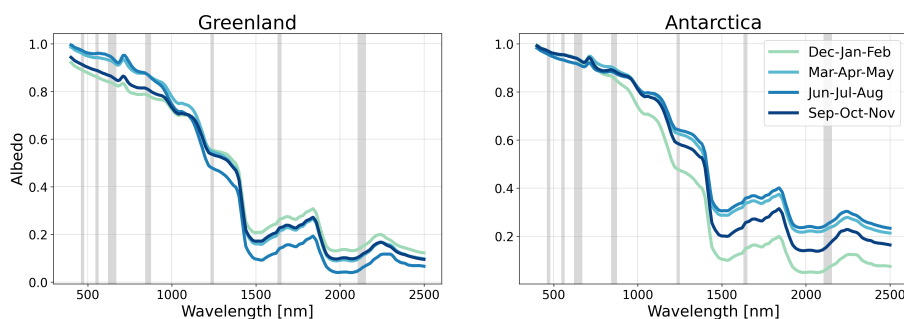
Figure 15. Spectra of different ice surfaces of the world obtained from the average over different pixels from the hyperspectral albedo maps. Seasonal variability is shown averaging the spectra over three-month periods, indicated with different colours. Gray bands represent the MODIS bandwiths.

spectral features of soils and vegetations expected to dominate the different surface types. For instance, different kind of forests all have the typical shape due to the VRE. However, the spectra of the various land types host much more information than the single spectrum of a tree or of a particular soil, and we can clearly see they are a linear combination of different spectra in the sample with varying weights. In fact, forests are the combination of trees with their typical spectral shape, but modulated by different soils reflectivities. As a result, the retrieved albedo of an entire forest is sensibly lower than the one of single trees in the dataset. This is in agreement with Jiang and Fang (2019), who generated different spectra for canopy tree radiative transfer

395
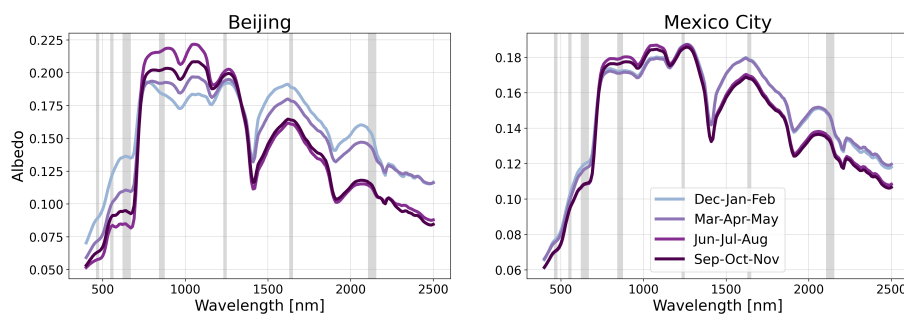
Figure 16. Spectra of two different cities (Beijing and Mexico City) of the world obtained from the average over different pixels from the hyperspectral albedo maps. Seasonal variability is shown averaging spectra over three-month periods, indicated with different colours. Gray bands represent the MODIS bandwidths.

simulations and studied the soils influence on the total reflectivity of the vegetated area. While typical vegetated features are
400   always present in the spectrum, they are sensibly modulated depending on the properties of the background soil.

4 Conclusions

In this work, we create hyperspectral albedo maps to study the wavelength-dependent characteristics of the albedo of the Earth's surface. We select various soils, vegetations and man-made materials spectra from three different datasets: the ECOSTRESS library, which has both soils and vegetations spectra, LUCAS dataset, which contained different soils of many countries in the
405   world, and the ISRIC dataset, a catalogue of thousands of soils of European Union countries. In total, we end up with 26635 spectra of different soils and vegetations from 53 countries of the world.

Due to the huge dimensionality of the final training dataset, we use a PCA regression algorithm to extract the principal components of the dataset. These principal components serve as eigenvectors to recover the albedo reflectivity of different pixels over Earth, starting from the MODIS land surface product. In particular, MODIS measures land surface properties in seven different
410   bands in the VIS/NIR wavelength range. These seven MODIS bands are used as the starting point to build the hyperspectral albedo maps. With the PCA, we extract six principal components as in Vidot and Borbás (2014), and, with the addition of a seventh constant eigenvector, we combine them with the seven bands of MODIS data, for which the albedo of all single pixels is known. From this computation, it is possible to extract the spectral albedo value in the entire wavelength range pixel by pixel.

415   To generate a climatological hyperspectral albedo map, we use the 8 days land surface product from MODIS, and we average every 8th DOY from 2013 to 2022. This allows us to get a climatological average of surface properties of the planet, to fill missing pixels which might be cloudy during a particular year, and to disentangle from the yearly variability patterns. As a final outcome, we obtain the HAMSTER hyperspectral albedo maps dataset with:

- a spectral resolution of 10 nm;

420   – a spatial resolution of 0.05° in latitude and longitude;

     – a temporal resolution of 8 day averaged in the time period between 2013 and 2022.

As demonstrated by Vidot and Borbás (2014) and Jiang and Fang (2019), PCA or SVD algorithms are powerful tools to combine a huge sample of soils and vegetations spectra and to reconstruct the albedo profile of different areas of the world. In our work, apart from generating hyperspectral albedo maps from the PCA as in Vidot and Borbás (2014), we also include
425   Jiang and Fang (2019) advice to train the PCA with a much larger dataset, accounting for different countries of the world. In addition, our hyperspectral albedo maps are given for 46 different DOY, thus making it possible to retain all the seasonal variability patterns present in MODIS's data.

Our MODIS climatological maps and hyperspectral albedo maps are validated against SEVIRI and TROPOMI land surface products. To perform this comparison, we adapted SEVIRI's dataset to MODIS resolution and projection, and we find that there
430   is a good agreement between both MODIS climatology and the HAMSTER hyperspectral maps with SEVIRI observations, up to discrepancies of 0.06, which is a typical order of magnitude for land surface product comparisons (Zhang et al., 2010; Shao et al., 2021). Similar results are found in the comparison with TROPOMI.

Already the MODIS climatological dataset displays interesting temporal and spatial patterns. Thanks to both high spatial and temporal resolution, we study Earth's temporal variability for different wavelengths, and we display the maximal albedo
435   difference of each pixel, highlighting regions with high temporal variability. The mean spectral albedo of the planet peaks at wavelegths larger than the VRE, while it shows a larger variability in the VIS wavelengths than the NIR ones, where the seasonal variation between snow covered high latitudes land in the Northern Hemisphere displays an increase of the surface albedo in the boreal winter.

We combine the information coming from temporal and spatial resolution of the MODIS climatology data with the possibility
440   to spectrally extend the information about different regions to create typical spectra of different land surface type. We find that:

     – forests, as expected, show typical vegetated spectral features, such as the VRE. Tropical rainforests do not undergo much seasonal change, while boreal forests have an increased reflectivity in the winter, when they get partially snow covered. Savanna regions experience a drying of the land after the end of the summer, which flatten the typical vegetation-induced spectral features.

445   – deserts show almost no seasonal variability, apart from those with occasional snow coverage. Depending on the properties of the soils, its colour and mineralogical composition, and the presence of sand, the overall reflectivity of the desert can greatly vary.

     – ice and snow covered surfaces, like Greenland's and Antarctica's ice sheets, reflect almost entirely in the VIS, with a steep decrease in the NIR. During summer months, their albedo is slightly lower than late winter or spring months due
450   to the melting of surface ice, which creates lakes on the top of the icy surface.

– urbanized areas, such as Tokyo and Mexico City, are the combination of many different man-made materials, soils and vegetations spectra and their spectral shape host features from all of them. The total reflectivity of a city is lower than 20%.

These hyperspectral albedo maps dataset can be used for many different applications, from improving climate models to Earth's remote sensing, and to correctly simulate the disk-integrated spectra of Earth (Emde et al., 2017), and correctly model Earthshine observations (Sterzik et al., 2012, 2019). Only using the full spectral variations of land surfaces, it is possible to correctly establish Earth's energy budget. Braghiere et al. (2023) studied the impact of assuming only two broadband albedo values, as done in Earth system models, versus using hyperspectral albedo maps. While the general radiative forcing is sensibly smaller than the one from doubling of CO_2, omitting the hyperspectral nature of Earth's surface causes deviation in many climatological patterns, such as precipitation and surface temperature, in particular over regional scales.

Data availability. The HAMSTER dataset is available on Zenodo at the following link: https://zenodo.org/records/10494404.

Video supplement. A video supplement of this work is available at https://av.tib.eu/media/66248, where we show the spectral and spatial evolution of HAMSTER for four different DOY.

Author contributions. GR designed the research. LB and UH performed the MODIS climatology. GR and FG trained the PCA from the soils and vegetations spectra. GR performed the analysis and made the plots. GR, CE, MFS and MM interpreted the results. GR wrote the draft. All of the authors contributed in improving the manuscript. CW and CE implemented the dataset in libRadtran.

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors thank NASA's MODIS/Terra+Aqua BRDF/Albedo Albedo Daily L3 Global 0.05 Deg CMG MCD43C3 dataset for providing the albedo maps used in this work. We also acknowledge ESA for providing SEVIRI and TROPOMI datasets. In addition, we thank the ECOSTRESS, LUCAS and ICRAF–ISRIC libraries for the surface spectra used in the PCA training.

References

Baldridge, A., Hook, S., Grove, C., and Rivera, G.: The ASTER spectral library version 2.0, Remote Sensing of Environment, 113, 711–715, https://doi.org/https://doi.org/10.1016/j.rse.2008.11.007, 2009.

Braghiere, R. K., Wang, Y., Gagné-Landmann, A., Brodrick, P. G., Bloom, A. A., Norton, A. J., Ma, S., Levine, P., Longo, M., Deck, K., Gentine, P., Worden, J. R., Frankenberg, C., and Schneider, T.: The Importance of Hyperspectral Soil Albedo Information for Improving Earth System Model Projections, AGU Advances, 4, e2023AV000 910, https://doi.org/https://doi.org/10.1029/2023AV000910, e2023AV000910 2023AV000910, 2023.

Carrer, D., Roujean, J.-L., and Meurey, C.: Comparing Operational MSG/SEVIRI Land Surface Albedo Products From Land SAF With Ground Measurements and MODIS, IEEE Transactions on Geoscience and Remote Sensing, 48, 1714–1728, https://doi.org/10.1109/TGRS.2009.2034530, 2010.

Cox, C. and Munk, W.: Measurement of the roughness of the sea surface from photographs of the sun's glitter, Journal of the Optical Society of America, 44, 838–850, 1954a.

Cox, C. and Munk, W.: Statistics of the sea surface derived from sun glitter, Journal of Marine Research, 13, 198–227, 1954b.

Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., and Bugliaro, L.: The libRadtran software package for radiative transfer calculations (version 2.0.1), Geoscientific Model Development, 9, 1647–1672, https://doi.org/10.5194/gmd-9-1647-2016, 2016.

Emde, C., Buras-Schnell, R., Sterzik, M., and Bagnulo, S.: Influence of aerosols, clouds, and sunglint on polarization spectra of Earthshine, Astronomy and Astrophysics, 605, A2, https://doi.org/10.1051/0004-6361/201629948, 2017.

Geiger, B., Carrer, D., Franchisteguy, L., Roujean, J.-L., and Meurey, C.: Land Surface Albedo Derived on a Daily Basis From Meteosat Second Generation Observations, IEEE Transactions on Geoscience and Remote Sensing, 46, 3841–3856, https://doi.org/10.1109/TGRS.2008.2001798, 2008.

Govaerts, Y. and Lattanzio, A.: Estimation of surface albedo increase during the eighties Sahel drought from Meteosat observations, Global and Planetary Change, 64, 139–145, https://doi.org/10.1016/j.gloplacha.2008.04.004, 2008.

Halko, N., Martinsson, P.-G., and Tropp, J. A.: Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, arXiv e-prints, arXiv:0909.4061, https://doi.org/10.48550/arXiv.0909.4061, 2009.

He, T., Liang, S., Yu, Y., Wang, D., Gao, F., and Liu, Q.: Greenland surface albedo changes in July 1981-2012 from satellite observations, Environmental Research Letters, 8, 044043, https://doi.org/10.1088/1748-9326/8/4/044043, 2013.

He, T., Liang, S., and Song, D.-X.: Analysis of global land surface albedo climatology and spatial-temporal variation during 1981-2010 from multiple satellite products, Journal of Geophysical Research (Atmospheres), 119, 10,281–10,298, https://doi.org/10.1002/2014JD021667, 2014.

ICRAF-ISRIC: ICRAF-ISRIC Soil VNIR Spectral Library, https://doi.org/10.34725/DVN/MFHA9C, 2021.

Jiang, C. and Fang, H.: GSV: a general model for hyperspectral soil reflectance simulation, International Journal of Applied Earth Observation and Geoinformation, 83, 101 932, https://doi.org/https://doi.org/10.1016/j.jag.2019.101932, 2019.

Juncu, D., Ceamanos, X., Trigo, I. F., Gomes, S., and Freitas, S. C.: Upgrade of LSA-SAF Meteosat Second Generation daily surface albedo (MDAL) retrieval algorithm incorporating aerosol correction and other improvements, Geoscientific Instrumentation, Methods and Data Systems, 11, 389–412, https://doi.org/10.5194/gi-11-389-2022, 2022.

Kurucz, R. L.: Synthetic Infrared Spectra, in: Infrared Solar Physics: Proceedings of the 154th Symposium of the International Astronomical Union, Tucson, Arizona, U.S.A., March 2-6 1992, edited by Rabin, D. M., Jefferies, J. T., and Lindsey, C., vol. 154, p. 523, Kluwer Academic Publishers, Dordrecht, 1992.

510 Liang, S., Wang, K., Zhang, X., and Wild, M.: Review on Estimation of Land Surface Radiation and Energy Budgets From Ground Measurement, Remote Sensing and Model Simulations, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3, 225–240, https://doi.org/10.1109/JSTARS.2010.2048556, 2010.

Liu, B., Guo, B., Zhuo, R., Dai, F., and Chi, H.: Prediction of the soil organic carbon in the LUCAS soil database based on spectral clustering, Soil and Water Research, 18, 43–54, https://doi.org/10.17221/97/2022-SWR, 2023.

515 Loarie, S. R., Lobell, D. B., Asner, G. P., Mu, Q., and Field, C. B.: Direct impacts on local climate of sugar-cane expansion in Brazil, Nature Climate Change, 1, 105–109, https://doi.org/10.1038/nclimate1067, 2011.

Lyons, E. A., Jin, Y., and Randerson, J. T.: Changes in surface albedo after fire in boreal forest ecosystems of interior Alaska assessed using MODIS satellite observations, Journal of Geophysical Research (Biogeosciences), 113, G02012, https://doi.org/10.1029/2007JG000606, 2008.

520 Mayer, B. and Kylling, A.: Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use, Atmospheric Chemistry and Physics, 5, 1855–1877, https://doi.org/10.5194/acp-5-1855-2005, 2005.

Meerdink, S. K., Hook, S. J., Roberts, D. A., and Abbott, E. A.: The ECOSTRESS spectral library version 1.0, Remote Sensing of Environment, 230, 111 196, https://doi.org/https://doi.org/10.1016/j.rse.2019.05.015, 2019.

Nakajima, T.: Effect of wind-generated waves on the transfer of solar radiation in the atmosphere-ocean system, , 29, 521–537,
525 https://doi.org/10.1016/0022-4073(83)90129-2, 1983.

Offerle, B., Jonsson, P., Eliasson, I., and Grimmond, C. S. B.: Urban Modification of the Surface Energy Balance in the West African Sahel: Ouagadougou, Burkina Faso., Journal of Climate, 18, 3983–3995, https://doi.org/10.1175/JCLI3520.1, 2005.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., and Fernández-Ugalde, O.: LUCAS Soil, the largest expandable soil dataset for Europe: a review, European Journal of Soil Science, 69, 140–153, https://doi.org/https://doi.org/10.1111/ejss.12499, 2018.

530 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Salomonson, V., Barnes, W., Maymon, P., Montgomery, H., and Ostrow, H.: MODIS: advanced facility instrument for studies of the Earth as a system, IEEE Transactions on Geoscience and Remote Sensing, 27, 145–153, https://doi.org/10.1109/36.20292, 1989.

535 Sánchez-Zapero, J., Martínez-Sánchez, E., Camacho, F., Wang, Z., Carrer, D., Schaaf, C., García-Haro, F. J., Nickeson, J., and Cosh, M.: Surface ALbedo VALidation (SALVAL) Platform: Towards CEOS LPV Validation Stage 4 – Application to Three Global Albedo Climate Data Records, Remote Sensing, 15, https://doi.org/10.3390/rs15041081, 2023.

Schaaf, C. and Wang, Z.: MODIS/Terra+Aqua BRDF/Albedo Albedo Daily L3 Global 0.05Deg CMG V061, https://doi.org/10.5067/MODIS/MCD43C3.061, 2021.

540 Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A.: AN INTRODUCTION TO METEOSAT SECOND GENERATION (MSG), Bulletin of the American Meteorological Society, 83, 977 – 992, https://doi.org/https://doi.org/10.1175/1520-0477(2002)083<0977:AITMSG>2.3.CO;2, 2002.

Scott, D.: Multivariate Density Estimation: Theory, Practice, and Visualization, A Wiley-interscience publication, Wiley, https://books.google.de/books?id=7crCUS_F2ocC, 1992.

545 Sellers, P. J., Meeson, B. W., Hall, F. G., Asrar, G., Murphy, R. E., Schiffer, R. A., Bretherton, F. P., Dickinson, R. E., Ellingson, R. G., Field, C. B., Huemmrich, K. F., Justice, C. O., Melack, J. M., Roulet, N. T., Schimel, D. S., and Try, P. D.: Remote sensing of the land surface for studies of global change: Models — algorithms — experiments, Remote Sensing of Environment, 51, 3–26, https://doi.org/10.1016/0034-4257(94)00061-Q, 1995.

Shao, C., Shuai, Y., Tuerhanjiang, L., Ma, X., Hu, W., Zhang, Q., Xu, A., Liu, T., Tian, Y., Wang, C., and Ma, Y.: Cross-Comparison of Global
550 Surface Albedo Operational Products-MODIS, GLASS, and CGLS, Remote Sensing, 13, https://doi.org/10.3390/rs13234869, 2021.

Shepherd, K. D., Palm, C. A., Gachengo, C. N., and Vanlauwe, B.: Rapid Characterization of Organic Resource Quality for Soil and Livestock Management in Tropical Agroecosystems Using Near-Infrared Spectroscopy, Agronomy Journal, 95, 1314–1322, https://doi.org/https://doi.org/10.2134/agronj2003.1314, 2003.

Sterzik, M. F., Bagnulo, S., and Palle, E.: Biosignatures as revealed by spectropolarimetry of Earthshine, , 483, 64–66,
555 https://doi.org/10.1038/nature10778, 2012.

Sterzik, M. F., Bagnulo, S., Stam, D. M., Emde, C., and Manev, M.: Spectral and temporal variability of Earth observed in polarization, , 622, A41, https://doi.org/10.1051/0004-6361/201834213, 2019.

Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Directionally dependent Lambertian-equivalent reflectivity (DLER) of the Earth's surface measured by the GOME-2 satellite instruments, Atmospheric Measurement Techniques, 14, 4219–4238,
560 https://doi.org/10.5194/amt-14-4219-2021, 2021.

Tilstra, L. G., de Graaf, M., Trees, V., Litvinov, P., Dubovik, O., and Stammes, P.: A directional surface reflectance climatology determined from TROPOMI observations, Atmospheric Measurement Techniques Discussions, 2023, 1–29, https://doi.org/10.5194/amt-2023-222, 2023.

Vidot, J. and Borbás, É.: Land surface VIS/NIR BRDF atlas for RTTOV-11: model and validation against SEVIRI land SAF albedo product,
565 Quarterly Journal of the Royal Meteorological Society, 140, 2186–2196, https://doi.org/10.1002/qj.2288, 2014.

Zhang, X., Liang, S., Wang, K., Li, L., and Gui, S.: Analysis of Global Land Surface Shortwave Broadband Albedo From Multiple Data Sources, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 3, 296–305, https://doi.org/10.1109/JSTARS.2010.2049342, 2010.

Zhu, X., Liang, S., Pan, Y., and Zhang, X.: Agricultural Irrigation Impacts on Land Surface Characteristics Detected From Satellite Data
570 Products in Jilin Province, China, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 4, 721–729, https://doi.org/10.1109/JSTARS.2011.2106152, 2011.