Earth System
Open Access
Science
Data

A database of global reference sites to support validation of satellite surface albedo datasets (SAVS 1.0)

Alexander Loew[1,2], Ralf Bennartz[3], Frank Fell[4], Alessio Lattanzio[5], Marie Doutriaux-Boucher[5], and Jörg Schulz[5]

[1]Department of Geography, Ludwig-Maximilians-Universität München, Munich, Germany
[2]Max-Planck-Institute for Meteorology, Hamburg, Germany
[3]Earth and Environmental Science Department, Vanderbilt University, Nashville, TN, USA
[4]Informus GmbH, Berlin, Germany
[5]European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), Darmstadt, Germany

Correspondence to: Alexander Loew (alexander.loew@lmu.de)

Abstract. Validating the accuracy and long-term stability of terrestrial satellite data products necessitates a network of reference sites. This paper documents a global database of more than 2000 sites globally which have been characterized in terms of their spatial heterogeneity. The work was motivated by the need for potential validation sites for geostationary surface albedo data products, but the resulting database is useful also for other applications. The database (SAVS 1.0) is publicly available through the EUMETSAT website (http://savs.eumetsat.int/, doi:10.15770/EUM_SEC_CLM_1001). Sites can be filtered according to different criteria, providing a flexible way to identify potential validation sites for further studies and a traceable approach to characterize the heterogeneity of these reference sites. The present paper describes the detailed information on the generation of the SAVS 1.0 database and its characteristics.

1 Introduction

Surface albedo is an essential climate variable (ECV) of major importance for Earth System Science (Bojinski et al., 2014). Global satellite-derived surface albedo datasets are used in applications such as numerical weather prediction, hydrology, agricultural monitoring or climate modeling (Houldcroft et al., 2009; Brovkin et al., 2013; Hagemann et al., 2013)

The first multidecadal data products of surface albedo have meanwhile been released from either polar-orbiting or geostationary satellites (Riihelä et al., 2010, 2013; Schaaf et al., 2011; Lattanzio et al., 2015). Geostationary satellite sensors provide a unique opportunity for the estimation of long-term surface albedo data records due to their multidecadal observational record. They provide spatial resolutions on the order of 1–10 km, later referred to as medium resolution. The validation of data products at these scales proves difficult, as a direct comparison with point-like (as compared to the size

of satellite pixels) in situ solar radiation flux measurements and derived local-scale surface albedo data is complicated through their often limited spatial representativeness.

Major efforts are therefore devoted within, for example, the CEOS (Committee on Earth Observation Satellites) Land Product Validation team (CEOS-LPV) as well as international projects such as the European FP7 project for Quality Assurance for Essential Climate Variables (QA4ECV, http://www.qa4ecv.eu/) to define protocols for best practice to (a) measure surface albedo at the ground, (b) develop traceable and quality-controlled algorithms for the retrieval of satellite surface albedo data and (c) devise methods for the comparison between these complementary datasets. Multidecadal records are typically derived from a series of sensors on subsequently operated observation platforms. A careful instrument intercalibration is required to avoid changes and drifts in the observing system causing spurious trends in the retrieved surface albedo data products themselves (Loew and Govaerts, 2010; Riihelä et al., 2013;

Loew, 2014). Fiducial reference sites are further required to estimate any systematic error in the satellite products.

The representativeness of a point-like surface albedo measurement contributes significantly to the overall error budget when comparing in situ measurements with medium-resolution surface albedo data (Román et al., 2009). A set of well-characterized reference sites is therefore needed, which has the potential to be used for the validation of surface albedo data products.

Cescatti et al. (2012) used 53 sites globally to validate the MODIS surface albedo data product. These are based on the network of global FLUXNET stations (Baldocchi et al., 2001) and have been chosen due to their well-characterized surface heterogeneity. Baret et al. (2006) have identified homogeneous sites from available in situ measurement networks for the validation of surface albedo data. These sites have been updated by homogeneous reference sites which were identified using high-resolution land cover information. This combined dataset is used in an automated online validation tool (Weiss et al., 2014) which was developed to provide a framework for validation of satellite products of terrestrial variables.

The present paper combines and enhances these previous activities: it introduces a new database of Surface Albedo Validation Sites (SAVS 1.0) providing a set of well-characterized global reference sites for the validation of terrestrial satellite observations with particular emphasis on the application for the validation of geostationary surface albedo (GSA) data products. SAVS 1.0 provides a traceable approach to characterize potential sites for Earth Observation data validation. It was developed for the validation of surface albedo data products derived from geostationary satellite data and contains information on more than 2000 potential reference sites globally. It provides a user-friendly interface that enables the users to efficiently filter potential validation sites according to different criteria. The database is accessible on the EUMETSAT website (http://savs.eumetsat.int/).

2   General approach

2.1   Validation site characteristics

The accuracy requirements for satellite surface albedo observations on the global scale have been identified by WMO (2011) as max(5 %; 0.0025), where the 5 % corresponds to the relative accuracy criteria and the albedo is defined as the fraction of the shortwave radiative fluxes and is not given in percent, as commonly done in the literature. The objective behind this accuracy requirement is to be able to detect changes in the Earth's radiative forcing up to an equivalent of 20 % of the total expected change in radiative forcing per decade. The derived temporal stability criteria [change per decade] are then max(1 %; 0.0001) (WMO, 2011).

The CEOS-LPV group is leading the generation of best-practice protocols for the validation of terrestrial ECV products derived from satellite data. An overview about existing in situ measurement networks, measurement standards and satellite datasets for different terrestrial ECVs is provided by the Global Terrestrial Observing System (GTOS). Details for surface albedo are provided by Schaaf et al. (2009).

Surface reference sites for space-based observations should have different characteristics. Most importantly, the spatial heterogeneity of the site should be small within the field of view of a particular observing instrument or compared to the spatial resolution of a particular data product. The availability of reference measurements at a particular site would further add to its usefulness for a robust evaluation. If a quantitative validation of a data product is envisaged, then the representativeness of reference data needs to be quantified (Román et al., 2009) as this influences the uncertainty in the data product evaluation. While reference data are only available for a limited number of sites on the globe, spatially representative sites which are known to be temporally invariant (e.g., deserts) can be also very useful to characterize the long-term temporal stability of a dataset.

Over the last decades, a number of terrestrial measurement networks have been established to measure a multitude of different variables such as surface fluxes, aerosols and terrestrial carbon fluxes. However, to our knowledge, none of the existing networks have been tailored to the specific needs for the evaluation of satellite data products and in particular a GSA dataset.

To identify potential reference sites for the purposes of this study, we have therefore used existing measurement networks as a starting point, assuming that already existing networks for global measurements of water and energy fluxes have been chosen to be representative for a surrounding region. The following strategy was then implemented for SAVS 1.0 (Fig. 1):

1. identification of potential validation sites based on existing network infrastructures;

2. characterization of the spatial homogeneity of these sites using ancillary information on topography, vegetation dynamics and land cover;

3. definition and application of criteria to identify sites suitable for surface albedo validation;

4. selection of sites considered to be most suitable for validation of satellite data products, in particular GSA data.

The evaluation of surface albedo data products typically requires that diurnal variations in surface reflectances be taken into account. These comprise, for example, terrain-induced shadowing, geometric navigation uncertainties as well as anisotropic effects caused by changing Sun position. These factors are particularly relevant for geostationary satellite-based surface albedo data products which have
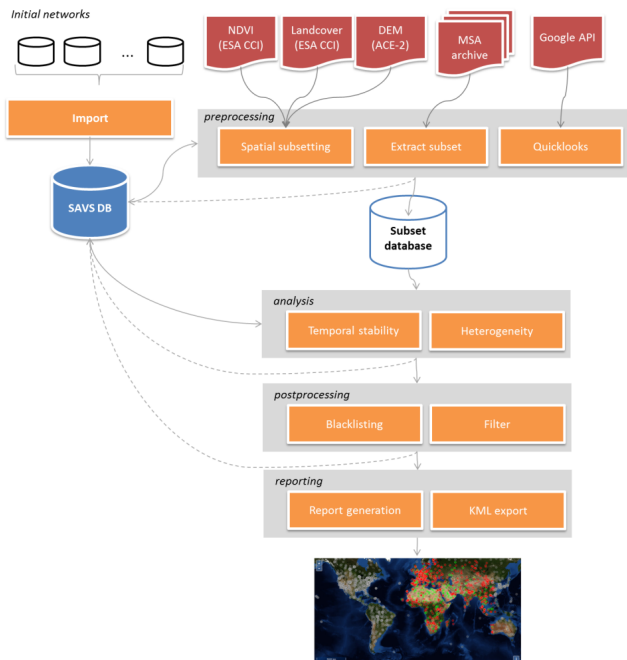
Figure 1. General work flow to identify and characterize potential GSA validation reference sites.

coarse spatial resolutions. Stringent requirements on the characteristics of a reference site to determine whether it is suitable for coarse-scale surface albedo are therefore required.

- Spatial homogeneity: due to the coarse spatial resolution and location uncertainties due to navigation uncertainties of GSA data, the spatial homogeneity of the reference site is important. The spatial homogeneity can be quantified by a number of proxies derived from vegetation and land cover information.

- Topographic homogeneity: topography can have a substantial effect on the diurnal course of surface-reflected directional radiances, which are the basic input into the generation of the GSA product. Thus a site which is located in an environment with steep terrain slopes might be affected by shadowing effects throughout the day.

While these homogeneity criteria are good indicators for the spatial variability around a particular site, they do not necessarily guarantee that the site is also representative of the surrounding areas. Landscapes with mosaic vegetation cover (e.g., shrublands, savannas) are heterogeneous by definition. A representative site would cover a larger area that represents the inherent spatial variability of these surface types. In addition, it is emphasized that the representativeness also has a temporal domain. A site might be representative of its surroundings for a certain period of the year (e.g., wet season), while it is not for other time periods.

The SAVS 1.0 database provides comprehensive metadata that enables the user to filter potential validation sites according to their needs. A set of reference sites suitable for surface albedo evaluation would also cover a wide range of possible albedo values, meaning that dark as well as bright reference sites should be identified and cover a wide range of different biomes to take into account different vegetation phenologies. The datasets used and metrics developed for SAVS 1.0 are described in the following sections and the selection criteria to identify potential GSA validation sites are laid out in Sect. 4.7. Thus, the SAVS 1.0 database provides a comprehensive inventory of potential validation sites, which contain a rich set of metadata that can be used to filter the database according to the users needs and application. While information on ground measurements is provided as well (Table 1), the database does not contain any reference data or information about whether a particular site is suited for a specific kind of application as this can only be judged by the user.

2.2 SAVS 1.0 processing workflow

The overall workflow for generating the SAVS 1.0 database is provided in Fig. 1. Site information from various networks is taken to populate the initial database. For all these sites, ancillary data are exploited to describe the spatial and temporal variability of different surface properties around the site location. General information and metrics are stored as site attributes within the database. At this step satellite surface albedo data can also be ingested to build a time series of satellite measurements at the location of the site.

These sites are then further used to calculate flags that a user might use to filter the database according to user-specific criteria. Additional quality flags are provided that are the result of a quality control procedure (blacklisting) that takes into account, for example, distance to coastlines or spurious ancillary data like errors in the used ACE-2 digital elevation model (DEM). User-friendly reports are finally generated for each site, allowing users to browse through the database content.

3 Data

3.1 Terrestrial site networks

The SAVS 1.0 database builds on already established monitoring sites which have been derived for a variety of different measurement purposes. The different networks included in SAVS 1.0 and their respective numbers of sites are provided in Table 1.

This first set of potential reference sites was screened to identify gaps with respect to geographic distribution, biome coverage and albedo value ranges. An additional 48 sites were subsequently identified by expert knowledge in an attempt to fill these gaps. Overall, this resulted in a total of 2220 potential sites worldwide. Some sites identified as part

of several networks were identified by their site keys and duplicates were subsequently removed from the database for the further analysis. However, duplicates can still remain in the database, as a site can have a different naming (key) in the different originating terrestrial site networks. As coordinates typically differ in these cases, close sites were not eliminated by default. An example is the site Fort Peck (105.1018° W, 48.308° N), which is part of the SURFRAD, BSRN and FLUXNET networks. The keys and coordinates differ in all networks. Thus, instead of removing two of the three occurrences, the SAVS 1.0 database contains results for all three sites. Additional information on the various source networks is, however, provided as a separate table on the SAVS 1.0 website, which also provides information on closest stations to allow potential users to easily identify duplicates. For some stations (e.g., MONGU), duplicates were identified according to their key, but the site coordinates differed by more than 10 km in distance. It was therefore not clear whether these sites actually correspond to different locations or whether there are large uncertainties in the specified coordinates. In such cases, the duplicates remained in the database as independent sites and were given unique keys. Figure 5 shows the spatial distribution of the remaining 2186 sites after removal of the duplicates. Details about the characteristics of these sites are provided in Sect. 5.

It should be emphasized that the networks used have very different characteristics and data availability and originate from different motivations. Some networks, like the Baseline Surface Radiation Network (BSRN), provide high-quality measurements of surface solar upward and downward fluxes (König-Langlo et al., 2013) which can be used as high-quality reference for surface albedo. The SURFRAD network has sites located only in the Unites States and corresponds to the US contribution to the BSRN network. Other site collections, like BELMANIP, do not necessarily provide any reference measurements suitable for a direct comparison with satellite surface albedo observations. The objective of the SAVS database is to characterize the general representativeness of the investigated sites, which is independent of availability of in situ surface albedo measurements as such. Details on availability of surface albedo reference data as well as references for the networks and their quality are provided in Table 1. The user of the SAVS 1.0 database should use these references to check for suitable reference measurements for their particular kind of application.

3.2 Elevation data

Topographic information is based on ACE-2 (Berry et al., 2008; ACE, 2014), a global DEM providing surface elevation data at a spatial resolution of 3 arcsec, or about 90 m. For each site coordinate, the surrounding topography within a 25 km radius was extracted from the ACE-2 dataset. Thus, an area of approximately 2000 km^2 of surface elevation data was extracted for each site for further analysis.

3.3 Land cover and vegetation data

High-resolution (300 m) land cover information was obtained from the ESA Climate Change Initiative (ESA CCI) Land Cover project (Bontemps et al., 2012), providing global coverage for 22 land cover classes together with ancillary information on vegetation, snow and fire dynamics. For each site, the ESA CCI land cover information was extracted for the same area as used for the topographic homogeneity analysis.

It should be emphasized that a land cover classification product has its own uncertainties, which implies that the attributes derived from the ESA CCI land cover dataset would also be uncertain. Tsendbazar et al. (2015) provide a comprehensive analysis of the accuracy of different globally available land cover datasets. The Normalized Difference Vegetation Index (NDVI) is a good proxy for the abundance and seasonality of vegetation, while it is limited to providing information on vegetation structure like differentiation between overstorey and understory vegetation. The CCI land cover dataset provides information on the mean seasonality of the NDVI with a temporal resolution of 7 days at pixel level, derived from SPOT Vegetation data for the period 1998 to 2012. For each pixel and week, 14 measurements are therefore available. The mean as well as 5 and 95 % percentile values were extracted from the land cover condition dataset around each site, similar to the land cover data.

These are further used to characterize the general horizontal homogeneity of the area surrounding a SAVS site. In that respect, the NDVI data and derived information give general information about the vegetation abundance, but they should not be used to characterize vegetation effects on the surface albedo as this is also affected by vegetation vertical structure which is not well approximated by NDVI data.

In addition, the ESA CCI land cover dataset provides information on snow and fire seasonality at pixel level. The probability for snow and fire occurrence is provided for 8-day periods and was extracted for the same area around each site.

As the probability is based on an analysis of a multiyear data record, the estimated snow and fire occurrence frequencies are only valid for the period 1998–2012 and represent the climatological mean occurrence of these disturbances. It is emphasized that the frequency of these occurrences would change with changing climate or land cover conditions. Snow and fire conditions inferred from ESA CCI land cover maps and NDVI are based on a fixed set of observations, which are static. If the land cover or vegetation density does not change, this information can be used to characterize the climatological mean surface conditions. However, under non-stationary conditions, when land cover or vegetation characteristics are changing, the ancillary data would need to be updated on a regular basis. The current version of SAVS needs to be therefore understood as a static characterization of the investigated sites at the time of generation of the SAVS database.

Earth Syst. Sci. Data, 8, 425–438, 2016

www.earth-syst-sci-data.net/8/425/2016/

Table 1. Terrestrial reference networks considered within SAVS 1.0 and details about availability of surface albedo reference data and literature references for the networks themselves as well as validation studies.

Network	Further details	No. of sites	Albedo data available	References
FLUXNET	http://fluxnet.ornl.gov/	252	yes	Baldocchi et al. (2001); Cescatti et al. (2012)
BSRN	http://www.bsrn.awi.de/	63	yes	König-Langlo et al. (2013)
AERONET	http://aeronet.gsfc.nasa.gov/	1176	no	Holben et al. (1998, 2001)
BELMANIP-2	http://calvalportal.ceos.org/web/olive/site-description	558	no	Baret et al. (2006); Weiss et al. (2014)
CEOS LandNet sites	http://calvalportal.ceos.org/ceos-landnet-sites	8	partly	
EOS core val. sites	http://landval.gsfc.nasa.gov/coresite_gen.html	41	partly	
SURFRAD	http://www.esrl.noaa.gov/gmd/grad/surfrad/index.html	7	yes	Augustine et al. (2000, 2005); Wang et al. (2012)
LTER	http://lternet.edu/	27	partly	Waide and Thomas (2012)
ALBEDOVAL-1		40	no	Fell et al. (2012)
Additional SAVS 1.0	Further sites identified by expert knowledge	48	no	NA
Sum		2220		
Duplicates		34		
Sum final	After removal of duplicates	2186		

3.4 Meteosat surface albedo (MSA) data

The SAVS 1.0 sites were then used for an initial evaluation of a surface albedo dataset derived from the EUMETSAT series of geostationary satellites (Meteosat). Details of the MSA data product are described in EUMETSAT (2014).

Observations from geostationary satellites allow for the retrieval of surface albedo information that complements retrievals from polar-orbiting instruments. The main advantage of geostationary observations lies in their high temporal resolution, which increases the likelihood for observations under cloud-free conditions. In addition, geostationary observations cover a long period (e.g., Meteosat observations are available since 1982), which makes them an important information source for climate studies.

A generic algorithm has been derived to retrieve surface albedo in a single broad visible (VIS) band from observations acquired by instruments on board geostationary satellites (Pinty et al., 2000a). It relies on a sophisticated algorithm for the joint retrieval of surface albedo and total atmospheric aerosol load, accounting for the anisotropy of the surface based on daily accumulation of VIS band data and the fast cloud detection method (Pinty et al., 2000b).

The algorithm has been applied to the VIS channel of the Meteosat Visible and InfraRed Imager (MVIRI) (Lattanzio et al., 2007) to provide the EUMETSAT MSA product. This includes a method for the estimation of the radiometric error and the propagation of this error in the retrieval scheme that specifically accounts for the differences in the performance of the various radiometers on board the Meteosat series and permits quantitative uncertainty estimates for each retrieval

result (Govaerts and Lattanzio, 2007). Loew and Govaerts (2010) have provided an update of calibration coefficients to maximize the temporal stability of the MSA multidecadal record. A subset of $50 \times 50 \text{ km}^2$ was extracted for each of the SAVS sites from the MSA dataset for inclusion into the database.

4 Methods

For each site, several statistical measures are calculated and stored as attributes within the SAVS 1.0 database to characterize the temporal stability and homogeneity of each site. At the same time ancillary land information, derived from an external dataset, is exploited to give a full site characterization. The statistical parameters that are used are briefly described in the following sections and a set of recommended filter criteria is used to identify potentially suitable sites for the validation of GSA datasets.

4.1 General information

General information like a unique site identifier, site coordinates and source network are stored for each site in the database. A quick look from high-resolution satellite imagery is provided as well.
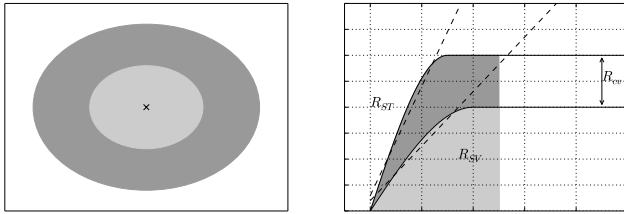
Figure 2. Geostatistical measures derived from variograms of footprints at different spatial scales according to Román et al. (2009). R_{cv}, R_{st} and R_{sv} quantify the relative coefficient of variation, the relative strength of the spatial autocorrelation and the structural variability respectively.

4.2 General geostatistical measures of spatial representativeness

Román et al. (2009) proposed several statistical parameters to express the representativeness of a location for the surrounding area. They are based on an omnidirectional semivariogram (γ) calculated as

$$\gamma(h) = \frac{1}{2} \frac{1}{N(h)} \sum_{i=1}^{N(h)} (z(x_i + h) - z(x_i))^2. \tag{1}$$

A spherical variogram model (Matheron, 1963) is then fitted to the empirical semivariogram with three parameters (a: range; c: sill; c_0: nugget). The semivariance can be estimated for areas of different sizes. Let us assume two areas X and Y, where $Y > X$, then different measures of representativeness can be estimated (Fig. 2). The first measure is the relative coefficient of variation R_{CV} that provides an estimate of the change of the variance compared to the mean value with changing spatial scale. It is defined as

$$R_{\mathrm{CV}} = \frac{\mathrm{CV}_Y - \mathrm{CV}_X}{\mathrm{CV}_X}, \tag{2}$$

where $\mathrm{CV}_{[X,Y]}$ are the coefficients of variation, defined by the ratio of the standard deviation to the mean, for areas at two different spatial scales and estimated at a distance $h > a$. The second parameter is the relative strength of the spatial autocorrelation (R_{ST}):

$$R_{\mathrm{ST}} = \frac{\mathrm{ST}_Y - \mathrm{ST}_X}{\mathrm{ST}_X}, \tag{3}$$

where

$$\mathrm{ST} = \frac{\gamma(a) - c_0}{\gamma(a) - \gamma(0)}. \tag{4}$$

A third parameter (R_{SV}) is used to quantify the relative change in structural variability. It is defined as

$$R_{\mathrm{SV}} = \frac{\mathrm{SV}_Y - \mathrm{SV}_X}{\mathrm{SV}_X}, \tag{5}$$
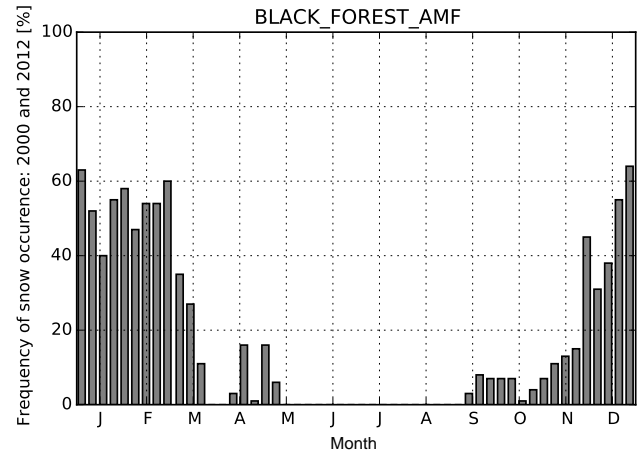

Figure 3. Example of frequency of snow occurrence for site in the Black Forest mountains.

where

$$\mathrm{SV} = \int_0^a \left(\frac{\gamma(h) - c_0}{c} \right) \mathrm{d}h. \tag{6}$$

An overall measure for the spatial representativeness is then defined as

$$\mathrm{ST}_{\mathrm{score}} = \left(\frac{|R_{\mathrm{CV}}| + |R_{\mathrm{ST}}| + |R_{\mathrm{SV}}|}{3} \right)^{-1}, \tag{7}$$

which is directly proportional to spatial representativeness. Thus, sites with high values should be more representative of their surroundings than sites with lower values. Further details are provided by Román et al. (2009), who define an additional measure for the representativeness of local in situ measurements which is not applied here as in situ reference data are not available for all sites within SAVS 1.0.

4.3 Topographic homogeneity

The topographic homogeneity is expressed in SAVS 1.0 using the geostatistical parameters defined in the previous section as well as the following parameters for circular areas around each site's center coordinates with radii (r) of 1, 2, 5, 10 and 20 km.

$$\text{Mean height (m)}: \bar{z} = \frac{1}{N(r)} \sum_i^N z_i. \tag{8}$$

Height standard deviation (m):

$$\sigma(r) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (z_i - \bar{z})^2}. \tag{9}$$

Earth Syst. Sci. Data, 8, 425–438, 2016
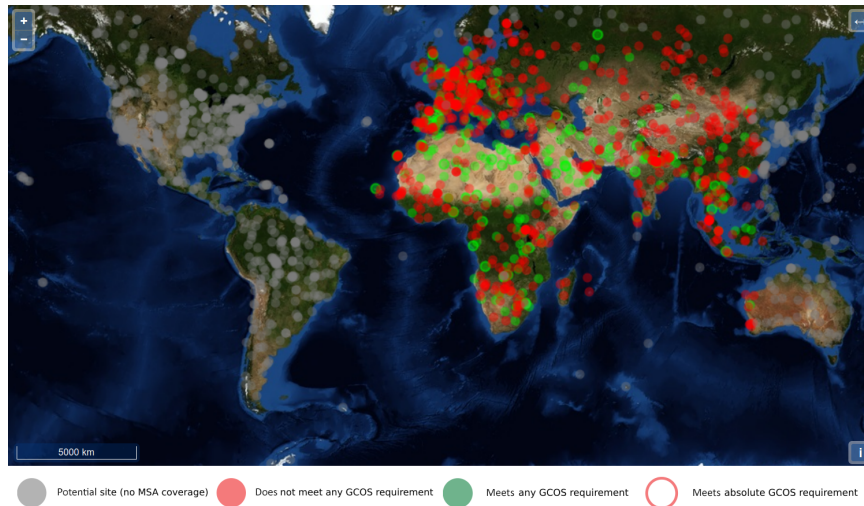
www.earth-syst-sci-data.net/8/425/2016/

Figure 4. Coverage of SAVS sites contained in the SAVS database (v1.0). Red and green dots represent sites within the footprint of Meteosat satellites, while grey sites represent sites outside of the Meteosat field of view.

Height range (m) :

$$\Delta z(r) = P_{95\,\%}(z(A[r])) - P_{5\,\%}(z(A[r])), \qquad (10)$$

where $A(r)$ corresponds to the area of the circle with radius r and $N(r)$ is the number of grid cells within that area. The height range Δz is estimated as the difference between the 95 and 5 % percentiles of the heights within area A to avoid unrealistic height ranges due to outliers.

4.4   Land cover homogeneity

To characterize the land cover homogeneity, the following parameters were derived from the extracted land cover subset:

- fractions of land cover classes within distances of 1, 2, 5, 10 and 20 km from the center coordinate;

- dominant land cover type within the same distances;

- distance to closest urban area (in kilometers);

- distance to closest open water bodies (in kilometers).

4.5   Vegetation homogeneity

Vegetated sites suitable for surface albedo validation require spatially and temporally homogenous vegetation conditions. Using the NDVI data provided with the ESA CCI land cover product (Bontemps et al., 2012), the spatial NDVI variability was characterized by the difference between the 5 and 95 % percentiles of the NDVI data. This was done for a variety of distances from the test site location ($r = 1, 2, 5, 10$ and 20 km) for a minimum annual NDVI map, representing conditions with lowest vegetation abundance, as well

as for conditions with maximum annual NDVI. In addition, the semi-variogram and the representativeness scores after Román et al. (2009) were calculated.
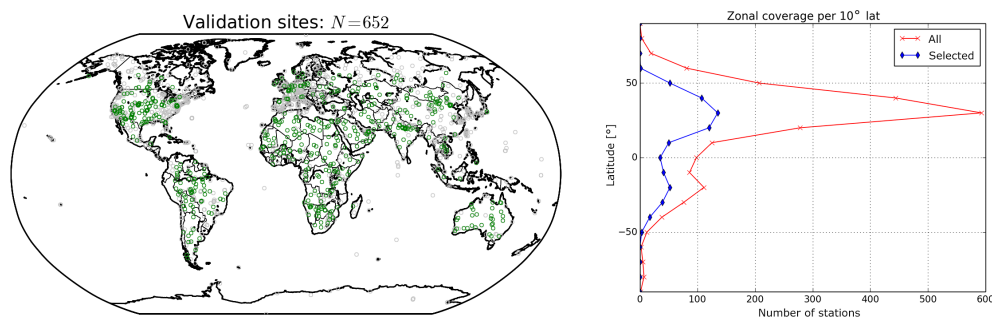
4.6   Disturbances

Occurrences of snow cover as well as disturbances such as fire complicate the validation of surface albedo data products as they induce abrupt changes of the surface albedo conditions. Thus, for many validation approaches one might want to filter the snow-covered areas or areas affected by fire as these show very different albedo conditions than normal. To provide a first mean to filter snow- or fire-affected sites, the SAVS 1.0 database also contains information on the probability of snow and fire occurrence, which allows the users to decide whether or not to include particular sites in their analysis. This information was again derived from the land cover condition information provided by the ESA CCI land cover product. The following disturbance attributes were derived for each site and are stored within SAVS 1.0.

- *Snow affected (true/false)*: true in the case that snow occurs at least once at any time during the year, meaning that at least one snow event was recorded in the observational record.

- *Snow probability*: the likelihood of snow occurrence within 8-day periods (t) derived from a multiannual time series, $P_{\text{snow}}(t) = N^{-1}\sum_{i=1}^{N} \text{snow}(t_i)$, where N is the number of years. An example is given in Fig. 3.

- *Fire affected (true/false)*: true in the case that fire occurs at least once at any time during the year, meaning that at least one fire occurred throughout the entire observational record.

Table 2. Criteria for identifying potential reference sites for validation of geostationary surface albedo data using SAVS 1.0.

Parameter	Threshold	Purpose
Latitude (°)	$\|\phi\| < 60$	Ensure coverage within geostationary observation domain.
Blacklisted	False	Ensure that blacklisted stations are not considered.
Land cover		
Distance to open water bodies (km)	10	Avoid open water bodies and their changing reflectance behavior with viewing geometry.
Minimum fraction of majority land cover type at 2 and 20 km distance	70 %	Avoid areas with heterogeneous land cover.
Topography		
Vertical range Δz (m) within a distance of 2 km	< 100	Avoid areas with significant terrain variability close to the investigated site.
NDVImax(5 km) − NDVImin(5 km) < x	$x = 0.1$	Avoid areas with high NDVI variability within a radius of 5 km. NDVImin and NDVImax correspond to the 5 and 95 % percentiles.

Figure 5. Selected GSA validation sites fulfilling the defined criteria (left) and zonal distribution in 10° latitude bands for all stations and selected stations (right).

– *Fire probability*: the likelihood of fire occurrence within 8-day periods derived from a multiannual analysis, $P_{\text{fire}}(t) = N^{-1} \sum_{i=1}^{N} \text{fire}(t_i)$, where N is the number of years.

4.7 Site preselection for geostationary surface albedo validation

A set of criteria was defined using the homogeneity criteria defined above to filter sites with a potential for mesoscale surface albedo stability analysis (Table 2). Several tests are then applied for each of the sites to check whether a site fulfills the given criteria. As a site might pass a test for some of the metrics but fail for another, the total number of passed tests is stored in the final database as well. This enables the user to easily filter the database in accordance to the number of successful tests and assign their own mechanisms to select suitable stations thereafter. A site was selected to be suitable for the validation of mesoscale surface albedo data products when it passed at least three of the different tests detailed in Table 2. It is emphasized that these are only a first recommendation by the SAVS 1.0 producers, but the database

provides all required information to the user to allow for a very flexible and task-specific filtering of the entire database.

4.8 Site reports

Results for each site are summarized in a comprehensive report. The report is based on a template which can be easily adapted according to the output format. Results can be viewed using any kind of web browser without the need for additional software. A summary page with all processed sites is provided which indicates whether a particular site matches one of the GCOS criteria on broadband surface albedo or not. The summary page also contains further information about the spatial site coverage which can be exported to various formats for usage in common Geographic Information Systems. All reports, as well as the SAVS 1.0 database itself, are accessible through the EUMETSAT website (http://savs.eumetsat.int/).
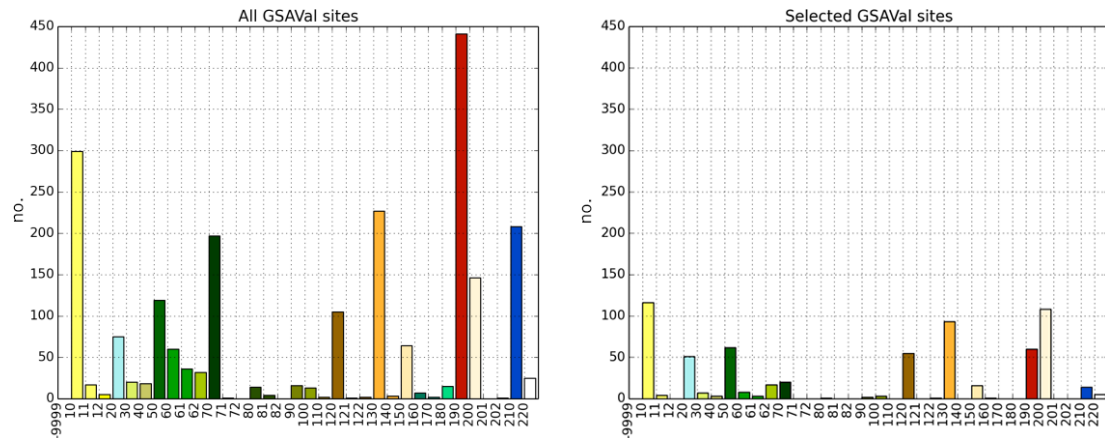
Figure 6. Land cover types of all sites identified (left) and selected sites (right). See Appendix A for an explanation of the land cover labels.

5   Results

The SAVS database comprises a total of 2186 sites which were all characterized in terms of their temporal and spatial homogeneity (Fig. 4). The sites cover a wide range of latitudes, land cover types and surface conditions and are therefore expected to provide a representative subset of surface conditions suitable for the evaluation of geostationary surface data products. Each site is characterized by a unique identifier. The database itself is provided in two simple text-based data formats which can be easily processed:

– JSON (JavaScript object notation; ECMA, 2013) allows for the storage of hierarchical data of any type in a simple text format. JSON is a text format that is completely independent of language but uses familiar programming conventions. It can be easily parsed by libraries available in different programming languages (http://www.json.org). These properties make JSON an ideal data-interchange format.

– CSV (comma-separated value) is used as an additional output to facilitate direct import into spreadsheet analysis programs or other analysis software.

A total of 652 sites were identified to fulfill at least three of the filter criteria for the validation of GSA data as defined in Table 2. They cover a large portion of the globe (Fig. 5, left). Figure 5 (right) shows the number of identified stations per 10° latitude band with maximum in the 30–40° N latitude band. Relatively few sites were identified for latitudes larger than 50°. Identifying a larger number of sites in the boreal area would be beneficial for the validation of surface albedo (and other) data products derived from polar-orbiting satellites.

The selected sites encompass a large variety of different land cover types. Figure 6 shows the histogram of different land cover types for all the sites within SAVS 1.0 as well as for the sites within the Meteosat footprint. The dominant
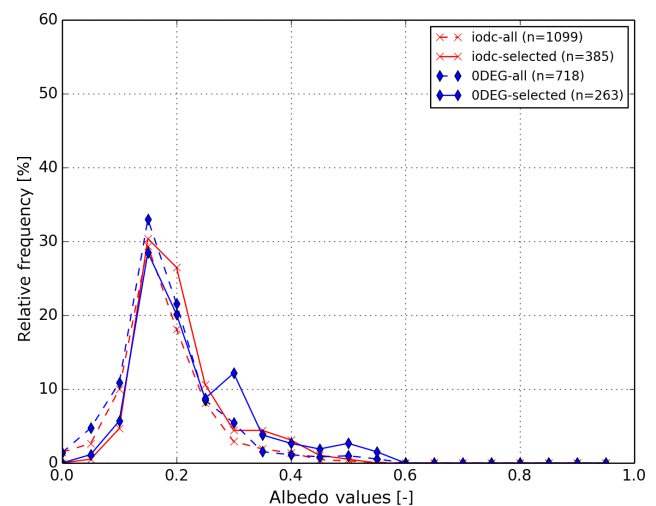

Figure 7. Histogram of surface albedo values covered by the SAVS database (dashed lines: all sites; solid lines: filtered sites).

land cover types of the GSA validation sites are cropland, grassland and bare areas. The different sites also cover a wide range of surface albedo conditions.

Figure 7 shows the frequency of the albedo values covered by the entire SAVS 1.0 sites as well as those identified by applying the filter criteria. It is, however, emphasized that the resulting subset of sites provides potential validation sites which then need to be carefully analyzed further with respect to regarding their temporal stability and the availability of in situ reference data.

6   Conclusions

The SAVS 1.0 database provides a comprehensive database for the characterization of potential sites for surface albedo validation. It consists of a statistical analysis of the site, using ancillary information from external land datasets. Ver-

sion 1.0 of the database is hosted by EUMETSAT (http://savs.eumetsat.int/) and contains 2186 sites where the spatial and temporal homogeneity was characterized using, in a traceable manner, a variety of statistical metrics. A set of recommended filter functions found to be most suitable for the evaluation of medium-scale GSA data products is proposed here. However, as all metrics are available to the user, the sites can be easily filtered according to user-specific criteria. The above notwithstanding, the user should be aware of the limitations of individual metrics, as these might be uncertain by themselves. Examples are the inherent uncertainties in the auxiliary datasets which are used, like land cover or NDVI maps. All auxiliary data are based on the time periods covered by the datasets used for the attribution of the SAVS sites. A regular update of the auxiliary datasets might therefore be desirable.

The SAVS database contributes to the CEOS-LPV activities and might also be of interest for validation stud-
ies beyond surface albedo applications. It is based on a traceable approach to characterize the individual sites using publicly available datasets. Further potential improvements of the SAVS database comprise the integration and cross-comparison of data from arbitrary surface albedo data products. As the processing scheme to characterize the SAVS sites is fully automated, a further improvement might be the development of a web-based user interface that allows the easy integration of new sites and datasets defined by a user.

7   Data availability

The SAVS 1.0 database is available through the EU-METSAT website under http://savs.eumetsat.int/. The current version has the following digital object identifier: doi:10.15770/EUM_SEC_CLM_1001.

Appendix A: Land cover labels

Table A1. Land cover types used within SAVS 1.0 as based on the ESA CCI land cover data product (Bontemps et al., 2012).

ID	Land cover type
10	Cropland, rainfed
11	Cropland, rainfed, herbaceous cover
12	Cropland, tree or shrub cover
20	Cropland, irrigated or post-flooded
30	Mosaic cropland (> 50 %), natural veg. (< 50 %)
40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (> 50 %)/cropland (< 50 %)
50	Tree cover, broadleaved, evergreen, closed to open (> 15 %)
60	Tree cover, broadleaved, deciduous
61	Tree cover, broadleaved, deciduous (closed > 40 %)
62	Tree cover, broadleaved, deciduous (open 15–40 %)
70	Tree cover, needleleaved, evergreen, closed to open (> 15 %)
71	Tree cover, needleleaved, evergreen, closed to open (> 40 %)
72	Tree cover, needleleaved, evergreen, closed to open (15–40 %)
80	Tree cover, needleleaved, deciduous, closed to open (> 15 %)
81	Tree cover, needleleaved, deciduous, closed (> 40 %)
82	Tree cover, needleleaved, deciduous, open (15–40 %)
90	Tree cover, mixed leaf type
100	Mosaic tree and shrub (> 50 %)/herbaceous cover (< 50 %)
110	Mosaic herbaceous cover (> 50 %)/tree and shrub (< 50 %)
120–122	Shrubland
130	Grassland
140	Lichens and mosses
150	Sparse vegetation (tree, shrub, herbaceous cover) (< 15 %)
160	Tree cover, flooded, fresh or brackish water
170	Tree cover, flooded, saline water
180	Shrub or herbaceous cover, flooded, fresh/saline/brackish water
190	Urban areas
200	Bare areas
201–202	Unknown
210	Water bodies
220	Permanent snow and ice
−9999	Unknown

Appendix B: List of attributes stored in database

Table B1. The following table gives an overview about all attributes stored for each SAVS site within SAVS 1.0.

Attribute	Type/unit	Range	Remark
Generic information			
ID	Char		Unique identifier for site
Latitude	Float/degree	$-90\ldots90$	
Longitude	Float/degree	$-180\ldots180$	
Source network	Char		Name of network the site originates from 0DEG/Indian Ocean coverage
Coverage	Bool	0/1	Specifies whether the site is located in the 0DEG/Indian Ocean coverage of the Meteosat satellites
Zenith angle	Float	$0\ldots90$	Specifies the nominal sensor zenith angle for the Meteosat satellites if the site is covered by those
gsa subset	char		Filename of extracted GSA long-term albedo dataset for site
Blacklisted	bool		Indicated if a site was blacklisted due to spurious data
Topography			
Heterogeneity parameters after Román et al. (2009): R_{st}, R_{cv}, R_{sv}, St			Román et al. (2009); see Eqs. (1)–(7)
Height difference (Δz) between 5 and 95 % percentiles at 1, 2, 5, 10, 20 km	(m)	–	See Eq. (10)
Land cover information			
Majority land cover type at 1, 2, 5, 10, 20 km	Char	–	
Area fraction of majority land cover type at 1, 2, 5, 10, 20 km	Float	$0\ldots1$	
Frequency distribution of land cover types within radius of 1, 2, 5, 10, 20 km			
Minimum distance to open water bodies	Float/km		
Minimum distance to urban areas	Float/km		
Vegetation homogeneity at 1, 2, 5, 10, 20 km			
Difference between NDVI extreme values (5, 95 % percentiles) within radius R for $\mathrm{NDVI_{MIN}}$	Float	$-2\ldots2$	
Difference between NDVI extreme values (5, 95 % percentiles) within radius R for $\mathrm{NDVI_{MAX}}$	Float	$-2\ldots2$	
Heterogeneity parameters after Román et al. (2009): R_{st}, R_{cv}, R_{sv}, St			Román et al. (2009); see Eqs. (1)–(7)
Disturbances			
Fire frequency: the probability of fire occurrence is stored for each day of the year	Float	$0\ldots1$	
Snow frequency: the probability of snow occurrence is stored for each day of the year	Float	$0\ldots1$	
$\mathrm{has_{snow}}$	Bool	0/1	Site is in general affected by snow: $\max(Pr(\mathrm{snow})) > 0$
$\mathrm{has_{fire}}$	Bool	0/1	Site is in general affected by fire: $\max(Pr(\mathrm{fire})) > 0$
Albedo dataset characterization			
Mean albedo value for area mean	Float	$0\ldots1$	
Temporal standard deviation of albedo within subset for area mean	Float	$0\ldots1$	
Mean albedo for site location	Float	$0\ldots1$	
Temporal standard deviation for albedo at site location	Float	$0\ldots1$	
Suitability for geostationary albedo validation			
Number and type of homogeneity tests passed			See Table 2
Albedo long-term stability analysis			
Linear regression parameters (slope, intercept for long-term albedo dataset using either weighted or ordinary least square approaches	Float		
Probability that at least one of the GCOS criteria for long-term stability of albedo ECV records is met	Bool	0/1	

Author contributions. Alexander Loew developed the technical part of the SAVS 1.0 database and wrote the major parts of the manuscript. Ralf Bennartz was responsible for the processing of the Meteosat albedo data. Frank Fell was a beta user of the SAVS 1.0 database for the analysis of the accuracy of the MSA data product. Alessio Lattanzio, Marie Doutriaux-Boucher and Jörg Schulz followed the SAVS development from EUMETSAT and gave very constructive feedback for the further improvement of the database and finally hosted the database on the EUMETSAT website. All co-authors contributed equally to the improvement of the manuscript.

References

ACE: available at: http://tethys.eaprs.cse.dmu.ac.uk/ACE2/ (21 September 2016), Tech. rep., 2014.

Augustine, J. A., DeLuisi, J. J., and Long, C. N.: SURFRAD – A National Surface Radiation Budget Network for Atmospheric Research, B. Am. Meteorol. Soc., 81, 2341–2357, doi:10.1175/1520-0477(2000)081<2341:sansrb>2.3.co;2, 2000.

Augustine, J. A., Hodges, G. B., Cornwall, C. R., Michalsky, J. J., and Medina, C. I.: An Update on SURFRAD – The GCOS Surface Radiation Budget Network for the Continental United States, J. Atmos. Ocean. Tech., 22, 1460–1472, doi:10.1175/jtech1806.1, 2005.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities, B. Am. Meteorol. Soc., 82, 2415–2434, doi:10.1175/1520-0477(2001)082<2415:fantts>2.3.co;2, 2001.

Baret, F., Morissette, J., Fernandes, R., Champeaux, J., Myneni, R., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S., and Nickeson, J. E.: Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: proposition of the CEOS-BELMANIP, IEEE T. Geosci. Remote Sens., 44, 1794–1803, doi:10.1109/tgrs.2006.876030, 2006.

Berry, P. A. M., Smith, R., and Benveniste, J.: ACE2: the new Global Digital Elevation Model, Tech. rep., Chania, Crete, 2008.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy, B. Am. Meteorol. Soc., 95, 1431–1443, doi:10.1175/BAMS-D-13-00047.1, 2014.

Bontemps, S., Herold, M., Kooistra, L., van Groenestijn, A., Hartley, A., Arino, O., Moreau, I., and Defourny, P.: Revisiting land cover observation to address the needs of the climate modeling community, Biogeosciences, 9, 2145–2157, doi:10.5194/bg-9-2145-2012, 2012.

Brovkin, V., Boysen, L., Raddatz, T., Gayler, V., Loew, A., and Claussen, M.: Evaluation of vegetation cover and land-surface albedo in MPI-ESM CMIP5 simulations, J. Adv. Model. Earth Syst., 5, 48–57, doi:10.1029/2012MS000169, 2013.

Cescatti, A., Marcolla, B., Santhana Vannan, S. K., Pan, J. Y., Román, M. O., Yang, X., Ciais, P., Cook, R. B., Law, B. E., Matteucci, G., Migliavacca, M., Moors, E., Richardson, A. D., Seufert, G., and Schaaf, C. B.: Intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network, Remote Sens. Environ., 121, 323–334, doi:10.1016/j.rse.2012.02.019, 2012.

EUMETSAT: Meteosat Surface Albedo Retrieval: Algorithm Theoretical Basis Document, Report, Doc.#EUM/OPS/SPE/12/3367, available at: http://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=PDF_MSA_ATBD&RevisionSelectionMethod=LatestReleased&Rendition=Web (last access: 21 September 2016), 2014.

Fell, F., Bennartz, R., Cahill, B., Lattanzio, A., Muller, J.-P., Schulz, J., Shane, N., Trigo, I., and Watson, G.: Evaluation of the Meteosat Surface Albedo Climate Data Record (ALBEDOVAL), Final Report, Tech. rep., EUMETSAT, available at: http://www.eumetsat.int/website/home/Data/ClimateService/index.html (last access: 21 September 2016), 2012.

Govaerts, Y. M. and Lattanzio, A.: Retrieval error estimation of surface albedo derived from geostationary large band satellite observations: Application to Meteosat-2 and Meteosat-7 data, J. Geophys. Res., 112, D05102, doi:10.1029/2006jd007313, 2007.

Hagemann, S., Loew, A., and Andersson, A.: Combined evaluation of MPI-ESM land surface water and energy fluxes, J. Adv. Model. Earth Syst., 5, 259–286, doi:10.1029/2012MS000173, 2013.

Holben, B., Eck, T., Slutsker, I., Tanré, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sens. Environ., 66, 1–16, doi:10.1016/s0034-4257(98)00031-5, 1998.

Holben, B. N., Tanré, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., Newcomb, W. W., Schafer, J. S., Chatenet, B., Lavenu, F., Kaufman, Y. J., Castle, J. V., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karneli, A., O'Neill, N. T., Pietras, C., Pinker, R. T., Voss, K., and Zibordi, G.: An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, J. Geophys. Res.-Atmos,, 106, 12067–12097, doi:10.1029/2001JD900014, 2001.

www.earth-syst-sci-data.net/8/425/2016/

Earth Syst. Sci. Data, 8, 425–438, 2016

Houldcroft, C. J., Grey, W. M. F., Barnsley, M., Taylor, C. M., Los, S. O., and North, P. R. J.: New Vegetation Albedo Parameters and Global Fields of Soil Background Albedo Derived from MODIS for Use in a Climate Model, J. Hydrometeorol., 10, 183–198, doi:10.1175/2008jhm1021.1, 2009.

König-Langlo, G., Sieger, R., Schmithüsen, H., Bücker, A., Richter, F., and Dutton, E. G.: The Baseline Surface Radiation Network and its World Radiation Monitoring Centre at the Alfred Wegener Institute, GCOS report 174, Alfred Wegener Institute, available at: http://www.wmo.int/pages/prog/gcos/Publications/gcos-174.pdf (last access: 21 September 2016), 2013.

Lattanzio, A., Govaerts, Y., and Pinty, B.: Consistency of surface anisotropy characterization with meteosat observations, Adv. Space Res., 39, 131–135, doi:10.1016/j.asr.2006.02.049, 2007.

Lattanzio, A., Fell, F., Bennartz, R., Trigo, I. F., and Schulz, J.: Quality assessment and improvement of the EUMETSAT Meteosat Surface Albedo Climate Data Record, Atmos. Meas. Tech., 8, 4561–4571, doi:10.5194/amt-8-4561-2015, 2015.

Loew, A.: Terrestrial satellite records for climate studies: how long is long enough? A test case for the Sahel, Theor. Appl. Climatol., 115, 427–440, doi:10.1007/s00704-013-0880-6, 2014.

Loew, A. and Govaerts, Y.: Towards Multidecadal Consistent Meteosat Surface Albedo Time Series, Remote Sensing, 2, 957–967, doi:10.3390/rs2040957, 2010.

Matheron, G.: Principles of geostatistics, Econ. Geol., 58, 1246–1266, doi:10.2113/gsecongeo.58.8.1246, 1963.

Pinty, B., Roveda, F., Verstraete, M. M., Gobron, N., Govaerts, Y., Martonchik, J. V., Diner, D. J., and Kahn, R. A.: Surface albedo retrieval from Meteosat: 1. Theory, J. Geophys. Res.-Atmos., 105, 18099–18112, doi:10.1029/2000JD900113, 2000a.

Pinty, B., Roveda, F., Verstraete, M. M., Gobron, N., Govaerts, Y., Martonchik, J. V., Diner, D. J., and Kahn, R. A.: Surface albedo retrieval from Meteosat: 2. Applications, J. Geophys. Res., 105, 18113–18134, doi:10.1029/2000jd900114, 2000b.

Riihelä, A., Laine, V., Manninen, T., Palo, T., and Vihma, T.: Validation of the Climate-SAF surface broadband albedo product: Comparisons with in situ observations over Greenland and the ice-covered Arctic Ocean, Remote Sens. Environ., 114, 2779–2790, doi:10.1016/j.rse.2010.06.014, 2010.

Riihelä, A., Manninen, T., Laine, V., Andersson, K., and Kaspar, F.: CLARA-SAL: a global 28 yr timeseries of Earth's black-sky surface albedo, Atmos. Chem. Phys., 13, 3743–3762, doi:10.5194/acp-13-3743-2013, 2013.

Román, M. O., Schaaf, C. B., Woodcock, C. E., Strahler, A. H., Yang, X., Braswell, R. H., Curtis, P. S., Davis, K. J., Dragoni, D., and Goulden, M. L.: The MODIS (Collection V005) BRDF/albedo product: Assessment of spatial representativeness over forested landscapes, Remote Sens. Environ., 113, 2476–2498, doi:10.1016/j.rse.2009.07.009, 2009.

Schaaf, C. B., Cihlar, J., Belward, A., Dutton, E., and Verstraete, M.: Assessment of the status of the development of the standards for the Terrestrial Essential Climate Variable: T8 – Albedo, Tech. rep., GTOS, Rome, Italy, available at: http://www.fao.org/gtos/doc/ECVs/T08/T08.pdf (last access: 21 September 2016), 2009.

Schaaf, C. L. B., Liu, J., Gao, F., and Strahler, A. H.: Land Remote Sensing and Global Environmental Change: NASA's Earth Observing System and the Science of ASTER and MODIS, chap. MODIS Albedo and Reflectance Anisotropy Products from Aqua and Terra, Springer, New York, 2011.

Tsendbazar, N.-E., de Bruin, S., Fritz, S., and Herold, M.: Spatial Accuracy Assessment and Integration of Global Land Cover Datasets, Remote Sensing, 7, 15804, doi:10.3390/rs71215804, 2015.

Waide, R. and Thomas, M.: Encyclopedia of Sustainability Science and Technology, chap. Long Term Ecological Research Network, Springer, New York, 2012.

Wang, K., Augustine, J., and Dickinson, R. E.: Critical assessment of surface incident solar radiation observations collected by SURFRAD, USCRN and AmeriFlux networks from 1995 to 2011, J. Geophys. Res., 117, D23105, doi:10.1029/2012jd017945, 2012.

Weiss, M., Baret, F., Block, T., Koetz, B., Burini, A., Scholze, B., Lecharpentier, P., Brockmann, C., Fernandes, R., Plummer, S., Myneni, R., Gobron, N., Nightingale, J., Schaepman-Strub, G., Camacho, F., and Sanchez-Azofeifa, A.: On Line Validation Exercise (OLIVE): A Web Based Service for the Validation of Medium Resolution Land Products. Application to FAPAR Products, Remote Sens., 6, 4190–4216, doi:10.3390/rs6054190, 2014.

WMO: Systematic observation requirements for satellite-based data products for climate modelling, Supplemental details to the satellite-based component of the "Implementation Plan for the Global Observing System for Climate in Support of the UN-FCCC (2010 Update)", GCOS-154, Geneva, 2011.