Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

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], 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. To validate the accuracy and longterm stability of terrestrial satellite data products, a network of reference sites is required. The present paper documents a 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 might be useful also for other applications. The publically available database (SAVS 1.0) is available through the EUMETSAT website (http://savs.eumetsat.int/) and allows to filter the sites according to different criteria. It provides a flexible mean to identify potential validation sites for further studies and a traceable approach to characterize the heterogeneity of these reference sites. The present manuscript 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) which is 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 (NWP), hydrology, agricultural monitoring, or climate modelling (e.g. Brovkin et al., 2013; Hagemann et al., 2013; Houldcroft et al., 2009).

First multi-decadal data products of surface albedo have been released in recent years from either polar orbiting or geostationary satellites (Lattanzio et al., 2015; Riihelä et al., 2010; Riihelä et al., 2013). Geostationary satellite sensors provide a unique opportunity for the estimation of long-term surface albedo data records due to their multi-decadal observational record. They provide spatial resolutions in the order of 1-10 km, further on 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 e.g. the CEOS (Committee on Earth Observation Satellites) Land Product Validation team (CEOS-LPV) as well as international projects such as e.g. 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 data sets. Multi-decadal records are typically derived from a series of sensors on subsequently operated observation platforms. A careful instrument inter-calibration is required to avoid that changes and drifts in the observing system cause spurious trends in the retrieved surface albedo data products themselves (Loew, 2014; Loew and Govaerts, 2010; Riihelä et al., 2013). 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 have 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 application for the validation of geostationary surface albedo (GSA) data products. SAVS 1.0 provides a traceable approach to characterize potential sites for EO 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

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

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

representativeness of reference data needs to be quantified (e.g. 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, terrestrial carbon fluxes. However, to our knowledge, none of the existing networks has been tailored to the specific needs for the evaluation of satellite data products and in particular geostationary surface albedo 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 (Figure 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 landcover;
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 geostationary surface albedo data.

The evaluation of surface albedo data products requires typically to take into account diurnal variations in surface reflectances like e.g. terrain induced shadowing, geometric uncertainties due to navigation uncertainties as well as anisotropic effects due to the change during the day of the Sun position. In particular for geostationary satellite based surface albedo data products which have coarse spatial resolutions, these factors are very relevant. Stringent requirements on the characteristics of a reference site suitable for coarse scale surface albedo evaluation 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.

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

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

following sections and the selection criteria to identify potential geostationary surface albedo validation sites are laid out in Sect. 4.7.

2.2 SAVS 1.0 processing workflow

The overall workflow for generating the SAVS 1.0 database is provided in Figure 1. Site information from various networks is taken to populate the initial database. For all these sites, ancillary data is 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 also satellite surface albedo data can be ingested to build a timeseries of satellite measurements at the location of the site.

The such obtained set of sites is 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 e.g. distance to coastlines or spurious ancillary data like e.g. errors in the used ACE-2 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 2.

This first set of potential reference sites was screened to identify gaps with respect to geographic distribution, biome coverage, and albedo value ranges. 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 coordinates and duplicates were subsequently removed from the database. For some stations (e.g. `MONGU`), duplicates were identified, but they differed by more than 10 km in distance. It was therefore not clear whether these sites correspond to different locations or if there are 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.

3.2 Elevation data

Topographic information is based on ACE-2 (ACE, 2014; Berry et al., 2008), a global digital elevation model (DEM) providing surface elevation data at a spatial resolution of 3 arcsec, or about 90 m. For each site coordinate, the surrounding

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

Open Access Earth System Science Data Discussions

topography within a 25 km radius was extracted from the ACE-2 dataset. Thus, an area of approximately 2000 km² of surface elevation data was extracted for each site for further analysis.

3.3 Landcover 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.

The Normalized Difference Vegetation Index (NDVI) is a good proxy for the abundance and seasonality of vegetation. The CCI land cover dataset provides information on the mean seasonality of the NDVI with a temporal resolution of eight days at pixel level, derived from SPOT Vegetation data for the period 1998 to 2012. For each pixel and day of year 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.

In addition, the ESA CCI land cover dataset provides information on the 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.

3.4 EUMETSAT geostationary surface albedo 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 this Meteosat Surface Albedo (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 consists 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 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 fast cloud detection method (Pinty et al., 2000b).

The algorithm has been applied to the visible (VIS) channel of the Meteosat Visible and InfraRed Imager (MVIRI) (Lattanzio et al., 2007) to provide the EUMETSAT Meteosat Surface Albedo (MSA). 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 allows to provide quantitative uncertainty estimates for each retrieval result (Govaerts and Lattanzio, 2007). Loew and Govaerts (2010) did provide an

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

update of calibration coefficients to maximize the temporal stability of the MSA multi-decadal record. A subset of 50x50 km² 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 external dataset, is exploited to give a full site characterization. The used statistical parameters are briefly described in the following sections and a set of recommended filter criteria is used to identify sites potentially suitable for the validation of geostationary surface albedo datasets.

4.1 General information

General information like a unique site identifier (ID), 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.

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

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

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, whereas Y>X, then different measures of representativeness can be estimated (Figure 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

$$R_{CV} = \frac{CV_Y - CV_X}{CV_X} \quad (2)$$

whereas $CV_{\{X,Y\}}$ 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})

$$R_{ST} = \frac{ST_Y - ST_X}{ST_X} \quad (3)$$

where

$$ST = \frac{\gamma(a) - c_0}{\gamma(a) - \gamma(0)} \quad (4)$$

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

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

$$R_{SV} = \frac{SV_Y - SV_X}{SV_X} \quad (5)$$

where

$$SV = \int_0^a \left(\frac{\gamma(h) - c_0}{c}\right) dh \quad (6)$$

An overall measure for the spatial representativeness is then defined as

$$ST_{score} = \left(\frac{|R_{CV}| + |R_{ST}| + |R_{SV}|}{3}\right)^{-1} \quad (7)$$

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 in Román et al. (2009) where they define also 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 km , 2 km, 5 km, 10 km, and 20 km.

Mean height [m]: $$\bar{z}(r) = \frac{1}{N(r)} \sum_{i=1}^{N(r)} z_i \quad (8)$$

Height Standard deviation [m] $$\sigma(r) = \sqrt{\frac{1}{N(r)-1} \sum_{i=1}^{N(r)} (z_i - \bar{z})^2} \quad (9)$$

Height range [m] $$\Delta z(r) = P_{95}\left(z(A(r))\right) - P_{05}(z(A(r))) \quad (10)$$

Whereas $A(r)$ 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 km, 2 km, 5 km, 10 km, and 20 km from the centre coordinate,
- Dominant land cover type within the same distances,
- Distance to closest urban area [km],
- Distance to closest open water bodies [km].

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

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 km, 2 km, 5 km, 10 km, 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. The SAVS 1.0 database therefore contains also information on the probability of snow and fire occurrence in order to support users when deciding 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*: Likelihood of snow occurrence within eight day periods (t) derived from a multiannual time series as $P_{snow}(t) = \frac{\sum_{i=1}^{N} has\ snow(t)}{N}$, where N is the number of years. An example is given in Figure 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.
- *Fire probability*: Likelihood of fire occurrence within eight day periods derived from a multiannual analysis as $P_{fire}(t) = \frac{\sum_{i=1}^{N} has\ fire(t)}{N}$, 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 3). Several tests are then applied for each of the sites to check if a site does fulfill the given criteria. As a site might pass a test for some of the metrics while it is failing 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 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 3. It is emphasized that these are only a first recommendation by the SAVS 1.0 producers, but that the database provides all required information to the user to allow for a very flexible and task specific filtering of the entire database.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

4.8 Site reports

Results for each site are summarized in a comprehensive report. The report is based on a template which allows to easily adapting 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 is matching 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/).

5 Results

The SAVS database comprises a total of 2186 sites which were all characterized in terms of their temporal and spatial homogeneity (Figure 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 (Java Script object notation; ECMA, 2013) allows for storage of hierarchical data of any type in a simple text format. JSON is a text format that is completely language independent but uses familiar programming conventions. It can 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 fulfil at least three of the filter criteria for the validation of geostationary surface albedo data as defined in Table 3. They cover a large portion of the globe (Figure 5, left). Figure 5 (right) shows the number of identified stations per 10° latitude band with maximum in the 30° N - 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 sites different land cover types for both, all sites within SAVS 1.0 as well as for the sites within the Meteosat footprint. The dominant 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.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

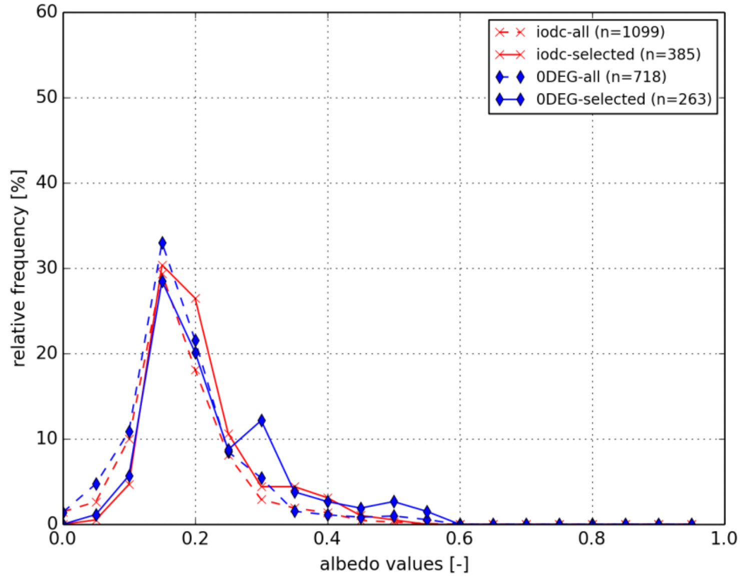

Figure 7 shows the frequency of the albedo values covered by the entire SAVS 1.0 sites as well as those identified by applying the filter criteria. It is however emphasized that the obtained subset of sites provides potential validation sites which then need to be further carefully analyzed regarding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

6 Data availability

The SAVS 1.0 database is available through the EUMETSAT website under http://savs.eumetsat.int/. The current version has the following Digital Object Identifier: doi: 10.15770/EUM_SEC_CLM_1001.

7 Conclusions

The SAVS 1.0 database provides a comprehensive database for the characterization of potential sites for surface albedo validation. It adds up a statistical analysis of the site to ancillary information from external land datasets. Version 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 using a variety of statistical metrics. A set of recommended filter functions found to be most suitable for the evaluation of medium scale geostationary surface albedo 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 SAVS database contributes to the CEOS-LPV activities and might be of interest also for validation studies beyond surface albedo applications.

The SAVS database is based on a traceable approach to characterize the individual sites using publically available datasets. Further potential improvements of the SAVS database comprise the integration and cross-comparison of data from arbitrary

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

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.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

8 Appendices

8.1 Appendix A: Land cover labels

Table 1: 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, decidous (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%), color_rgb(41,101,0)}})
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	Shrubland
121	Shrubland
122	Shrubland
130	grassland
140	Lichens and mosses

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

Open Access Earth System Science Data Discussions

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	unknown
202	unknown
210	water bodies
220	permanent snow and ice
-9999	unknown

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

Open Access Earth System Science Data Discussions

8.2 Appendix B: List of attributes stored in database

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 … 90	
longitude	float / degree	-180 … 180	
Source network	Char		Name of network the site originates from
0DEG / IODC coverage	Bool	0/1	Specifies if the site is located in the 0DEG or Indian ocean coverage (IODC) of the Meteosat satellites
Zenith angle	Float	0 … 90	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): Rst, Rcv, Rsv, St			Román et al. (2009); see eq. 1-7
Height difference (Δz) between 5% and 95% percentiles @ 1,2,5,10,20 km	[m]	-	See eq. 10
	Landcover information		
Majority landcover type @ 1,2,5,10,20 km	char	-	

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

Area fraction of majority landcover type @ 1,2,5,10,20 km	float	0 … 1	
Frequency distribution of landcover 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 @ 1,2,5,10,20 km			
Difference between NDVI extreme values (5%,95% percentiles) within radius R for NDVI_MIN	Float / -	-2 … 2	
Difference between NDVI extreme values (5%,95% percentiles) within radius R for NDVI_MAX	Float / -	-2 … 2	
Heterogeneity parameters after (Román et al., 2009): Rst, Rcv, Rsv, St			Román et al. (2009); see eq. 1-7
Disturbances			
Fire frequency: the probability of fire occurrence is stored for each day of the year	float	0 … 1	
Snow frequency: the probability of snow occurrence is stored for each day of the year	float	0 … 1	
has_snow	Bool	0/1	Site is in general affected by snow: max(Pr(snow)) > 0.
has_fire	Bool	0/1	Site is in general affected by fire: max(Pr(fire)) > 0.
Albedo dataset characterization			
Mean albedo value for area mean	Float / -	0 … 1	
Temporal standard deviation of albedo within subset for area mean	Float / -	0 … 1	

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

Open Access Earth System Science Data Discussions

Mean albedo for site location	Float / -	0 … 1	
Temporal standard deviation for albedo at site location	Float / -	0 … 1	
Suitability for geostationary albedo validation			
Number and type of homogeneity tests passed			See Table 3
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	

9 Author contribution

Alexander Loew developed the technical part of the SAVS1.0 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 SAVS1.0 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.

10 Acknowledgements

The authors thank EUMETSAT for funding the Albedoval-2 project in the context of which the SAVS database was established. We acknowledge in particular all agencies and individuals of the networks listed in Table 2 who contribute with their work to the evaluation and further improvement of terrestrial satellite data products. Work of the main author was in addition partly funded by the Cluster of Excellence CliSAP (EXC177), University of Hamburg, funded through the German Science Foundation (DFG).

11 References

ACE: http://tethys.eaprs.cse.dmu.ac.uk/ACE2/, 2014.

Baldocchi, D.: ‘Breathing’ of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems, Australian Journal of Botany, 56, 1-26, 2008.

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, Bulletin of the American Meteorological Society, 82, 2415-2434, 2001.

Baret, F., Morissette, J. T., Fernandes, R. A., Champeaux, J. L., Myneni, R. B., 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, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1794-1803, 2006.

Berry, P. A. M., Smith, R., and Benveniste, J.: ACE2: the new Global Digital Elevation Model, Chania, Crete2008.

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, Bulletin of the American Meteorological Society, 95, 1431-1443, 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, 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, Journal of Advances in Modeling Earth Systems, 5, 48-57, 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 Sensing of Environment, 121, 323-334, 2012.

EUMETSAT: Meteosat Surface Albedo Product: Users Manual, EUMETSAT, 2014.

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, Journal of Geophysical Research: Atmospheres, 112, n/a-n/a, 2007.

Hagemann, S., Loew, A., and Andersson, A.: Combined evaluation of MPI-ESM land surface water and energy fluxes, Journal of Advances in Modeling Earth Systems, 5, 259-286, 2013.

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, Journal of Hydrometeorology, 10, 183-198, 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, WCRP, 2013.

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, 2015.

Lattanzio, A., Govaerts, Y. M., and Pinty, B.: Consistency of surface anisotropy characterization with meteosat observations, Advances in Space Research, 39, 131-135, 2007.

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, 2014.

Loew, A. and Govaerts, Y.: Towards Multidecadal Consistent Meteosat Surface Albedo Time Series, Remote Sensing, 2, 957, 2010.

Matheron, G.: Principles of geostatistics, Economic Geology, 58, 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, Journal of Geophysical Research: Atmospheres, 105, 18099-18112, 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, Journal of Geophysical Research: Atmospheres, 105, 18113-18134, 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 Sensing of Environment, 114, 2779-2790, 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, 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., Goulden, M. L., Gu, L., Hollinger, D. Y., Kolb, T. E., Meyers, T. P., Munger, J. W., Privette, J. L., Richardson, A. D., Wilson, T. B., and Wofsy, S. C.: The MODIS (Collection V005) BRDF/albedo product: Assessment of spatial representativeness over forested landscapes, Remote Sensing of Environment, 113, 2476-2498, 2009.

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 Sensing, 6, 4190, 2014.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

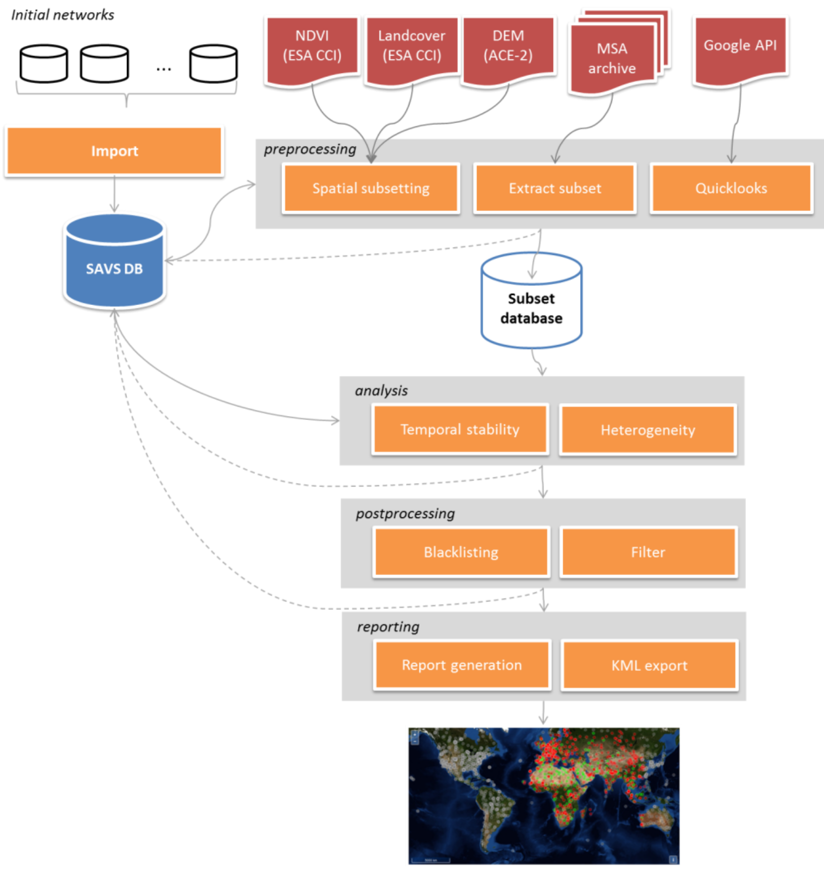

Figure 1: General workflow to identify and characterise potential GSA validation reference sites.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

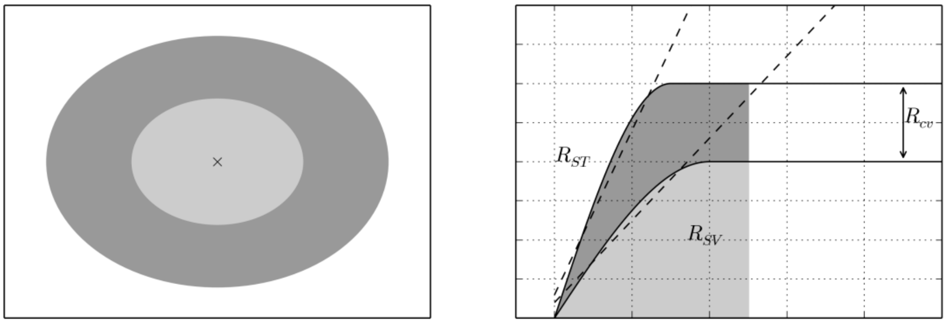

Figure 2: Geostatistical measures derived from variograms of footprints at different spatial scales according to Román et al. (2009). R_{cv}, R_{st}, and R_{sv} quantify the relative coefficient of variation, the relative strength of the spatial autocorrelation, and the structural variability respectively

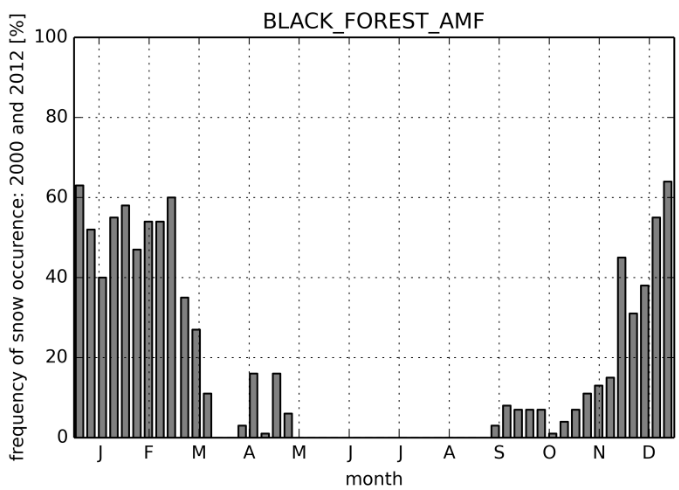

Figure 3: Example of frequency of snow occurrence for site BLACK_FOREST_AMF.

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

Open Access Earth System Science Data Discussions

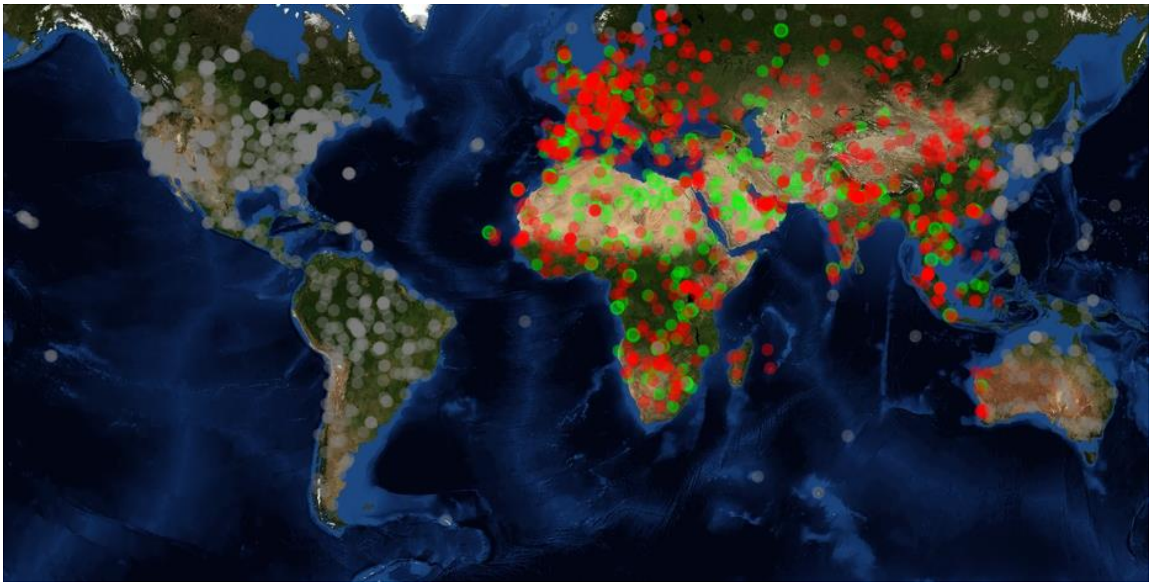

Figure 4: Coverage of SAVS sites contained in the SAVS database (v1.0). red and green dots represent sites within the footprint of Meteosat satellites, while grey sites represent sites outside of the Meteosat Field of View.

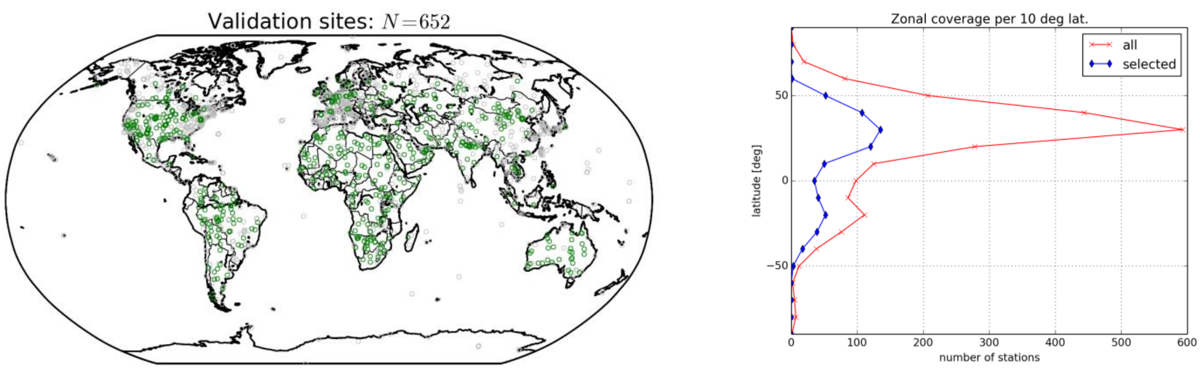

Figure 5: Selected GSA validation sites fulfilling the defined criteria (left) and zonal distribution in 10° latitude bands for all stations and selected stations (right).

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

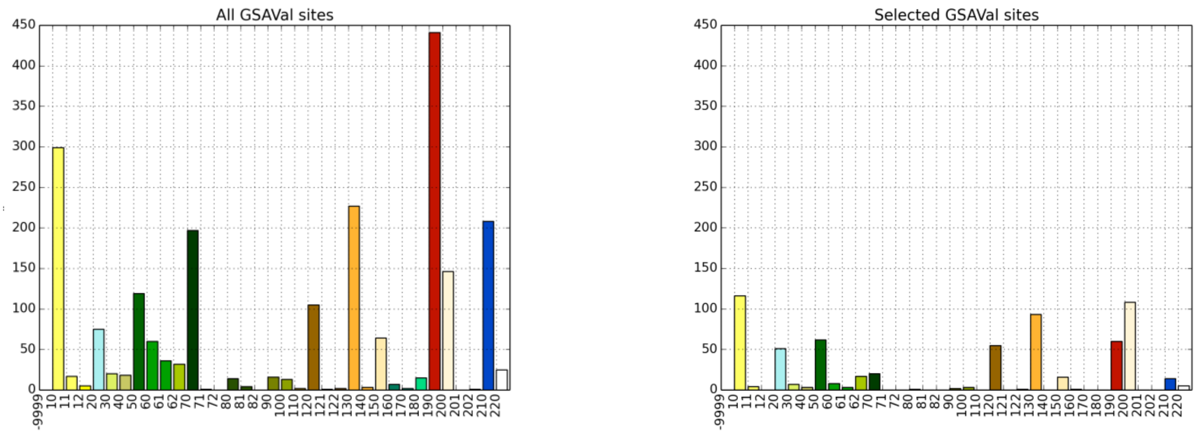

Figure 6: Land cover types of all sites identified (left) and selected sites (right). See Appendix B for an explanation of the land cover labels.

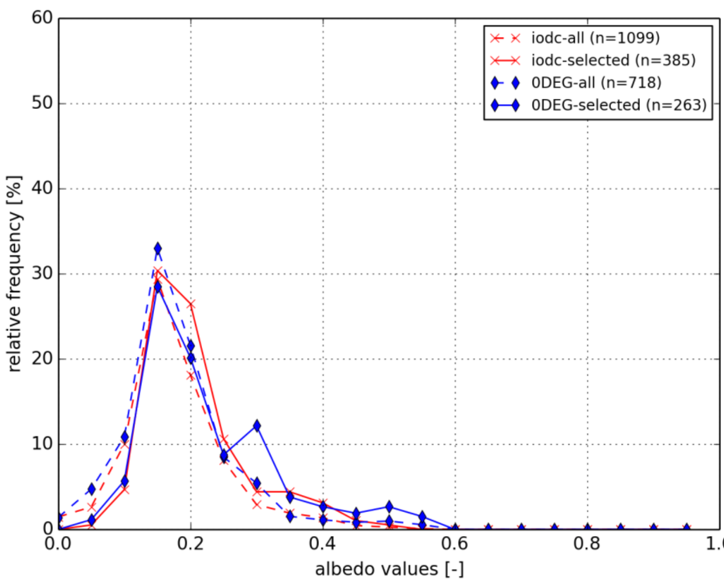

Figure 7: Histogram of surface albedo values covered by the SAVS database (dashed lines: all sites, solid lines: filtered sites).

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

Open Access Earth System Science Data Discussions

Table 2: Terrestrial reference networks considered within SAVS 1.0.

Network	Reference / Remark	# of sites
FLUXNET	(Baldocchi, 2008; Baldocchi et al., 2001) Initial study by (Cescatti et al., 2012) for 53 sites http://fluxnet.ornl.gov/	252
BSRN	König-Langlo et al. (2013) http://www.bsrn.awi.de/	63
Aeronet	http://aeronet.gsfc.nasa.gov/	1176
BELMANIP-2	(Baret et al., 2006; Weiss et al., 2014) http://calvalportal.ceos.org/web/olive/site-description	558
CEOS LandNet sites	http://calvalportal.ceos.org/ceos-landnet-sites	8
EOS core val. sites	http://landval.gsfc.nasa.gov/coresite_gen.html	41
Surfrad	http://www.esrl.noaa.gov/gmd/grad/surfrad/index.html	7
LTER	http://lternet.edu/	27
ALBEDOVAL-1	Incl. SAFARI2000, http://daac.ornl.gov/S2K/safari.shtml	40
Additional SAVS 1.0	Further sites identified by expert knowledge	48
SUM		**2220**
DUPLICATES		**34**
SUM final	After removal of duplicates	**2186**

Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-11, 2016
Manuscript under review for journal Earth Syst. Sci. Data
Published: 29 April 2016

Open Access Earth System Science Data Discussions

Table 3: Criteria for identifying potential reference sites for validation of geostationary surface albedo data using SAVS 1.0.

Parameter	Threshold	Purpose
Latitude	abs (lat) < 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 m	Avoid areas with significant terrain variability close to the investigated site.
NDVI		
NDVImax(5 km) – NDVImin(5 km) < x	x = 0.1	Avoid areas with high NDVI variability within a radius of 5km, whereas NDVImin and NDVImax correspond to the 5% and 95% percentiles.