Response to Reviewers

We thank the two reviewers very much for their time spent on the review and their constructive comments. We will respond to each reviewer question in the following. Responses will be indicated in blue. Changes in the revised manuscript are indicated in **bold font**.

Reviewer #1

With the recent advent of multidecadal timeseries of satellite-based surface albedo, the availability of appropriate reference data for the dataset validation is a topical issue. Here the authors present a new database which attempts to gather together information about the various in situ measurement networks of surface albedo and to present that information to interested users in a collated and standardized fashion. The effort behind the paper has clearly been substantive and the creation of a publicly available database itself is commendable. While there are some minor issues in the manuscript where the authors should provide some more relevant information and some points of caution to the user/reader, on the whole the paper is well written and does a good job of explaining the idea behind the database. As such, I recommend that the manuscript be approved or publication after minor revision

Thanks for this general judgment.

General comments on the manuscript:
1. The availability of multiple criteria for site selection is a good thing. However, some of the criteria presented may have weaknesses of their own, which should be mentioned in the manuscript. Specifically, the land cover classification used in the landcover homogeneity test can be uncertain; global hit rates of LU classification for CCI Land cover has been reported as around 75% [Tsendbazar et al., 2015], implying that the results of the land cover test may not always be trustworty. Secondly, NDVI is also not automatically a good proxy, as it does react to vegetation abundance and seasonality, as the authors note, but it does not react well to vegetation structure (understoryvs. overstory), which does influence the BRDF behaviour of the validation site area and thus affects the representativeness of the measurements. Some caveat emptor information for the reader is recommended

We agree that it is sensible to manage the expectation of the reader and raise the awareness on the limitations of the auxillary datasets used. We have therefore adopted the manuscript as follows:

1.) Added some additional information on accuracy and limitations of land cover classification into section 3.3. However, as we could not find the 75% accuracy for the CCI LC in the reference given by the reviewer (Tsendbazar et al., 2015) we used this reference, but kept the accuracy description more generic.
2.) Added additional information on limitations of NDVI data usage in section 3.3
3.) Added a few critical remarks on the limitations of certain attributes in the database in the conclusions section.

2. Similarly, the quality of the in situ measurement data is not equal between all networks. While I do not expect the authors to be capable of providing robust measurement accuracy numbers as a selection criteria – as such data is approximate at best - I do expect the authors to include some general summary of the quality evaluation literature of at least the largest measurement networks included in the database (Aeronet, BELMANIP, BSRN), least inexperienced readers assume that all data is created equal.

We agree that the in situ measurement quality and availability are not similar or comparable between the different networks. As we had already outlined in the original version of the manuscript, it is not the purpose of the SAVS database to provide only sites with reference data. Rather, we wish to document potential sites for validation. In that sense SAVS is first of all completely independent of any reference measurements. However, we agree it might be useful for the reader to have a clear understanding where they can expect reference data to be available. As this highly depends on the time periodwe have not tried to list available in situ measurements on

a site level. Instead, we decided to provide general information on data sources and potential available albedo reference data as additional information in Table 1. We have further added additional references on the networks that also contain information about the quality of available measurements and serve as a starting point for the reader. Last, we have also revised the text in the manuscript to describe the above subjects more clearly. Changes can be found in section 3.1 of the manuscript.

Minor comments:
1. pg 3, line 7: "in particular a geostationary..."
Done, thanks!

2. pg 3, line 20: "due to the change during the day of the Sun position" - English words, German grammar. Please revise.
Sounded odd, we agree. Sentence has been changed.

3. pg 5, lines 8-11: Some of the text refers to the NDVI data as day-of-year based, some as 8-day means. Which is correct
The temporal sampling of the CCI landcover conditions is actually 7 days. We have adapted the manuscript accordingly.

4. Although not relating to the manuscript itself, I encourage the authors to keep developing the SAVS database and its web access methods. It took me a while to find the actual database on the pages, and I would welcome a web interface allowing the user to filter the whole database with the criteria mentioned in the manuscript before downloading only that part of the database matching the given criteria. Of course, direct data access through the search results would be excellent, but probably outside the range of feasibility.

Thanks for this feedback! We agree that the data access could be facilitated through a better web interface. We will discuss what we can do in that respect. Regarding your suggestion to search and filter the data we agree in principle as well. The way the SAVS database was originally set up did not foresee a dedicated web user interface. The basic idea is that users can download the database themselves (different formats provided) and then either use their own software or simple spreadsheet applications to filter the database according to their needs. This is the reason why so far no advanced user access methods have been provided. As we agree that this might be useful nevertheless, we would discuss the options we have to improve the interface.

References provided by the reviewer:
Tsendbazar, N. E., de Bruin, S., Fritz, S., & Herold, M. (2015). Spatial Accuracy Assessment and Integration of Global Land Cover Datasets. Remote Sensing, 7(12),
15804-15821

Reviewer #2

General comments: The article "A database of global reference sites to support validation of satellite surface albedo data sets (SAVS 1.0)" by Loew et al., describes a new tool to identify surface field locations that may be best suited for satellite derived albedo product inter-comparisons and evaluations. The database provides information on homogeneity of land surface type and an estimation of the variation of minimum/maximum annual NDVI (both relying on the 2010 300m ESA CCI land cover), and an estimation of the surface topographic variation all within 1, 2, 5, 10, and 20km of the site center. This database was specifically developed for evaluation of geostationary data and therefore is relatively lacking in high latitude sites. Overall, however, this data set represents a highly useful contribution to the field. The description is appropriate with only three major caveats which need to be addressed

We thank the reviewer for this constructive initial statement

Specific comments:
Firstly, the data set presumes that homogeneity of land cover (within 1-20km) is the most important attribute in satellite derived albedo product evaluation and inter-comparison. While this is important and interesting information, evaluation can not only be restricted to homogeneous regions (so-called pure pixels) but must also be applied to heterogeneous regions (mixed pixels) to fully characterize a satellite product.

We fully agree with the statement that reference datasets need to be built also for heterogeneous cases. These can be either real sites where proper upscaling approaches are applied or even synthetic cases where the upscaling kernels are known by definition. The first is currently done within the CEOS-LPV group where protocols for albedo validation are currently under development (personal communication: Gabriela Schaepman-Strub). The latter, synthetic references sites, are currently explored by e.g. the JRC within the frame of the FP7-QA4ECV project.

Therefore it is important to not only identify field sites situated within homogeneous land covers but also should identify sites which lie within heterogeneous land cover combinations that are similarly heterogeneous within the 1-20km range. One immediately thinks of mixed forests or savanna locations that would represent a mix of classes in the close vicinity of a tower with high quality in situ measurements but which might or might not be very representative of the same mixed forest over the greater region due to species mix, structural variability, canopy cover, and timbering. One must also caution that heterogeneous covers like mixed forests can be quite uniformly homogeneous over large areas during the growing season but can become quite heterogeneous during the shoulder and winter seasons – yet the site may still (or may not) be quite representative of that heterogeneity present during these seasons. This difference between site homogeneity and site representativeness needs to be much better discussed.

We fully agree and have added further explanations on the site representativeness in both, spatial and temporal context, to the manuscript to raise awareness for the reader (in particular section 2.1).

Secondly, there is little assistance given to the user to reach the in situ data associated with each site other than the list of links in Table 2 of this document – more information should be including in the site information. This is particularly egregious with respect to the Baseline Surface Radiation Network (BSRN) sites which are explicitly identified in GTOS and GCOS documentation as the gold standard of in situ (tower) data for evaluation of satellite data products – but these sites (assuming they are representative) are not flagged as achieving GCOS criteria and standards in this database (note: Albedo and Reflectance Anisotropy, ECV-T8: GTOS Assessment of the status of the development of standards for the Terrestrial Essential Climate Variables, 2009 and GCOS document 154, pages 42 and 81).

The reviewer is asking to provide further information for availability of in situ data. As the SAVS database does not aim at providing any reference measurements, we believe it is best to document for the user where they can in principle get access to reference data. This is given by the links in Table 2. We do not provide explicit links to reference data for each of the sites (where available), as this information would need to be maintained, as references (websites) are changing and the reference data availability also varies over time and for each station. We therefore believe it is more appropriate to provide the necessary information for the user where the reference data could be accessed (e.g. for BSRN this would require two mouseclicks from the user with the information given in Table 2). We nevertheless further clarified in Table 2 for which sites albedo reference data would be available and also added further references to literature describing the accuracy of the ground measurements.

The reviewer is then referring to the GCOS criteria and mentions that the BSRN sites are not flagged to meet the GCOS criteria. We believe that this is a misunderstanding. GCOS provides „best principles“ for ground measurements as well as requirements for the accuracy and long-term stability of different ECVs (GCOS 2011). The SAVS database does **not** provide any information if a station has established best practice „golden standards“ as e.g. defined by GCOS. The SAVS sites were used to analyze the long-term stability of the EUMETSAT surface albedo data record. If a site was compliant with the temporal stability criteria defined in GCOS (2011), this was marked to match the GCOS requirement (green in Figure 1). Thus, this flagging is not related in any kind to the measurement practices on the ground.

References to the GCOS2011 (=GCOS #154) and GTOS-ECV-T8 have been added as suggested by the reviewer and the work of GCOS and GTOS is properly acknowledged. We could however not identify what the reviewer meant with the references to pages #42 and #81 in GCOS2011. Page #42 is related to Ozone and aerosol retrievals, while page #81 is related to faPAR.

References:

GCOS 2011: SYSTEMATIC OBSERVATION REQUIREMENTS FOR SATELLITE-BASED DATA PRODUCTS FOR CLIMATE 2011 Update. Supplemental details to the satellite-based component of the “Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update)”, December 2011. GCOS – 154

Furthermore there is little discussion of the in situ data associated with each site anywhere (other than the links in Table 2). The Ft Peck site is illustrative of this. It is listed as a member of the SURFRAD network. Nowhere is it mentioned that it is therefore also a member of the BSRN network (as SURFRAD sites are the US contribution to BSRN) and as a BSRN site, this indicates that in situ instrumentation are carefully and frequently maintained and calibrated and the in situ data are of the highest quality. Is this a case of a “duplicate” site which is mentioned in the text?? In the case of duplicates, all associated networks should be listed in the dataset descriptions – not just one network. Now, Ft Peck not only provides high quality data of a relatively homogeneous site but the FOV of the in situ instrumentation is also spatially representative of the greater area.

We see the need to raise the awareness for the reader that the different networks used are not independent and have also provide different kind of measurements as well as different kind of measurement qualities. We agree that it is useful to provide additional information on the source networks for the duplicates. This is straightforward for the sites that were recognized as obvious duplicates. For other sites, and Ft. Peck is one of these, it is actually not straight forward. The reason is that the matching for duplicates was done based on a key provided by the original networks. If the site key was the same, then it was identified as duplicate. If the key was different, then the site was NOT identified as duplicate. To give an example for Ft. Peck: This is actually part of three used networks with all different keys:

FLUXNET:	US-FPE
SURFRAD:	FORT_PECK
BSRN:	FPE

As a consequence, this particular site is contained three times in the current SAVS database. This means, that the source network traceability requested by the reviewer is actually already given. The disadvantage of matching by a given site key is that obviously some of the duplicates were not recognized.
In principle it would be better to use the station coordinates for resolving duplicates. However, this is not straight forward as coordinates differ. The Ft. Peck has actually three different sets of coordinates in the three networks involved. Thus, an elimination of stations by coordinate was not done for SAVS 1.0.

The text in the previous version of the manuscript was perhaps imprecise in that sense. We have therefore modified the text to better document the actual approach and consequences. We provide further information for the reader on how different networks are related and that there are a) duplicates that are removed (due to same key), but that there are b) duplicates that are kept due to different keys (Ft. Peck is an example for this). We also provide an additional table with information on source networks for each of the SAVS sites which also includes information on closest sites to resolve potential duplicates. We believe that this maximizes the traceability of source networks.

Changes can be found in section 3.1 of he revised manuscript.

Tower based albedo data (from a 10 meter tower) represents a field of view of 127 meters on the ground. Using the same Román et al., 2009 indices on higher resolution data (e.g. Landsat), the Ft Peck in situ data has been found to be spatially representative of the surrounding 2km and therefore one can safely perform a point to pixel comparisons with medium resolution satellite data products. On the other hand, the Bondville site is also a SURFRAD site (and thus a BSRN site). However it is only listed as an Aeronet site in the database (note that by not listing the other networks you may be implying that there is no in situ tower based albedo data available). It is situated within a large region of cropland. However, despite the homogeneity of the cropland land cover type surrounding the site, it is perhaps one of the least spatially representative sites for satellite evaluation in the BSRN system. This is because the tower is placed on a grassy area in the middle of the croplands and thus the types of crops surrounding the tower frequently change. Furthermore the greater region is sporadically intersected by roads and drainage ditches, thus greatly challenging the ability of the tower data to capture the spatial representativeness of the region. Thus, while this site is a long term site with high quality instrumentation within a large area of homogeneous cropland, it is not a particularly good site to use for satellite-derived intercomparions and evaluations. Now while acknowledging that the SAVS database has not undertaken to directly provide in situ data, some links to ALL the data networks associated with a site would seem warranted in each site description and certainly some further discussion (and references) on the caveats in using the associated in situ data should be provided to the user.

We totally agree with the reviewer that high quality measurements as such do not guarantee that a site is suitable for satellite data validation. The examples given by the reviewer are very illustrative for this. The reviewer has given the example of the Bondville site and has emphasized the need for providing information on the source networks. As discussed in the previous answer, the different source networks are already traceable and we now provide additional information on closest stations for each of the site. The Bondville site is a typical case where the problem of different coordinates occurs. The reviewer is correct that Bondville is part of SURFRAD and thus of BSRN. However, the coordinate between the SURFRAD and BSRN networks differ by approximately 2 km! The station keys are different as well. As discussed before, we believe that it is therefore a better approach to provide the information on the closest station and let the user decide if the sites with different coordinates should be considered as a single location or as two different locations.

Information on ALL associated data networks is now provided on the SAVS website, as outlined before. Further referees and critical discussion on ground data availability and quality have been added to section 3.1 of the manuscript.

Thirdly, the dataset relies on the 2010 ESA CCI land cover for homogeneity measures. This land cover is becoming dated and thus the land covers surrounding a site may well be changing. A GoogleEarth cutout is also provided for each site to aid in visual inspection of the site. A signal of historical snow fall potential and of fire potential is also provided –but there are of course many additional types of disturbance which may have occurred and things may have changed quite a lot since the 2010 CCI (and there should be some mention of the overall quality associated with the CCI). It is not clear how frequently the information associated with the sites in this database will be updated and some discussion on this point should be provided

We agree that the attributes derived from the ESA CCI LC dataset are limited to the time period currently provided by this dataset. The basic idea is to provide a general overview about snow/fire occurrences for each of the sites and therefore allow the user to filter easily e.g. all sites without any snow or fire. We also agree that the conditions will/might change over time.
We have therefore added further details on the limitations of the ESA CCI LC conditions in section 3.3 and also a short critical discussion on the need for data updates in the conclusion section.

Technical corrections

Page 1, line 25: Why are only geostationary and AVHRR albedo datasets mentioned and not the 16 years of MODIS and MISR products????

The reason why we did not add a reference to the MODIS record is that we speak of multi-decadal (>1 decades) dataset. As MODIS has not reached the 20-year threshold we did not mention it. However, given the fact that the

MODIS record is really unique we agree that it would make sense to add a reference to MODIS as well. We added the reference to Schaaf et al. (2011) to the manuscript.

Page 2, Line 10: In situ data representativeness is briefly mentioned here but perhaps a better discussion of site data representativeness - as opposed to only site land cover homogeneity - is warranted. This is also where the BSRN network, with its extremely high quality of in situ measurements, should be acknowledged (as well references to both GTOS (GTOS document ECV-T8) and GCOS documentation (GCOS document 154)).

Characterizing the representativeness of a particular site for the validation of satellite albedo data products is difficult to generalize, as it not only depends on the site characteristics, but also on the properties of the observing system. For SAVS1.0 we had therefore decided to leave the decision whether a particular site is usefull to the users. The database therefore provides only information on the heterogeneity of different parameters, which were derived in a traceable manner.

We agree that BSRN provides highly accurate surface radiation measurements. However, BSRN was not designed to allow for comparison with medium resolution EO data. BSRN sites are typically on grassland patches, which are not necessarily representative for the surrounding area covered by a satellite pixel. Within SAVS we have therefore explicitly decided to NOT provide any specific recommendation on the usefulness of a site for a particular application, like outlined before.

The issue of site representativeness is even explicitly mentioned on the BSRN website under the „Stations“ listing. There it says:

„Notice to data users concerning site representativeness
Please be advised that although BSRN initially was tasked to develop a network of sites that were spatially representative of the surrounding region, the reality for some of the sites deviates from that original intent for a variety of operationally practical reasons. Depending on your data application, it may be important for you to examine the specific geographic, topographic, and surface type distribution associated with the sites for which you will utilize the data. This is particularly true for coastal, mountain, and some island sites where local effects can be significant.”

We have further emphasized the overall motivation for SAVS in the revised version of the manuscript and also properly acknowledge the GTOS work which was missing in the original manuscript (thanks for this!). With this, we hope to clarify for the reader what should be expected from the database as such.

Page 4, Line 17-23 Awkward phrasing …perhaps: “The evaluation of surface albedo data products typically requires that diurnal variations in surface reflectances be taken into account …” and “Stringent requirements on the characteristics of a reference site to determine whether it is suitable for coarse scale surface albedo are therefore required:”

The sentence has been changed according to the reviewers suggestion. Thanks!

Page 6, line 25 – Duplicates should have all of their associated networks identified.

We agree that it might be useful for the reader to have traceable information on the individual source networks. As we cannot change the database content as such as this would require to issue a new DOI for the database related to this publication, we will provide a matrix, mapping each site to the corresponding source networks as ancillary information on the SAVS website. The text has been changed accordingly.

Page 7, Line 14 - A few words are required here – indicating how important it is to verify that you are not looking at snow (or a recent fire scar) before using a particular site at a particular time

Some further explanation has been added as suggested.

Page 10, sections 4.5 and 4.6 - caution readers that all evaluations are based on the aging and static CCI

We agree that the state is static. In principle we don’t expect that this will change the results too much, unless the land cover type changes drastically. The reason is, that for the used land cover conditions we look on climatological means anyway and do not cover any inter-annual or intra-annual variability.
We followed the reviewer’s suggestion and added a note of caution for the reader in section 3.3.

Page 11, Line 30 – Not all required data information is currently provided – again the duplicate sites should have ALL of their associated networks identified so readers can locate the appropriate in situ data.

As mentioned above, the additional information on source networks will be provided on the SAVS website.

There are some minor awkward English phrasings throughout that should be addressed by a copy editor

The revised manuscript has been proof-read by two native speakers from the scientific community.