REVIEWER #1

Comments

L166: It is not clear whether the authors themselves performed the ERAI-LG simulation, or if this is an available reanalysis product from ECMWF? If the latter, please provide a reference. If the former, please provide much more information about how the experiments were configured and performed. This is pivotal to be able to assess whether the differences between ERAI-L and ERAI-LG are, in fact, explained by the land cover, and not some other confounding variable(s).

A: Thank you very much for this valuable comment. Indeed, a more detailed explanation about this experiment was missing. We now include a new paragraph, starting in line 166.

L197: by "attributed to ocean areas" I assume that the authors mean that they were coastal sites, and that the predominant land cover type in the corresponding ERA grid cell was ocean. Is that correct? Perhaps a clarification is required here.

A: Thanks for the comment. Indeed that was what was meant. We changed the text accordingly. See line 210 in the new document.

L210: Given that observational in-situ data are available 1964-2015, and reanalyses are reliable at least over the satellite era, some justification is required here for why the period 2000-2013 was selected for the study (especially since 2009 data are missing almost everywhere, so n~13).

A: Thanks for the comment. Data availability of this in-situ radiation dataset is limited to 2000-2013, at least that is the timerange that was supplied by the institute. We added this information in the text. See line 225 in the new document.

L232: Is there any sensitivity to the grid cell extraction method? For example, another approach would be to use a "nearest-neighbor" remapping; would this change any answers?

A: We checked for nearest-neighbor remapping and in fact the results did not change in any meaningful form.

L253: The use of a local T2m is non-standard, and does not correspond to the feedback quantification model by Cess and Potter 1988. Perhaps the authors could offer some explanation here, and a description of what impact this change has on the results, and their interpretation?

A: Thank you for your comment on using 2m temperature. Using 2m temperature has multiple reasons. The original definition is using "surface air temperature" which is also remarked by Cess and Potter 1988: "Here, and in the remainder of this paper, we employ surface temperature, rather than surface air temperature, as an indicator of surface climate. The reason for this is that in their study of climate feedback processes Cess et al. [1985] found that these processes are more appropriately defined in terms of surface temperature. A further benefit, with respect to GCM inter-comparisons, is that GCMs explicitly calculate surface temperature, whereas they differ in their definitions of surface-air temperature." However, Cess and Potter 1988 already point out a crucial point, which is comparability. In a perfect world, near surface temperature would be just the very thin first layer above the surface, however comparing a broader network of stations with reanalyses, this variable is not available. Therefore, we used the closest possible to "near surface" temperature available for us. That said, using T2m is rather standard by now, as studies by Fletcher et al. 2015, Xiao et a. 2017 and Kevin et al. 2017 show. Nevertheless, we added an interpretation of the results concerning using 2m temperature. See line 280 in the new document.

Xiao, L., Che, T., Chen, L., Xie, H., & Dai, L. (2017). Quantifying Snow Albedo Radiative Forcing and Its Feedback during 2003–2016. *Remote Sensing*, *9*(9), 883.

Kevin, J. P. W., Kotlarski, S., Scherrer, S. C., & Schär, C. (2017). The Alpine snow-albedo feedback in regional climate models. *Climate Dynamics*, *48*(3-4), 1109-1124.

L259: Surely a major limitation of estimating alpha_land using MAMJ when Sc=0% is that there are many locations for which Sc is always > 0 in MAMJ. What do the authors use for alpha_land in those cases? And how much "more realistic" do the authors find that using MAMJ is, compared to August? My suspicion is that the values should be very similar.

A: Thanks for the comment. Maybe the text is a little bit unclear at this point, but since we have daily data, we look for days where snow is zero. And in June there is always at least one day without snow. If there are multiple days without snow, we use an average value of snow free albedo. That said, we tested using August values, and the results are very similar. Nevertheless, taking values out of MAMJ seems less artificial, which we mean by saying "more realistic". We adjusted the wording. See line 289 in the new document.

L289: Perhaps the authors have some additional evidence (spatial maps, for instance) to support the claim that the higher correlation for MERRA2 is due to aerosol deposition? If so, then I think it needs to be shown, because on its own Figs. 2c-d do not really allow us to draw any meaningful conclusions about physical processes. Also, on L396 the authors state that it is the vegetation schemes in MERRA2 and ERAI-L that decrease the snow albedo; is this contradictory to the point about aerosol deposition?

A: Thank you for this interesting comment. Indeed, we probably put too much emphasis on this point and toned down our wording. Vegetation schemes are responsible for a longterm albedo reduction but not so much on a day to day scale. We know include a small paragraph where we investigated the day-to-day variability in combination with aerosols. See line 322 in the new document.

L306: The issue of grid vs point comparisons is a very common problem. I wonder if anyone has attempted to use spatial interpolation (e.g. kriging) on the 40+ station observations to produce a "gridded" snow depth product?

A: We are working on a gridded product (with 400 stations as input) and it will hopefully be ready for research in 2018. Stay tuned for follow up paper by the first two authors.

L308: I am not sure where the evidence is presented to support the claim about snow- free albedo?

A: We deleted this statement.

L399: I am confused by Figs.4-5. In Fig.4b it is shown that the mean SNC term (alpha_snow - alpha_snowfree) is similar for the stations and ERAI-LG, and in Fig.4f the mean alpha_snow values are also similar. Yet, in Fig.5a, the alpha_snowfree values are hugely different (for which I could find no explanation), so how can Fig.5a be cor- rect, and yet still produce similar SNC in Fig.4b?

A: Thank you for your comment. SNC is not only albedo contrast (see equation 1), however SNC is a product defined by albedo contrast (Figure 4e) and snow melt sensitivity (4d). Albedo contrast is "so" similar because the differences in the y-axis of snow free albedo are actually relatively huge, but absolutely rather small in the grand scheme of things. That said, albedo contrast between stations and ERAI-LG is still roughly 0.05

L402: If the observed snow-free albedo is similar to that for grass, why does the ERAI- LG simulation still do so badly in this quantity (Fig.5a)?

A: Thank you for this comment. An explanation can be found in L574. Changing the vegetation scheme only helps to make the radiation characteristics over snow covered grid points more realistic. Snow free albedo is still as seen from satellite and is not dynamic in reanalyses. We made sure to underline that point.

L424: The sentence ending "overestimated complete snow cover albedo cancel each other out." seems to be highly important; however, it was not clear which panels of Fig.4/5 are supposed to show this cancellation? Also, what is "complete snow cover albedo"?

A: Indeed, additional information was missing, the sentence was not complete. We changed the wording and added information. See line 460 in the new document.

Supplement Figs.5-6: I recommend centering the colorbar labels in the bins, so that it is clear which color corresponds to which vegetation type.

A: We improved the Supplement Figures 5&6

REVIEWER #2

General comments:

1) The derivation of total SAF differs slightly from prior studies, and although it won,Äôt drastically impact results, a comment on the reasoning behind this should be added. Motivating studies (Fletcher et al., 2012; Fletcher et al., 2015) calculated NET SAF as independent of SNC and TEM (whereas here NET = SNC+TEM). Instead these components are calculated to show that they can explain most of NET. Also, note that Fletcher et al. (2015) found that the additivity of SNC and TEM was not well satisfied on regional scales, perhaps due to observational uncertainty.

A: Thank you very much for your comment. Indeed the difference to previous computations was not highlighted, so we added a comment and a motivation about that in Chapter 3. See line 238 in the new document.

2) The article is difficult to follow at times because of readability issues and typos. Several examples are listed below.

Specific comments:

Abstract: A comment should be added regarding the difficulty of comparing point and gridded data. Similar to what is on L579-582.

A: We added a similar statement to the abstract. See line 31 in the new document.

L58: remove “the” before Arctic warming.

A: Removed

L60: remove “of the global warming signal”.

A: Removed

L63: Pithan and Mauritsen 2014 (Nature Geoscience) would be a good citation to add
here.

A: We added this citation to the references. See line 63 in the new document.

L66-68: awkward wording, please address.

A: Thanks for the comment. We simplified the wording. See line 68 in the new document.

L69-70: change to “. . .an initial warming is strengthened over time. . .”.

A: Changed. See line 70 in the new document.

L72: change to “Snow can cause such a feedback because in its absence the surface absorbs more . . .” or similar.

A: Thanks for the comment. We simplified the wording.See line 72 in the new document.

L74: remove “This”.

A: Removed

L89: add “between models” after SAF variability.

A: Added. See line 89 in the new document.

L97: change “an” to “a”.

A: Changed. See line 96 in the new document.

L93-104: clarify that these studies are referring to the average SAF across the NH extratropics, not the entire NH.

A: Clarified. See line 88 and following in the new document.

L107: define CMIP at first use.

A: Defined. See line 106 in the new document.

L109: “From a large set of SAF estimates for individual models” - reword this.

A: Reworded. See line 108 in the new document.

L111-117: Fletcher et al. (2015) only used the different snow cover and temperature datasets from reanalyses, not their albedos. This is an important difference from the current study.

A: Highlighted this difference. See line 117 in the new document.

L119: Satellite products of what? Snow cover, temperature, albedo, etc. Please clarify.

A: Clarified. See line 120 in the new document.

L163: change “local” to “site measurements” or similar.

A: Reworded.See line 163 in the new document.

L178-184: awkward wording – repetitive use of "diagnose".

A: Reworded.vSee line 193 in the new document.

L191: I don’t think Solar Radiation and Radiation Balance Data should be capitalized here.

A: Corrected. See line 205 in the new document.

L194: Fix “containes”.

A: Corrected. See line 208 in the new document.

L194: Remove “Of these”.

A: Removed.

L197: change “to ocean areas, so” to “as ocean areas, meaning”.

A: Reworded. See line 211 in the new document.

L201: change to snow cover fraction.

A: Reworded. See line 216 in the new document.

L210: Why limit the study period to 2000-2013? I assume this may be related to the availability of satellite (i.e. MODIS) data used in previous studies, but this should be explicitly stated.

A: This circumstance is now explained in line 225 in the new document.

L218: Change to “for the MAM period and for 3 stations also June values are missing” to “during MAM and at 3 stations in June.”

A: Thanks for the comment. We addressed this issue. See line 233 in the new document.

L230-231: Some comment on the resolution of the reanalyses is needed, and the difficulties associated with a point to gridbox comparison.

A: We added comments on the grid box comparison. See line 248 in the new document.

L235: Change “for the long-term climate change signal are highly correlated” to “under long-term climate change are highly correlated”.

A: Reworded. See line 255 in the new document.

L240: fix “decreaseof” and “theearlier”.

A: corrected. See line 259 in the new document.

L241: change “exposition” to “exposure”.

A: Deleted.

L256: See general comment #1, and address this.

A: Thanks for the comment. We addressed this issue. See line 259 in the new document.

L264: Can you provide a brief comment on what those previous studies found?

A: Added a brief comment. See line 296 in the new document.

L265: “We” shouldn’t be capitalized.

A: Corrected. See line 294 in the new document.

L267: remove “involved in the SAF computations”.

A: Removed.

L278: change to “better represents”.

A: Corrected. See line 311 in the new document.

L289: Is there any evidence linking this directly to aerosols? Why isn’t there a larger disparity between MERRA2 and ERAI-land in Fig 2a?

A: Thank you for this thoughtful comment. We investigated this relationship now a little bit more in detail. You can find the new information in line 322.

L294-295: repetitive, remove “Considering the representation of day-to-day variability”. Figure 3 caption: should say “station data”. Also, I’m not sure what the difference is between TEM and snow melt sensitivity here. On L252 it is stated that TEM will be referred to as snow melt sensitivity. Is this the snow cover sensitivity (snow cover change per degree warming)?

A: We clarified the context and added additional information to the explanation of what we mean with snow melt sensitivity. See line 274 in the new document.

L335-343: The similar nature of these results implies that the vegetation types at most of the sites must be similar, can you comment on this?

A: Yes, we expect them to be WMO standard, that means observations are done over cut grass everywhere. See line 164 in the new document.

L353: Change to “put the station data in context”.

A: Changed. See line 389 in the new document.

L357: ,“Changing the vegetation to short grass adds about 1K to the responses” –

the correct interpretation is that it adds an additional 1% albedo decrease per degree of warming.

A: Reworded. See line 394 in the new document.

Fig 5 A: Why doesn’t the ERAI-LG case have a snow-free albedo that resembles the stations if 0.2 is the albedo of grass?

A: Thanks for the comment. We now explain this feature in line 573.

Fig 5e: This looks the same as Fig 4f, is it? Why is one called "snow albedo" and the other "mean albedo"?

A: Thanks for your comment. Mean albedo is averaged over both, snow and snow free albedo.

L424: “For ERAI-LG, the effect of the underestimated snow-free albedo and overestimated complete snow cover albedo cancel each other out” I don’t understand what this is referring to, please clarify. Wouldn’t an overestimated snow albedo and underestimated snow-free albedo create a larger albedo contrast, and thus stronger SAF?

A: Thanks for the comment. It was very unclear before, we rephrased and clarified this point. See line 460 in the new document.

L426: remove "season"

A: Removed.

L450: change "for both" to "when it comes to"

A: Changed. See line 485 in the new document.

L504: remove "properties"

A: Removed.

L506: remove "the"

A: Removed.

L515-518: I find it unlikely that day-to-day variability in albedo is strongly influenced by changing vegetation, as these processes occur on much longer timescales. Are you referring to the different vegetation states between the tower location (i.e., in a clearing) and the larger grid cell (mixture of vegetation types)? If so, please clarify, as this would impact the maximum surface albedo and thus the variability. Also, I don,Äôt see "flooding" as a major factor for spring albedo, clarify or remove this.

A: We removed the mentioning of vegetation and flooding. See line 554 in the new document.

L536: Fix "databecause"

A: Corrected. See line 572 in the new document.

L550: Should say "CMIP3/CMIP5".

A: Corrected. See line 582 in the new document.

L579-582: I think this is a very important statement that should be emphasized in the abstract.

A: Thanks for the comment. It is now highlighted in the abstract.

All Figures: Increase the font size for axis labels.

A: We increased the font size for axis labels.

Figure 1: I recommend adding some latitude/longitude labels.

A: We added latitude longitude labels.

Figure 2: Make the axis range for correlation plots the same (c,d) to allow for easier comparison.

A: We adjusted the axis range.

Figure 7: Caption says "Figure 4", correct this.

A: Corrected

Table 1: Capitalize "lon" in the table heading.

A: Corrected

Figure 5-6: The text on these figures is very grainy, please fix.

A: We don't know how that happened, but we tried to fix it now for the new version.

LIST OF RELEVANT CHANGES

Since the paper was accepted with minor changes, here is a short list of significant changes we made:

1. Clarified the methodology concerning calculating the SAF
2. Added a more detailed description of the experimental reanalysis set up
3. Added a paragraph in investigating the impact of aerosols on albedo reduction
4. Improved readability of Figures

Spring snow albedo feedback over Northern Eurasia: Comparing in-situ measurements with reanalysis products

Martin **Wegmann**[1], Emanuel **Dutra**[2], Hans-Werner **Jacobi**[1] and Olga **Zolina**[1,3]

[1]Institute for Geosciences and Environmental Research (IGE), Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, Grenoble, France*

** Institute of Engineering Univ. Grenoble Alpes*

[2]Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

[3]Shirshov Institute of Oceanology, Moscow, Russia

ABSTRACT

This study uses daily observations and modern reanalyses in order to evaluate reanalysis products over Northern Eurasia regarding the spring snow albedo feedback (SAF) during the period from 2000 to 2013. We used the state of the art reanalyses ERA-Interim land and the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA2) as well as an experimental setup of ERA-Interim land with prescribed short grass as land cover to enhance the comparibility with the station data, while underlining the caveats of comparing in-situ observations with gridded data. Snow depth statistics derived from daily station data are well reproduced in all three reanalyses, however day-to-day albedo variability is notably higher in stations compared to any reanalysis product. The ERA-Interim grass setup shows an improved performance in representing albedo variability and generates comparable estimates for the snow albedo in spring. We find that modern reanalyses show a physically consistent representation of SAF, with realistic spatial patterns and area-averaged sensitivity estimates. However, station-based SAF values are significantly higher than in the reanalyses, which is mostly driven by the stronger contrast beween snow and snow-free albedo. Switching to grass-only vegetation in ERA-Interim land increases the SAF values up to the level of station-based estimates. We found no significant trend in the examined 14-year timeseries of SAF, but inter-annual changes of about 0.5% K^{-1} in both station-based and reanalysis estimates were derived. This inter-annual variability is primarily dominated by the variability in the snow melt sensitivity, which is correctly captured in reanalysis products. Although modern reanalyses perform well for snow variables, efforts should be made to improve the representation of dynamic albedo changes.

1. Introduction

Global warming is enhanced at high northern latitudes, where the Arctic near-surface air temperature has risen at twice the rate of the global average in recent decades – a feature called Arctic amplification **(Serreze and Barry 2011).** Climate model experiments for the 21st and 22nd centuries show that Arctic warming will continue and intensify under all emission scenarios (**Collins et al. 2013**). Arctic amplification results from several processes interacting with each other such as the albedo feedback due to a reduction in snow and ice cover, enhanced poleward atmospheric and oceanic heat transport, and changes in humidity **(Serreze and Barry 2011, Pithan and Mauritsen 2014).**

Being one of the critical factors of the Arctic amplification, the surface albedo feedback implies a decrease of reflected shortwave radiation at the top of the atmosphere in conjunction with decreasing surface albedo and increasing near-surface temperature (**Thackeray and Fletcher 2016**). It is considered to be a positive feedback in the sense that an initial warming is strenghtened over time, quantified through the change in surface albedo per unit change of temperature **(Robock 1983, Cess et al. 1991, Qu and Hall 2007**). Snow melt triggers this feedback via surface absorption of shortwave radiation followed by conversion to longwave radiation, warming the lower layers of the troposphere (**Curry et al. 1996**). Snow albedo feedback (SAF) and its impact on climate have been studied for several decades (**Wexler et al. 1953, Budyko 1969, Schneider and Dickinson 1974, Lian and Cess 1977).** It got further attention in the wake of anthropogenic global warming accompanied by the reduction of snow and ice cover over the Northern Hemisphere (NH) **(Bony et al. 2006, Qu and Hall 2007, Fernandes et al. 2009, Flanner et al. 2011, Qu & Hall 2014, Fletcher et al. 2015, Thackeray and Fletcher 2016).**

During 1979–2011, the Arctic snow cover extent in June decreased at a rate of -21% per decade (**Derksen and Brown 2012**). Climate model projections for the end of the 21st century show an even more reduced Arctic cryosphere and, thus, the SAF will continue to modulate Arctic warming **(Brutel-Vuilmet et al. 2013).** The SAF is especially effective over the NH since most of it is covered by snow during boreal wintertime **(Groisman et al. 1994). Hall (2004)** found that 50% of the total NH extratropics SAF caused by global warming occurs during spring, while **Qu and Hall (2014)** estimated that the SAF variability between models accounts for 40-50% of the spread in the warming signal over the continents of the NH extratropics.

Several studies investigated spring NH extratropic SAF based on satellite, reanalysis and model datasets **(Fernandes et al. 2009, Fletcher et al. 2012, Qu and Hall 2014, Fletcher et al. 2015).** Satellite-based estimates of SAF vary within ±10% depending on the analysed data set. **Hall et al. (2008)** used the International Satellite Cloud Climatology Project (ISCCP) data (**Schiffer and Rossow 1983**) to calculate a SAF strength of -1.13% K^{-1}, whereas **Fernandes et al. (2009)** using Advanced Very High Resolution Radiometer (AVHRR) data (**Justice et al. 1985**) found a slightly weaker SAF of -0.93% K^{-1}. **Qu and Hall (2014)** determined the SAF using Moderate Resolution Imaging Spectroradiometer (MODIS) data (**Hall et al. 2002**) and found a value of -0.87% K^{-1} for springtime. Considering different spatial and temporal domains as well as the variety of methods applied, the SAF estimates around -1% $K^{-1}$from satellite data can be considered as quantitatively consistent.

Model- and reanalysis-based estimates are somewhat higher compared to those derived from satellite data. **Fletcher et al. (2015)** investigated Coupled Model Intercomparison Project 3 and 5 (CMIP3/CMIP5) ensembles to estimate the SAF for an assortment of Global Climate Models (GCMs). The authors found a SAF ensemble model mean of -1.2% K^{-1} for the NH extratropics, which is in fair agreement with MODIS values, but is higher compared to ISCCP- and AVHHR-based estimates. Within this comparison **Fletcher et al. (2015)** also investigated SAF computations based on ERA-Interim **(Dee et al. 2011),** Modern-Era Retrospective Analysis for Research and Applications (MERRA) **(Rienecker et al. 2011)** and NCEP-2 **(Kanamitsu et al. 2002)** reanalyses, thus, providing the most up to date assessment of SAF in reanalysis datasets. While MERRA data resulted in a slightly weaker SAF of -1.17% K^{-1} compared to ERA-Interim (-1.23% K^{-1}), both reanalyses show similar SAF values compared to MODIS.

That said, most studies use satellite derived albedo data in conjunction with temperature and snow cover data from reanalyses.

Although satellite products of snow cover and albedo cover large parts of the NH, they exhibit low temporal resolution and significant uncertainties for high solar zenith angles as well as complex terrains (eg. **Wang et al. 2014**). **Thackeray and Fletcher (2016)** compared CMIP3/CMIP5 model families and found that the models represent the SAF process rather accurately. However, there are still inherent biases likely related to the use of outdated parameterizations. In this respect the use of in-situ observations would provide an opportunity for evaluating SAF estimates in different gridded datasets and especially among reanalyses. However, estimating SAF in the Arctic using in-situ data is challenging, mostly because of the lack of reliable, relevant observations, both in the temporal and spatial domain. Furthermore, the lack of in-situ SAF estimates hampers the understanding of SAF in high latitude climates (**Graversen and Wang 2009, Gravesen et al. 2014**).

In this study we use a unique dataset of daily observations and modern reanalyses over Northern Eurasia in order (1) to evaluate reanalysis products with respect to radiation and snow properties and (2) to determine the SAF in spring between 2000–2013 based on in-situ measurements. We compare different land-reanalysis products with modified vegetation settings. Specific questions to be addressed in this study are the following: How well do the modern reanalyses reproduce snow and radiation features on a daily resolution? What are realistic estimates of the SAF from the station data over Northern Eurasia and how well do they compare to the gridded reanalyses data? What are the major characteristics of space-time variability of the SAF in station and reanalysis data?

The paper is organized as follows. After describing the different datasets and the methods in sections 2 & 3, we evaluate the daily output for snow, radiation fluxes and temperature within these datasets in section 4.1. In section 4.2 we assess the results of the SAF computations and the differences between products including also an analysis of the spatial and temporal variability. Section 5 discusses the results and considers potential implications for future studies.

2. Data

2.1 Reanalysis Data

To investigate the SAF processes in reanalyses, we evaluated two products: the ERA-Interim-land (ERAI-L, **Balsamo et al. 2015**) and Modern-Era Retrospective analysis for Research and Applications, Version 2((MERRA2) (**Gelaro et al. 2017**). ERAI-L is a land-surface only simulation driven by the near-surface meteorology and fluxes from the ERA-Interim atmospheric reanalyses (**Dee et al. 2011**). The land-surface model in ERAI-L (HTESSEL) has several enhancements compared with the land-surface model used in ERA-Interim including the snowpack representation (**Dutra et al. 2010**). ERAI-L considers the prognostic evolution of snow mass and density, and for exposed areas there is also a prognostic evolution of snow albedo. For shaded snow, i.e. snow under high vegetation, the albedo is considered constant and dependent on vegetation type (see **Dutra et al. 2010** for more details). Since the in-situ measurements in this study are observed over clear cut vegetation, idealized simulations prescribing grassland everywhere were carried out with the ERAI-L configuration (hereafter ERA-Interim land grass only (ERAI-LG)). The ERAI-LG simulation was carried out with the same model and setup as ERAI-L, differing only in the land cover used. The land-surface model used in ERAI-L, HTESSEL, accounts for sub-grid scale land cover variability by representing several land tiles, namely: low vegetation, high vegetation, bare ground, exposed snow (snow on top of bare ground or low vegetation), shaded snow (snow under high vegetation) and interception. The land cover is prescribed with four maps: low and high vegetation cover (cvl and cvh) and low and high vegetation types (tvl and tvh). The bare ground fraction is computed as cvb= 1 - cvl - cvh, the snow fraction is a function of the mean grid-box snow depth and the interception fraction as a function of the mean interception reservoir water content. For the ERAI-LG simulation, the high vegetation cover was set to zero (cvh=0), the low vegetation cover to one (cvl=1) and the low vegetation type to grassland. In this idealized simulation the entire globe was covered in grass land so that only the low vegetation and exposed snow (when snow is present) tiles were active. The main goal of this simulation is to evaluate the role of land cover when comparing point observations with gridded reanalysis and to evaluate pathways to improve reananalyses in representing albedo processes.

MERRA2 also includes a dedicated land module for surface variables. Furthermore, it applies an updated Goddard Earth Observing System (GEOS) model and analysis scheme and assimilates more observations than its predecessor MERRA (**Rienecker et al. 2011**). Finally, MERRA2 uses observation-based precipitation data to force its land-surface parameterizations, similar to what formerly was known as MERRA-land. Unlike ERAI-L, MERRA2 consists of a full land-atmosphere reanalysis. Its incremental analysis update (IAU) scheme improves upon 3D-Var by dampening the analysis increment. In IAU, a correction is applied to the forecast model gradually, limiting precipitation spinup in particular.

For near-surface temperature we use 2m air temperature for both the reanalyses and observations. Moreover, we do not use albedo computed by the reanalysis, but calculate it from the radiative flux components consistent with the observed albedo. For this purpose, we use upward and downward shortwave radiation at the surface as diagnosed by ERA-Interim and MERRA2 as well as surface net and surface incoming radiation from the station observations. Snow depth is used as inferred by reanalyses and, if needed, converted to cm. More information about general characteristics of reanalysis products in the Arctic can be found in **Lindsay et al. (2014), Dufour et al. (2016)** and **Wegmann et al. (2017).**

2.2 Observational in-situ data

To evaluate reanalysis perfomance, we used newly assembled in-situ radiation observations from Russian meterological stations. This dataset includes 4-hourly solar radiation and radiation balance data from the World Meteorological Organisation (WMO) World Radiation Network of the World Radiation Data Center (WRDC) at the Voeikov Main Geophysical Observatory, Saint Petersburg, Russia. The original WRDC data contains time series from 65 locations. We selected 47 stations for this study because they overlap with daily snow depth and 2m temperature observations (see Supplement Table 1). Of these 47 stations three were attributed by ERAI-L to ocean gridpoints and we decided to remove the three coastal stations from the initial dataset, so that the final dataset consists of 44 stations. Temperature and snow depth observations were taken from the All-Russian Research Institute of

Hydrometeorological Information World Data Centre (RIHMI-WDC), Obninsk, Russia. A detailed description of this dataset is provided by **Bulygina et al. (2010)**. This dataset includes snow depth as well as snow cover fraction around meteorological stations. Snow cover information in this data set is not stored in percentages, but rather in a scale of integers from 0 to 10 (for example, 50% is assigned a value of 5, but so is 53%). This makes these data hardly applicable for precise SAF calculations. Snow depth information is measured in centimeters with the precision of 1 cm. This might lead to an underestimation of snow depth in case of shallow snow (between 0 and 1 cm). All variables (temperature, snow depth and snow cover, surface LW radiation budget and surface SW radiation, the sum of the surface short-wave and long-wave radiation budgets) were represented as daily time series for the period 2000–2013, which is the time period available for the radiation observations by the Voeikov Main Geophysical Observatory.

Figure 1 shows the location of the stations together with the climatological 2000–2013 MAMJ snow depth as computed by ERAI-L. The distribution of stations is quite heterogeneous, with very few stations located in Eastern Siberia and in the Far East. Moreover, some stations have prolonged periods of missing values; six stations have more than 50% missing values in the daily timeseries for MAMJ. For monthly means, the total number of missing values generally decreases from 2000 to 2013 (see Supplementary Figure 1). However, data for the year 2009 are missing at 44 out of 47 stations during MAM period and at 3 stations in June. Nevertheless, spatial and temporal coverage of this data set is exceptional for the analysis of albedo in this region. It is also important to note that neither snow nor radiation from these stations were assimilated in the reanalysis datasets and, therefore, our inter-comparisons are completely independent.

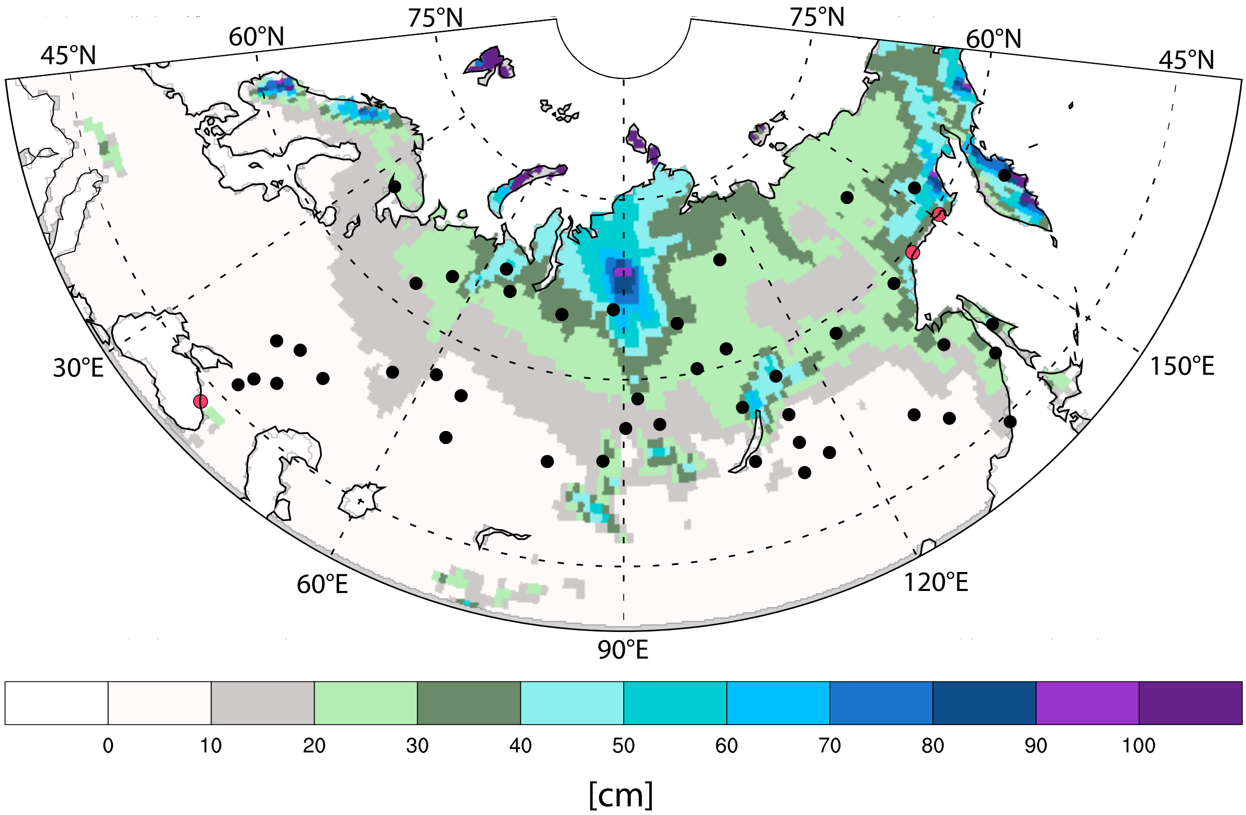


Figure 1: Station location and snowdepth [cm] for the 2000–2013 MAMJ average taken from ERAI-L. Red colored stations are excluded by the land-sea mask of ERAI-L.

3. Methods

To evaluate the climatic variables needed for the SAF computation, we first compared daily values of snow depth, albedo and 2m temperature from the meteorological stations with those from the reanalyses. To co-locate observations with reanalyses, we extracted the information of the gridcell from the reanalysis, in which the station is located. In case of ERA-Interim land, horizontal resolution is 0.75° x 0.75° degrees, whereas MERRA2 has a horizontal resolution of 0.5° x 0.625° degrees. That said, the extracted values of the gridcell are expected show less variability and lower peak values, since they are integrated over a larger spatial domain, which dampens extreme values. We then derived long-term differences, performed a correlation analysis and also compared the variability among the datasets for the MAMJ period.

Since the SAF signals for the seasonal cycle and under long-term climate change are highly correlated **(Hall and Qu 2006)**, we focus here on the evaluation of the seasonal cycle. Snow cover is converted from snow depth following a logarithmic equation according to which 2.5 cm of snow depth was defined as equivalent to 100% snow cover (**Fletcher et al. 2015**). We split SAF into a snow cover component (SNC) and a temperature/metamorphosis component (TEM). SNC relates to the decrease of the albedo linked to the earlier melting of snow. TEM concerns the reduction of snow albedo due to enhanced metamorphism and larger grain sizes at warmer temperatures. In this study we focus on these two components of the feedback process, rather than the general classic term for net SAF ($\Delta\boldsymbol{\alpha}/\Delta\boldsymbol{T}$), since our goal is to evaluate differences in the more intricate terms of SAF. In the following, we assume that SAF=SNC+TEM, which was shown to be true in nearly all cases for the NH (**Fletcher et al. 2012, Fletcher et al. 2015**). Therefore, we compute the two terms as

$$\boldsymbol{SNC} = (\overline{\boldsymbol{\alpha}_{snow}} - \boldsymbol{\alpha}_{land})\, \Delta\boldsymbol{S}_c/\Delta\boldsymbol{T}_{2m} \qquad \textbf{(1)}$$

and

$$\boldsymbol{TEM} = \overline{\boldsymbol{S}_c}\, \Delta\boldsymbol{\alpha}_{snow}/\Delta\boldsymbol{T}_{2m} \quad , \qquad \textbf{(2)}$$

where $\boldsymbol{\alpha}_{snow}$ is the snow-covered surface albedo, $\boldsymbol{\alpha}_{land}$ is the snow-free surface albedo, $\boldsymbol{S}_c$ is the snow cover fraction and $\boldsymbol{T}_{2m}$ is the 2 m temperature. The first term of SNC $(\overline{\boldsymbol{\alpha}_{snow}} - \boldsymbol{\alpha}_{land})$ is also known as albedo contrast, whereas the second term $(\Delta\boldsymbol{S}_c/\Delta\boldsymbol{T}_{2m})$ will be referred to as snow melt sensitivity. In (1) and (2) deltas indicate month-to-month changes and the overbars indicate means over the two adjacent months. Note that $\Delta\boldsymbol{T}_{2m}$ does not represent a hemispheric mean but rather the difference at an individual location. It was found that the contribution of SNC and TEM to the overall SAF is between 60 to 70% and 30 to 40 % for the NH (**Fletcher et al. 2015**).

In our SAF assessment, we use 2 m temperature as a surrogate for near surface air temperaure, since the latter variable is not represented by stations. Using 2m temperature introduces some uncertainty to the results since atmospheric temperature advection can play a role in local temperature evolution. However, by now multiple studies (**Fletcher et al. 2015, Xiao et al. 2017, Kevin et al. 2017**) deal with 2 m temperature in their SAF assessment, mainly also due to the same comparability issues.

Since daily data are available, we define $\boldsymbol{\alpha}_{snow}$ as the monthly mean over all daily estimates during the specific month when $\boldsymbol{S}_c = \mathbf{100}\%$. Moreover, we define $\boldsymbol{\alpha}_{land}$ as the mean over all daily estimates during MAMJ (in some stations this might only occur in June) when $S_c = 0\%$. This allows for a less artifical estimation of α_{land} than conventionally using summer (e.g. August) albedo.

4 Results

4.1 Daily data evaluation

Since 2m air temperature in reanalyses has been comprehensively evaluated in previous studies (eg. **Schubert et al. 2014, Lindsay et al. 2014**), we only perform a general comparative asssement of the daily values of albedo and snow depth in the SAF computations. That said, **Lindsay et. al 2014** found that 2m temperatures show slight negative biases over Russia in Winter for both ERA-Interim and MERRA1, whereas in summer ERA-Interim shows basically no bias and MERRA1 shows slight positive biases. Improvements in this regard from MERRA1 to MERRA2 are to be expected.

Figure 2 shows an overall comparison between station data and reanalyses in terms of correlations, differences and magnitude of variability quantified by the standard deviation for the albedo and snow depths. On a day-to-day basis MERRA2 and ERAI-L are underestimating average albedo values compared to observations by about 0.1 during MAMJ (Figure 2a). On the other hand, ERAI-LG shows a much smaller average deviation from the station data with differences close to zero. However, the overall range of the boxplot for ERAI-LG is similar to the other two reanalyses resulting in only slightly less absolute deviations from the observations.

For snow depth (Figure 2b), all three reanalysis datasets show an overestimation of daily values for MAMJ. Interestingly, ERAI-LG shows the largest deviations from observed values, although the grass better represents the conditions at the observational sites. This can be caused by biases in the observations due to surrounding higher vegetation creating a snowfall shadow or negative instrumental biases (**Rasmussen et al. 2012**). Moreover, positive biases in particular for precipitation can occur in reanalysis products (**Brun et al. 2013**).

The analysis of daily correlations (Figure 2 c and d) demonstrates that the correlations for the albedo are generally low among all three experiments, whereas for some stations they can reach correlation coefficients higher than 0.8. Surprisingly, the correlations between MERRA2 and station data are the highest for albedo and the lowest for snow depth. The observed difference between MERRA2 and the ECMWF experiments regarding the correlation for albedo can likely be explained by the introduction of aerosols (and their respective deposition) in MERRA2. Using the MERRA2 aerosol product, we find a few days per station that show a co-existence between days with constant day-to-day snow depth (no snowfall or melt event), albedo decrease and strong (>75% percentile event for a location timeseries) aerosol deposition, both in stations and MERRA2 (not shown). We realize however, that there are other drivers for a local albedo decrease, which we are not able to isolate. Therefore, aerosols can modulate the albedo variability during periods of constant snow depth and are a good addition in reanalysis datasets. How big the quantitativ impact in the reanalysis really is, remains an open question. Further studies are needed to investigate the impact of aerosols on snow albedo representation. For snow depth, the correlation values are dominated by snowfall and melting events. Also in this case, the grass-only experiment shows no increased performance compared to the classic ERAI setup.

All reanalyses severely underestimate the day-to-day variability of the albedo (Figure 2 e and f). MERRA2 and ERAI-L show similar means, but reach the overall station level only in specific grid cells. A clear improvement is observed in ERAI-LG, which shows the smallest deviation from station estimates. Nevertheless, all modern reanalyses fail to adequately reproduce daily varability in the observed albedo. On the other hand, for snow depth the agreement is very good. The means of all four products are around the values of 8 to 10 cm, with the grass-only experiment being the closest to the average station variability.

In summary, the boxplot analysis (Figures 2) reveals that there is a general improvement in agreement between stations and ERAI-L if vegetation is set to grass only. However, none of the reanalysis products can accurately reproduce day-to-day albedo variability. This is likely explained by the comparison of grid versus point observations, where small-scale variations are averaged out.

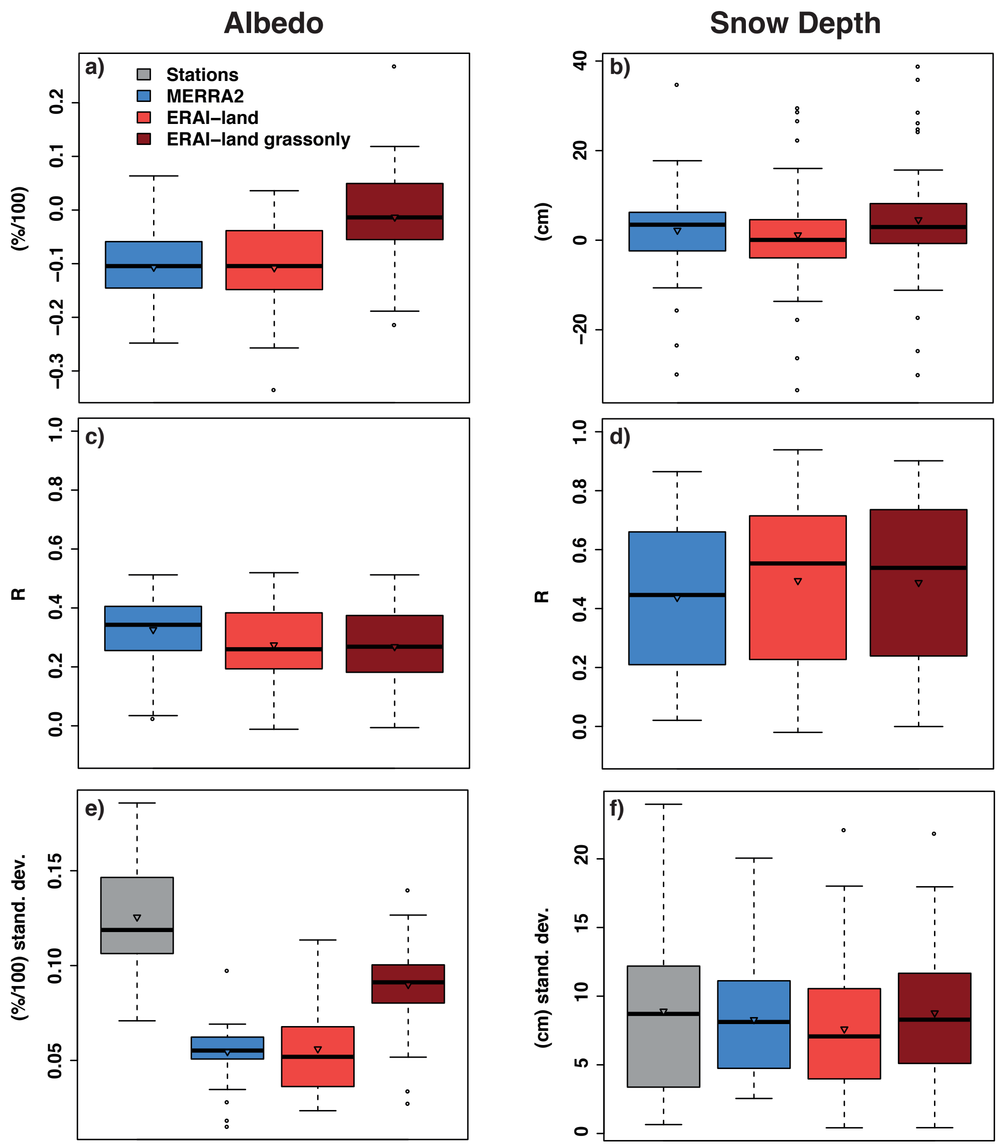


Figure 2: Boxplot analysis for daily albedo (a, c, e) and snow depth (b, d, f) estimates using data from 44 locations over 2000–2013 MAMJ period. (a) and (b) Difference between station and reanalysis, (c) and (d) linear correlation between station and reanalysis, (e) and (f) standard deviation. Triangle indicates the mean value.

4.2 Analysis of feedback components

To assess regional patterns of key SAF components, we show their spatial distribution over Russia as revealed by the observations in Figure 3 (See Supplement Figures 2-4 for the respective distribution from the reanalyses data).

Strong SNC (Figure 3a) responses in the station data are observed in Southern European Russia and Western Siberia as well as over the Far East. The weaker responses are observed in Southern Eastern Siberia. TEM (Figure 3b) follows a similar distribution but is more homogeneously distributed with most negative values in Central Siberia and towards the Arctic coastline. Snow melt sensitivity (Figure 3c) is strongest in the mid-latitudinal and subpolar regions north of 50° N, such as Finland to the southeast, west and north of Lake Baikal and along the Pacific Coast. Here the temperatures react most strongly to seasonal snow melt. While there is a broad agreement between the stations and ERAI-LG in this region, stations show a somewhat stronger snow melt sensitvity (not shown). Snow melt sensitvity is a key factor for the SNC calculations and, thus, shapes the spatial variability of SNC.

The other key factor in the SNC calculations is the contrast in albedo between snow-covered and snow-free periods (Figure 3d). The observed albedo contrast is characterized by a relatively homogeneous patttern with somewhat smaller values in the southern regions, especially over Southern Eastern Siberia east of the Lake Baikal. In general, a north-south gradient is visible with similar patterns as in SNC. Mean albedo for spring (Figure 3e) shows that highest values are found closer to the Arctic coastline, in Central Siberia and towards the western border. Lower mean albedo values are mostly located east of Lake Baikal. This distribution is in general agreement with the reanalyses datasets, especially for the lower values in the south east.

Finally, since TEM follows closely the general MAMJ snow distribution, we show average snow depth in Figure 3f. A clear north-south gradient is visible with hotspots at the Pacific coast and towards the Barents-Kara sea. Moreover, snow depths from stations follow closely the ERA-L snowdepth distribution shown in Figure 1.

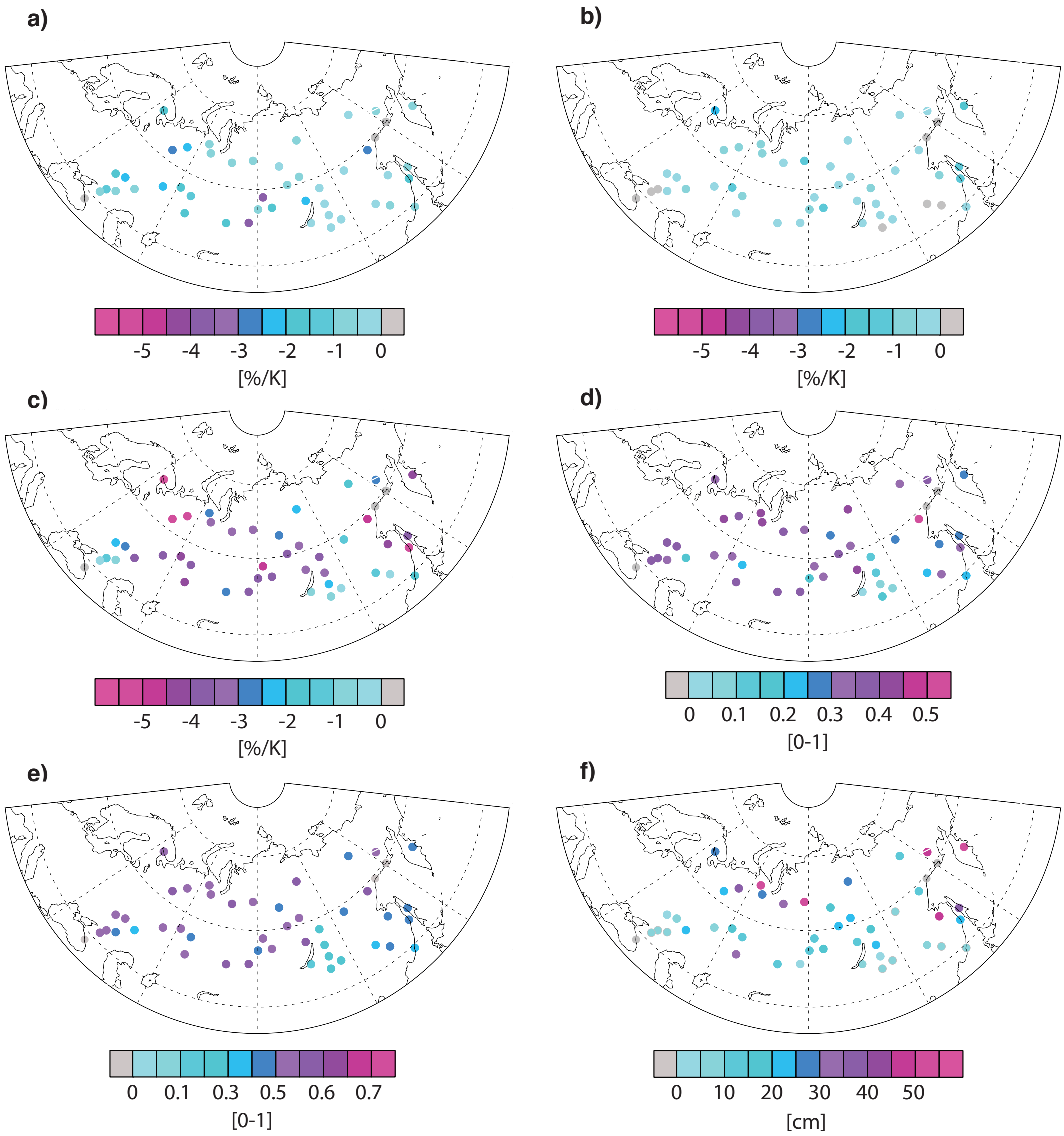


Figure 3: Mean SAF components in station data for 2000–2013 MAMJ. a) SNC, b) TEM, c) snow melt sensitivity, d) mean albedo contrast, e) mean albedo, f) snow depth.

To analyse the differences between the datasets and to put the station data in context, Figure 4a shows the response for SAF computed for the entire period 2000-2013 and all 44 locations. Stations show much stronger SAF (-2.5% K^{-1}) compared to MERRA (-1.6% K^{-1}) and ERAI-L (-1.8% K^{-1}). At the same time ERAI-LG shows SAF estimate close to that derived from the station data (-2.8% K^{-1}). Thus, changing the vegetation to short grass adds an additional 1% albedo decrease per degree of warming to the feedback process. The further analysis of the two components of SAF (SNC and TEM,

Figure 4 b and c) shows that ERAI-LG reproduces well the SNC signal derived from the station data (-1.6% K^{-1} mean for stations and -1.7% K^{-1} mean for ERAI-LG), whereas the other two reanalyses show much weaker SNC values. The lowest value of -0.56% K^{-1} was obtained from the MERRA2 data. In general, SNC responses largely explain differences in SAF (Figure 4a).

For TEM values (Figure 4c), all three reanalyses are in a good agreement with the observations with MERRA2 showing the best agreement. Changing the vegetation to grass in ERA-Interim results in a TEM component, which is 0.4-0.5% K^{-1} stronger compared to the standard version of ERA-Interim. Given that TEM represents the response to snow metamorphosis, good performance of MERRA2 is in agreement with findings implied by Figure 2. However it is worth noting that for the station network as well as for the ECMWF experiments, locations with positive TEM are calculated. This is due to snow albedo changes being positive in some instances (Figure 4c).

To further investigate the nature of the SNC and TEM responses we show in Figure 4d the results for snow melt sensitivity, which is one of the two key components in the SNC response (1). This component is barely influenced by the underlying vegetation. All three reanalysis datasets agree very well with the station network, with ERAI-LG showing the closest agreement for both mean and median. This indicates an accurate representation of this relationship in both NASA and ECMWF land surface modules.

Figure 4d implies that the changes in the SNC should stem from the albedo contrast, the second key component expressed as the average difference between albedo values for a complete snowcover and snow-free conditions (Figure 4e). Indeed, MERRA2 shows the lowest albedo contrast among all datasets, resulting in very low SNC values. Albedo contrast in ERAI-L is higher than MERRA2, but is on average still lower compared to the observations, which show average values around 0.35. ERAI-LG shows the strongest albedo contrast, which is twice as large compared to the experiment with classic vegetation cover. These striking differences among the datasets mainly drive the SNC results.

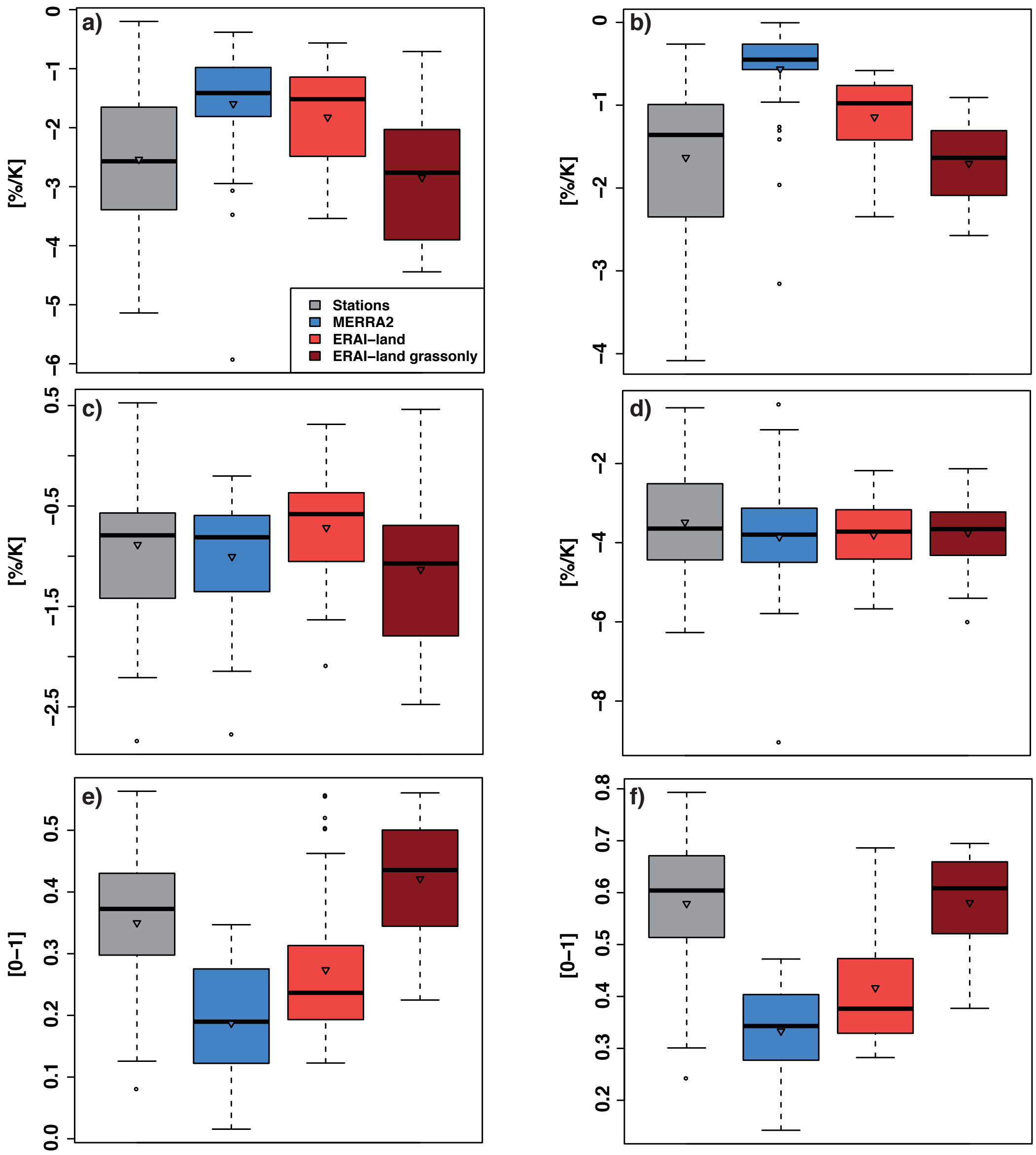


Figure 4: Boxplot analysis for MAMJ 2000–2013 a) SNC+TEM, b) SNC, c) TEM, d) snow melt sensitivity, e) albedo contrast and f) snow albedo. Triangle indicates the mean value.

Snow albedo is well captured by the grass-only experiment showing the same average value around 0.6 as determined from the observations (Figure 4f). The standard vegetation schemes used in MERRA2 and ERAI-L reduce the snow albedo in the analyzed grid cells to 0.33 and 0.37. The differences in snow albedo between the products is the main driver for the differences in the albedo contrast since the snow-free albedo values are remarkably similar for all reanalysis products (Figure 5a). Nevertheless, they strongly deviate from the snow-free albedo determined from the observations, which is roughly twice as large compared to the reanalyses with a mean value of about 0.21 and which is very close to albedo values for grass (see e.g. **Betts and Ball 1997**, **Wei et al. 2001**).

To explore the impact of different factors on the TEM estimates, we show in Figure 5 mean values of temperature, snow cover and albedo, as well as the average change of snow albedo during spring. Also, to underline the crucial role of in-situ snow depth information, mean snow depth is shown. Mean station snow depth lies within the range of reanalyses values, with higher values reported by ERAI-LG. Moreover, stations have the lowest snow cover among all datasets (Figure 5 b and c). This difference is likely due to the conversion of snow depth to snow cover as well as from the precision (in centimeters) of the Russian snow depth measurement. Precision of snow depth diagnosed by reanalysis is much finer and the logarithmic conversion here can be performed more accurately. As a result, TEM values diagnosed by stations are probably too low. If we consider instead in-situ snow cover information from stations, the average snow cover is quite similar to reanalyses (ca. 55%), and the average TEM value gets stronger. However, replacing converted snow cover with observed snow cover in Eq. (2) is a questionable procedure, as the remaining terms were computed using snow depth conversion. Thus, for consistency we show lower values of TEM in Figure 4.

Temperature is well represented by all datasets with MERRA2 being about 1 K colder compared to stations, which is quite notable for such a robust varaiable. However, absolute values of temperature do not have a strong impact on the computation of TEM, since month-to-month changes in temperature affect both TEM and SNC computations. For ERAI-LG albedo contrast, the effect of the underestimated snow-free albedo and overestimated snow albedo cancel each other out. Finally, the snow albedo change during spring (Figure 5f) is very similar in station data and in MERRA2 (-0.09 average in both datasets), which points towards an adequate representation of snow metamorphosis and aerosol deposition in MERRA2. The ERAI-LG experiment shows a stronger change of snow albedo during spring than the standard version. ERAI-L potentially keeps the temperature and therefore snow metamorphosis more constant throughout spring due to a more stable local temperature climate induced by the vegetaiton. Note also, that some stations show an increase of snow albedo during spring. This can be caused by fresh snow accumulation in late spring in some locations.

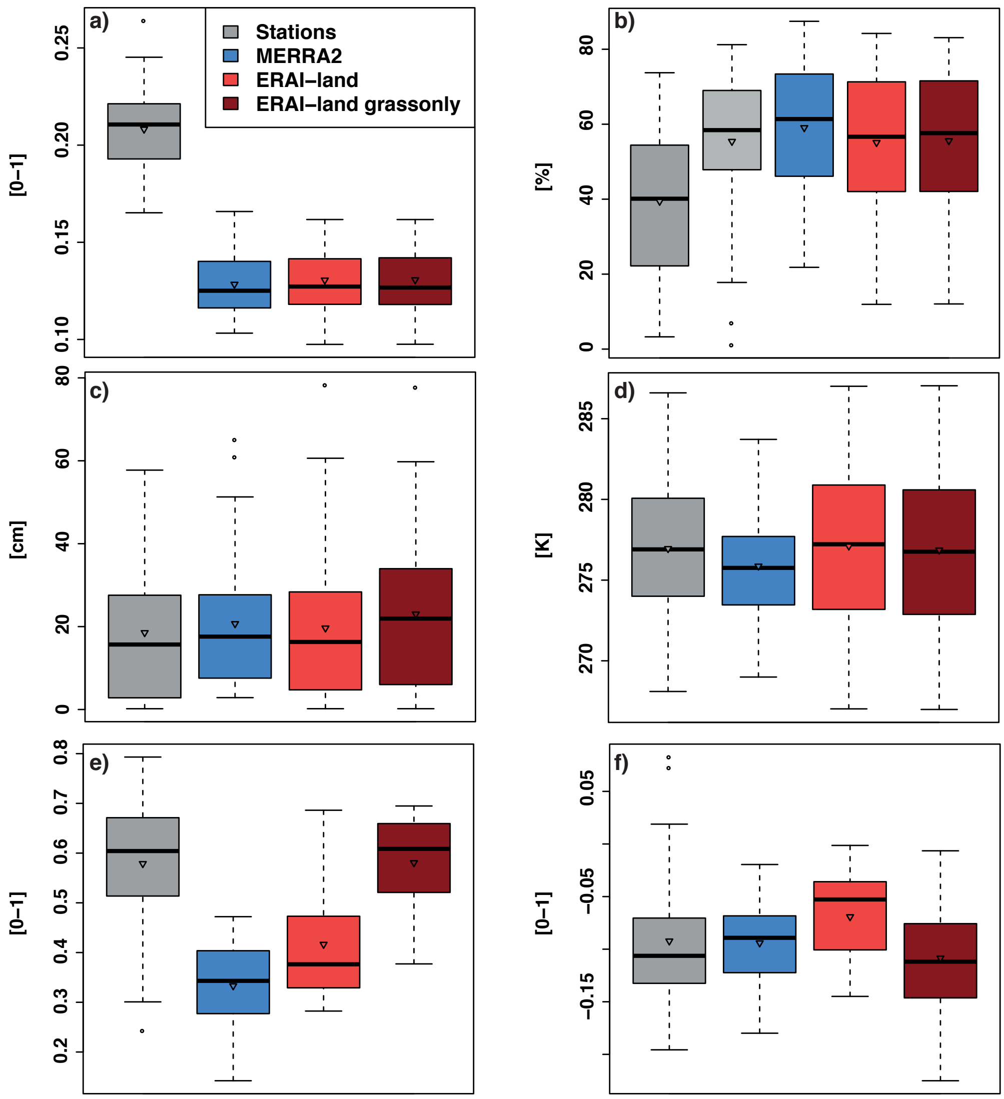


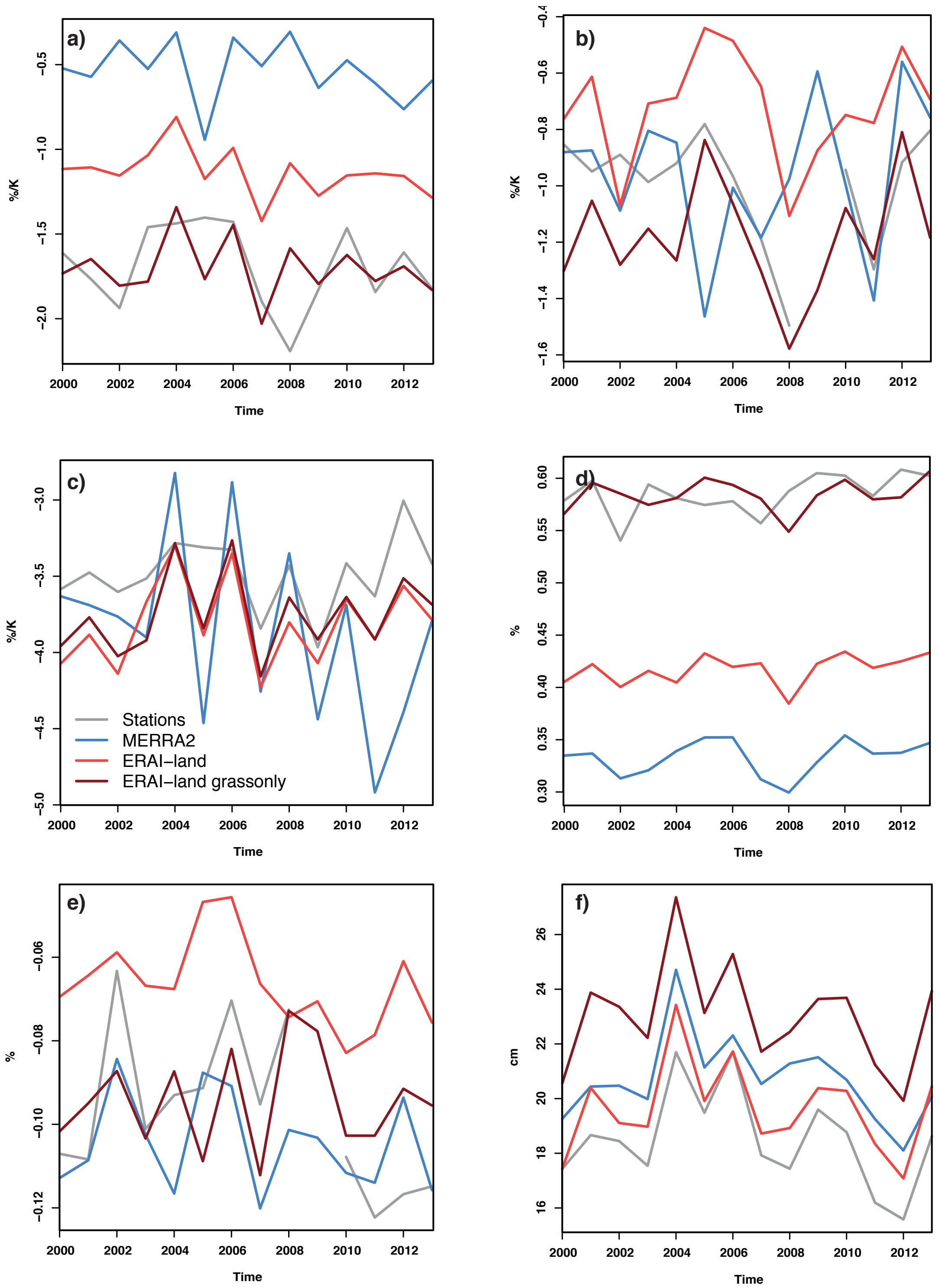
Figure 5: Boxplot analysis for MAMJ 2000–2013 a) snow free albedo, b) snow cover fraction, where the light grey boxplot is the originally observed snow cover from stations, c) snow depth, d) 2m temperature, e) mean albedo and f) snow albedo change within the season. Triangle indicates the mean value.

Figure 6 shows timeseries (2000–2013) for the mean values for SAF-related variables. Timeseries for SNC (Figure 6a) and TEM (Figure 6b) show that inter-annual variations of up to 0.5% K^{-1} are possible for both stations and reanalyses. Moreover, for both SNC and TEM, ERAI-LG seems to reproduce well the overall baseline and the magnitude of variability.

For snow melt sensitivity (Figure 6c) the agreement among the datasets is very good when it comes to magnitude and interannual variability, with MERRA2 showing an amplified inter-annual variability (up to 1.5% K^{-1}), which is beyond the magnitudes observed at stations. As already noted above, snow melt sensitivity seems to be a rather well reproduced process in modern reanalyses. Since snow-free albedo is quite constant over time in the reanalyses, the albedo contrast is dominated by the snow albedo (Figure 6d). ERAI-LG and the station network agree very well on the magnitude of snow albedo, whereas ERAI-L and MERRA2 fail to reproduce such high values. Magnitudes of inter-annual variability can reach up to ±0.05 in stations, with slightly weaker response in reanalyses. Correlation between stations and reanalyses is rather low, only individual years are captured correctly by ERAI-LG (see Supplement for correlation values).

Snow albedo change within spring (Figure 6e) is well captured by MERRA2 and ERAI-LG. Furthermore, ERAI-LG captures well the inter-annual varability for this metric. Specifically, variability during 2001–2004 and 2005–2008 periods is quite well represented. On the other hand, ERAI-L seems to lack the consistency with observations. Finally, as it was mentioned in section 4.1, snow depth variability (Figure 6f) is very well captured by all reanalyses. Again, ERAI-LG overestimates snow depth by up to 5 cm, with the other two reanalyses being on average 1-2 cm above the station values.

501



502 **Figure 6: Yearly timeseries of selected MAMJ SAF components averaged over all**
503 **44 locations. a) SNC, b) TEM, c) snow melt sensitivity, d) snow albedo, e) snow**
504 **albedo change within the season, f) snow depth.**

505

To further demonstrate the effect of the vegetation changes in the ERA-Interim land reanalysis, Figure 7 shows anomalies between ERAI-L and ERAI-LG. The structure follows Figure 6, with SNC and TEM shown in Figure 7a&b. As is clearly visible both variables are generally less negative in ERAI-L, a fact already known from timeseries and boxplot analysis. The largest impact of the vegetation changes is found for Northern Russia, the Pacific coast and the western region between Black and Caspian Sea. Interestingly, but as expected, snow melt sensitivity (Figure 6c) is not the key driver behind this distrubution. Since snow melt sensitivity is not directly linked to vegetation changes, the anomaly distribution is very heterogenous, with positive and negative anomalies over the whole domain. As known from the timeseries plot, snow sensitivity in ERAI-LG is overall slightly weaker than in ERAI-L, probably due to positive feedbacks such as reduction of nighttime cooling over higher vegetation types. The main driver behind the distribution of SNC is albedo contrast (Figure 7d). Albedo contrast is overall higher in ERAI-LG, especially along the borders of the domain, highlighted already for SNC.

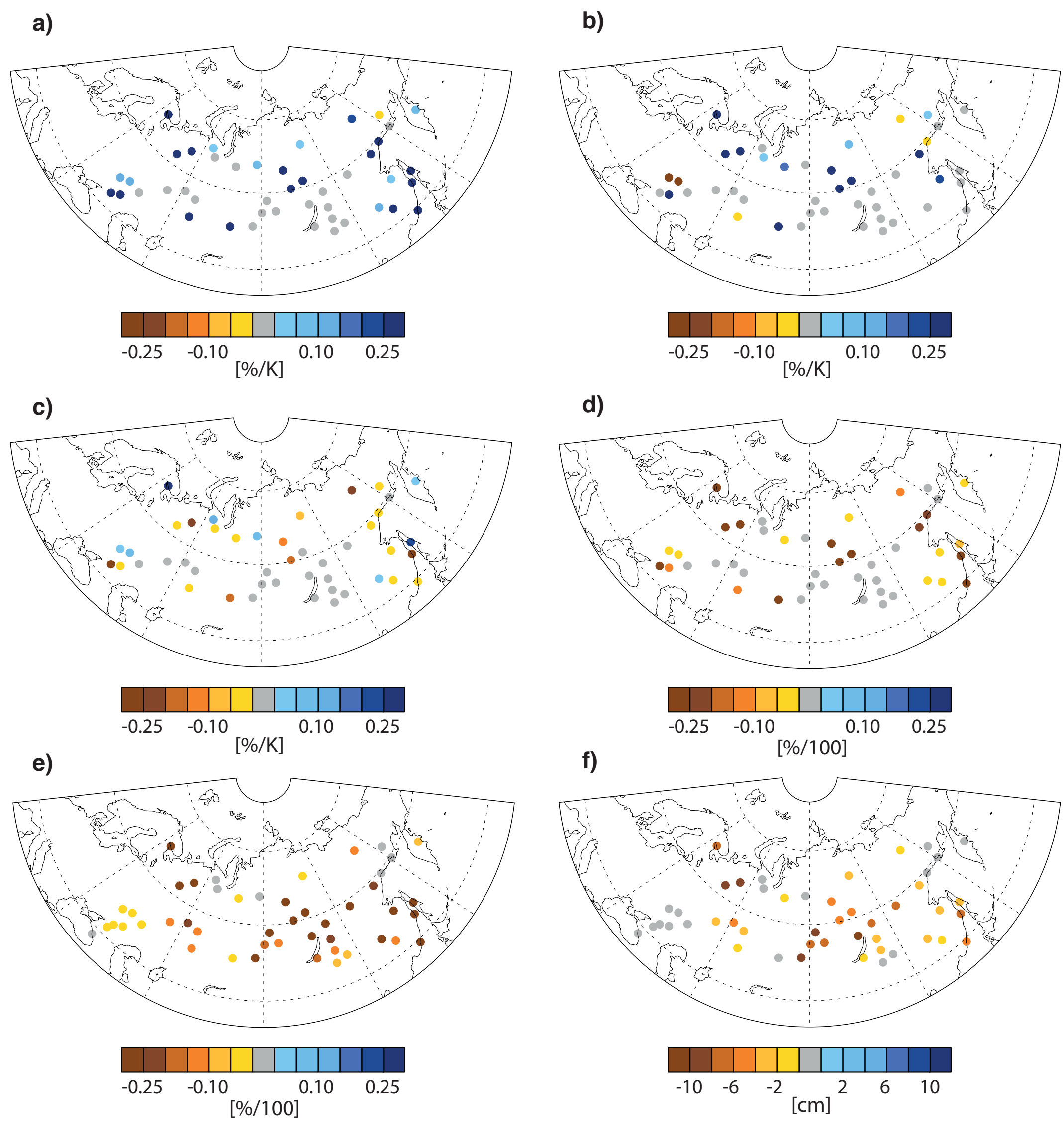


Figure 7: Mean SAF components in anomalies of ERAI-L minus ERAI-LG for 2000-2013 MAMJ. a) SNC, b) TEM, c) snow melt sensitivity, d) mean albedo contrast, e) mean albedo, f) snow depth.

5. Discussion

We compared spring SAF and its components determined from in-situ measurements over Russia for the period 2000–2013 with data derived from three modern reanalysis products restricted to the grid cells including the observational sites. This was achieved by using a unique collection of station measurements of radiation and snow characteristics investigating for the first time observed SAF over this broad spatial and temporal domain. Besides ERAI-L we also used a customized version of ERAI-L (ERAI-LG), in which vegetation was set to grass in all concerned grid cells.

All three reanalysis datasets are completely independent from the analyzed station data. While a direct comparison of point measurements with grid cell output always introduces uncertainties due to the spatial varibailty of the surface, this is for now the only way to evaluate reanalyses data using in-situ observations. An alternative option would be satellite data, which come with their own uncertainties (e.g. **Romanov et al. 2002, Foster et al. 2005, Wang et al. 2014).**

Snow depth statistics derived from daily station data are reasonably well reproduced in all three modern reanalyses, which is in agreement with **Wegmann et al. (2017)** who investigated April snow depth in ERAI-L. While snow depth differences between ERAI-L and ERAI-LG are small, ERAI-LG shows slightly higher deviations from the station data than ERAI-L that might be caused by the higher vegetation in station surroundings and by underestimation of snowfall due to instrumentation used at the Russian station network (**Rasmussen et al. 2012**).

Day-to-day variability of albedo is notably higher in station data compared to any reanalysis product. Besides spatial averaging over the reanalyses grid cells, this is potentially caused by land surface changes due to weather (e.g. soil moisture change, aerosol deposition), which are not represented in the reanalyses. However, ERAI-LG demonstrates increasing albedo variability, nearly doubling the standard deviations diagnosed by ERAI-L with the standard vegetation scheme.

The limitations of the station data imply some constraints for comparisons with reanalysed data. As near-surface temperature is unavailable in station data, we used for both stations and reanalyses 2m air temperature, which reduces the strength of the SAF feedback. Secondly, snow cover is underestimated in station data due to the measurement precision of 1cm, which reduces the strength of the TEM component. The snow albedo and the snow-free albedo are substantially higher in station data than in the reanalyses with classic vegetation boundary conditions (MERRA2 and ERAI-L). Compared to other observation-based studies, spring snow albedo and grass albedo derived from our station network is quite realistic **(Roesch et al. 2009, Stroeve et al.**

2006). Thus, the difference revealed by reanalyses is likely due to averaging over grid cells.

Results from ERAI-LG clearly demonstrate that SAF and its components are very close to those in the station data. The largest improvement was found for albedo contrast and for snow albedo, which both are more realistic in ERAI-LG. At the same time snow-free albedo in all three reanalyses (including ERAI-LG) was found to be lower than in the station data, because snow-free albedo in all reanalysis data sets is prescribed as a monthly climatology from MODIS data. As MODIS mostly registers albedo from Taiga and Tundra vegetation, a stark difference to the grass albedo from the stations occurs.

MERRA2 shows the lowest SAF values resulting from a very low albedo contrast, which is probably a consequence of the vegetation scheme in the MERRA2 land module. On the other hand, MERRA2 represents TEM reasonably well most likely due to the accurate representation of the intra-seasonal snow albedo changes. Thus, relative snowpack changes appear to be well represented in MERRA2, probably also due to a more accurate representation of aerosols.

In general, we found higher SAF values in ERAI-L than in the recent CMIP3/CMIP5 analyses of NH SAF by **Fletcher et al. (2015)**. This disagreement results from a variety of factors. First, our domain is limited to Russia only, thus excluding considerable parts of Eurasia as well as North America. In this respect our domain is set within a high SAF region, which may explain higher SAF values compared to the NH average by **Fletcher et al. (2015**). On the other hand, MERRA2 shows good agreements with the NH CMIP4/5 SAF results, however mostly because the albedo contrast is very low. Furthermore, as we pointed out above, in-situ observations used here tend to slightly overestimate SAF, mainly due to higher snow albedo values. This is because in-situ snow albedo is typically measured by a sensor installed over a vegetation-free snow pack. The vegetation scheme used in reanalyses gives lower snow albedo values implying realistic vegetation cover such as taiga or tundra. However, our MERRA2 results agree fairly well with the findings of **Fletcher et al. (2015)**. Moreover, mean values of the albedo independent variable snow melt sensitivity are very close to the "observational" snow melt sensitivity computed by **Fletcher et al. (2015).**

We also found agreements with **Fletcher et al. (2015)** in the representation of the spatial pattern of the SAF components. **Fletcher et al. (2015)** as well as **Fernandes et al. (2009)** have shown maxima in SAF over northern Canada, northern Siberia and southwestern Eurasia. The relation of 60:40 between SNC and TEM, which is found in modeled, satellite and reanalysis data, was replicated by our station network. We found similar spatial patterns for SAF and its components in both stations and gridded data specifically for Southern Russia, while the pattern of station responses is less homogenous compared to the gridded data. Also consistent with **Fletcher et al. (2015)**, we found higher snow melt sensitivity north of 50° N. Finally, albedo contrast distribution, which closely follows the snow albedo pattern, is in very good agreement with the gridded analysis of snow albedo by **Fletcher et al. (2015)**.

6. Conclusions

Reanalyses including land surface modules show a physically consistent representation of SAF with realistic spatial patterns and area-averaged sensitivity estimates. ERAI-LG shows a better performance in representing station-based estimates considering the uncertainty associated with "point to grid cell" comparisons. Accounting for aerosol-related processes would likely improve this performance in future reanalysis releases. Thus, for the analysis and validation of large-scale temporal and spatial averages of SAF modern reanalyses seem to be an appropriate tool.

However, for analysing processes on smaller scales and high temporal resolution studies, a healthy dense station network is required. The idealized ERAI-LG simulation also highlights the caveats of comparing in-situ observations with gridded model data. In this study, we show these discrepancies in terms of albedo and snow depth. Other variables, in particular 2m temperature, can be expected to have a similar signal arising from the differences between the model's gridcell land cover and the actual station conditions. Our findings show that the experimental approach in ERAI-LG allows for an enhanced use of in-situ observations to diagnose the SAF in not-forested areas.

Considering future studies, the extension to other regions and use of other regional in-situ data might give further insights into regional hotspots of SAF. Cross-validation efforts employing model, reanalysis, satellite and station data may help to generate blended products to investigate radiation and albedo feedbacks in the changing Arctic, a region where SAF is especially strong. Regional modelling, including a variety of multi-layer land surface models over areas with a relatively dense observation network can provide a quantitative estimation of uncertainties among complex variables such as snow depth, albedo or SAF.

Acknowledgements. This study was suported by the ARCTIC-ERA project funded by the Belmont Forum Fund through the ANR. OZ also benefited from the support by the Russian Ministry of Education and Science (project no. 14.613.21.0083, ID RFMEFI61317X0083). ED was supported by the Portuguese Science Foundation (FCT) under project IF/00817/2015.

References

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T. & Vitart, F. 2015: ERA-Interim/Land: a global land surface reanalysis data set. Hydrology and Earth System Sciences, 19, 389-407

Betts, A. K., & Ball, J. H. 1997: Albedo over the boreal forest. Journal of Geophysical Research: Atmospheres, 102, 28901-28909.

Bony, S., Colman, R., Kattsov, V.M., Allan, R.P., Bretherton, C.S., Dufresne, J., Hall, A., Hallegatte, S., Holland, M.M., Ingram, W., Randall, D.A., Soden, B.J., Tselioudis, G. and Webb M.J. 2006: How Well Do We Understand and Evaluate Climate Change Feedback Processes?. J. Climate, 19, 3445–3482

Brun, E., Vionnet, V., Boone, A., Decharme, B., Peings, Y., Valette, R., Karbou, F., and Morin, S 2013: Simulation of Northern Eurasian Local Snow Depth, Mass, and Density Using a Detailed Snowpack Model and Meteorological Reanalyses, J. Hydrometeorol., 14, 203–219

Brutel-Vuilmet, C., Ménégoz, M., & Krinner, G. 2013: An analysis of present and future seasonal Northern Hemisphere land snow cover simulated by CMIP5 coupled climate models. The Cryosphere, 7, 67

Budkyo, M. I. 1967: The effect of solar radiation on the climate of the earth. Tellus, 21, 611-19

Bulygina, O. N., Groisman, P. Y., Razuvaev, V. N., & Radionov, V. F. 2010: Snow cover basal ice layer changes over Northern Eurasia since 1966. Environmental Research Letters, 5, 015004.

Cess, R. O., & Potter, G. L. 1991: Interpretation of snow-climate feedback as produced by 17 general circulation models. Science, 253, 888

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., Francis, J., Dethloff, K., Entekhabi, D. & Overland J. 2014: Recent Arctic amplification and extreme mid-latitude weather. Nat. Geosci., 7, 627–37

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A.J. & Wehner M. 2013: Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Curry, J. A., Schramm, J. L., Rossow, W. B., & Randall, D. 1996: Overview of Arctic cloud and radiation characteristics. Journal of Climate, 9, 1731-1764.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bid- lot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., & Vitart, F. 2011: The ERA–interim reanalysis: Configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597

Derksen, C. & R. Brown, 2012: Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections. Geophysical Research Letters, 39.

Dufour, A., Zolina, O. and Gulev, S.K., 2016: Atmospheric moisture transport to the Arctic: Assessment of reanalyses and analysis of transport components. Journal of Climate, 29, 5061-5081.

Dutra, E., Balsamo, G., Viterbo, P., Miranda, P. M., Beljaars, A., Schär, C., & Elder, K. 2010: An improved snow scheme for the ECMWF land surface model: description and offline validation. Journal of Hydrometeorology, 11, 899-916.

Fernandes, R., H. Zhao, X. Wang, J. Key, X. Qu, & A. Hall 2009: Controls on

Northern Hemisphere snow albedo feedback quantified using satellite earth observations, Geophys. Res. Lett., 36, L21702

Flanner, M. G., Shell, K. M., Barlage, M., Perovich, D. K., & Tschudi, M. A. 2011: Radiative forcing and albedo feedback from the Northern Hemisphere cryosphere between 1979 and 2008. Nature Geoscience, 4, 151.

Fletcher, C. G., Thackeray, C. W., & Burgers, T. M. 2015 Evaluating biases in simulated snow albedo feedback in two generations of climate models. Journal of Geophysical Research: Atmospheres, 120, 12-26.

Fletcher, C. G., Zhao, H., Kushner, P. J., & Fernandes, R. 2012: Using models and satellite observations to evaluate the strength of snow albedo feedback. Journal of Geophysical Research: Atmospheres, 117

Foster, J. L., Sun, C., Walker, J. P., Kelly, R., Chang, A., Dong, J., & Powell, H. 2005: Quantifying the uncertainty in passive microwave snow water equivalent observations. Remote Sensing of environment, 94, 187-203

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M. & Zhao B. 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). J. Climate, 30, 5419–5454

Graversen, R. G., & Wang, M. 2009: Polar amplification in a coupled climate model with locked albedo. Climate Dynamics, 33, 629-643

Graversen, R. G., Langen, P. L., & Mauritsen, T. 2014: Polar amplification in CCSM4: Contributions from the lapse rate and surface albedo feedbacks. Journal of Climate, 27, 4433-4450.

Groisman, P. Y., Karl, T. R., Knight, R. W., & Stenchikov, G. L. 1994: Changes of snow cover, temperature, and radiative heat balance over the Northern Hemisphere. Journal of Climate, 7, 1633-1656.

Hall, A. 2004: The role of surface albedo feedback in climate. Journal of Climate, 17, 1550-1568.

Hall, A., & Qu, X. 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. Geophysical Research Letters, 33(3).

Hall, A., Qu, X., & Neelin, J. D. 2008 Improving predictions of summer climate change in the United States. Geophysical Research Letters, 35

Hirahara, S., Ishii, M., & Fukuda, Y. 2014: Centennial-scale sea surface temperature analysis and its uncertainty. Journal of Climate, 27, 57-75.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M., & Potter, G. L. 2002: Ncep–doe amip-ii reanalysis (r-2). Bulletin of the American Meteorological Society, 83, 1631-1643.

Kevin, J. P. W., Kotlarski, S., Scherrer, S. C., & Schär, C. 2017: The Alpine snow-albedo feedback in regional climate models. Climate Dynamics, 48, 1109-1124.

Kobayashi, S., Yukinari, O.T.A., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Miyaoka, K. & Takahashi, K., 2015: The JRA-55 reanalysis: General specifications and basic characteristics. Journal of the Meteorological Society of Japan. Ser. II, 93, 5-48.

Lian, M. S., & Cess, R. D. 1977: Energy balance climate models: A reappraisal of ice-albedo feedback. Journal of the Atmospheric Sciences, 34, 1058-1062.

Lindsay, R., Wensnahan, M., Schweiger, A., & Zhang, J. 2014: Evaluation of seven different atmospheric reanalysis products in the Arctic. Journal of Climate, 27, 2588-2606.

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. 2015: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. Geoscientific Model Development, 8, 1339-1356.

Pithan, F. & Mauritsen, T. 2014: Arctic amplification dominated by temperature feedbacks in contemporary climate models. Nature Geoscience, 7,181.

Qu, X., & Hall, A. 2007: What controls the strength of snow-albedo feedback?. Journal of Climate, 20, 3971-3981.

Qu, X., & Hall, A. 2014: On the persistent spread in snow-albedo feedback. Climate dynamics, 42, 69-81

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A.P., Black, J., Thériault, J.M., Kucera, P., Gochis, D., Smith, C., Nitu, R., Hall, M., Ikeda, K., & Gutmann E. 2012: How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. Bull. Amer. Meteor. Soc., 93, 811–829

Reichle, R. H., Draper, C. S., Liu, Q., Girotto, M., Mahanama, S. P., Koster, R. D., & De Lannoy, G. J. 2017: Assessment of MERRA-2 land surface hydrology estimates. Journal of Climate, 30, 2937-2960.

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F.R., Ruddick, A.G., Sienkiewicz, M. & Woollen, J. 2011: MERRA: NASA’s Modern-Era Retrospective Analysis for Research and Applications. J. Climate, 24, 3624–3648

Robock, A. 1983: Ice and snow feedbacks and the latitudinal and seasonal distribution of climate sensitivity. Journal of the Atmospheric Sciences, 40, 986-997.

Roesch, A., Gilgen, H., Wild, M., & Ohmura, A. 1999: Assessment of GCM simulated snow albedo using direct observations. Climate dynamics, 15, 405-418.

Romanov, P., Gutman, G., & Csiszar, I. 2002: Satellite-derived snow cover maps for North America: accuracy assessment. Advances in space Research, 30, 2455-2460.

Schiffer, R. A., & Rossow, W. B. 1983 The International Satellite Cloud Climatology Project(ISCCP)- The first project of the World Climate Research Programme. American Meteorological Society, Bulletin, 64, 779-784.

Schneider, S. H., & Dickinson, R. E. 1974: Climate modeling. Reviews of

Geophysics, 12, 447-493.

Serreze, M. C., & Barry, R. G. 2011: Processes and impacts of Arctic amplification: A research synthesis. Global and Planetary Change, 77, 85-96.

Stroeve, J. C., Box, J. E., & Haran, T. 2006: Evaluation of the MODIS (MOD10A1) daily snow albedo product over the Greenland ice sheet. Remote Sensing of Environment, 105, 155-171.

Thackeray, C. W., & Fletcher, C. G. 2016: Snow albedo feedback: Current knowledge, importance, outstanding issues and future directions. Progress in Physical Geography, 40, 392-408.

Wang, Z., Schaaf, C.B., Strahler, A.H., Chopping, M.J., Román, M.O., Shuai, Y., Woodcock, C.E., Hollinger, D.Y. & Fitzjarrald, D.R. 2014: Evaluation of MODIS albedo product (MCD43A) over grassland, agriculture and forest surface types during dormant and snow-covered periods. Remote Sensing of Environment, 140, 60-77.

Wei, X., Hahmann, A. N., Dickinson, R. E., Yang, Z. L., Zeng, X., Schaudt, K. J., Schaaf, C.B. & Strugnell, N. 2001: Comparison of albedos computed by land surface models and evaluation against remotely sensed data. Journal of Geophysical Research: Atmospheres, 106, 20687-20702.

Wegmann, M., Orsolini, Y., Dutra, E., Bulygina, O., Sterin, A., & Brönnimann, S. 2017: Eurasian snow depth in long-term climate reanalyses. The Cryosphere, 11, 923.

Wexler, H. 1953: Radiation balance of the Earth as a factor in climatic change. In: Shapley H (ed) Climatic Change. Cambridge: Harvard University Press,73-105

Xiao, L., Che, T., Chen, L., Xie, H., & Dai, L. 2017: Quantifying Snow Albedo Radiative Forcing and Its Feedback during 2003–2016. Remote Sensing, 9, 883.