REVIEWER 1

General comments: Overall, this a informative and relevant paper with some issues, which need to be resolved. The paper investigates the performance of different reanalyzes products in representing snow depth in the NE part of Eurasia. The authors use daily snow depth measurements from 820 Russian meteorological stations to compare climatologies and 13 long-term stations to analyze temporal differences. The topic of the investigation fits very well into the journal's scope. It is one of the very few studies that thoroughly evaluates the snow depth represented in different reanalyzes products. As such, I consider the work as being relevant for the scientific community. For most parts, the methods are described appropriately and the conclusions are well-based on the results obtained. The paper however suffers from a simple overview (look up table) of the underlying datasets. Please see the listing below for further details. These issues should be improved before publication of the paper. For this purpose, only very few new analyses are required and the basic structure of the paper does not have to be changed. I'd therefore suggest returning the manuscript to the authors for minor revisions.

Response:

Major issues:

- For readers not familiar with reanalysis products a paragraph is missing where it is explained which snow variables are provided in such products and how they are calculated.

R : Indeed, this information was missing. We added a paragraph about snow computation in the reanalyses at the end of section 2.1

- A table is missing where the characteristics of the different reanalysis products are listed. Such a table should contain which product belongs to which of the two families, what are the differences in regard to the assimilated data, what are the differences in spatial and temporal resolution, etc.

R : Thanks for this suggestion. A good overview table was needed. We added Table 1 in the manuscript with details concerning the differences in the reanalyses.

- I miss a kind of uncertainty assessment. Could you please mention that there is some uncertainty due to the elevation differences between the grid cell and the station. Did you also try to use the neighboring grid cell with smallest elevation difference

instead? The temporal resolution of the reanalysis products may also not fit the snow observation time. Do products with finer spatial or higher temporal resolution perform better?

R : Thank you for pointing that out. We mention this uncertainty now in line 209. We did not include gridboxes with the smallest elevation difference since especially in the case of 20CR, topography is rather coarse in the model and we wanted to keep the procedure the same in all gridded datasets. We also added in line 228 the information about temporal resolution. Indeed, observation time does not fit 100 % the daily (or 6-hourly resolution) in the reanalyses. We used a finer grid in ERA20C than was used for 20CR and we only see minor improvements. Assimilated data and model physics play a more important role.

- In order to be able to properly assess the different errors measures for the 15 long-term stations presented in different figures the reader needs to have an idea about the mean and standard deviation of the different analyzed snow depth values of each individual station. I suggest to add this information to table 1 or to add a new table. The

information of the percentage of missing values is currently hard to read and could be easily combined with the climatological information of each station.

R : For better assessment of missing data we averaged the missing data for all three months and changed Table 2. Standard deviation can be seen in the Taylor diagrams. For mean value investigation, we initially had Figure 1 in mind. However, we see the point Reviewer #1 makes and added additional standard deviation and mean value analysis boxplots for the 13 long-term stations in the supplement, so it is easier for the reader to access these values.

- In order to test if the relatively poor hitrate is influenced by temporal issues between reanalysis and observation, I suggest to also calculating the hitrate when +/- 1 day shift in the reanalysis is allowed.

R : Very good point. We exchange this analysis in the manuscript and mention the results for the fixed date just briefly.

Minor issues:

L30: On order to prevent misunderstanding, replace “data sets” with “reanalysis products”

R : Changed

L66: Why “slowly”? Often the state of the snow cover changes very fast.

R : Changed to “corresponding”

L116-117: The last sentence in this paragraph cannot be understood by readers unfamiliar with reanalysis products.

R : Clarified L116

L163-174: What is the difference to the “Historical Soviet Daily Snow Depth (HSDSD)

product [Armstrong, 2001]”? Would there be more long-term data series than only 15?

R : The dataset we used contains overall more stations, but the long-term stations are mostly the same. Therefore, unfortunately no more long-term data series than just 15.

L163-174: Please add some information how snow depth was measured. Point measurement on a stake or mean snow depth from snow courses? Just out of personal interest: What did change in the measurement procedure after 1965?

R : We added that information in section 2.2.

The procedure of snow observations changed in the past:

- **Changed the size of the stake (1924,1939)**
- **Changed the rules for the use of stake (1935, 1939),**
- **Changed the requirements for observation platform (1940, 1954),**
- **Changes in the rules archiving (1966)**

L192: I cannot find “red marked” stations?

R : That was an artifact, deleted

L199: daily accumulated snow depth

R : Changed

L213-214: To be able to better follow your explanations, the meridians should be indicated in Figure 1.

R : For clearer assessment of Figures 1&2, we added meridians.

L222-225: Please add a sentence mentioning that the depicted snow depth represents the mean maximum snow depth for each shown month.

R : Added that information

L252-253: Please explain why you compare Northern Russia (and e.g. not Eastern Russia) in this step.

R : We use this area based on the climatology maps of snow cover. In our view this is the region with the highest snow depths. We added that explanation to the text.

Figure 3: Is there any argument not to use the same scale on all three graphs?

R : No there is not. We now use the same scale in all three graphs.

Figure 4: Is there any argument not to use the same scale on all three graphs?

R : No there is not. We now use the same scale in all three graphs.

L293: “Daily” still means monthly maximum snow depth?

R : Daily in this case means « as measured », on a day to day basis. These are used for all following analysis procedures, like correlation, hitrate etc. This allows us to have a very strong statistic.

Figure 7: Is there any argument not to use the same scale on all six graphs? Are these hitrates based on 1981-2010 or on the longest period available?

R : No there is not. We now use the same scale in all three graphs. Hitrates are based in the longest period possible. We added that statement to the text.

L388-390: Please add links to the table and figure where these results can be seen. What are the arguments to call the correlations "very" high? They are mostly below 0.8.

R : Thanks for pointing that out. If we have the daily resolution and spatial grids in mind, the results are quite remarkable. However, our wording here was wrong. We changed the wording accordingly.

L391: I don't understand what you mean with "although dealing with a large sample size"?

R : Again, thanks for pointing that out. Wording was not correct. Is changed.

L400-402: To which period does this statement apply?

R : Longest period possible

L403-404: I guess the RMSE is smallest in October because absolute values are smallest in October!

R : We agree! This point is made in L406.

L449: Crutemp: Please add version and reference.

R : Reference and version is added.

Supplementary Table 1: What period do these numbers refer to?

R : Longest period possible, except for ERA-Interim where they refer to 1981-2010

Supplementary Figure 3-5: Is there any argument not to use the same scale on all three graphs?

R : For the boxplot graphs in the supplement we decided to keep different

scales since metrics change the scale quite a bit between different months and we want reader to see the maximum amount of details since the Taylor Plots show only median values.

Should the median values of Supplementary Table 1 not be found in Figure 3 and 4?

R : Thank you for pointing that out. Yes, they should be found in these figures. However, we found that the numbers for ERA-Interim were not up to date (wrong time window selected) and there was an error in one entry for ERA20c-land. We updated all numbers accordingly.

Supplementary Figure 5: The unit “cm2“ for the variance seems strange? Why not use
the more common measure standard deviation?

R : We added a boxplot for standard deviation

Supplementary Figure 6: Is there any argument not to use the same scale on all three
graphs? What is Hadsipr2?

R : No there is not. We now use the same scale in all three graphs. We added the information about the SLP reconstruction

Supplementary Figure 7: Is there any argument not to use the same scale on all three
graphs?

R : No there is not. We now use the same scale in all three graphs.

REVIEWER 2

The manuscript addresses an important topic, which fits very well to the scope of the journal. There has been a lot of uncertainty in the recent trends in Siberian snow cover in autumn, and the manuscript to some degree reduces this uncertainty, by showing that the observed trends strongly vary in space (Figure 2a). Moreover, interesting results are presented on the centennial time scale, showing major differences between the U.S. and European reanalyses until about 1940. The manuscript has, however, also weaknesses, and I suggest that major revisions should be made before publication.

Major comments:

1. A lot of results are presented on the performance of reanalyses in various months and regions, evaluated using various skill scores. The manuscript is, however, lacking analysis on the reasons for the better of worse performance of reanalyses. For example, major differences are found for the period 1901-1940 (Figures 3 and 4), and a reader is certainly interested in understanding the reasons for the differences. The differences can originate from (a) different data assimilated or different methods applied in assimilation of the same data, (b) different model results for precipitation and its phase, (c) different model results for snow melt, and possibly (d) different parameterizations (if any) applied for snow metamorphosis causing changes in snow density and, accordingly, thickness. The authors should pay at least some attention on these issues. If it is too difficult to find answers to issues (b) to (d), at least the snow schemes applied in the models should be compared. There may be major differences in the schemes for snow thermodynamics, which may explain the different results in early years when the role of data assimilation was probably smaller.

R : Thank you for pointing out some key elements. In the discussion part we investigate several options why the difference might occur, namely temperature and sea level pressure differences between the datasets. However, as you rightly pointed out, an outline of differences among the snow scheme was missing, which is added now at the end of section 2.1 . Moreover, we added Table 1, where it is more apparent what data assimilation is used and what boundary conditions are used. That said, we

can dig only so far into technical details. Our investigations still show that assimilation and snow schemes are very similar, and we still support the idea of dynamical reasons for the changes in snow. We added a plot for vertical integrated mass of atmosphere, which points out a problem in ERA20C, namely to much high pressure over the Arctic in the first half of the 20th century. With this we hope to give enough initial ideas as to why the snow states diverge. Future studies need to check this feature in more detail.

2. The arguments for conclusions presented in Sections 5 and 6 are not clear. Why do you write in the beginning of Section 5 that the results indicate a good performance of reanalyses (change “datasets” to “products”) and that climatologies are well represented? All figures presenting comparisons against observations include considerable errors, and Figure 3 only comparing different reanalyses includes huge differences. Also, most of the correlation coefficients presented are not “very high”. A correlation of 0.6 only explains 36% of the variance. If you consider the results good, did you have reasons (in addition to Khan et al. 2008) to expect worse results? Do you have arguments to set relevant thresholds for “good performance”?
R : Indeed, the wording here is not correct. We clarified the section and added arguments as why we see the performance as “good”

3. In general, the text is not particularly clearly written. See Minor comments below.
Minor comments:
Lines 31-34: unclear text
R : clarified

Line 51: alter . . . modulate
R : changed

Line 59: has severely impacted
R : changed

Line 60: “From 1979 to 2011” or “Between 1979 and 2011”

R : changed

Lines 62-63: I am not sure, if Park et al. (2013) also report regional snow cover increase associated with low sea ice concentration. The main message of their study is, however, the opposite, given by the title of the paper: “The role of declining Arctic sea ice in recent decreasing terrestrial Arctic snow depths”.

R : Indeed, they only report regional specifics. Deleted the citation at this point

Line 76: climate models

R : changed

Lines 79-81: Global reanalyses have at least equally large spatial coverage as satellite

products. So, the work “compromise” is perhaps not the best.

R : Clarified

Lines 85-86: not all reanalyses listed here extend further back in time.

R : Clarified

Line 98 and analogously in many other places: Brun et al. (2013)

R : corrected

Line 124: Medium-Range

R : corrected

Line 130: assimilating synoptic observations of atmospheric surface pressure

R : corrected

Line 144 delete “model”

R : deleted

Line 146: tell the resolution also in km.

R : We added resolution information in Table 1

Line 150: “follows exactly the CMIP5 proposal” is unclear

R : Not sure what is unclear at this point. Added explanation as to what is CMIP5.

Line 186: perhaps “exceeding”

R : changed

Lines 279-284: the text is unclear and appear contradicting. Be clearer to which seasons you refer to in the beginning. On lines 283-284 the ECMWF is considered excellent in 1901-1940, but in Figure 4 the ECMWF appear excellent only in 1901-1910 and 1980.

R : We tried to clarify that part. However note that each point represents a 30 year long climatology, which is shifted by 10 years from point to point rather than a 10 year long climatology.

Lines 419-421: Snow drift may indeed generate differences between observations and reanalysis products. In addition to resolution, however, the differences may simply originate from a lack of snow drift parameterization in the reanalysis snow scheme

(see Major comment 1).

R : We added information about snow schemes in Section 2.1, see above.

Lines 427-428: The differences in input data should be quantified in Section 2.1.

R : Input data is now part of Table 1

Lines 438-443: The cause and consequence related to sea ice melt remains unclear. Without clarifying this, the processes at play in the pre-1950s sound very speculative.

R : We clarified this part. We do not want to tackle sea ice feedbacks here. This example should just be used as a dynamical reason as to why high pressure can lead to less snowfall.

Line 449: Why do you think that ERA20CA is most probably much too warm in

April?

R : Our best guess are dynamical reason, like temperature advection, due to pressure differences.

List of relevant changes

- We responded to the specific remarks of the two reviewers
- Wording changed in the abstract
- Added Table 1 for an overview of the used reanalyses
- Added a description of the snow schemes in the reanalyses
- Added meridians in Figure 1&2
- Timeseries plots are now consistent in terms of Y axis
- Changed hitrate analysis to ±1 day
- Changed wording and structure of discussion

Changes in the manuscript are marked in red

Eurasian snow depth in long-term climate reanalyses

Martin **Wegmann**[1,2,3], Yvan **Orsolini**[4], Emanuel **Dutra**[5,6], Olga **Bulygina**[7], Alexander **Sterin**[7] and Stefan **Brönnimann**[2,3]

[1] *Institut des Géosciences de l´Environnement, University of Grenoble, France*

[2] *Oeschger Centre for Climate Change Research, University of Bern, Switzerland*

[3] *Institute of Geography, University of Bern, Switzerland*

[4]*NILU—Norwegian Institute for Air Research, Kjeller, Norway*

[5] *ECMWF European Centre for Medium-Range Weather Forecasts, Reading, UK*

[6] *Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Portugal*

[7] *All-Russian Research Institute of Hydrometeorological Information—World Data Centre, Obninsk, Russian Federation*

Corresponding Author:

Martin Wegmann, martin.wegmann@univ-grenoble-alpes.fr

Abstract

Snow cover variability has significant effects on local and global climate evolution. By changing surface energy fluxes and hydrological conditions, changes in snow cover can alter atmospheric circulation and lead to remote climate effects. To document such multi-scale climate effects, atmospheric reanalysis and derived products offer the opportunity to analyze snow variability in great detail far back to the early 20th century. So far only little is know about their quality. Comparing snow depth in four long-term reanalysis datasets with Russian in situ snow depth data, we find a moderately high daily correlation (around 0.6-0.7), which is comparable to correlations for the recent era (1981-2010), and a good representation of sub-decadal variability. However, the representation of pre-1950 inter-decadal snow variability is questionable, since reanalysis products divert towards different base states. Limited availability of independent long-term snow data makes it difficult to assess the exact cause for this bifurcation in snow states, but initial investigations point towards representation of the atmosphere rather than differences in assimilated data or snow schemes. This study demonstrates the ability of long-term reanalysis to reproduce snow variability accordingly.

1. Introduction

Snow is an important component of the climate system over the mid- and high-latitude regions of the Earth. Its high shortwave albedo and low heat conductivity modulate heat and radiation fluxes at the Earth´s surface and thus directly modulates regional temperature evolution and ultimately atmospheric circulation patterns (Barnett et al. 1988, Cohen and Rind 1991, Callaghan et al. 2011, Cohen et al. 2014). Moreover, because snow acts as a temporary water reservoir, snow variability impacts soil moisture, evaporation and ultimately precipitation processes (Yasunari et al. 1991).

As a result, snow cover has an essential influence on ecological (Jonas et al. 2008, Peñuelas et al. 2009) and economical systems (eg. Agrawala 2007). Vice versa, snow cover itself is determined by climate variations. Recent Arctic warming has severely impacted spring snow cover. From 1979 to 2011, Arctic April snow cover extent decreased at a rate of -17.8% per decade (Derksen and Brown 2012). In contrast, regional snow cover increase in autumn over Eurasia was found in connection with low Arctic sea ice concentration (Honda et al. 2009, Wegmann et al. 2015), indicating the complexity of global and regional processes leading to snow cover changes.

Reciprocally, as a corresponding component of the climate system, the snow cover influences large-scale climate patterns, and has been tapped as a source of predictability at the subseasonal-to-seasonal scale, especially over Eurasia in autumn and winter (Cohen and Entekhabi 1999, Jeong et al. 2013, Orsolini et al. 2013, Wu et al. 2014, Ye et al. 2015,).

Therefore, large-scale monitoring and quantifying of snow cover is crucial for assessing climate change and its representation in climate models (eg. Frei and Gong 2005, Brown and Mote 2009, Brown and Robinson 2011, Liston and Hiemstra 2011, Ghatak et al. 2012, Zuo et al. 2015) and for analyzing cryosphere-climate feedbacks (eg. Flanner et al. 2011, Orsolini and Kvamstø 2009, Zhang et al. 2013). Here we analyze snow depths in climate reanalyses in comparison to in-situ data, with the aim to better assess cryosphere-atmosphere coupling processes in the context of the 20th century climate evolution.

To this end, reanalysis products provide a compromise between the high temporal resolution and length of in-situ observational datasets (eg. Bulygina et al. 2010) and the large spatial, but relatively short-term coverage of satellite products (Siljamo and Hyvärinen 2011, Frei et al. 2012, Hüsler et al. 2014). Comprehensive reanalyses datasets are well suited to investigate processes and mechanisms, and a variety of reanalyses are now routinely produced by meteorological prediction centers such as (but not limited to) NCEP-DOE, ERA-40 and ERA-Interim, and JRA-25 and JRA-55 (e.g. Uppala et al. 2005, Onogi et al. 2007, Compo et al. 2011, Dee et al. 2011, Rienecker et al. 2011, Poli et al. 2013).

However, so far only a few studies analyzed snow representation in reanalysis products. Khan et al. (2008) compared measured snow data with snow water equivalents and snow depth in the NCEP-DOE (Kanamitsu et al. 2002), ERA-40 (Uppala et al. 2005) and JRA-25 (Onogi et al 2007) reanalysis products over Russian river basins. They found that the ERA-40 outperformed the NCEP-DOE and JRA25 in terms of correlations and mean values. Despite reproducing well the seasonal variability, all reanalysis products struggled with snowmelt season values. Brown et al. 2010 compared ERA-40 and NCEP/NCAR snow cover extent to satellite and in-situ datasets. They found that for the period 1982-2002 ERA-40 shows higher correlations and smaller root mean squared errors (RMSE) than the NCEP reanalysis, and that May values were considerably better approximated than June values. Brun et al. (2013) forced the CROCUS snow model with atmospheric conditions from ERA-INTERIM (1970-1993) and found very high agreements with Eurasian in-situ snow measurements. However, no snow output from the reanalysis directly was evaluated.

In addition, climate reanalyses extending back to the beginning of the 20th century or earlier have now been produced for multi-decadal climate studies. Contrarily to the above-mentioned reanalyses, these climate reanalyses, namely the 20th Century Reanalysis (20CRv2) (Compo et al. 2011) and ERA-20C (Poli et al. 2016), solely rely on assimilation of surface data. Even fewer studies have tried to quantify snow cover extent and depth and their potential impact on climate in such centennial reanalyses. Recently, Peings et al. (2013) compared in-situ snow measurements over Russia with 20CRv2 for the whole 20th century, and found that it consistently and realistically represents the onset of Eurasian snow cover. However, the authors only investigated the snow dataset in a binary fashion (snow/no snow).

Given the lack of inter-comparison studies of snow depth between reanalyses products, we evaluate snow depth in four centennial state-of-the-art reanalyses. The goal of this study is to assess the consistency between in-situ observations and reanalyses estimation of snow depths. To assess this performance, we focus on early snowfall season (October, November) and early snow melt season (April). This assessment also includes specialized reanalyses for land surface processes, driven by input from the atmosphere.

This article is structured as follows. Section 2 gives an overview of the various datasets analyzed, whereas Section 3 defines the methods used in the comparison. Section 4 presents the results for the evaluation. After discussing the results in Section 5, conclusions are drawn in Section 6.

2. Data

In this study, we use six different climate reanalysis datasets, which can be divided into two families, namely the European Centre for Medium-Range Weather Forecasts (ECMWF) products and the NOAA-CIRES Twentieth Century Reanalysis products. These datasets are compared with Russian in-situ snow depth measurements.

2.1 Reanalysis Datasets

The Twentieth Century Reanalysis Version 2 (20CRv2) dataset allows retrospective 4-dimensional analysis of climate and weather between 1871 and 2012 (Compo et al. 2011). It was achieved by assimilating synoptic observations of surface pressure into the NCEP GFS model using an Ensemble Kalman Filter variant. Prescribed boundary conditions are HadISST1.1 (Rayner et al. 2003) monthly sea-surface temperature (SST) and sea ice cover data as well as forcing of CO_2, volcanic aerosols and solar radiation.

526 **Table 1: Reanalysis product characteristics**

Reanalysis	**Assimilated data**	**Spatial resolution**	**Data assimilation method**	**Type**	**Time Interval**	**Sea ice and SST**
ERA-Interim	Surface, upper air, satellite	T255	4D-Var	Spectral	1979-present	NCEP prescribed
ERA-Interim land	none, HTESSEL land model nudged to ERA-Interim atmosphere	T255	none, HTESSEL land model nudged to ERA-Interim atmosphere	Spectral	1979-present	
ERA-20C	Surface pressure and marine surface winds	T159	4D-var	Spectral	1900-2010	HadISST2
ERA-20C land	none, HTESSEL land model nudged to ERA-20C atmosphere	T159	none, HTESSEL land model nudged to ERA-20C atmosphere	Spectral	1900-2010	
20CRv2	Surface pressure	T62	Ensemble Kalman Filter	Spectral	1871-2012	HadISST1.1
20CRv2c	Surface	T62	Ensemble	Spectral	1851-	COBE-

	pressure		Kalman Filter		2014	SST2

* Here NCEP refers to changing suite of operational sources from National Centers for Environmental Prediction.

The 20th Century Reanalysis Version 2c (20CRv2c) uses the same model as version 2 with new sea ice boundary conditions from the COBE-SST2 (Hirahara et al. 2014), new pentad Simple Ocean Data Assimilation with sparse input (SODAsi.2, Giese et al. 2015) sea surface temperature fields, and additional observations from ISPD version 3.2.9 (Cram et al. 2015). SODAsi2c is generated by tapering SODAsi.2 at 60° N/S to COBE-SST2 SSTs, which makes the Arctic sea ice and SSTs consistent. For both products, we use the mean of the 56-member ensemble, at a 6-hourly temporal resolution. The spatial resolution corresponds to a Gaussian T62 grid.

The ERA-20C (ERA20C) reanalysis (Poli et al. 2016) uses the Integrated Forecast System (IFS) as a framework to assimilate observations of surface pressure and marine surface winds. It is a global atmospheric reanalysis for the period 1900 – 2010 with a 3-hourly temporal resolution and a horizontal resolution of T159 with 91 vertical levels, reaching from the surface up to 1 Pa. Sea – ice cover and SST forcing come from an ensemble of realizations (HadISST.2.0.0.0), where the variability in these realizations is based on the uncertainties in the observational sources used for this forcing. The radiation scheme follows exactly the Climate Model Intercomparison Project (CMIP5) proposal, including aerosols, ozone and greenhouse gases (Hersbach et al. 2015).

In addition to the ERA20C reanalysis, the ERA-20C and ERA-Interim (1979-2015) (Dee et al. 2011) land versions (Balsamo et al. 2015) (ERA20CL & ERA-INTERIM-land) are used in our assessment. These land reanalyses consist of off-line runs of the ECMWF land surface model, driven by the atmospheric forcing from the respective reanalysis. When calculating the correlation and root-mean-square error, both the corrected (with GPCP) and uncorrected version of ERA-INTERIM-land are used (referred to ERAINTL-d and ERAINTL-e, respectively). For spatial plots, we only show the corrected version. ERA20C was analyzed in 0.5° resolution, and ERA-INTERIM-land in 1° resolution. It is important to note that none of the atmospheric or land reanalyses used in this study assimilated snow measurements. Moreover, all products are available on 6-hourly resolution but were used in daily resolution for comparison with stations.

In ERA20C, ERAINTL-d and ERAINTL-e snow is represented as an additional layer on top of the upper soil layer, with independent prognostic thermal and mass contents (Dutra et al. 2010). The snow pack is represented by a single layer with an evolution of snow temperature, snow mass, snow density, snow albedo, and a diagnostic formulation for the snow liquid water content. The snow mass evolves following a water balance equation coupled to the energy budget via snow phase changes. In 20CRv2 and 20CRv2c snow is also represented as an independent layer on top of the soil layer with independent prognostic thermal and mass content (Ek et al. 2003, Koren et al. 1999), but there is no account for liquid water content. The parameterizations used for snow density, albedo and fractional coverage are different in the two snow schemes. These constraints might impact the snow depth evolution since there is no constrain by surface data assimilation. However, there are no major differences between the snow models and their complexity is comparable.

2.2 Snow depth observations

This study uses time series of daily snow depths for 820 Russian meteorological stations (distributed as shown in the supplementary Figure 1). The time series are prepared by RIHMI-WDC (All-Russian Research Institute of Hydrometeorological Information—World Data Centre). Meteorological data sets are automatically checked for quality control. Since the procedure of snow observations changed in the past, particular attention was given to the removal of all possible sources of inhomogeneity in the data. However, there have been no changes in the observation procedures since 1965. Daily observations are measured on three stakes at the weather station, where the average of all three is registered in the time series. When using monthly data, we use the maximum snow depth during that month instead of mean value, because it reflects the process of snow accumulation (snow depth is a cumulative and highly inertial characteristic of climate system). It is especially essential for autumn months when the main processes of snow accumulation occurs over the territories of Russia.

3. Analysis procedure

3.1 Choice of long-term daily snow observations

Out of the over 800 stations, 15 stations were selected with a record extending back to the beginning of the 20th century on a daily basis. Stations with records extending into the 19th century were shortened to start from 1901. All time series end in 2011. Stations with different starting years are indicated in Table 2. Furthermore, Table 2 displays the location of the 15 stations, including the elevation above sea level. To correlate daily measurements with daily reanalysis values, values from the closest grid cell to the station location were chosen. The results therefore include uncertainties concerning the surrounding topography of the stations. Moreover, the relative amount of missing data is shown for the average of all three months. As can be seen, data availability differs considerably between months and stations. However, one station (ID 35108) exceeding 20% missing data in all three months was excluded from further analysis. We also excluded one station (ID 32098) for which the related grid box was classified as ocean. This results in a final selection of 13 stations.

Table 2: 15 long-term snow stations taken out of the Russian snow station data pool. Listed are WMO ID, name, coordinates, elevation as well as starting year and missing values. Missing values are indicated relative to the whole sample size of each individual station as average of April, October and November.

WMO ID	Station Name	Coordinates	Elevation above sea level	Starting year if not 1901	Missing values in %
22550	Arhangel`sk	64°30` N 40°44` E	8		9.6
23405	Ust`-Cil`ma	65°26` N 52°16` E	78	1914	6.3
23711	Troicko-Pecherskoe	62°42` N 56°12` E	135		6.1
24641	Viljujsk	63°47` N 121°37` E	110	1903	17.3
24966	Ust`-Maja	60°23` N 134°27` E	169		16.8
26063	St. Petersburg	59°23` N 30°18` E	3	1902	11.3
27199	Kirov	58°36` N	157		11.7

		49°38` E			
27675	Poreckoe	55°11` N 46°20` E	136		17.5
27955	Samara (Bezencuk)	52°59` N 49°26` E	45	1904	7.5
28275	Tobol`sk	58°09` N 68°15` E	49	1907	19.2
28440	Ekaterinburg	56°50` N 60°38` E	281		3.8
30758	Chita	52°05` N 113°29` E	671	1926	8.9
32098	Poronajsk	49°13` N 143°06` E	7	1908	4.5
35108	Urals (Kazakhstan)	51°15` N 51°17` E	37		25.5
35121	Orenburg	51°41` N 55°06` E	115		8.8

3.2 Calculation of extreme event detection

To evaluate the detection rate of extreme daily snow depth events, we calculate the 98th percentile values in all reanalysis products in two different ways. Extreme events were calculated for both absolute daily snow depth and accumulated daily snow depth, the later being the snow depth difference between two consecutive days. The selected dates in the reanalyses are then compared to the station dates. Based on the number of dates selected using station data, a percentage hit-rate is calculated, namely the amount of extreme events in station data divided by the amount of correctly selected dates in reanalyses. Snow observations were performed at 8 am local time, which is different to any of the available reanalysis output. To allow some margin of error, we also perform this hitrate analysis for ±1 day shift.

4. Results

4.1 Spatial features and magnitude

While quantitative estimates of how the reanalysis products differ from station data will be shown later, we first show multi-decadal climatology and tendency maps for a more qualitative inspection of the snow representation in reanalyses. Starting with the recent period, Figure 1 shows the snow depth climatology over 1981-2010 for April, October and November. Unsurprisingly, April displays the overall highest values.

Highest snow depths over Eurasia are located in northern Siberia along the 90° E meridian. Elevated snow depths are also found over the Russian Far East and over Kamchatka in particular. Both of the features displayed in the station data are also represented by all reanalysis products. Overall, there is a broad agreement in the position of high snow depth areas as well as the snow region boundaries. However, ERA20C shows notably lower snow depths in northern Siberia, compared to ERA-INTERIM-land and 20CRv2c, but the latter shows generally higher snow depth than station data, especially in April and November.

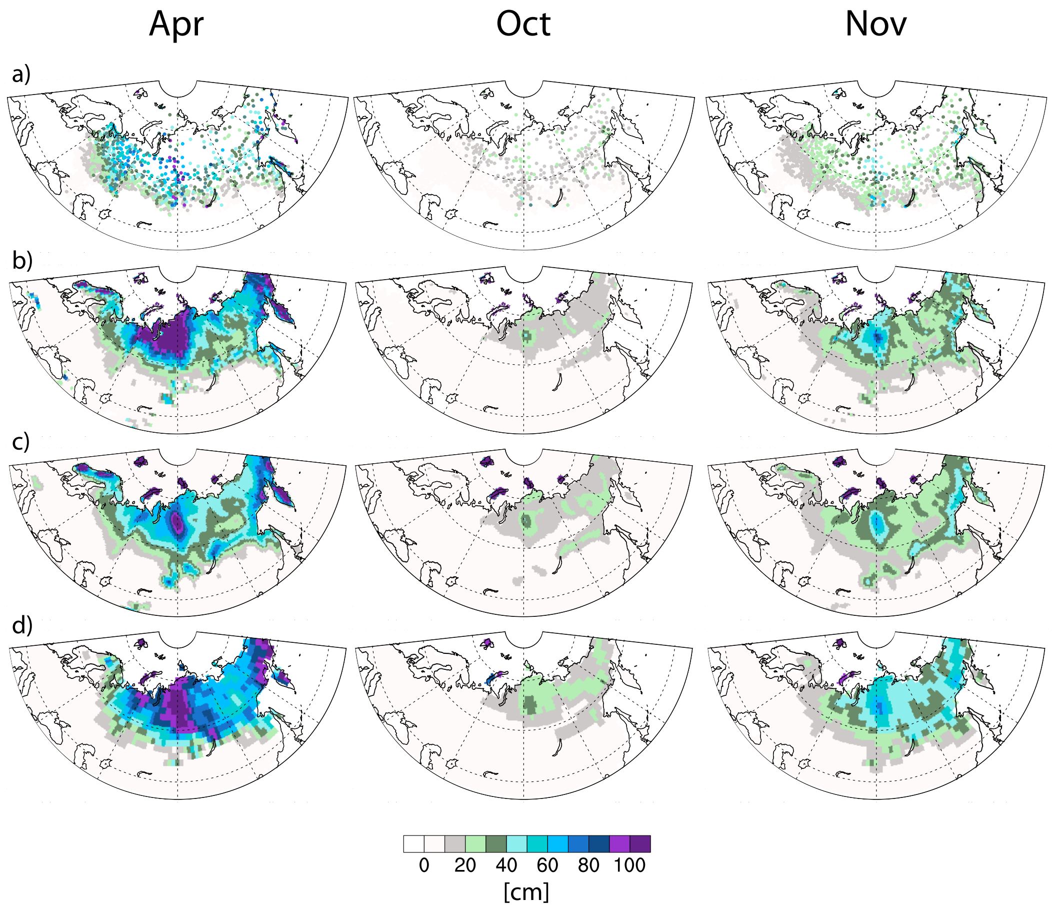

Figure 1: 1981-2010 mean maximum snow depth climatology of (from left to right) April, October and November in a) observations, b) ERA-INTERIM land-d c) ERA20C and d) 20CRv2c. ERA20CL, ERA-INTERIM land-e and 20CRv2 are not displayed due to insubstantial differences to ERA20C, ERA-INTERIM land-d and 20CRv2c.

The decadal tendency in the recent era is shown in Figure 2, as snow depth anomalies between the 1996-2010 period minus those in the 1981-1995 period. In April, the region with strongest snow depth decrease is the western, European part of Russia, west of the Urals and between the Barents and Caspian Sea. This feature is clearly underestimated by all reanalyses, best represented by 20CRv2, followed by ERAINT-l. However, the sign of the tendency is not homogenous over the region in the reanalyses, and local snow depth increases can be found. A second region of snow decrease, which is broadly captured by the reanalyses is the Russian Far East, with ERA20C displaying poorer agreement. A pronounced positive anomaly is found in reanalyses north of Lake Balkhash and extending toward the coasts of the Bara and Laptev Seas, a region where the station coverage is poor though. Towards southern Russia, the observed signal is more complex with snow depth increase towards the border to Kazakhstan, but with snow depth decrease further east on the western side of Lake Baikal, which the gridded products fail to capture, both in terms of extend and magnitude. In autumn, and especially in November, the in-situ data reveal a broad longitudinal dipolar pattern with decrease (increase) of snow depths in the eastern (western) part of Russia, reproduced by the reanalyses.

Overall, 20CRv2c captures the observed patterns slightly better than ERA-Interim-land, while ERA20C shows the poorest agreement.

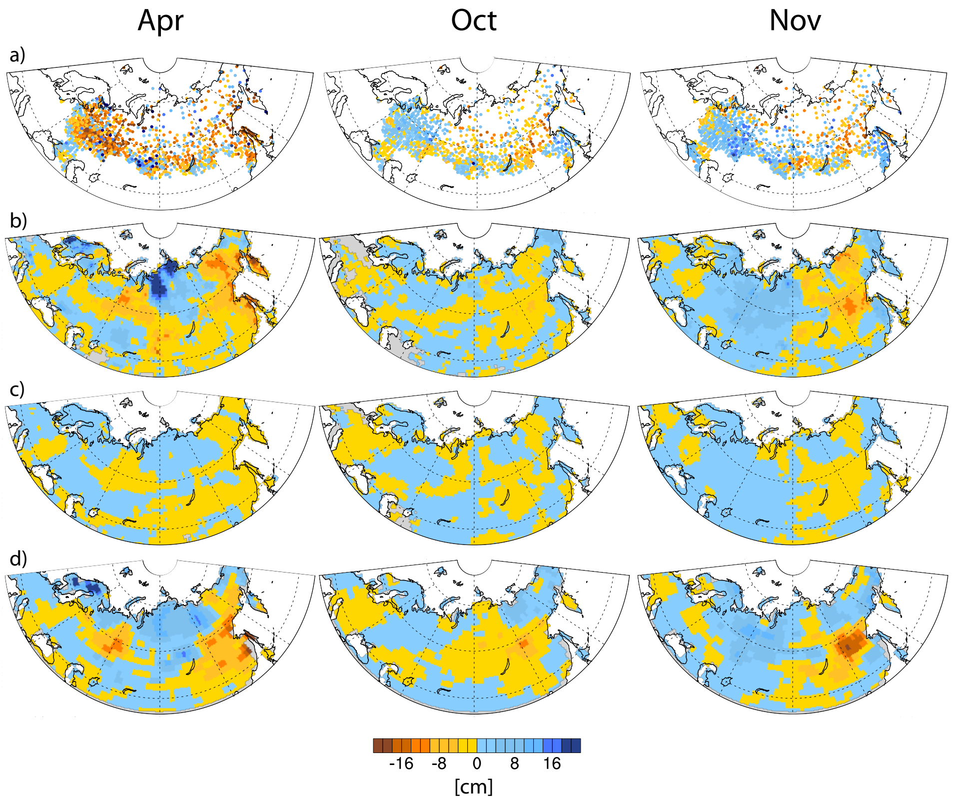

Figure 2: 1996-2010 minus 1981-1995 snow depth anomalies of (from left to right) April, October and November in a) observations, b) ERA-INTERIM land--**d**, c) ERA20C and d) 20CRv2c. ERA20CL, ERA-INTERIM land-e and 20CRv2 are not displayed due to insubstantial differences to ERA20C, ERA-INTERIM land-d and 20CRv2c.

4.2 Inter-decadal performance

Figure 3 shows the long-term decadal changes over the Northern Russia snowpack (averaging between 50°-150° E and 60°-75° N) in the different climate reanalyses, the region of highest snow depths in the selected months. Series of 30-year climatological anomalies were computed with a moving window of 10 years, using 1981-2010 period as a reference climatology. From the 1941-1970 period onward, all four products show similar tendencies. Further back in time however, the gridded products diverge: ERA20C & ERA20CL continue a downward tendency (mean anomalies decrease) whereas the 20CRv2 & 20CRv2c reanalyses show an overall increase in snow depth, resulting in a notable difference by the early 20th century. This evolution is, despite minor differences, true for all three months. For all months, the 20CR family of reanalyses show strong positive anomalies for the 1911-1940 period, the main period of the Early Twenty Century Arctic Warming (ETCAW).

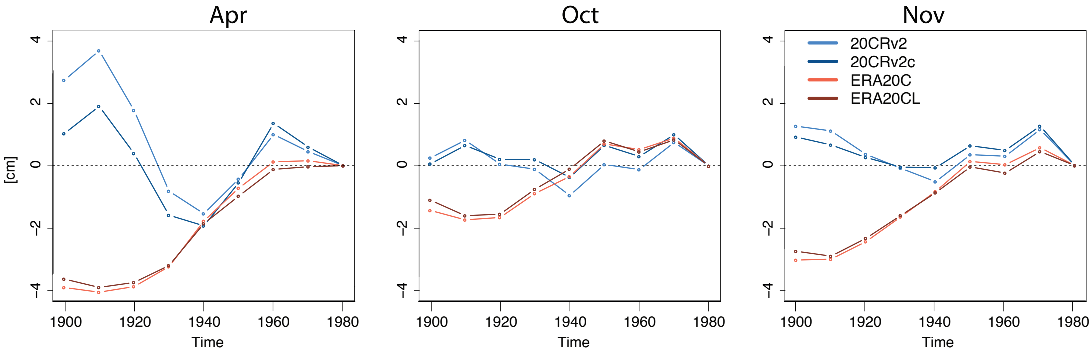

Figure 3: Time series of snow depth anomalies in (from left to right) April, October and November averaged over the main northern Russia snow pack (50°-150° E, 60-75° N). Each data point represents a 30-year long climatology, starting from 1901-1930 until 1981-2010 with 10 year shifts. Anomalies are calculated relative to the 1981-2010 climatology.

Unfortunately, none of the 13 selected stations with a long record is located in that northern Russia region. A similar behavior emerges however if the comparison is made between the 13 stations and the collocated reanalysis data, as shown on Figure 4. Again, comparing to the 1981-2010 reference climatology disregards differences in snow depth magnitude and helps focusing on long-term tendencies. All three months show a divergence of the two reanalysis families towards the beginning of the 20th century. Going backward in time from the recent era, tendencies are similar until the 1941-1970 period but, afterwards, the ECMWF reanalyses show a declining mean snow depth whereas the 20CR reanalyses favor an increase in snow depth. Interestingly, snow station data agrees very well with the 20CR reanalyses until the 1951-1980 climate for all three months. In comparison, the ECMWF reanalyses show much more pronounced deviations from the station data anomalies. Towards the beginning of the century, the station data agrees more and more with the ECWMF reanalyses in autumn. The ECMWF reanalyses achieve an excellent representation for the 1901-1930 and 1911-1940 periods in autumn (for the 1901-1930 spatial anomalies see Supplementary Figure 2). This however is not the case for April, where 20CRv2 data is closest to in-situ observations.

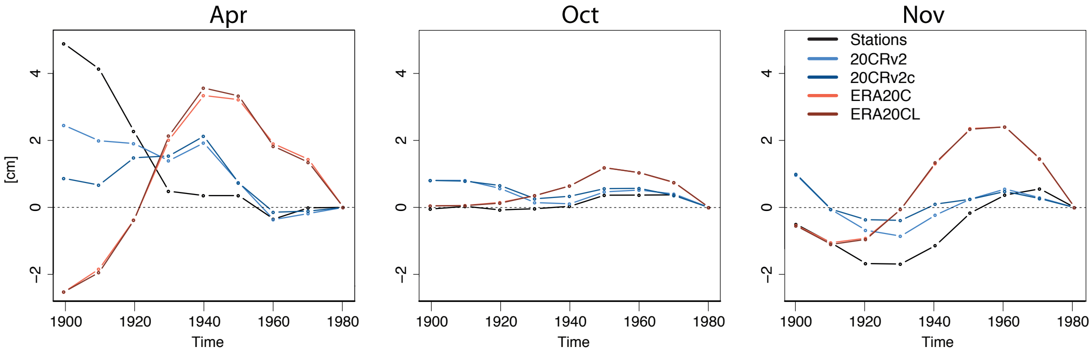

Figure 4: Top: Time series of snow depth anomalies in (from left to right) April, October and November for the average of the 13 station locations. Each data point represents a 30-year long climatology, starting from 1901-1930 till 1981-2010 with 10 year shifts. Anomalies are calculated relative to the 1981-2010 climatology.

4.3 Sub-decadal and daily performance

Moving away from decadal tendencies, we now evaluate the daily and the inter-annual snow variability over the 13 selected stations with records extending back to the early days of the 20th century. Figure 5 presents the daily performance between station data and the reanalyses over the recent period (1981-2010).

The melting season (April) generally exhibits the weakest correlation between grid and station, with slightly better values for October and highest values for November. However, this ranking can differ for individual station locations. For the period 1981-2010, the ERA20C reanalysis achieves better results than the 20CR reanalyses, especially so in April, indicating that melting and temperature evolution is somewhat more accurate in the ECMWF reanalyses. November and even more so October correlations are very similar in all four long-term reanalysis products. As to be expected, the ERA-INTERIM-land reanalysis, given the higher quality of atmospheric forcing in the recent era and the finer spatial resolution, generally scores the highest when compared to the respective station with medians above 0.8 in all three months. Note that in the correlation analysis ERA-INTERIM-land-d achieves higher averaged correlation coefficients than the uncorrected version.

Looking at long-term correlations (Figure 6), the ECMWF reanalyses slightly outperform the 20CR in April, but less so than in the 1981-2010 period. The opposite is now true for October, where the 20CRv2 and 20CRv2c achieve slightly higher averaged correlation coefficient values, whereas in November, all long-term reanalyses have comparable correlations with station data with slightly higher values for the 20CR family. In two out of three months, the ERA20C-land version does not realize higher accuracy than the parent product ERA20C. The same is true for the new 20CRv2c, which outperforms 20CRv2 only in November.

We note that long-term daily correlation coefficients for individual northern stations repeatedly exceed 0.7 (see Supplement Table 1). Only two stations (ID 30758 & ID 35121) consistently show very low correlations across the seasons and reanalyses, probably because of their southern positions. In general terms, the linear correlation performance decreases from northern to more southern stations. This reflects the sensitivity of snowfall in relatively mild environments, resulting in short periods of snow availability. Such small-scale snowfall events are hardly captured by the reanalyses.

.

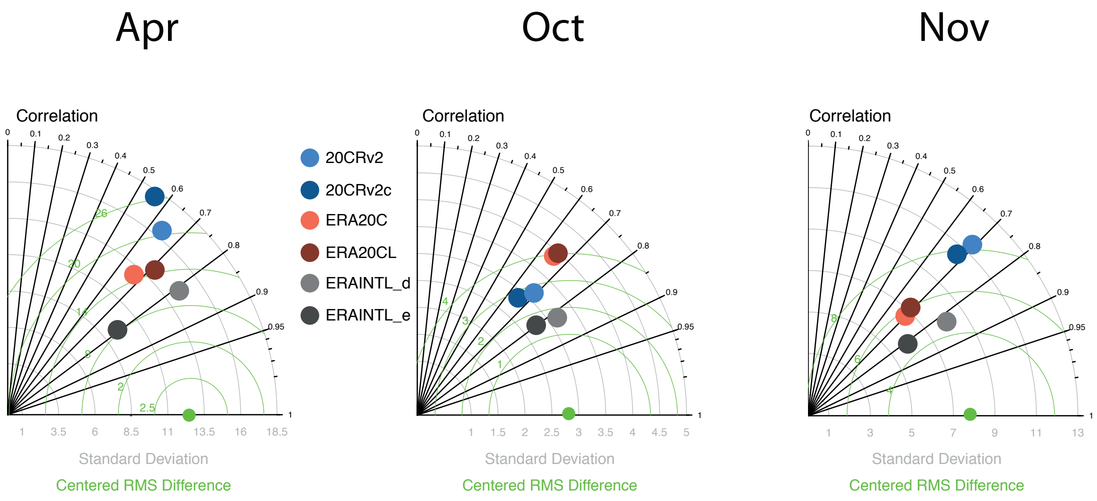

Figure 5: Taylor diagrams showing the median of the 13 station locations using daily data for the period 1981-2010. The X-axis and Y-Axis indicate the standard deviation, the radians indicate correlation values and the green circles indicate centered RMSE. The green dot shows the observed variability. For more details concerning the datasets statistics, see Supplementary Figures 3-6.

Root mean square error (RMSE) values obviously differ from location to location (see supplement Table 1). Averaging over all stations reanalyses products were found to produce the absolute largest deviations from the *true* station timeseries in April, followed by November and lastly October. The low October RMSE is influenced by the relatively small absolute snow depth values during that month. Thus, even deviations from zero (e.g. incorrect event of snowfall) will be small. Again, as expected the ERA-INTERIM land produces the smallest RMSE over all reanalyses. The ERA-INTERIM land version without the precipitation correction has lower RMSE in April and November than the version with the precipitation correction. This could be due to the scarcity and uncertainty of rain-gauge observations in the region, which would deteriorate the GPCP-based correction. The pair of ERA20C reanalyses clearly outperforms the 20CR pair in April and November, but is on equal terms in October.

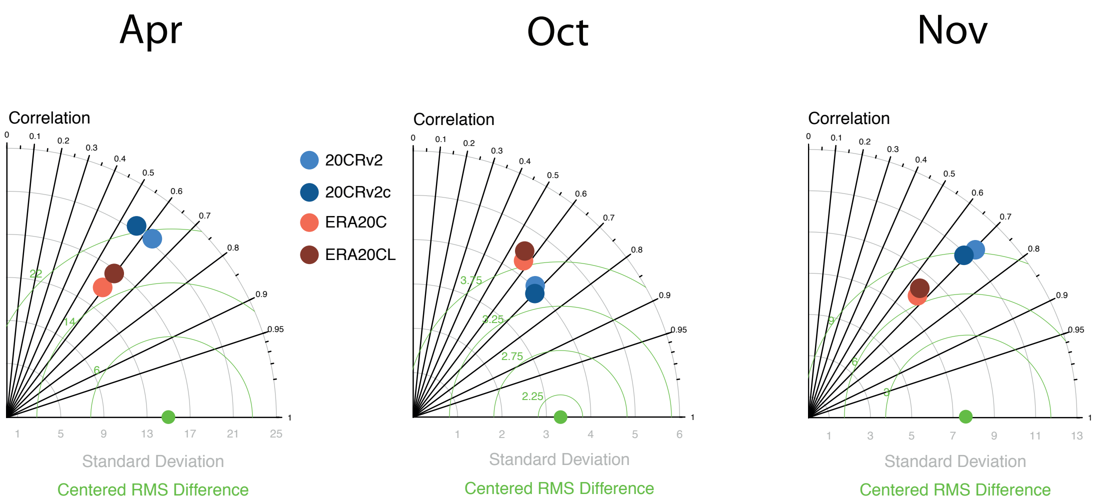

Figure 6: Taylor diagrams showing the median of the 13 station locations using daily data for the longest period available (see Table 1). The X-axis and Y-Axis indicate the standard deviation, the radians indicate correlation values and the green circles indicate centered RMSE. The green dot shows the observed variability. For more details concerning the datasets statistics, see Supplementary Figures 3-6.

Finally, to address variability characteristics of the reanalysed snow depth values, Figure 5&6 (X-axis) also show the median standard deviation of anomaly time series averaged over the 13 stations. As expected, April and November show much higher variability than October. All ECMWF products show a good representation of the station standard deviation. The uncorrected ERA-INTERIM land version apparently suppresses a certain amount of variability with lower median values than the rest of the ECMWF family products. On the other side, both 20CR reanalyses overestimate the variability. October values for 20CRv2 and 20CRv2c are very much influenced by one outlier location, so that the median is still well within the range of the station median.

Assessment of variability is especially important in the framework of extreme events. Since the replication of variability and daily correlation seems promising, an extreme event hit-rate is computed to measure how well the reanalysis products can detect the exact dates of extreme events. Figure 7a shows the hit-rate of days with extreme absolute snow depth values whereas Figure 7b shows the hit-rate of days with extreme accumulation of snow depth for the 13 station locations. Since in-situ data snow depth and snow depth in reanalyses are not exactly measured at the same time, we allow the reanalysis to be off by ± 1 day. Better daily correlations in April (Fig. 5) seem to help the ERA20C reanalyses to capture slightly more dates correctly than the two 20CR products. The opposite is true for autumn months, especially for absolute snow depth maxima. Interestingly, changing from absolute to accumulation extremes helps ERA20C to achieve a higher hit-rate, whereas the 20CR products show a slightly worse hit-rate for the latter metric. Moreover, ERA20C land, which shows a very similar if not better performance for absolute snow depth extremes, shows a slightly poorer performance for detecting accumulation extremes. Overall though, mean hit-rates stay well below 50%; only for single locations did the hit-rates exceed this threshold. If we remove flexibility to be off by one day, the amount of correct hits is reduced even further (over all by ca. 10%, no shown)

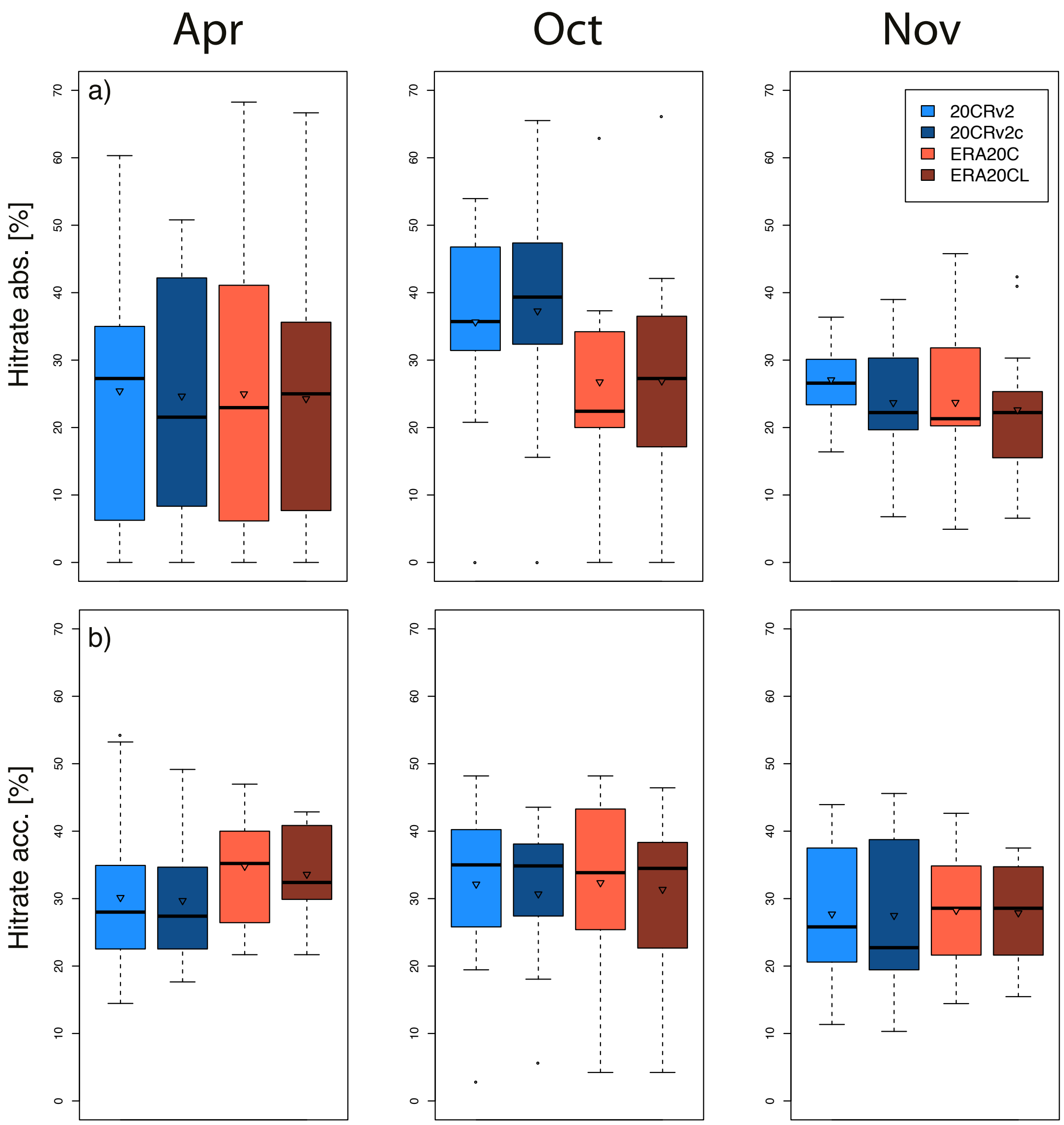

Figure 7: Boxplots graphs for the extreme events hitrate analysis of the 13 snow depth station locations, where the triangle denotes the mean, the bold black line denotes the median, the box denotes the 25-75% percentile range (or interquartile range), the whiskers show the upper and lower end or at most the 1.5 x interquartile range and the dots denote outlier. a) shows boxplots for absolute snow extreme events the longest possible time period, b) same as a) but for snow accumulation. Hitrates are computed for the longest period possible.

5. Discussion

Comparing snow depths in multiple long-term, centennial reanalyses with in-situ measurements over Russia, our results indicate ambivalent performances of the reanalysis products. Climatologies are well represented spatially, but overestimate the mean snow depth in most parts of the analyzed domain. Long-term daily correlations revealed decent coefficient values for most of the station locations. Snow depths from surface input-only reanalyses consistently show linear correlations of 0.6 and higher, although dealing with fluctuating daily data, including rapid changes in weather patterns. Moreover, due to spatial averaging and shortcomings in model topography relatively low correlation coefficients are expected. Khan et al. 2008 found best case basin-wide correlations of around 0.65 in ERA-40 and JRA-25, with much worse correlations for the NCEP-DOE reanalysis. All these reanalyses assimilated a variety of input data, not only surface data as is the case with the centennial reanalyses examined in this study. We found that reanalyses with less assimilated data do perform equally or better for a substantially longer time period.

Moreover, Khan et al. (2008) state that all evaluated reanalysis snow products showed the worst matching in April. The same result was found in our analysis, where April values showed the smallest correlation and highest absolute error (RMSE). Therefore, it can be assumed that models used for creating the reanalysis datasets still struggle with properly representing melting season (Slater et al. 2001). Looking at the RMSE, it could be shown that the 20CRv2 & 20CRv2c generally overestimate snow depth, and that ERA20C & ERA20CL are closer to the station data. This is true for the recent past, as for the centennial analysis. The same applies to the variability comparison. Interestingly, the snow depth RMSE in October is smaller than in the other months, but day-to-day variability (correlation) appears to be better in November. This indicates that the initial snowfall in October, if occurring, is harder to capture than in November, but also generates only small snow depths. Therefore, even if completely missed by the reanalysis, it produced only small RMSEs.

Peings et al. (2013) found that 20CRv2 displays a good performance in detecting the daily advance of October and November snow (between 80-100% hitrate). We found that 20CRv2 shows good long-term daily correlations in October and November, even higher than ERA20C. That said, binary snow information as well as correlation analysis masks the details of snow amount, which is better seen in anomaly or climatology maps. Moreover, our hit-rate analysis of dates for extreme snow depths and snow accumulation showed that for the 13 station locations only about 45% of the dates were correctly computed when compared to station data. Among the explanations for this underwhelming performance are a) the assimilation of only surface data in the reanalyses (which challenges the computation of the complex conditions for extreme snowfall), b) the long time frame in which assimilated data quantity is decreasing back in time and c) spatial resolution of the reanalyses which can not resolve features like small scale uplift or orographic precipitation, or at even smaller scale, snowdrift. With these deficiencies in mind, the achieved correlation coefficients for the centennial timeseries are even more remarkable.

However, analysis of inter-decadal tendencies of snow depth revealed a peculiar evolution, even though snow schemes and assimilated data are comparable. Generally, the ECMWF datasets compute a stronger snow depth decrease before the 1940s than the 20CR products for the main Russian Arctic snow field. Since climatological maps do not show substantial differences, origin of the large disagreements must emerge in the pre-1950s period. The assimilated input data is near identical between ERA20C and 20CRv2c, and thus model biases seem to be the source of divergence.

One reason for the snow depth evolution could be the overestimation of Arctic SLP (sea level pressure) during the pre-1950s in ERA20C (Belleflamme et al. 2015). Indeed we found that ERA20C shows high (higher than 20CR or reconstructed values) positive SLP anomalies for the beginning of the 20th century over Central Russia (see Supplementary Figure 7) together with a peculiar increase of atmospheric mass towards the beginning of the 20th century (not shown). Such a high pressure anomaly over the high latitudes might lead to reduced poleward moisture transport, as well as decreased cloud cover and downward long wave radiation, which is very efficient in melting snow. Moreover, stable atmospheric conditions prevent vertical motion and therefore condensation. Knudsen et al. (2015) showed that, in the recent era, Arctic anti-cyclonic circulation patterns also promote low snowfall in summer over the Russian sector of the Arctic, and a similar association with (too) high pressure could be at play in ERA20C in the pre-1950s. On the other hand, if compared to station data, the ERA20C snow depths show a good agreement for anomalies early in the 20th century.

Furthermore, near-surface temperatures influence snow depth evolution. The new 20CRv2c dataset uses alternative sea ice and SSTs representations as boundary conditions, which improves the 2m temperature performance over the Arctic compared to 20CRv2. Nevertheless, it is generally still colder than ERA20C or CRUTEMP4.4 (Jones et al. 2014). However, ERA20C is most probably much too warm during April, whereas the 20CR reanalyses seem to be too cold during November and December, thus they might be overestimating snow depths (see Supplementary Figures 8 and 9). Ultimately, there is no clear and simple answer to this issue and our analysis can only provide an initial assessment of the discrepancy between the two families of reanalyses.

The results of the snow climatologies hint towards heterogeneous dataset issues. Decadal tendencies in the second half of the 20th century are better represented by the 20CR datasets (relative to their baseline), whereas tendencies for the first half of the century are better represented in ERA20C. Unfortunately, only 13 stations could be used to verify long-term evolution in snow depth. Data recovery from a higher density network with better spatial coverage is needed to really constrain the diverging snow states in these long-term reanalyses. Moreover, future reanalysis or model comparisons might be needed. The CERA (ERA20C plus coupled ocean) and GSWP3 could give further insight into this topic. Model inter-comparisons concerning snow representation might reveal necessary qualities to compute a realistic snow depth.

6. Conclusion

Snow depth and its evolution from a variety of centennial reanalyses have been tested against in-situ observations over the Russian territory. Long-term reanalyses are able to reproduce daily and sub-decadal snow depth variability very well however generally overestimate snow depths. Moreover, computing the exact day of extreme snow accumulation is still a difficult task for these datasets. Spatially, the region of high and low snow, and the snow cover boundaries are well represented. However, inter-decadal comparison of snow depth revealed some issues with pre-1950s snow climates over northern Russia. The ECMWF and NOAA reanalyses show diverging snow states (low or high, respectively), most probably likely a consequence of assimilation schemes or model biases rather than input data.

To further understand and quantify changes during the current and future Arctic warm periods, it is imperative to maintain and expand a dense network of (Arctic) snow measuring stations (including their meta data). Reproducing observed snow (depth) in climate models is a difficult challenge since many environmental factors determine snowfall amount and ultimately snow depth. In-situ snow depth measurements and reanalyses are important tools to evaluate the performance of climate models.

Acknowledgments. YO was supported by the Norwegian Research Council (project SNOWGLACE # 244166 and EPOCASA #229774/E10). AS and SB were supported by the EU-FP7 project ERA-CLIM2 (607029). MW, YO, SB, AS and OB acknowledge funding by the European ERAnet.RUS programme, especially within the project ACPCA. MW also benefitted from the ARCTIC-ERA project funded by Agence Nationale de la Recherche (ANR) through the Belmont Fund initiative.

References

Agrawala, S., 2007: Climate change in the European Alps: adapting winter tourism and natural hazards management. Organisation for Economic Cooperation and Development (OECD).

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

Barnett, T. P., L. Dümenil, U. Schlese, and E. Roeckner, 1988: The effect of Eurasian snow cover on global climate. *Science*, **239**, 504–507

Belleflamme, A., X. Fettweis, and M. Erpicum, 2015: Recent summer Arctic atmospheric circulation anomalies in a historical perspective. *The Cryosphere*, **9**, 53–64.

Brown, R. D. and P. W. Mote, 2009: The response of Northern Hemisphere snow cover to a changing climate*. *Journal of Climate*, **22**, 2124–2145.

Brown, R., C. Derksen, and L. Wang, 2010: A multidata set analysis of variability and change in Arctic spring snow cover extent, 1967–2008. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **115**, D16111.

Brown, R. D. and D. A. Robinson, 2011: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty. *The Cryosphere*, **5**, 219–229.

Brun, E., V. Vionnet, A. Boone, B. Decharme, Y. Peings, R. Valette, F. Karbou, and S. Morin 2013: Simulation of Northern Eurasian Local Snow Depth, Mass, and Density Using a Detailed Snowpack Model and Meteorological Reanalyses, *J. Hydrometeorol.*, **14** , 203–219, doi:10.1175/JHM-D-12-012.1.

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

Callaghan, T. V., M. Johansson, T. D. Prowse, M. S. Olsen, and L. O. Reiersen, 2011: Arctic cryosphere: changes and impacts. *Ambio*, **40**, 3–5

Cohen, J. and D. Entekhabi, 1999: Eurasian snow cover variability and Northern Hemisphere climate predictability. *Geophysical Research Letters*, **26**, 345–348.

Cohen, J. and D. Rind, 1991: The effect of snow cover on the climate. *Journal of Climate*, **4**, 689–706

Cohen, J., J. A. Screen, J. C. Furtado, M. Barlow, D. Whittleston, D. Coumou, J. Francis, K. Dethloff, D. Entekhabi, and J. Overland, 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nature Geoscience*, **7**, 627–637.

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P. and S. Brönnimann, 2011: The Twentieth Century Reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, **137**, 1–28.

Cram, T. A., G. P. Compo, X. Yin, R. J. Allan, C. McColl, R. S. Vose, J. S. Whitaker, N. Matsui, L. Ashcroft, R. Auchmann, et al., 2015: The international surface pressure databank version 2. *Geoscience Data Journal*, **2**, 31–46.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, et al., 2011: The ERA–interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**, 553–597.

Dutra, E., G. Balsamo, P. Viterbo, P.M.A. Miranda, A. Beljaars, C. Schar, and K. Elder 2010: An improved snow scheme for the ECMWF Land Surface Model:

Description and offline validation. *Journal of Hydrometeorology*, **11,** 899-916

Ek, M.B., K.E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno and J.D. Tarpley 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research: Atmospheres*, **108**, 8851

Flanner, M. G., K. M. Shell, M. Barlage, D. K. Perovich, and M. A. Tschudi, 2011: Radiative forcing and albedo feedback from the Northern Hemisphere cryosphere between 1979 and 2008. *Nature Geoscience*, **4**, 151–155.

Frei, A. and G. Gong, 2005: Decadal to century scale trends in North American snow extent in coupled atmosphere-ocean general circulation models. *Geophysical Research Letters*, **32**, L18502.

Frei, A., M. Tedesco, S. Lee, J. Foster, D. K. Hall, R. Kelly, and D. A. Robinson, 2012: A review of global satellite-derived snow products. *Advances in Space Research*, **50**, 1007–1029.

Ghatak, D., C. Deser, A. Frei, G. Gong, A. Phillips, D. A. Robinson, and J. Stroeve, 2012: Simulated Siberian snow cover response to observed Arctic sea ice loss, 1979–2008. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **117**, D23108.

Giese, B. S., H. F. Seidel, G. P. Compo, and P. D. Sardeshmukh, 2015: An ensemble of historical ocean reanalyses with sparse observational input. *Journal of Geophysical Research: Oceans*, **submitted**.

Hersbach, H., C. Peubey, A. Simmons, P. Berrisford, P. Poli, and D. Dee, 2015: ERA-20CM: a twentieth-century atmospheric model ensemble. *Quarterly Journal of the Royal Meteorological Society*.

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

Honda, M., J. Inoue, and S. Yamane, 2009: Influence of low Arctic sea-ice minima on anomalously cold Eurasian winters. *Geophysical Research Letters*, **36**, L08707.

Hüsler, F., T. Jonas, M. Riffler, and J. P. Musial, 2014: A satellite-based snow cover climatology (1985-2011) for the European Alps derived from AVHRR data. *The Cryosphere*, **8**, 73.

Jeong, J.H., H.W. Linderholm, S-H. Woo, C. Folland , B-M. Kim, S-J. Kim and D.

Chen 2013: Impact of snow initialization on subseasonal forecasts of surface air temperature for the cold season. *Journal of Climate*, **26**, 1956-1972.

Jonas, T., C. Rixen, M. Sturm, and V. Stoeckli 2008: How alpine plant growth is linked to snow cover and climate variability. *Journal of Geophysical Research: Biogeosciences (2005–2012)*, **113**, G03013.

Jones, P.D., D.H. Lister, T.J. Osborn, C. Harpham, M. Salmon and C.P Morice 2014: Hemispheric and large-scale land surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research*, **117**, D05127

Kanamitsu, M., Ebisuzaki, W., Woollen, J. and Shi-Keng, Y., 2002. Ncep-doe amip-ii reanalysis (r-2). *Bulletin of the American Meteorological Society*, **83**, 1631.

Khan, V., L. Holko, K. Rubinstein, and M. Breiling, 2008: Snow cover characteristics over the main Russian river basins as represented by reanalyses and measured data. *Journal of Applied Meteorology and Climatology*, **47**, 1819–1833.

Knudsen, E. M., Y. J. Orsolini, T. Furevik, and K. I. Hodges, 2015: Observed anomalous atmospheric patterns in summers of unusual arctic sea ice melt. *Journal of Geophysical Research: Atmospheres*, **120**, 2595–2611.

Koren, V., J. Schaake, K. Mitchell, Q.Y. Duan, F. Chen and J.M. Baaker 1999: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research: Atmospheres*, **104**, 19569-19585

Liston, G. E. and C. A. Hiemstra, 2011: The changing cryosphere: Pan-Arctic snow trends (1979-2009). *Journal of Climate*, **24**, 5691–5712.

Onogi, K., J. Tsutsui, H. Koide, M. Sakamoto, S. Kobayashi, H. Hatsushika, T. Matsumoto, N. Yamazaki, H. Kamahori, and K. Takahashi, 2007: The JRA-25 re-analysis. *気象集誌. 第 2 輯*, **85**, 369–432.

Orsolini, Y. J. and N. G. Kvamstø, 2009: Role of Eurasian snow cover in wintertime circulation: Decadal simulations forced with satellite observations. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **114**, D19108.

Orsolini, Y. J., R. Senan, G. Balsamo, F. J. Doblas-Reyes, F. Vitart, A. Weisheimer, A. Carrasco, and R. E. Benestad, 2013: Impact of snow initialization on subseasonal forecasts. *Climate Dynamics*, **41**, 1969–1982.

Peings, Y., E. Brun, V. Mauvais, and H. Douville, 2013: How stationary is the relationship between Siberian snow and Arctic Oscillation over the 20th century? *Geophysical Research Letters*, **40**, 183–188.

Peñuelas, J., T. Rutishauser, and I. Filella, 2009: Phenology feedbacks on climate change. *Science*, **324**, 887.

Poli, P., H. Hersbach, D. P. Dee, A. J. Simmons, F. Vitart, P. Laloyaux, D. G. H. Tan, C. Peubey, J.-N. Thépaut, Y. Trémolet, E. V. Hólm, M. Bonavita, L. Isaksen, and M. Fisher 2016: ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, **29**, 4083-4097.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **108**, 4407.

Rienecker, M. M., M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D. Schubert, L. Takacs, and G.-K. Kim, 2011: MERRA: NASA’s modern-era retrospective analysis for research and applications. *Journal of Climate*, **24**, 3624–3648.

Siljamo, N. and O. Hyvärinen, 2011: New geostationary satellite-based snow-cover algorithm. *Journal of Applied Meteorology and Climatology*, **50**, 1275–1290.

Slater, A. G., C. A. Schlosser, C. Desborough, A. Pitman, A. Henderson-Sellers, A. Robock, K. Y. Vinnikov, J. Entin, K. Mitchell, F. Chen, et al., 2001: The representation of snow in land surface schemes: Results from pilps 2 (d). *Journal of Hydrometeorology*, **2**, 7–25.

Uppala, S. M., P. W. Kållberg, A. J. Simmons, U. Andrae, V. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, and G. A. Kelly, 2005: The ERA-40 reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2961–3012.

Wegmann, M., Y. Orsolini, M. Vázquez, L. Gimeno, R. Nieto, O. Bulygina, R. Jaiser, D. Handorf, A. Rinke, and K. Dethloff, 2015: Arctic moisture source for Eurasian snow cover variations in autumn. *Environmental Research Letters*, **10**, 054015.

Wu, R., G. Liu, and Z. Ping, 2014: Contrasting Eurasian spring and summer climate anomalies associated with western and eastern Eurasian spring snow cover changes. *Journal of Geophysical Research: Atmospheres*, **119**, 7410–7424.

Yasunari, T., Kitoh, A. and Tokioka, T., 1991: Local and remote responses to excessive snow mass over Eurasia appearing in the northern spring and summer climate—a study with the MRI GCM. *J. Meteor. Soc. Japan*, **4**, 473-487.

Ye, K., R. Wu, and Y. Liu, 2015: Interdecadal change of Eurasian snow, surface temperature, and atmospheric circulation in the late 1980s. *Journal of Geophysical Research: Atmospheres*, **120**, 2738–2753.

Zhang, X., J. He, J. Zhang, I. Polyakov, R. u. d. Gerdes, J. Inoue, and P. Wu, 2013: Enhanced poleward moisture transport and amplified northern high-latitude wetting trend. *Nature Climate Change*, **3**, 47–51.

Zuo, Z., S. Yang, R. Zhang, D. Xiao, D. Guo, and L. Ma, 2015: Response of summer rainfall over China to spring snow anomalies over Siberia in the NCEP CFSv2 reforecast. *Quarterly Journal of the Royal Meteorological Society*, **141**, 939–944.