Reviewer 1

This is a unique temporal data collection of aufeis data in the Indigirka river basin, Russia. Aufeis or naleds deposits are thick accumulations of ice that form during winter along stream and river valleys in arctic and subarctic regions impacting hydrology and geomorphology of these regions. The authors compiled and standardized historical data on aufeis deposits in the Eastern Siberian Indigirka river basin from a historical Russian National cadastre complementing data using historical topographical maps and added a new data set on aufeis derived from Normalized Differential Snow Index (NDSI) index calculation using Landsat 8 OLI sensor data. The authors cross-referenced the historical and the present-day data collection. The data collection is organised as a Geographic Information System GIS data base including data on location, area coverage, elevation, time stamp, source of data in form of attribute tables and the aufeis objects in the data format of GIS point and polygonal vector layers. The Indigirka aufeis catalogue is published on PANGAEA in the form of a GIS data base with a helpful and detailed read-me description of the attribute tables. The data collection will be of interest to hydrologists, climatologists, geomorphologists, cryologists and social science. The authors document in the manuscript the generation of the historical and the modern date data sets and its meta data characteristics. The authors also discuss the validity of data, the cross referencing between historical and nowadays aufeis deposits and reasons for mismatches in areal coverage and locations and possible changes due to climate.

Comment: The paper is in general clearly written with many details provided. However, the article including the title, the PANGAEA data publication including title, abstract and the metadata description need to be carefully edited for English before acceptance of the paper. The data compilation process and metadata is not thoroughly and clear enough shown and explained and the GIS data require further standardization and optimization to make them reusable.

Technical issues, GIS data: 1) the GIS shape files contain different projections: The GIS data catalogue is published in PANGAEA as an ARCGIS project data base. The downloaded data base is user friendly readible and usable using the proprietary GIS software ARCGIS. ARCGIS licenses are costly and many user groups may use open source GIS or other geodata software packages. Using ARCGIS software the shape-files are automatically but only virtually brought to the same projection. The GIS shape files are also readable and reusable using open source geodata software packages – however the 2 data collections have different projections (the aufeis kadastr shape file contains the projection "Asia_North_Lambert_Conformal_Conic" the aufeis Landsat shape file not). This requires users of these datasets who are using free software packages to reproject the shape files to a common projection prior to being able to use the data sets together. Please standardise the shapefiles using one projection

Response: We prepared the data according to the comments. The GIS database contains the data of aufeis in two forms: ArcGIS 10.1/10.2 and Qgis 3* projects. All data and projects have WGS 1984 coordinate system (without projection). ArcGIS and Qgis projects contain two layers, such as Aufeis_kadastr (historical aufeis data collection, point objects) and Aufeis_Landsat (satellite-derived aufeis data collection, polygon objects).

Comment: 2) the GIS attribute files do not contain self-explanatory attribute names: The Indigirka aufeis data collection is a highly valuable data set, specifically also because the authors are using cross reference indices to link the data sets. This needs to be made more clear in the naming and cross-referencing of the attribute names. E.g., the cross reference index should be also named accordingly, e.g. as cross index similarly in both attribute tables, not named ID in the aufeis_Landsat data set and named PolygonID in the aufeis kadastr data set. Naming of similar attributes should be standardized between the data sets, e.g. the attribute area in sqkm. Suggestions on attribute naming is attached as supplement. Please consider to change attribute names to more self-explanatory names.

Response: We followed the suggestions on enhancing attribute naming as much as possible. Though due to the limited length of the name we could not do it in all namings. See the Tables 1 and 2 in the paper. The PANGAEA database is updated accordingly.

Comment: The data set can also be uploaded in Google Earth with visualization of the data objects and the metadata and will be by this very easy re-usable if attribute naming and cross-referencing between the 2 data sets will be made as self-explanatory as possible.

Response: We uploaded the database into Google Earth and added the files to PANGAEA database. Additionally the watershed borders which are mentioned in the analysis in the paper added in Google earth format.

Comment: 3) consistency of published GIS data with manuscript content: Authors show in the manuscript assessments of both data sets – cadastre derived and satellite derived related to elevation. The attribute elevation is however missing in the attribute table of aufeis_Landsat. Consider to add information on elevation into the attribute table of the aufeis Landsat data set.

Response: The attribute Elevation is added to Landsat data set (See also Table 2 in the paper).

Comment: Issues, data publication on PANGAEA: Title: aufeis is the plural form of aufeis, the plural form aufeises does not exist.

Response: We changed the title of the database to “Aufeis (naleds) of the North-East of Russia: GIS catalogue for the Indigirka River basin (Russia)”

Comment: Abstract: The abstract should be extended to contain more technical information on the data. Authors should inform the users that the data download will consist of a complete ARCGIS project containing 2 different feature GIS shape files with historical and the nowadays aufeis data collection. The authors can add short information in the abstract on how the data were generated. Very useful for future users of the GIS data is to provide in the abstract text information on the projection of the GIS data collection – this is sometimes handy for reading data in in some open source geodata software packages.

Response: We extended the abstract as the following.

The GIS database contains the data of aufeis (naleds) in the Indigirka River basin (Russia) from historical and nowadays sources, and complete ArcGIS 10.1/10.2 and Qgis 3* projects to view and analyze the data. All data and projects have WGS 1984 coordinate system (without projection). ArcGIS and Qgis projects contain two layers, such as Aufeis_kadastr (historical aufeis data collection, point objects) and Aufeis_Landsat (satellite-derived aufeis data collection, polygon objects). Historical data collection is created based on the Cadastre of aufeis (naleds) of the North-East of the USSR (1958). Each aufeis was digitized as point feature by the inventory map (scale 1:2 000 000), or by topographic maps. Attributive data was obtained from the Cadastre of aufeis. According to the historical data, there were 896 aufeis with a total area 2063.6 km2 within the studied basin. Present-day aufeis dataset was created by Landsat-8 OLI images for the period 2013-2017. Each aufeis was delineated by satellite images as polygon. Cloud-free Landsat images are obtained immediately after snowmelt season (e.g. between May, 15 and June, 18), to detect the highest possible number of aufeis. Critical values of Normalized Difference Snow Index (NDSI) were used for semi-automated aufeis detection. However, a detailed expert-based verification was performed after automated procedure, to distinguish snow-covered areas from aufeis and cross-reference historical and satellite-based data collections. According to Landsat data, the number of aufeis reaches 1213, with their total area about 1287 km2. The difference between the Cadastre (1958) and the satellite-derived data may indicate significant changes of aufeis formation environments.

Comment: The authors could add an overview figure of the data set as additional information.

Response: We uploaded the database into Google Earth and added the files to PANGAEA database. Additionally the watershed borders which are mentioned in the analysis in the paper added in Google earth format. We also added overview figure to the database.

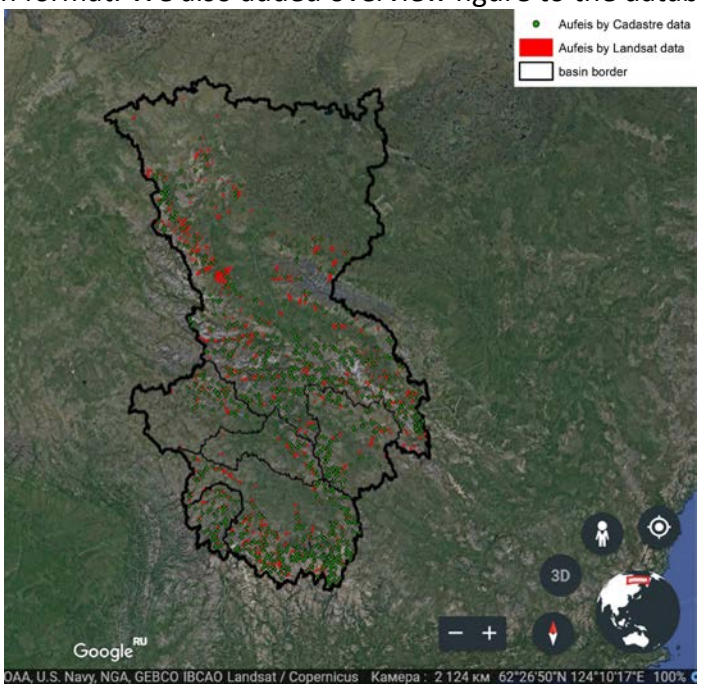

Fig. Google Earth aufeis database overview

Comment: Published data: the authors published the GIS project with 2 feature layer data and the 2 data collections also in form of ASCII files and a detailed read me word file documenting

the attribute tables. Information on the GIS project itself in the read-me file is missing: e.g., information on the format (ARCGIS) and projection.

Response: We added the missing information.

Comment:- Issues, manuscript:

General: aufeis is the plural form of aufeis, the plural form aufeises does not exist. Authors could also consider to sometimes refer to aufeis deposits in the manuscript if this fits.

Response: We fixed wrong plural form through the text and figures.

Comment: Authors could refer to the cadastral map instead of map throughout the text, also to better distinguish for the reader the cadastral map from topographic map forms.

Response: The expression "Cadastral map" has been introduced starting from Line 133 after the description of the Cadsatre.

Comment: Abstract: The authors should enrich the abstract with much more information on the technical generation and technical contents of the data set and with less discussion on changed areas and potential reasons that would be kind out of scope and not the focus of this ESSD publication. A great meta data information in this data collection is the cross-reference index enabling users of this data set to link and compare these very different 2 data set types: the historical and the nowadays aufeis data sets.

Response: Short information on Landsat-based aufeis detection and cross-reference index is added in the abstract.
Lines 13-16: Identification of aufeis by late-spring Landsat images was performed with a semi-automated approach according to Normalized Difference Snow Index (NDSI) and additional data. Then, a cross-reference index was set for each aufeis, to link and compare historical and satellite-based aufeis data sets.

Comment: keywords: reconsider the keywords, e.g., aufeis, Indigirka, Bolshaya Momskaya, Land-sat, NDSI, cadastre, cadastral map;

Response: We changed the keywords according to the comment.
Line 34-35. Keywords: aufeis, Indigirka, Landsat, NDSI, Cadastre, Cadastral map, Bolshaya Momskaya aufeis

Comment: Introduction: authors should provide an explanation what is aufeis in the first sentences of the introduction. That aufeis are thick accumulations of ice that form during winter along stream and river valleys in arctic and subarctic regions.

Response: We provided the explanation.

Lines 38-40. Aufeis (naleds in Russian, icings in English) are the accumulations of ice that are formed by freezing underground, surface and atmospheric waters on the surface of the earth or ice along streams and river valleys in arctic and subarctic regions.

Comment: 2 Research objective: this subtitle is misleading as the motivation of this study and data set compilation is already well introduced by the authors in the introduction chapter. This chapter describes the study region. Please add an overview figure with the geographical setting of the Indigirka river basin and the extent of the data set in relation to Eastern Siberia. E.g., Figure 6 is already to zoomed in to provide this information.

Response: We changed this subtitle to Study region (line 105).An overview figure with the geographical location of the Indigirka river basin is added (Line 484).

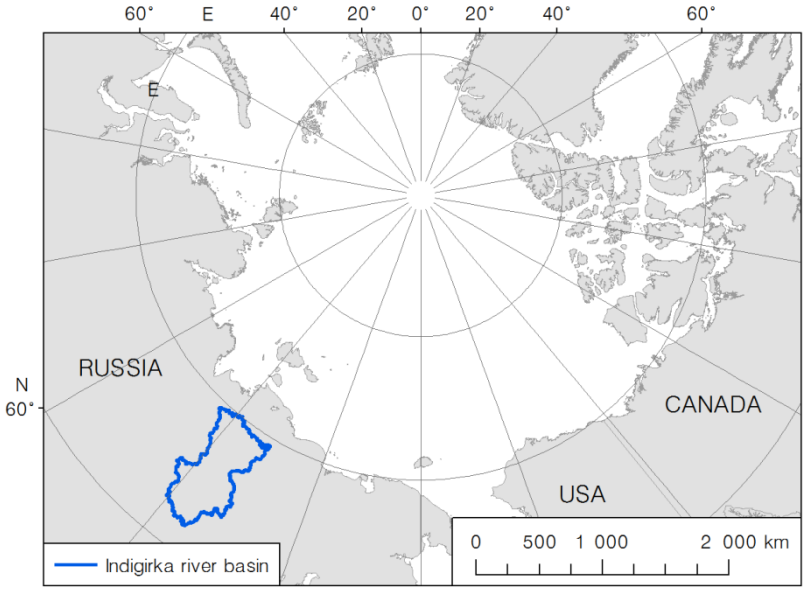

Fig. 1 Geographical location of the Indigirka river basin

Comment: 3 Material and Methods: The authors should add the tables from the published read me file in the respective subsections 3.1 and 3.2.

Response: Table 1 and 2, which contain the structure of the GIS database of aufeis according to Cadastre and Landsat images has been added.

Comment: The authors should add flow chars to make their data processing steps more clear in the in the respective subsections 3.1 and 3.2. For example the role of the thalweg creation remains unclear to the reader.

Response: Thalweg creation was an essential step of semi-automated separation of the aufeis from snow-covered areas by late-spring Landsat images. Indeed, almost all aufeis are located either at streams or thalwegs, or in immediate proximity to them. On the contrary, the snow cover in late spring mainly remains on mountains ridges and other areas with high altitude, e.g. relatively far from thalwegs. Based on the preliminary analysis of aufeis location in relation to created network of thalwegs, we estimated, that 1.5 km wide buffer zone around the thalwegs covers almost all aufeis. So, snow and ice covered areas, which are located outside this buffer, are excluded from the further analysis.

The explanation has been added.
Line 213-223: Aufeis detection algorithm was realized in ArcGIS with the help of the ModelBuilder application. Apart from the Landsat images, the digital terrain model (DTM) GMTED2010 (Danielson and Gesch, 2011) with the spatial resolution of 250 m was used to build a network of thalwegs within the study basin. This is essential for semi-automated separation of the aufeis from snow-covered areas by late-spring Landsat images. Indeed, almost all aufeis are located either at streams or thalwegs, or in immediate proximity to them. On the contrary, the snow cover in late spring mainly remains on mountains ridges and other elevated locations, e.g. relatively far from thalwegs. Based on the preliminary analysis of aufeis location in relation to created network of thalwegs, we found, that 1.5 km wide buffer zone around the thalwegs covers almost all aufeis. So, snow and ice covered areas, which are located outside this buffer, are excluded from the further analysis.

The ASTER GDEM data set needs to be introduced and explained as the meta data information on elevation is taken from this digital data set. Also for the Land-sat derived dataset? This does not become clear to the reader.
We added the information about DEM.
Line 214-216: Apart from the Landsat images, the digital terrain model (DTM) GMTED2010 (Danielson and Gesch, 2011) with the spatial resolution of 250 m was used to build a network of thalwegs within the study basin.

Comment: 3.2. The level of the USGS Landsat data product that was used remains unclear. The authors did not use the Landsat T1 Level2 (L2) that is the surface reflection coefficient already? Did the authors use the Landsat T1 Level1 data products that are terrain-corrected (T1) and Top-of-Atmosphere radiances (L1)? Because authors refer to brightness?

Response: We used Landsat 8 collection 1 Level1T (terrain-corrected) data products. The explanation has been added. Line 204-205: We used Landsat 8 collection 1 level-one terrain-corrected product (L1T) with radiometric and geometric corrections

Comment: The authors describe: Preprocessing of the images (transformation brightness into reflection coefficient) was performed with the use of Semi-Automatic Classification Plugin module in QGIS 2.18. Does it mean that an atmospheric correction was performed to surface reflection coefficient? Which type of atmospheric correction was performed to come to the surface reflection coefficient / surface reflectance?

Response: Preprocessing of the images was performed with the use of Semi-Automatic Classification Plugin module (QGIS 2.18). It includes the calculation of surface reflectance and atmospheric correction by Dark Object Subtraction (DOS1) image-based algorithm, described by (Chavez, 1996). The explanation has been added.
Line 209-212: Preprocessing of the images was performed with the use of Semi-Automatic Classification Plugin module (QGIS 2.18). It includes the calculation of surface reflectance and atmospheric correction by Dark Object Subtraction (DOS1) image-based algorithm, described by (Chavez, 1996).

Comment: 3.3 A good description of the cross reference between the aufeis deposits in the historical aufeis data collection and the nowadays data collection is missing. Authors can consider to add a short sub-paragraph 3.3. It would be helpful for re-using the data set if authors put some details here, e.g. highlight that there is the cross reference ID in both attribute tables.

Response: The sub-paragraph is added
Line 256-271: Cross-verification of aufeis data collections by the Cadastre (1958) and satellite imagery was performed in two steps. At the first step, we found closest aufeis in the Landsat-derived dataset for each aufeis from the Cadastre data, if the distance between them was less than 5000 m. The determination of search radius is based on a preliminary analysis of the aufeis locations by the Cadastre in relation to Landsat-based dataset. As a result, the cross index (identifier of the closest aufeis in the Landsat-derived dataset) and minimum distance (m) to the closest aufeis were determined for aufeis from Cadastre. For Landsat-based dataset, the cross index is the key field for the reference to the dataset from Cadastre. At the second step, a full manual verification was performed to found the mistakenly interrelated aufeis. For example, if the closest aufeis from Cadastre and from Landsat-based dataset were at a distance of less than 5000 m, but in different thalwegs, they were considered as different (unrelated) aufeis. In total, 260 aufeis from Cadastre were not verified by Landsat images. For them, the NoData value (–9999) was set in the Cross Index and Distance fields of attributive table (see Table 1 with the structure of GIS dataset from Cadastre).

Comment: 4 Results and verification: The chapter does not seem to describe or focus on verification?

Response: We changed to subtitle “Results” (line 273)

Comment: In the first section of 4 Results the authors very interestingly assess the linkages and differences between the data sets – this could become a subchapter 4.1. with a title relating to the comparison of the historical to the modern data collection. All of the above points can be addressed with minor corrections, just a few sentences or less.

Response: We corrected the title to “Comparison of the historical and modern data collection “(line 274)

Comment: consider adding a Discussion chapter with a short discussion about the usability of this data set on aufeis area growth or decline, could be one outcome of your study on the variability to assign higher variability and lower accuracy to the extraction of the aufeis area at lower elevation? Would it be possible to assign different reliability (consistency of measurement) levels for the representativeness of the derived aufeis area ? e.g. a coding of robustness 0 to 3 or a type of error code based on the authors regional and thematic expertise, related to elevation (as the authors describe that too low elevation not as good because early aufeis melt and higher variability, too high not as good because too late snow melt?).

Response: We added the Discussion section. We do not think we may assign relative reliability; instead some general analysis of the data limitations (lines 367-421) is presented.

Reviewer 2 Anna Liljedahl

The comments were sent in the form of doc document. So we have combined them here.

Comment: Suggestion to change the title to "Historical and recent aufeis, Indigirka River basin, Russia"

Response: Accepted. We changed the title.

Comment: In the abstract to specify present or historical aufeis are located in the elevation band of 1000-1300 m.

Response: Specified. Line 25: Most present and historical aufeis are located in the elevation band of 1000 – 1200 m.

Comment: Suggestion to the reference. I think you need to list the last name of the author here (not the title) or alternatively, the publisher, and the publication year.

Response: We changed the reference as the following: (Aufeis of Siberia…, Nauka, 1981) – Line 46

Comment: The question to the reference (Alekseev, 2016) - Would be good to include which region this study represents.

Response: Line 74-77 Expanded the sentence as the following: However, the same author (Alekseev, 2016) states a general tendency to the decrease of aufeis volume for the last 50-60 years in some aufeis-affected areas of Russia such as Baikal region, South Yakutia, Kolyma region, Eastern Sayan Mountains, following the increase of global and local air temperature.

Comment: Unclear. Is it 896 or 808 aufeis in your database?

Response: Lines 156-157. Clarified. Our compilation contains data on 896 aufeis. The aufeis are presented as point objects in our database. The areas are specified only for 808 aufeis.

Comment: Confusing. Do you mean that there was no recording date provided in the 1958 map, but only in the Cadastre (the catalog)?

Comment: Unclear. Do you mean that the 1958 Cadastre/catalog was not solely based on aerial photos, but also through other sources that may not necessarily reflect aufeis coverage in ~1958?

Response: Lines 162-166. Clarified. The dates of ice recording for the remaining 34 % of the aufeis were not described, meaning that aufeis detection could be carried out based not on the visible ice presence at the aerial images but on geomorphological features of river valleys. Therefore, the Cadastre might as well contain data on old aufeis glades, where the aufeis themselves were absent.

Comment: Clarify. Did you use these maps or did Grosse use these maps?

Response: Lines 167-172. Clarified. Spatial positioning of the Cadastral Map of aufeis was conducted using the location description by Russian topographic maps with the scale of 1:200 000. Grosse and Jones (2011) used the same set of maps for compiling the dataset of pingos (frost mounds) in northern Asia and described those maps in details therein. The maps of 1:200 000 scale were based on more detailed maps of 1:50 000 and 1:100 000 scale, which were derived from aerial photography acquired in the 1970–1980's.

Comment: Please provide some information on how many basins were included in correlation analysis.

Response: Lines 328-330. Clarified. Among 6 basins, the Spearman rank correlation coefficients between the basin average elevation and aufeis percentage are 0.71 and 0.77, the aufeis percentage assessed with the Cadastre and satellite data respectfully.

Comment: Can you make a stronger conclusion? For example that the total aufeis area have decreased over time, while simultaneously it appears that additional aufeis may have formed over time.

Response: We added the discussion section (lines 367-421) which shows the limitations of the datasets. Also we added the analysis of area reduction of large and giant aufeis.

We do not think we have the complete evidence for strong conclusion. In Conclusion section we are rather cautious:

Lines 436-439: The analysis of large and giant aufeis seems to indicate that there has been a significant decrease in aufeis area over the period of last 70 years. Additional analysis of historical aerial photography data could help to clarify the issue of aufeis area decline trend since the middle of the 20th century to the present.

Comment: Spelling of aufeis through the text and in the figures

Response: Corrected in the text and figures.