

## ***Interactive comment on “Aufeis of the Indigirka river basin (Russia): the database from historical data and recent Landsat images” by Olga Makarieva et al.***

### **Anonymous Referee #1**

Received and published: 9 December 2018

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 lo-

C1

cation, 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. 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 standardisation 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 userfriendly readable and usable using the proprietary GIS software ARCGIS. ARCGIS licences are costly and many user groups may use open source GIS or other geodata software packages. Using ARCGIS software the shapefiles 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

2) the GIS attribute files do not contain self-explanatory attribute names: The Indigirka

C2

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. The data set can also be uploaded in Google Earth with visualisation 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.

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.

— Issues, data publication on PANGAEA: Title: aufeis is the plural form of aufeis, the plural form aufeises does not exist. 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. The authors could add an overview figure of the data set as additional information. Published data: the authors published the GIS project with 2 feature layer data and the 2 data collections also in

C3

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.

— 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. 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.

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.

keywords: reconsider the keywords, e.g., aufeis, Indigirka, Bolshaya Momskaya, Landsat, NDSI, cadastre, cadastral map;

1 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.

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.

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. The authors should add flow charts

C4

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. 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 Landsat derived dataset? This does not become clear to the reader. 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? 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? 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 subparagraph 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.

4 Results and verification: The chapter does not seem to describe or focus on verification? 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 morden data collection. All of the above points can be addressed with minor corrections, just a few sentences or less.

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

C5

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?).

Please also note the supplement to this comment:

<https://www.earth-syst-sci-data-discuss.net/essd-2018-99/essd-2018-99-RC1-supplement.pdf>

---

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2018-99>, 2018.

C6

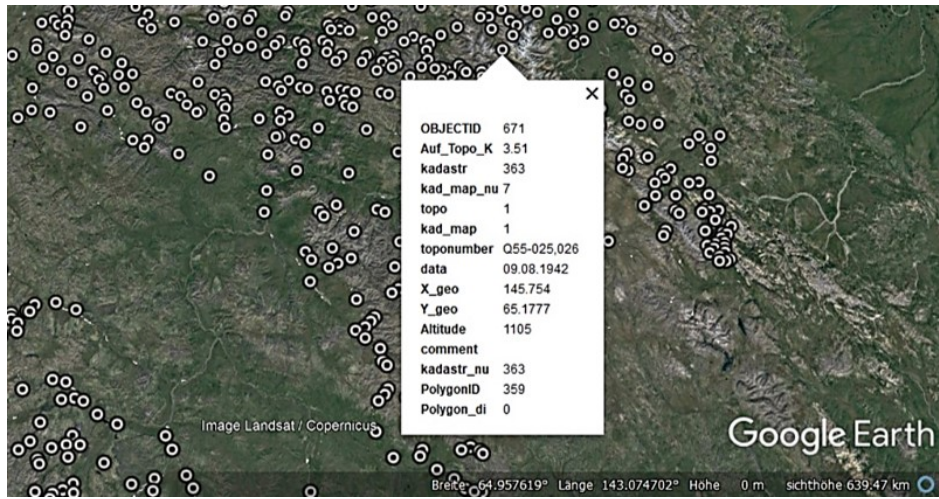


Fig. 1. example visualisation of aufeis data set and attribute names in google earth