Review comments on ESSD-2021-83

Youhua Ran et al. present a relevant raster data collection of permafrost-related ground quantities, these are the GCOS Essential Climate Variables, ECVs: mean annual ground temperature MAGT at zero annual amplitude (ZAA), and active layer thickness (ALT) representative for a thermal state of permafrost for the time window from 2010 to 2016. In addition, the authors also derived permafrost probability and what is the novelty: aridity-index related permafrost regions.

Despite the high value of providing these mapped permafrost related quantities and the very interesting novel approach integrating the aridity, the manuscript still lacks clarity and accuracy in describing products and methods and the published permafrost map products are not consistent.

The training data set as it is described in this manuscript using a large data collection of MAGT at ZAA lacks transparency and is not publicly available. The major revision requirements are better product descriptions and higher transparency on the training data collection on MAGT in ZAA as most important issue.

These are the main points of concern that should be solved by providing more details and discussion on MAGT in ZAA:

i)   The depth of ZAA is stongly changing throughout the Northern hemisphere: e.g. at higher latitudes minimum and maximum air temperature span a much large temperature range than at mid latitudes. In case of this large temperature range the ZAA depth is only reached at deeper ground depths of 10 to 15 m. This is in contrast to ZAA at more shallower depths in discontinuous permafrost and mid latitude regions. This means the depth of MAGT varies considerably in this map product, please add this to discussion, Is it possible to add the depth of ZAA as an additional metadata raster in the product? Please expand on this in the discussion chapter. This is also relevant for comparison with other products because mapped regional, circumarctic, global MAGT products and simulations in other communities refer to MAGT in specific depths always.

ii)   As the authors state the MAGT at ZAA training data is based on the most comprehensive field data collection by Alto et al. 2018. However, this higher level data collection derived from various sources is not publicly available. Alto et al. 2018 describe in their comprehensive manuscript in detail the methods and the sources of the data. The authors describe how for extracting MAGT at ZAA or close to ZAA they manually calculated these data from ground temperature depth profiles from the different data providers (GTN-P data base, Roshydromet, national PIs). However, in context of this MAGT at ZAA map product there are open questions: for example Roshydromet temperature depth profiles have a standardized maximum depth of 3.20 m. What value exactly represents MAGT at ZAA in regions with considerably deeper ZAA depths then 3 meters? Please show transparency on this issue and discuss. Please also provide the detail on how you averaged the different temporal resolutions (e.g., hourly, daily etc measurements) of the ground temperature input data sets.

iii)   The majority of the MAGT sites of this data collection are not within permafrost zones (continuous, discontinuous, isolated) and do not represent 'permafrost' temperatures. Please show the share of 'permafrost' vs non permafrost MAGT at ZAA training data and could you add an estimate of different accuracies in deriving 'permafrost' vs non permafrost MAGT, at least the authors should make readers aware of this issue and discuss it.

iv)   The MAGT at ZAA data collection in Alto et al. 2018 refers to the time span 2000 – 2014. The presented state of permafrost in the raster layer is from 2010 to 2016. Eventually the authors have explained the temporal representativeness of the training data set related to the time span from 2010 to 2016 in their manuscript. If the authors did they should describe it more clearly, if not the authors should add this information.

This referee comments do not implement that the produced raster sets are not valid – they are of value and should be used in several communities - but the accuracy of these products stated in this manuscript is unrealistic already by the nature and the noise of the MAGT at ZAA input data.

In summary, an additional raster or other form of meta data information on the depth of MAGT is required for a good usage of the mapped permafrost products in other, also permafrost-not experienced communities: discussions and details on i) to iv) should be provided in the manuscript.

The other input data should be described more clearly, stating data sources, exact product names, native spatial resolution, temporal resolution and time stamps of the products, e.g. also in the form of a table.

Examples are the source of the lake data set is unclear, its native spatial resolution, also the sentence 'small lakes were filtered out by majority statistical processes' remains unclear. Still other large surface water bodies, such as the large Arctic rivers are not excluded. This data treatment does not seem to be consistent. Please discuss.

When showing the permafrost extent of the Northern hemisphere, could the authors add also the values including the lake area for a comparison with other permafrost map products that have lake areas included?

The inspection of the published map products Ran et al. 2021 https://data.tpdc.ac.cn/en/data/5093d9ff-a5fc-4f10-a53f-c01e7b781368/ shows that large lakes are not excluded, e.g. the area of the deep Lake Baikal in Siberia contains MAGT at ZAA values. The authors need to correct their product masking surface waters and upload a new version.
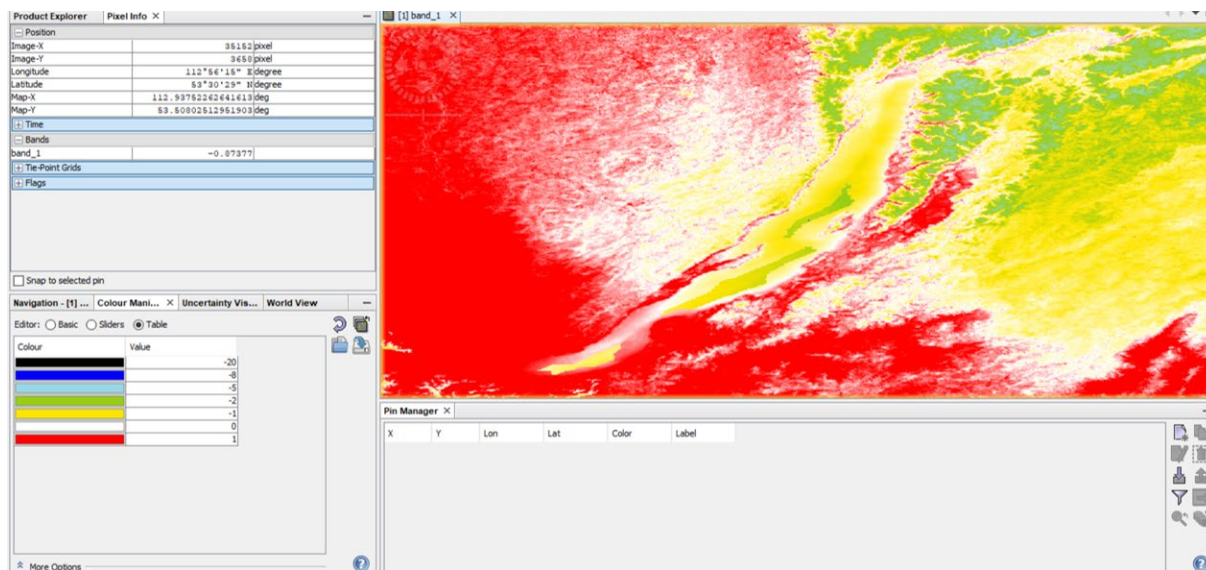
The figure shows a snapshot of the MAGT product with the area of Lake Baikal containing MAGT patterns looking similar to bathymetry related features?

It would be user-friendly to convert the GIS no data value of – 9999 into a more user-friendly no data value, e.g. NaN.