

Earth Syst. Sci. Data Discuss., referee comment RC2
<https://doi.org/10.5194/essd-2021-439-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on essd-2021-439

Ingmar Nitze (Referee)

Referee comment on "Retrogressive thaw slumps along the Qinghai–Tibet Engineering Corridor: a comprehensive inventory and their distribution characteristics" by Zhuoxuan Xia et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-439-RC2>, 2022

Summary and general comments

The manuscript/preprint "Retrogressive thaw slumps along the Qinghai-Tibet Engineering Corridor: a comprehensive inventory and their distribution characteristics" by Xia et al., describes a geospatial vector dataset of retrogressive thaw slumps along the QT Engineering corridor. The dataset is an important piece in the scope of a recently started effort to create pan-arctic/global datasets and inventories of retrogressive thaw slumps (IPA action group) for the training and validation of machine/deep-learning models.

The manuscript describes the data collection, data processing and final dataset thoroughly, however with some points that can be improved. In addition to the description of the dataset and its creation, the authors present a further analysis of the dataset and its relation to geographical parameters.

The manuscript has a good quality. However, I see some potential for improvement in the language. Although I am not a native English speaker, I noticed that some paragraphs did have some language issues. Therefore, I would recommend to put some emphasis on language editing.

Please find specific comments regarding the dataset and manuscript below.

Data

The data are easily accessible through the PANGAEA data archive and are citable with a DOI. The authors are using the standard "ESRI Shapefile" format. Although this is the quasi standard, this data format has its drawbacks limiting the attribute name length or having multiple files. As this is a comparably small dataset, the authors may provide data also in OGC compatible "GeoJSON" or the more robust "geopackage/GPKG" format, which is a bit more flexible and often a bit easier to use. However, this is just a minor/optional suggestion.

Detailed questions and comments

Dataset

If you have the possibility to make edits to the dataset, I would be happy if you could check the following suggestions. I understand that updating a published dataset is maybe not the most straightforward task.

"Probabilit": This manually assigned (kind of arbitrary) attribute is not very intuitive to me. I would rather understand it as a calculated output from e.g. the DL model. This is not a major point, but you may find a better naming. However, if not that is also alright.

"Source Image": PlanetScope Scene or OrthoTile?

Year-Month: I think it might be better to split this attribute into (1) Year, (2) Months. However, tracing back to the original image, before mosaicking, would be even better. If this is possible just use some standard data format, e.g. ISO format or YYYY-MM-DD, and add the original filename. If that is not possible, just leave as is.

Do you have DL model versions? This information may help with reproducibility.

Manuscript

l.35 – remove “normally”

l.66 – “lack” typo?

l.68 – “combing” typo?

l.79ff – is the vegetation cover destruction/disturbance evenly distributed or limited to certain areas? Is that somehow linked to the presence of RTS?

L85: Figure 1a. What are the white spots? Glaciers?

Figure 1a: Perhaps you could use a different projection as EPSG:4326 often creates some distortions (squeezes Latitude). This is just a little suggestion, perhaps at $\sim 30^\circ\text{N}$ it is not too bad compared to high latitudes.

L92: which data product (Scenes or Orthotiles)?

L95: Which DEM did you use? Absolute elevations?

l100: “in several local sites”. Can you be a bit more specific? How many, how much area? Are they representative?

Do you think these spots can be added to Fig1. without “overfilling” it. If it is not possible that’s no problem, just a suggestion.

110 ff: I understand that this is a data paper, where the processing has been done, but why did you only use RGB and not the NIR band, which from my experience helps a lot (at least in the Arctic)?

Did you test other combinations training/inference year combinations, e.g. train on 2020?

115ff Paragraph 4.2. Did you try to add more information to the deep learning model? Which are the input bands, was it only RGB?

134 Typo "PlanetScpoe" (Same in the flowchart, lower green parallelogram)

137 What do you mean with sub-images? How did you create them, what is their size? Please provide more information what they are.

139 Please change "changes yearly" to "annual changes"

140 Do you automatically retrieve the headwall? This part is somehow unclear to me? I assume you are inferring the footprints, is that correct? So the headwall position is interpretation, right? Then it is logical and (1) can be omitted.

160/Fig4. I think you can still add a scalebar to the World Imagery I think, it has the same extent, so we can safely assume the same scale. Alternatively you could use only one north arrow and scale bar for the entire figure, as it is the same for each map.

165/Table1: It is not clear what "negative polygons" are for training. In a binary classification/segmentation I would assume to only have positive polygons (target class) OR a raster mask with positive and negative (background) values.

Table1: better use "Prediction" or "Inference" instead of "Predicting" as a heading

166ff: How did you handle inaccurate polygons? I sometimes experience, that thaw slumps are perhaps correctly identified, but the polygon might not outline the RTS correctly. How did you handle these cases? Please provide more information.

166ff: In the same sense, did you do some fine-tuning on the DL model output? E.g. I suppose the model has some kind of probability output, where a threshold (0-100%) can be set to (1) impact the number of detected RTS and (2) impact the polygon shape in the polygonization (raster to vector) process. Could you perhaps provide a little bit of insight

either here, or even better in the detailed workflow description.

175: Do you have a specific minimum mapping unit (MMU)? Is 0.05ha you MMU? If yes, please mention that.

176: It would be nice if you could mention the study area size again in relation to 1700ha.

182: Does soil texture correlate to excess ground ice in these regions or is it independent?

190ff/Fig5. The red colors are really hard to pick up on all three maps. Please check if you can find a better color with a lot more contrast to the background. I guess for colorblind people, they might be just invisible.

For panel (a) you may just want to use the background of Figure 1, as the Planet Mosaic does not add significant information.

Panel (a): Please use some slightly different styling for the location indicators of b and c, e.g. letter in a box or so. At the moment it has the same styling as the main panel descriptor (a).

194: Fig 6. I like the comparison of histograms (b,d,e)! It clearly shows the differences to the overall area.

Fig 6c: The red numbers are very hard to read. Please check your color scheme. Perhaps you can use lighter shades of grey for the bars. Think about changing the color of the numbers. You could provide information what the numbers mean in the caption.

Fig 6f/g From looking at the diagrams it's not immediately clear which one is inside and which one outside. I recommend to use either barcharts with or keep the round diagrams but use different signatures for the inner vs outer part and add a legend. I may go with barcharts as the proportions are easier to pick up visually.

201: I think it would be nice to shortly repeat the "close to road" definition for better readability.

Figure 7a: It would be helpful if you could normalize the values by area. E.g. percent or ha/km². Then values become comparable to other studies.

241ff: this paragraph may need some more language editing

243: better use "novelty" instead of newness

246 multi-time -> multi-temporal

254: "some false positives" I would use a bit stronger wording, as FP vastly outnumbered true positives.

263/264: here you mention that you have some kind of MMU, you did not state that above (see comment further above).

267 ff: in this paragraph you use the terms "thaw slumps" and "retrogressive thaw slumps". Before you used the abbreviation RTS. Please be more consistent.

271: "or area receives less solar radiation". That somehow does not read well.