Point-by-point response, Review II - Anonymous

Reviewer Comments
Author response
Changes made in manuscript

The combination of available infrastructure data sets to increase the semantic content is a good idea that has the potential to improve future analysis regarding contaminated sites and related hazards. Especially the increased number of usage categories is promising for many applications. The usage examples are well done and will be helpful for future users of the data set.

Below I outlined several points that would improve the manuscript in my opinion.

Figure 2 may be valuable for users of the data but could go into the appendix. Instead, a figure like a flowchart or decision tree on how the input data sets were combined may be beneficial for the reader for a better understanding of the process.
Thank you for recommending these changes. We moved the figure to the appendix (Figure A1) and added a flowchart (Figure A2), describing the data harmonization and subsequent decision tree process.
--

Regarding OpenStreetMap: As I understood from the text, the authors use OSM as an alternative source of linear infrastructure instead of the satellite derived Bartsch et al dataset. Previous publications including Bartsch et al 2021, which the authors heavily cite throughout the text, point out that OSM has some major drawbacks, including inconsistencies, and lacking some more recent infrastructure developments. In line 147-148 the authors note that they use OSM for the linear infrastructure, not SACHII, for buildings they describe a decision tree of combining the data sets (from line 207). Is there a reason a similar approach was not used for linear infrastructures, especially given the concerns previous publications have raised about OSM. Maybe just some additional discussion could be included on this.

Thank you for this valuable comment. Indeed, Hjort et al. (2018) report that OSM shows regional inconsistencies (with better coverage in North America and Eurasia compared to Asia) and presumably omits parts of isolated, smaller communities.
In terms of the OSM road network, however, Barrington-Leigh and Millard-Ball (2017) report that approximately 83% percent of the global road network is represented in OSM, providing a valuable data basis.

The decision to use OSM instead of SACHI for linear infrastructure was based on two major concerns:

1. Road features in the SACHI database show a low accuracy (producer's accuracy of the final product ranging from 26-57 % compared to buildings (79.5-95 %)) (Bartsch et al., 2021).
2. Due to the relatively coarse spatial resolution of 10 m, SACHI also
    a. does not represent small/ unpaved roads
    b. struggles with distinguishing between road adjacent areas (buildings, paved areas in general, etc.) and roads itself, leading to a blurred representation of the road network (visible in Figure 2 b).

From a technical standpoint it is also beneficial to provide the road network as line features (OSM default) to enable routing.
We included these arguments in the section "Materials & Methods" and "Discussion":

Page 6, line 148:
When examining the linear transport infrastructure, we observed some gaps in the data, particularly in settlements: extracting narrow paths or distinguishing between a linear gravel road and other human-impacted areas, such as driveways or exploration pads, were difficult with the limited spatial resolution of the Sentinel sensors (10 m). In addition, the "road" class showed a particularly low mapping accuracy compared to the "building" class (Bartsch et al., 2021). As OSM on the other hand is estimated to represent 83 % of the global road network (Barrington-Leigh and Millard-Ball, 2017), we decided to use OpenStreetMap data to represent the linear transport infrastructure.

Page 22, line 453:
This limitation may be attributed to the peripheral status of Arctic environments within the global OSM mapping network, primarily due to their sparse population. Hjort et al. (2018) report such a limitation for isolated, smaller communities and with regional variability (e.g. better coverage in North America and Eurasia compared to Asia).

Page 24, line 483:
Nonetheless, 78 % of the true road grid cells were accurately represented in the SIRIUS dataset. When accounting for the offset grid cells, this value increases to 92 %. This value underlines the effectiveness of OSM for representing linear infrastructure opposed to SACHI. OSM allows not only a clear distinction between roads and adjacent infrastructure areas but also the inclusion of narrow roads and footways. Looking ahead, it could prove beneficial to integrate official data from local or federal agencies (e.g. Alaska Department of Transportation) to evaluate the comprehensiveness of the OSM linear infrastructure data. Further, incorporating the Trans-Alaskan-Pipeline would provide a spatial context for contamination, oil exploration and transportation data.

--
Figure 6: interbal – internal?
Yes, thank you. That was a typo.
--
Figure 10 could go into the appendix.
We agree that this is - although useful for the data user - better placed in the appendix. We moved the figure accordingly (Figure A3).
--

Line 470: Is this issue not shown in Figure 8? Maybe include a reference here to this figure, just for improved understanding for the reader.
Yes, thank you for pointing this out. This is shown in the figure. We referenced it in the text.
––
Maybe I missed it in the text, but your input data sets have different extents? So would the accuracy of your output not differ between areas where both/all inputs were available and those that rely solely on OSM data?
Thank you, and yes, you are right we did not address this in the text. We added a statement in the "Discussion" section:

Page 24, line 501:
It is important to note that the high mapping accuracy of 78 % (92 %) for linear and 94 % for polygonal infrastructure in our test area of Shishmaref can likely be expected for most of the coastal regions (until 100 km inland). For inland areas (beyond the extent of SACHI), the infrastructure data relies solely on OSM, which may show the above mentioned limitations. Once again, integrating further official data sources could improve its quality.

Citation: https://doi.org/10.5194/essd-2023-393-RC2