

Comment on **essd-2022-251**

Christopher Brown (Referee)

Referee comment on "High-resolution and Multitemporal Impervious Surface Mapping in the Lancang-Mekong Basin with Google Earth Engine" by Genyun Sun et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-251-RC3>, 2022

The authors present a method for extrapolating annualized annotations for binary impervious surface (IS) / non-impervious surface (NIS) into future years when NIS -> IS transitions are assumed to be insignificantly probable. The authors also present a method for converting a short sequence of annual IS / NIS binary classifications into a sequence congruent with the previous assumption regarding transitions. The authors combine these methods in a standard supervised Sentinel-1 / Sentinel-2 random forest classification workflow in Google Earth Engine to produce annual IS / NIS maps of the Lancang-Mekong basin from 2016 to 2021. These maps are compared to a number of existing global LULC products with an IS class, or a class the authors considered comparable, using a dataset created specifically for this purpose. The results of the author's assessment indicate good overall agreement with their sample.

I would begin by asking the authors to carefully copy-review their manuscript: I find that there are grammatical and/or punctuation errors on nearly all of the proof pages. In particular I suggest paying special attention to extraneous usage of "the" before nouns, and general comma usage. I also suggest replacing verbs used in a subjective context such as "excellent" or "state-of-the-art" with the less strongly worded "popular" or "modern." There are also opportunities to unify language e.g. "dynamic monitoring of IS" is more commonly known as "change monitoring," and the authors are not consistent in their wording. I find the structure of the manuscript's narrative generally well paced and direct.

The author's characterization of the Sentinel-1 and Sentinel-2 missions was imprecise. Line 62 asserts Sentinel-1 has a 10m "spatial" resolution, however Sentinel-1's IW spatial resolution is actually 5m x 20m with the GRD product specifically resampled to a pixel resolution of 10m (<https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-1/overview/mission-summary>). Line 103 discusses Sentinel-2 "image blocks," which is not a Sentinel-2 data product I'm aware of, and lines 103-104 discuss Sentinel-1 and SRTM "scenes," which are not part of the ARD scheme used in delivering either of these products. On line 123, the authors state Sentinel-1's revisit to be 6 days, and while this is

true of the instrument, the availability of the data used by the authors (GRD IW) is highly variable between orbits, location, and polarization (https://www.researchgate.net/figure/Sentinel-1-constellation-observation-revisit-and-coverage-scenario-April-2021-10_fig2_351632665).

I have concerns with the expert system introduced by the authors to perform sample migration, that which formed the basis of their strategy to extend their training data. Besides a lack of ablation analysis on the many parameters, Table 4 (which would benefit from NIS column totals) raises doubts about the author's procedure. Considering that a fundamental assumption in the author's framework is that IS -> NIS changes are not permitted, one would expect points migrated to successive years to decrease with time. This is not the case, and e.g. points that were not migrated to 2017 (15,649 total) are in fact migrated to 2018 (16,075), calling into question the validity of this approach since it violates the IS -> NIS assumption between 2017-2018. At the extremes, it would be possible for a point to exist exclusively in 2016 and 2021. I am also concerned with the use of pixel-counting statistics used in Equations 2-5, as it is not stated whether or not the authors account for processing baseline updates in the underlying Sentinel-2 collection. Failing to do so would lead to double-counting and so erroneous results.

I appreciate that the procedural diagrams in Figures 3, 4, and 5 were easy to follow, which is somewhat of a rarity.

I also have concerns with the temporal consistency check introduced by the authors. From first principles, the authors are, in my opinion, suggesting a rather complicated procedure to post-process a length 6 vector with only 64 possible values, 57 of which are actually invalid. The authors do not state the boundary considerations for the sliding window, which is important as it effectively defines the IS change sequence semantics in ambiguous cases like "101010." Furthermore, that such ambiguous sequences are collapsed to "valid" sequences seems like a fallacious assumption in the overall methodology when a "transition" or "low-confidence" classification may be more appropriate.

Little information is given regarding the stratification performed to collect validation data.

I find that Figures 10 and 11 and the corresponding section offer little insight or value into the author's data.

The author's compute area statistics from their maps yet do not offer any confidence intervals or uncertainties. These statistics are all but the bare minimum required to publish estimations of area based on remotely sensed data.

I recommend the authors discuss the potential for change dynamics to effect their

agreement statistics with the external data products outside of the sampling year. The authors conclude on Lines 348-349 that similar accuracies to WorldCover-10 2020 in 2021 indicate that the sample migration strategy was successful, but since no baseline was established, it is unclear whether similar performance can be attributable to a combination of class-shift and definition penalizing WorldCover-10, extra training data introduced in re-using points across years, or the nuances of the algorithm introduced by the authors.

One of the biggest potential contributions from this effort would have been the underlying training data sampled from 2016, and I encourage the authors to make this available.

While there is good work in this research effort, I believe the authors should intensely proofread their writing and try to keep their language exact and consistent. I also believe the novelty, the suggested sample migration and temporal smoothing, demands more methodological consideration and emphasis in the manuscript. The algorithms introduced also require more justification to contextualize their effects on the accuracy assessment. The authors should seek existing literature on area estimation to ensure confidences are published with the computed statistics. My overall recommendation is for reconsideration after major revisions.