

General comments on the validation using 'authoritative' datasets

1. Surface Hydrology vector dataset

We have compared major waterways of the SH dataset to our L8 water mask in a similar way to OSM. Firstly, we have obtained very confusing results related to positional differences between SH and L8 centerlines, they looked too large for SH. Further investigations revealed that this happened because SH includes many short water polylines (Figure 1, 2), making it much harder to apply Goodchild's method to determine positional differences (influence of the buffers along the river channel becomes significant). To overcome this problem we actually had to implement a butt cap buffer algorithm within Google Earth Engine, to exclude water pixels at the ends of the segments as much as possible:

<https://code.earthengine.google.com/9f6ecfdad25b6e4f10d652a7ae77029a>.

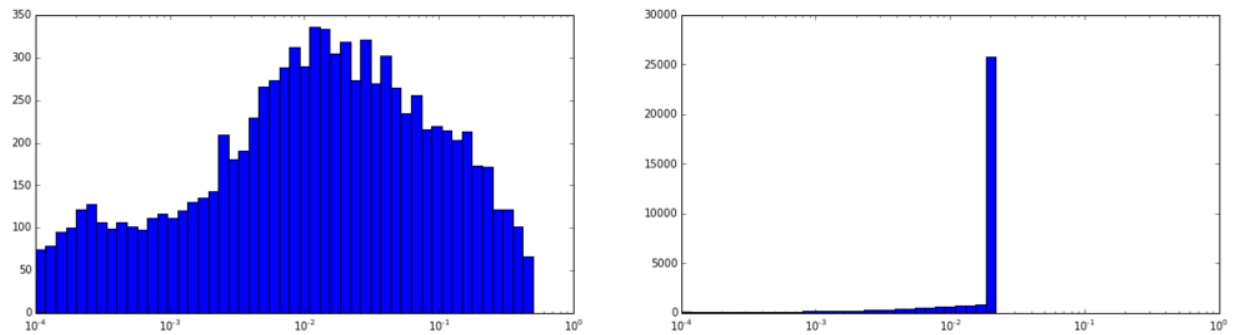


Figure 1. Histogram showing lengths of the OSM river polylines before and after the split_segments procedure

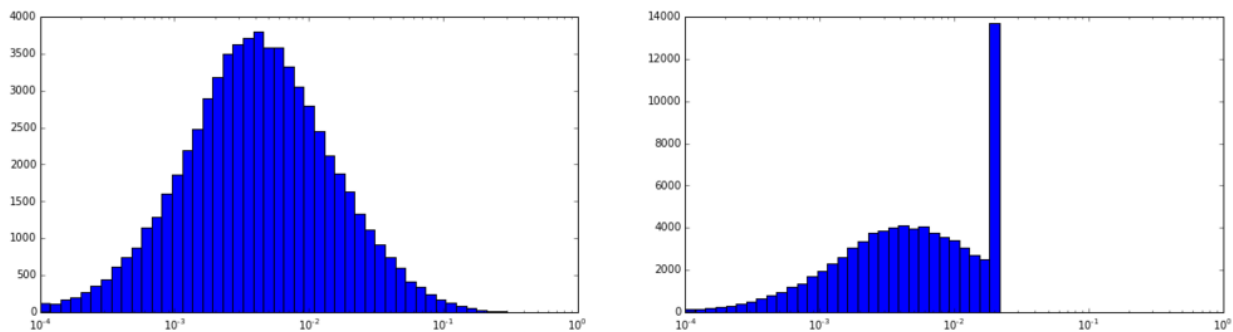


Figure 2: Histogram showing lengths of the SH river polylines before and after split procedure. Note that there are many short polylines left (right chart).

After improving the algorithms, we have estimated positional differences between Landsat 8 water mask and both OSM and SH datasets once again, limiting maximum buffer size to 100m, and score threshold = 0.75, which covers most of the local offset errors.

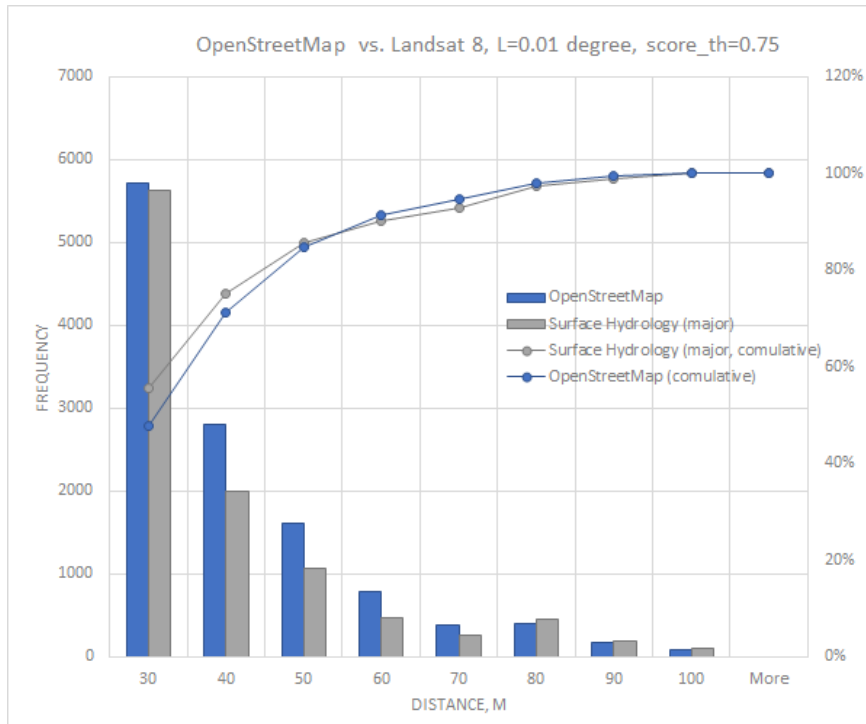


Figure 3. Positional differences between Landsat 8 water mask and two vector datasets: SH and L8.

From the Figure 3 it can be seen that OSM dataset (on average) is very similar to SH in the areas where they overlap with L8.

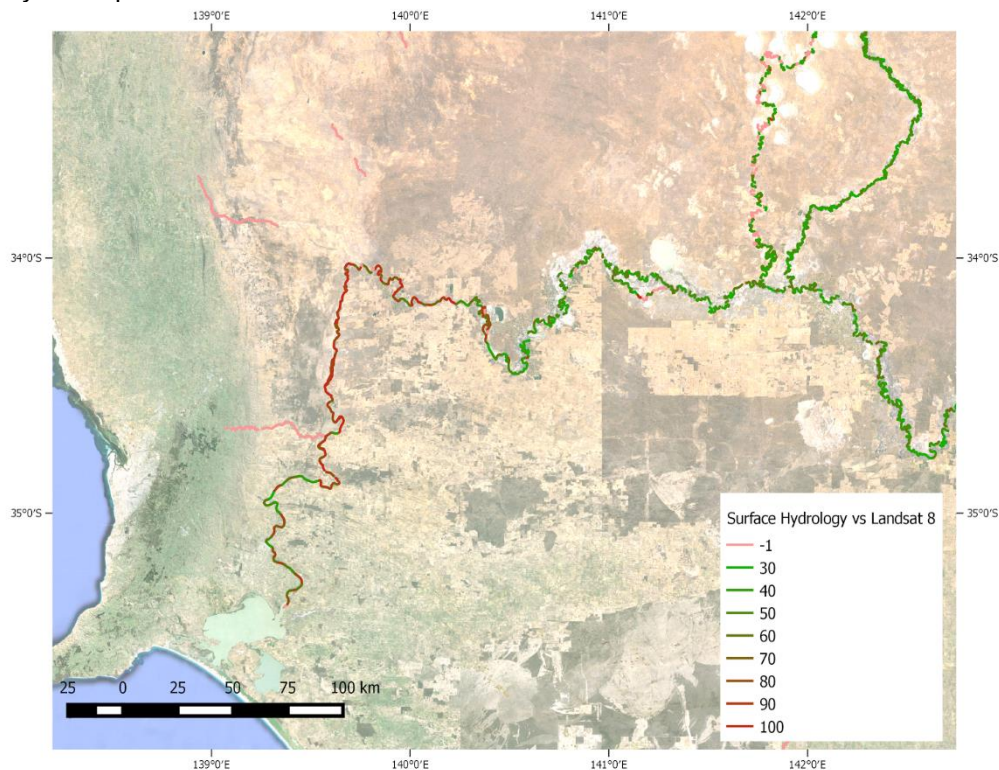


Figure 4: Downstream area where large horizontal bias is observed in SH (Landsat, OSM and Google background maps seems to match there so it is very probable that the error is in SH)

Even though we found Surface Hydrology to have a better coverage when compared to OpenStreetMap. In several places OpenStreetMap seems to provide a better local quality, especially downstream of (see Figure 4).



Figure 4: Location where a horizontal bias (~100m) was observed in Surface Hydrology rivers, mainly downstream of the Lake Victoria, see also <https://code.earthengine.google.com/e07404ea402a3a8b8c43bd411f9096a6>. Thick line on the Figure show segments where the differences are larger than 90m.

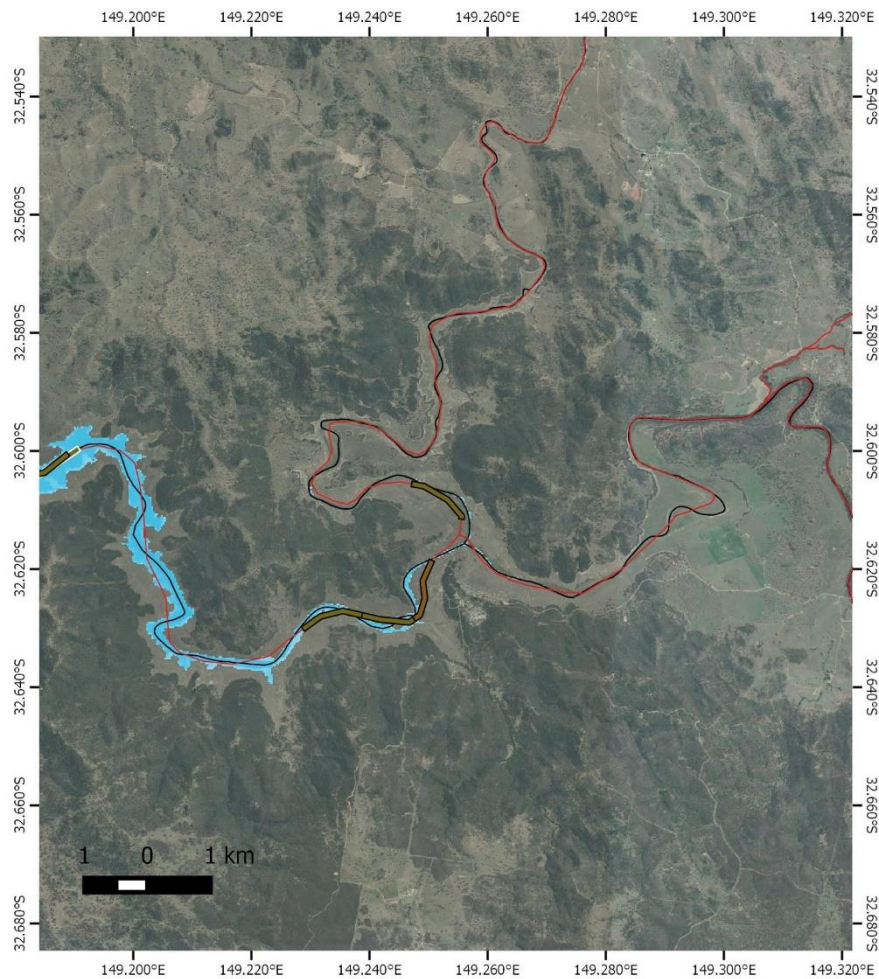


Figure 5: Location where OpenStreetMap error were observed. Thick line shows segments where offsets larger than 90m were observed (and where Landsat 8 water mask was available). See also <https://code.earthengine.google.com/2571bfa67595bc69e2bcf116ef6925d7>

2. 5m DEM for Australia

This dataset would be indeed very useful to improve the quality of the final water mask for the Murray-Darling basin. We have noticed that the dataset covers mainly flat areas. Even though this dataset sounds very useful for validation, it is actually not critical for our study, since our focus was to develop method which can be applied globally. It is also a bit unrealistic to make use of that DEM within the time frame allocated to prepare a review. The most time consuming part would be to generate LDD and HAND to estimate inundation areas, like in <http://onlinelibrary.wiley.com/doi/10.1002/hyp.10581/abstract>.

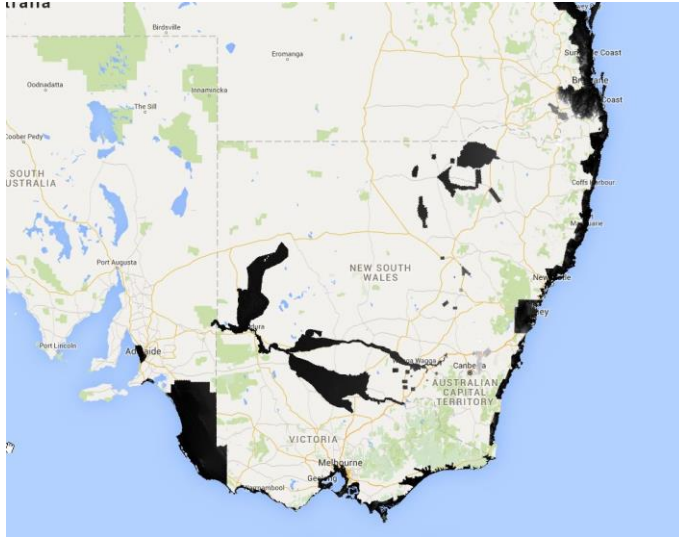


Figure 6: 5m AU DEM dataset in Google Earth Engine.

Additional effort would be required to perform a detailed comparison using 5m AU DEM. The main reason is that the current approach (DEM > LDD > HAND) will have to be evaluated to produce good results. Visual inspections show that in many cases OSM local quality is relatively good, but SH is frequently more accurate, when 5m AU DEM is available.



Figure 7: Screenshot from the Google Earth Engine showing hillshaded 5m AU DEM (green), OSM (yellow), major SH rivers (blue) and skeleton of our Landsat 8 water mask (light blue).

<https://code.earthengine.google.com/fc19a83e9579afd9107b9dc083c3a85e>

Our general impression is that SH provides a better quality when 5m DEM is available when compared to OSM. However, in some cases, rivers in SH are not assigned major attribute, like can be seen on the above figure, where water is detected using Landsat but major SH waterline is not available. Also, additional, more sensitive parameters need to be used to detect smaller details, like small zoom-in area on the above figure.

Systematic horizontal shift between Landsat and high-resolution imagery is another problem that would require additional method development to be corrected automatically.

3. Water Observations from Space

Quick comparison of WOfS and our water mask datasets shows a very good match between permanent water bodies. But we are also not sure if a detailed quantitative analysis would be correct, mainly because WOfS is based on *many years of observation, starting from 1984* and our water mask focuses mainly on the permanent water mask observed from 2013. Some of the water bodies have already disappeared and some others appeared, like Lake Mokoan (Figure 7).

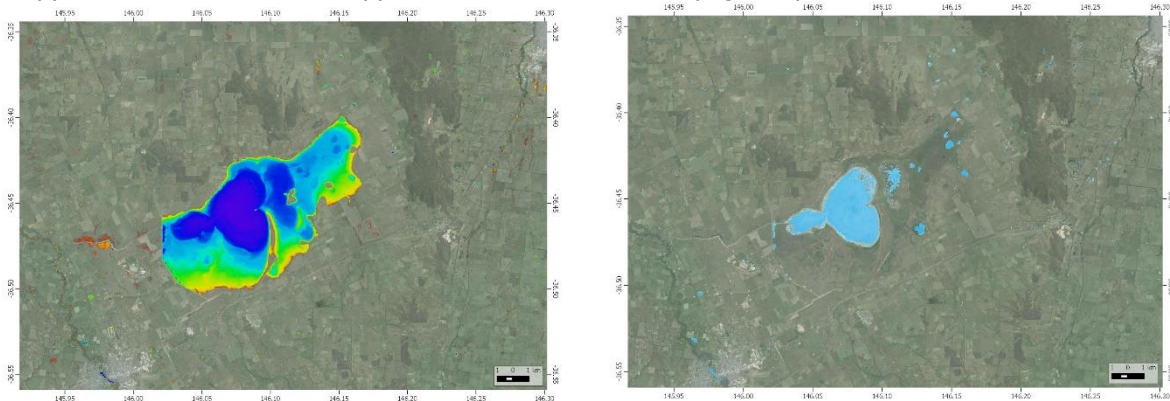


Figure 8: Lake Mokoan as seen on WOfS (left) and Landsat 8 water mask (right, our study).

Additionally, WOfS seems to include both permanent and flood water masks, making the analysis even more difficult.

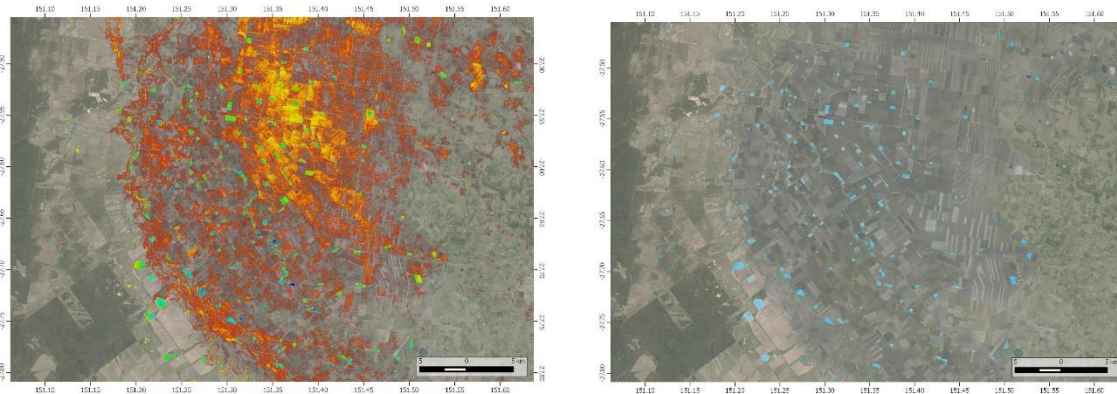


Figure 9: Agricultural reservoirs as observed using WOfS (left) and our water mask (right)

Furthermore, the dataset seems also to suffer from comission errors in case of dark vegetated areas, as can be seen on the Figure 9.

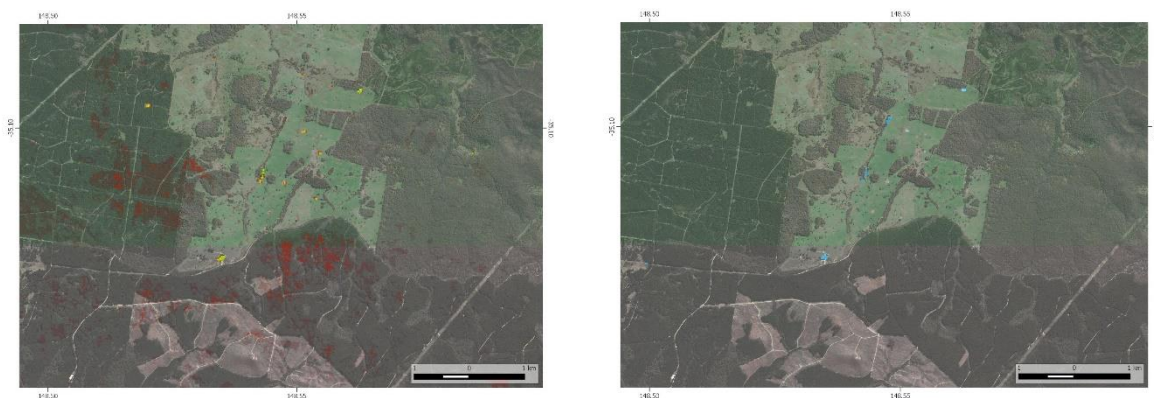


Figure 11: Commission error present in WOFS dataset (left) and our water mask (right)