Article Review: *Beach-face slope dataset for Australia*

Floris Calkoen

Preprint of article reviewed here is available at
https://essd.copernicus.org/preprints/essd-2021-388/

Summary

The authors present a novel dataset of beach slopes for the coastline of Australia, which they derive by combining satellite-derived shorelines, from Landsat satellite imagery, with tidal levels, from the global tide model FES2014. The method assumes that variability in shoreline positions can be explained two components: 1) a tidal component, driving high-frequency variability; 2) and, a sediment-transport component, driven by processes characterized by lower frequencies. The authors isolate the higher-frequency fortnightly tidal component from the other sediment-transport processes by using the Lomb-Scargle transform. This method to estimate beach-slopes from Landsat satellite imagery was introduced in Vos et al. (2020). The dataset and article currently under review are essentially a a product of the method introduced in that paper.

The following relationship is assumed to relate beach slopes to shorelines positions and tidal heights:

$$\Delta x_{corrected} = \Delta x + \frac{z_{tide}}{tan\beta} \tag{1}$$

, where $\Delta x_{corrected}$ is the tidally-corrected cross-shore position,Δx is the instantaneous cross-shore position, z_{tide} is the corresponding tide level and $tan\beta$ is the beach-face slope.

The Lomb-Scargle transformation is used to identify the peak of the fortnightly tidal component in series of tidal levels, sampled according to the timestamp of the satellite images. Following eq. 1, series of shoreline-positions are tidally corrected for a range of beach slope values. For each series in the collection of corrected shoreline-positions the tidal energy is calculated by taking the integral for a set window around the identified peak frequency. Finally, the beach slope is estimated by taking the beach-slope value that minimizes the tidal energy. The dataset was created by applying this method to all OSM-identified sandy beaches, along the Australian coastline, at 100-m alongshore resolution. Uncertainty in beach-slope estimates is expressed by computing the width of 5% upper and lower beach-face slope confidence bounds. The dataset is made available at transect level as well as averaged per beach.

The dataset is an important milestone for making coastal models more data-driven. The authors successfully apply their recently-introduced technique to derive beach-slopes from Landsat satellite imagery. Although their earlier work (Vos et al., 2020) already provided a proof-of-concept for beach-slope derivation from satellite imagery, this study shows that such method can be applied at continental scale. The article and dataset are understandable and straightforward to use. The dataset is significant (continental scale, high resolution), unique, usuful (science, coastal management, floodrisk modeling and beach safety) and complete (to the best of my knowledge). Nevertheless the manusscript can be substantially improved by addressing some limitations; and, the datasat could be presented more computer-friendly. More importantly, although the temporal frequency of shoreline-position deriviations is essential to the used methodology, the sample data indicate that the repeat cycle is often not 7 or 8 days; this might be an important inconsistency and should be addressed. Overall, I would therefore recommend to return it to the authors for minor revisions.

1 General comments

1. In Vos et al. (2020) the authors explain that Landsat data is used because this archive has observations at the same point every eight days. This 8-day revisiting period (or sampling interval) is critical for deriving fortnightly tidal cycles (14.26 days) from the shoreline-position series as in a signal with less frequent observations this tidal cycle would be indistinguishable. Strictly speaking the Landsat repeat cycle is already too long to meet Nequist requirements, and therefore the frequency periodogram shows an aliased pattern.

   Figure 1 shows that at least for Cable beach (toy dataset provided with Vos et al. (2020)) most of the intervals between image acquisition are not eight days. For about 18% the interval is greater than 8 days. When using the algorithm `SDS_slope.reject_outliers()` to remove outliers from the shoreline-position series this figure increases to 39%. Figure 1 shows that the interval between image acquisitions (Cable beach) is typically seven days. It also shows shows that often a revisiting time of zero days is found. In line with these figures, the dataset presently under consideration, has, for some transects 1286 observations between 1999 and 2019—which is 373 more than you would expect from `365.25 (days per year) * 20 (year) / 8 (revisiting period) = 913`.

   Considering the unequally-spaced image-acquisition intervals:

   (a) What causes this highly-unequal temporal distribution of shoreline-position series? This cannot be solely clouds right? In general this data seems to contradict the statement made in Vos et al. (2020) about a target sampling period of no more than 8 days.

   (b) The notebook example provided with Vos et al. (2020) shows that for Cable beach dominant tidal-energy peak is observed at 14.8 days, which is in accordance with the observation that the most frequent image-acquisition interval for this beach is 7 days—enough to observe fortnightly tidal cycles according to the Nequist theorem. However, for many sites the dominant sampling interval might be 8 days (or something else), which results in a peak at 17.5 days (or something else). When creating this dataset, was there a method to automatically determine the peak tidal frequency? If so, how were situations with about an equal amount of 7- and 8 day image acquisition intervals handled? The extraction and use of peak tidal frequency could be addressed in more detail. Presumably some is explained in line 112-114, but it is not entirely clear if this refers to the methods used in Vos et al. (2020) or in this paper.

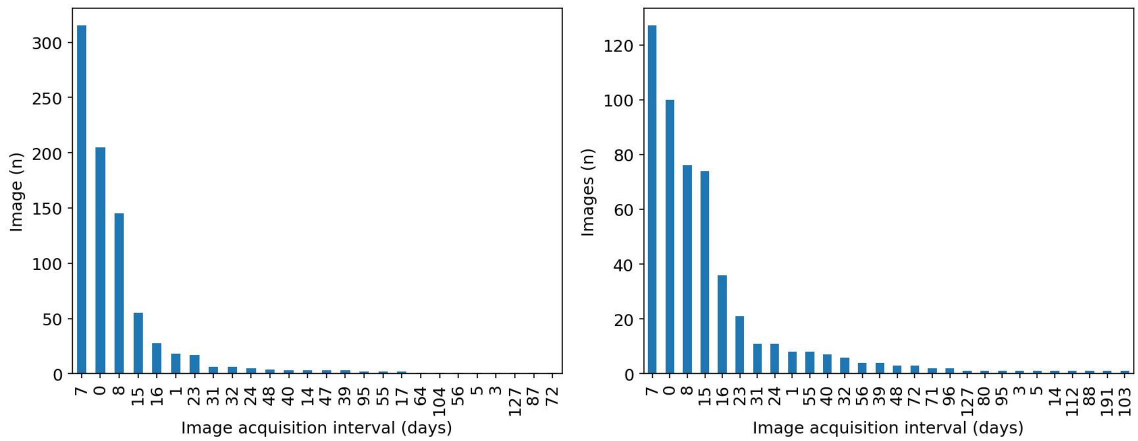

Figure 1: Interval between image acquisition (days) for time-series with all shoreline positions (left) and time series with outliers removed from the shoreline-position series (right)

(c) Landsat 5, 7 and 8 repeat cycles are all 16 days while Landsat 5 retired when 8 was introduced (*Landsat—Earth Observation Satellites*, 2016), so how can there be so many image-acquisitions with an interval of zero days?

(d) Does the method still work when removing outliers from the series of shoreline positions? If not, this would be useful information to include.

2. The authors state that their method paves the way for the generation of large-scale beach-face slope datasets (line 54-55). Although the beach-face dataset is an important milestone, I think it would be fair to stress a few limitations associated to using Landsat satellite imagery.

   (a) The spatially unequal distribution of images in the Landsat archive (Wulder et al., 2016) will hamper derivation of beach slopes in many areas around the globe. For example, many areas in Europe and Africa only have about a quarter of the images as Australia.

   (b) Cloud cover has a negative impact on deriving shorelines from satellite imagery, which is also acknowledged by the authors (Line 111). The dataset presents beach slopes for Australia, a continent which is in particular characterized by low mean cloud frequency (Wilson & Jetz, 2016).[1]

   Australia can be considered the most favorable continent for beach-slope derivation from Landsat satellite imagery because of its low cloud frequency and large collection of Landsat satellite images. Unfortunately, for many other areas, it will be quite a challenge to derive shorelines from satellite imagery at the same temporal resolution as Australia. As the shoreline-position signal might be too weak to observe forthnightly tidal cycles, I am not sure if it will be feasable to create similar datasets for all other continents around the globe. Maybe it would be fair to note such limitations when discussing the results.

3. Considering the data:

[1]Viewer available at http://www.earthenv.org/cloud

(a) To ease programmatic access, I recommend to rename identifiers to names without spaces and special characters (e.g., "Average # of shoreline points" is often quite clunky to work with"). I would suggest to use rather short and concise identifiers while providing an extensive description with the metadata. Also try to be consistent in naming[2]: currently some identifiers contain capitals, others do not; most contain spaces, yet others are separated by underscores or hyphens. I would consider just using underscores or PascalCase.

(b) Attribute names in Table 1 do not match identifiers used in the dataset. Make sure the attribute names match the identifiers in the dataset exactly.

(c) In general, when presenting a dataset try to stick as much to the FAIR guiding principles for scientific data management (Wilkinson et al., 2016). For instance, it could be considered to make the data accessible (index-able) according to STAC[3] specification.

2 Specific comments

1. Line 84: Not of critical importance, but why not use OSM 2021?

2. Line 112: I believe only the FFT algorithm is not able to deal with unequeally spaced time series, but a regular fourier transformation is able to deal with unequally spaced data. Why not use that one?

3 Technical corrections

Overall length of the article is appropriate, although some sections can be written more to the point: paragraph about the relevance of a beach-face slope dataset in the introduction can be written more concisely; and, some sentences the in the conclusion could be moved to the discussion/introduction.

3.1 Textual

Following are a few textual suggestions and corrections:

1. Line 37: I would suggest to change "coarser/finer" with "steeper/flaat" to "coarser (finer)" and "steeper (flatter)"; seems to be more in accordance with existing literature.

2. Line 50: "In particular" *and* "specifically" is a bit over the top in this sentence. I would suggest to pick one of these. For example, "Recently, Vos et al. (2020) introduced a method to specifically.."

3. Line 55 - 68: I would rewrite this paragraph more concisely, probably leaving the quoted text out; the importance of a beach-slopes datasat can be stressed in a few sentences.

4. Line 75: "Inundation forecasting" to "flood risk modeling"?

[2] For example, see https://stackoverflow.com/questions/7662/database-table-and-column-naming-conventions
[3] https://stacindex.org

5. Line 92: "This method is described in detail in Vos et al. (2020) *and* combines.." Conjunction feels a bit weird here. Maybe better to "This method, which is described in detail in Vos et al. (2020), combines .."
6. Line 107-8: Would either leave "open-source" or "publicly" out because they imply each other. Maybe something like "Time-series were obtained with CoastSat, a toolbox publicly available at.."
7. Line 111: I would suggest to also include a reference to the founders of the Lomb-Scargle method (Lomb, 1976; Scargle, 1982) as well as to the comprehensive overview provided by VanderPlas (2018).
8. Line 116: Maybe change "from Bujan et al. (2019)" to "following Bujan et al. (2019)"?
9. Line 205: "which means" to "which implies"?
10. Line 258-262: I would leave the results of Vos et al. (2020) out of the conclusion.
11. Line 268-271: Consider to rewrite this sentence. For example "low-confidence areas, in the south-western sector are characterized by . . . , while..". Further, I would suggest to rephrase the sentence so that references are no longer required—would be more appropriate to include such information in introduction or methodology.
12. Line 255-289: The conclusion contains a lot of information which might be more appropriate in a section discussing the results. Consider to use the conclusion to list the most important findings of this study provide and outlook for the future. For example, the effects of aliasing and the reference to Eleveld et al. (2014), Bishop-Taylor et al. (2019) should be discussed earlier. Similarly, the importance of a beach-slope dataset was already discussed extensively in the introduction.
13. Table 1: "id" should be capitalized.
14. Table 1: In the column "Values", I would leave "confidence" out in "High/Medium/Low confidence".
15. Table 1: Consider to describe the ranges in the "Values" section mathematically instead of with words.

References

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, *39*(2), 447–462. https://doi.org/10.1007/BF00648343

Scargle, J. D. (1982). Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, *263*, 835. https://doi.org/10.1086/160554

Eleveld, M. A., van der Wal, D., & van Kessel, T. (2014). Estuarine suspended particulate matter concentrations from sun-synchronous satellite remote sensing: Tidal and meteorological effects and biases. *Remote Sensing of Environment*, *143*, 204–215. https://doi.org/10.1016/j.rse.2013.12.019

Landsat—Earth observation satellites (USGS Numbered Series No. 2015-3081). (2016). U.S. Geological Survey. Reston, VA. https://doi.org/10.3133/fs20153081

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Wilson, A. M., & Jetz, W. (2016). Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions (M. Loreau, Ed.). *PLOS Biology*, *14*(3), e1002415. https://doi.org/10.1371/journal.pbio.1002415

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., & Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment*, *185*, 271–283. https://doi.org/10.1016/j.rse.2015.11.032

VanderPlas, J. T. (2018). Understanding the Lomb–Scargle Periodogram. *The Astrophysical Journal Supplement Series*, *236*(1), 16. https://doi.org/10.3847/1538-4365/aab766

Bishop-Taylor, R., Sagar, S., Lymburner, L., Alam, I., & Sixsmith, J. (2019). Sub-Pixel Waterline Extraction: Characterising Accuracy and Sensitivity to Indices and Spectra. *Remote Sensing*, *11*(24), 2984. https://doi.org/10.3390/rs11242984

Bujan, N., Cox, R., & Masselink, G. (2019). From fine sand to boulders: Examining the relationship between beach-face slope and sediment size. *Marine Geology*, *417*, 106012. https://doi.org/10.1016/j.margeo.2019.106012

Vos, K., Harley, M. D., Splinter, K. D., Walker, A., & Turner, I. L. (2020). Beach Slopes From Satellite-Derived Shorelines. *Geophysical Research Letters*, *47*(14), e2020GL088365. https://doi.org/10.1029/2020GL088365
_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL088365