Enabling coastal data science at continental scale

Calkoen, Floris
floris.calkoen@deltares.nl

Baart, Fedor
fedor.baart@deltares.nl

January 9, 2023

Abstract

The coastal community widely anticipates that in the next years data-driven studies are going to make essential contributions to bringing about long-term coastal adaptation and mitigation strategies at continental scale (e.g., [7, 8, 9]). This expectation is likely triggered by several coastal analyses (e.g., [3, 4, 5]) that showed in recent years how Earth Observation data can be used to monitor and track historical change in the coastal zone at unprecedented spatial scales - suddenly coastal science can be considered as a data-rich research field. These analyses have in common that they use publicly-available and analysis-ready data, hosted at major cloud providers, who simultaneously offer the scientists' computer instances altogether with distributed processing tools that are designed to handle the size of these datasets (e.g., [1]). However, we note that when analyses become more complex, i.e., require specific algorithms, (geospatial) pre- and post-processing or require data products that are not hosted by the cloud providers, the cloud-native, distributed processing workflows are broken, which, in effect, make the analyses impractical at continental scale (e.g., [6]). We believe that the next generation of data-driven coastal models that target continental scales — which are expected to take advantage of the ever-increasing amount of Earth Observation and coastal data by using algorithms that thrive on data — can only be built when: 1) processing workflows are scalable; 2) computations are run proximately to the data; 3) data are available in cloud-optimized formats; 4) and, data are described following standardized metadata specifications.

In this study, we introduce these practices to the coastal research community and showcase the advantages of scalable coastal data analysis by two case studies. In the first example we map building footprints in areas prone to coastal flooding and estimate the assets at risk. For this analysis we chunk a coastal flood-risk map into several tiles and incorporate those into a coastal STAC catalog. The STAC catalogs are then used to index the data assets from providers and are combined using several scalable geospatial operations. In the second example we benchmark instantaneous shoreline mapping using cloud-native workflows against conventional methods. With data-proximate

computing we are able to process xx mb per second, or, in other words, map xx km shoreline per hour, whereas conventional methods [6] are able to map 500 meter per 2 hours. In practice this means that a highly-specialized coastal mapping expedition can be upscaled from regional to global level. Most of the used software packages used in these analyses are part of the community-driven Pangeo project [2], although some additional packages were used for scalable geospatial data analysis and cloud I/O.[1] We will publish our code, including self-explanatory Juypter notebooks, at `https://github.com/openearth/coclico-workbench/notebooks/coastal_mask_with_building_footprints.ipynb`. Such notebooks may enable coastal scientists, researchers and engineers to deliver interoperable and reproducible coastal data products and/or workflows.

To conclude, we foresee that in next years several coastal data products are going to be published, of which some may be considered "big data". To incorporate these data products into the next generation of coastal models, it is urgently required to agree upon protocols for coastal data management. With this study we do not only want to show the advantages of scalable coastal data analysis; we mostly want to encourage the coastal community to start a dialogue on how to manage our coastal data in an era of exponential data growth.

References

[1] Noel Gorelick et al. "Google Earth Engine: Planetary-scale Geospatial Analysis for Everyone". In: *Remote Sensing of Environment* 202 (2017), pp. 18–27. DOI: 10.1016/j.rse.2017.06.031. URL: https://doi.org/10.1016/j.rse.2017.06.031.

[2] Joseph Hamman, Matthew Rocklin, and Ryan Abernathy. "Pangeo: A Big-data Ecosystem for Scalable Earth System Science". In: (Apr. 1, 2018), p. 12146. URL: https://ui.adsabs.harvard.edu/abs/2018EGUGA..2012146H (visited on 09/26/2022).

[3] Arjen Luijendijk et al. "The State of the World's Beaches". In: *Sci Rep* 8.1 (Dec. 2018), p. 6641. ISSN: 2045-2322. DOI: 10.1038/s41598-018-24630-6. URL: http://www.nature.com/articles/s41598-018-24630-6 (visited on 09/29/2021).

[4] Lorenzo Mentaschi et al. "Global Long-Term Observations of Coastal Erosion and Accretion". In: *Sci Rep* 8.1 (Dec. 2018), p. 12876. ISSN: 2045-2322. DOI: 10.1038/s41598-

[1] The Planetary Computer Python container contains all necessary software packages, which are listed at `https://github.com/microsoft/planetary-computer-containers/blob/main/python/environment.yml`

018-30904-w. URL: http://www.nature.com/articles/s41598-018-30904-w (visited on 09/29/2021).

[5] Nicholas J. Murray et al. "The Global Distribution and Trajectory of Tidal Flats". In: *Nature* 565.7738 (2018), pp. 222–225. DOI: 10.1038/s41586-018-0805-8. URL: https://doi.org/10.1038/s41586-018-0805-8.

[6] Kilian Vos et al. "CoastSat: A Google Earth Engine-enabled Python Toolkit to Extract Shorelines from Publicly Available Satellite Imagery". In: *Environmental Modelling & Software* 122 (Dec. 1, 2019), p. 104528. ISSN: 1364-8152. DOI: 10.1016/j.envsoft.2019.104528. URL: https://www.sciencedirect.com/science/article/pii/S1364815219300490 (visited on 12/08/2021).

[7] Roshanka Ranasinghe. "On the Need for a New Generation of Coastal Change Models for the 21st Century". In: *Scientific Reports* 10.1 (1 Feb. 6, 2020), p. 2010. ISSN: 2045-2322. DOI: 10.1038/s41598-020-58376-x. URL: https://www.nature.com/articles/s41598-020-58376-x (visited on 12/07/2021).

[8] Ian L. Turner et al. "Satellite Optical Imagery in Coastal Engineering". In: *Coastal Engineering* 167 (Aug. 1, 2021), p. 103919. ISSN: 0378-3839. DOI: 10.1016/j.coastaleng.2021.103919. URL: https://www.sciencedirect.com/science/article/pii/S037838392100079X (visited on 01/17/2022).

[9] Sean Vitousek et al. "The Future of Coastal Monitoring through Satellite Remote Sensing". In: *Cambridge Prisms: Coastal Futures* (Nov. 28, 2022), pp. 1–43. ISSN: 2754-7205. DOI: 10.1017/cft.2022.4. URL: https://www.cambridge.org/core/product/identifier/S275472052200004X/type/journal_article (visited on 12/07/2022).