A novel data ecosystem for coastal analyses

Calkoen, Floris
floris.calkoen@deltares.nl

Baart, Fedor
fedor.baart@deltares.nl

Kras, Etienne
etienne.kras@deltares.nl

Luijendijk, Arjen
texttt arjen.luijendijk@deltares.nl

January 10, 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 view is also supported by CoCliCo[1], a Horizon 2020 project, where coastal data form the fundamental building block for the primary deliverable, an open-web portal that aims to improve decision making on coastal risk management and adaptation. The promise of data is likely triggered by several coastal analyses (e.g., [4, 5, 6]) that showed how Earth observation data can be used to monitor the coastal zone at unprecedented spatial scales - suddenly coastal science can be considered as a data-rich research field. Analyses akin typically use analysis-ready data, hosted at major cloud providers, who simultaneously offer the scientists' computer instances with distributed processing tools to handle the size of the data sets (e.g., [2]). However, we note that when analyses become more complex, i.e., require specific algorithms, pre- and post-processing or include data that are not hosted by the cloud provider, the cloud-native, distributed processing workflows are often broken, which, in effect, make the analyses impractical at continental scale.

We believe that the next generation of data-driven coastal models that target continental scales can only be built when: 1) processing workflows are scalable; 2) computations are run in proximity 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 by showcasing the advantages of cloud-native workflows [1] 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

[1]https://coclicoservices.eu/about/

into a coastal SpatioTemporal Asset Catalog (STAC). The second example benchmarks instantaneous shoreline mapping using cloud-native workflows against conventional methods. With data-proximate computing, processing time is reduced from the order of hours (e.g., [10]) to seconds per shoreline km, which means that a highly-specialized coastal mapping expedition can be upscaled from regional to global level.

The analyses mostly rely on "core-packages" from the Pangeo project, [3], with some additional support for scalable geospatial data analysis and cloud I/O, although they can essentially be run on a standard Python Planetary Computer instance. We will publish our code, including self-explanatory Juypter notebooks, at `https://github.com/floriscalkoen/egu2023`.

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 stewardship. With this study we do not only want to show the advantages of scalable coastal data analysis; we mostly want to encourage the coastal research community, beyond CoCliCo, to adopt FAIR data management principles [11] and workflows in an era of exponential data growth.

References

[1] Ryan P. Abernathey et al. "Cloud-Native Repositories for Big Scientific Data". In: *Computing in Science & Engineering* 23.2 (Mar. 1, 2021), pp. 26–35. ISSN: 1521-9615, 1558-366X. DOI: `10.1109/MCSE.2021.3059437`. URL: `https://ieeexplore.ieee.org/document/9354557/` (visited on 07/18/2022).

[2] 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`.

[3] 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).

[4] 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).

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

[6] 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.

[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).

[10] 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).

[11] Mark D. Wilkinson et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship". In: *Sci Data* 3.1 (Dec. 2016), p. 160018. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18. URL: http://www.nature.com/articles/sdata201618 (visited on 09/08/2021).