Dynamic Coastal Mapping Using Sentinel-1 and Sentinel-2 Data Through Digital Earth Africa

Qingxiang Liu [1], Fang Yuan[1], Caitlin Adams [1], Lisa-Maria Rebelo[2], Michael Wellington[3]

[1] FrontierSI, 727 Collins St. Melbourne VIC, 3008, Australia - lliu, fyuan,cadams@frontiersi.com.au
[2] Digital Earth Africa, Pretoria, South Africa - lisa.rebelo@digitalearthafrica.org
[3] Geoscience Australia, Canberra, Australia - Michael.wellington@digitalearthafrica.org

Keywords: Coastal Mapping, Digital Earth Africa, Sentinel-1, Sentinel-2, Image Classification.

Abstract

Coastal erosion poses a continuous threat to ecosystems, infrastructure, and property. To address these challenges and mitigate the effects of coastal changes, effective and current monitoring is essential. It is particularly important to monitor coastlines and coastal changes in Africa, where a significant portion of the population resides in coastal regions. While optical satellite imagery has been used for large-scale annual coastlines and change monitoring for Africa, its availability and quality are largely limited by the presence of cloud and cloud shadow. In comparison, using radar satellite observations such as Sentinel-1 data can provide consistent coastal mapping and change detection regardless of cloud presence.

This paper outlines a fully automated supervised machine learning workflow using Sentinel-1 data and training samples extracted from Sentinel-2 data. It also explores the performance of the workflow for different coastal morphology types across the African coast. The workflow has proved to perform better and produced results that were visually more consistent with Sentinel-2 data compared to thresholding methods. While challenges exist to distinguish between land and water over smooth sandy beaches and rough near-shore water surfaces, our workflow provides an alternative method for coastal change mapping where optical satellites provide insufficient observations free from clouds. Python code of the proposed methodology has been made publicly available.

1. Introduction

1.1 Coastal Changes and SDGs

Coastal changes, especially erosion, are an ongoing threat to ecological habitats, infrastructure, and properties. The Sustainable Development Goals (SDGs) 14 and 15 outline the importance of protecting coastal ecosystems and halting coastal land degradation and biodiversity loss. To achieve the goals and help minimise the impacts of coastal changes, efficient and up-to-date coastal change monitoring is needed. Monitoring coastlines and coastal changes for Africa is important as a large population is hosted by coastal areas.

1.2 Earth Observation for Coastal Mapping

Optical satellite imagery, typically from Landsat and Sentinel-2, has been used for large-scale annual coastlines and changes monitoring for Africa (Bishop-Taylor et al., 2021; Digital Earth Africa, 2021a). However, the presence of cloud and cloud shadow largely limited the availability and quality of data. In comparison, radar satellite observations such as those from Sentinel-1 provide valid observations regardless of cloud presence. By developing a workflow to implement land/water classification with Sentinel-1 data, it is possible to identify shorelines and detect coastal changes even in cloudy regions.

Currently, most of the existing research work on water detection and shoreline mapping using Sentinel-1 data employed an unsupervised threshold-based method (Li et al., 2020; Liang and Liu., 2020; Pelich et al., 2020; Chen and Zhao, 2022). However, the performance of the method depends largely on the histogram distribution, or the proportion of land/water pixels within the defined area of interest. Besides, thresholding-based methods mostly only use single band of the Sentinel-1 data.

Supervised image classification including stochastic methods (Tan, 2023) and deep learning based methods (Philipp et al., 2022; Zhang et al., 2022) have also been proposed and employed. However, the methods mostly require in-situ or manually digitised training data, which can be challenging to obtain, especially at a continental scale. In addition, current studies are mostly focused on a local study area without analysis at a continental scale.

Aiming for a more robust fully automated method for coastal classification and change mapping, this paper outlines a supervised machine learning workflow using Sentinel-1 data and training samples extracted from Sentinel-2 data. It also explores the performance of the workflow for different coastal morphology types across the African coast.

1.3 Contribution of This Work

In summary, this work makes the following contributions to the community:

- A supervised workflow of annual coastal mapping using Sentinel-1 as the main data source is proposed. Instead of requiring ground-surveyed training data, it automatically extracts training data from Sentinel-2 data. It has shown better performance than a traditional unsupervised thresholding method.

- The workflow accounts for the dynamic nature of the shorelines through tidal filtering and the use of backscatter temporal variability as features.
- We explored the performance of the workflow for different coastal geomorphology types across African coast and identified advantages and challenges of the workflow relevant to the types.
- The Python code for the workflow is available to the public. It is based on Digital Earth (DE) Africa Analysis Sandbox, a free cloud computing environment.

The structure of the remainder of the paper is as follows. Section 2 introduces our proposed workflow and experiment data used. Results, comparisons and discussions are presented in Section 3. Lastly, Section 4 concludes the paper and outlines future work.

2. Data and Methodology

2.1 Experiment Locations and Data

150 locations (50 for each of the three coastal geomorphology types, i.e. Wetland, Beach and Bedrock) were randomly selected across the African coast (Figure 1), using the coastal geomorphology data (Mao et al., 2022). The spatial extent for each selected location is a square of approximately 2 km by 2 km.

Sentinel-1 Radiometrically Terrain Corrected (RTC) Gamma-0 normalised radar backscatter data (Digital Earth Africa, 2021b; Yuan et al., 2022) at 20 m spatial resolution from 2018 to 2022 were acquired for all the locations as the main data for the workflow. Sentinel-2 Level 2A Surface Reflectance bands acquired within the same period and locations were also used in the workflow, and were resampled to the same spatial resolution as the Sentinel-1 data. Both datasets were sourced from DE Africa platform which provides various satellite imagery and derived products for Africa. The workflow was developed on the DE Africa Analysis Sandbox, a free cloud computing environment.

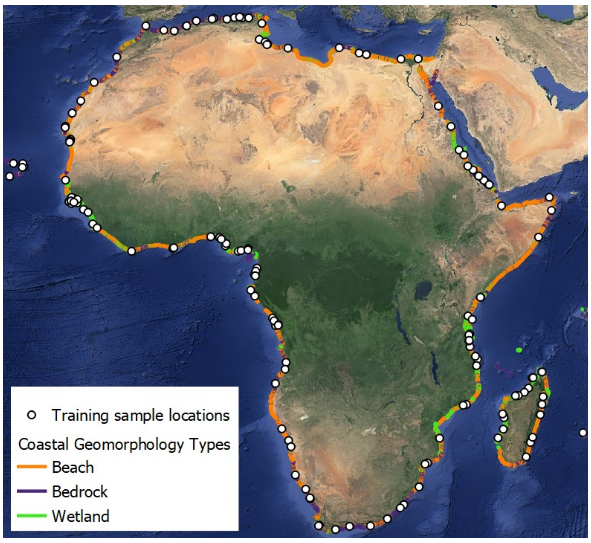


Figure 1. Experiment locations map. Basemap: Google Satellite.

2.2 Annual Coastal Mapping Workflow

The annual coastal mapping workflow is illustrated in Figure 2. Python scripts and Jupyter Notebooks of the methodology have been made publicly accessible on GitHub: https://github.com/frontiersi/DEAfrica_coastlines_s1. The workflow can be divided into five major steps as follows.

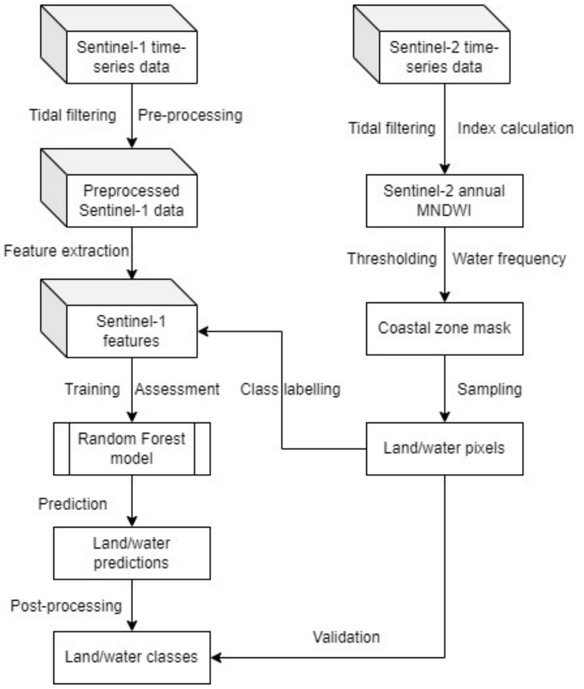


Figure 2. Flowchart of proposed annual coastal mapping workflow.

2.2.1 Sentinel-1 Data Pre-processing

A few pre-processing steps were applied to Sentinel-1 data: tidal filtering, orbit track filtering, and conversion to decibel (dB). Firstly, a per-pixel filtering was implemented to keep only observations from the dominant orbit direction, i.e. either ascending or descending. This was expected to minimise the effects of inconsistent looking angles and obit directions. Secondly, to reduce effects of extreme tides, only Sentinel-1 observations within 25 percent of entire tidal range below or above Mean Sea Level were kept (Bishop-Taylor et al., 2019). The FES 2014 global tide model (Carrere et al., 2015) was employed. The filtered observations were then converted to dB.

2.2.2 Land/Water Training Samples Identification

Training samples of the two classes of interest, i.e., water and land, were identified from Sentinel-2 data. The tidal filtering applied to Sentinel-1 was also applied. To further improve reliability of the training samples, experiment locations with median tidal height difference between Sentinel-1 and -2 observations higher than 0.5 m were excluded from the analysis. The Modified Normalised Difference Water Index (MNDWI) (Xu, 2006) was calculated for each Sentinel-2 observation to distinguish between land and water, based on the band calculation equation:

$$MNDWI = \frac{Green - SWIR1}{Green + SWIR1}, \quad (1)$$

where Green and SWIR1 represent green and short-wave infrared 1 bands. Annual median composites of the MDNWI were then calculated to reduce cloud/cloud shadow effects of single-time acquisitions. A pixel was identified as water when the annual median MNDWI was higher than zero, and land class otherwise.

Additionally, a simple coastal zone mask was calculated based on annual frequency of being classified as water and morphological processing. Only training samples within the mask were sampled as they were expected to be more helpful in building the classification model than pixels from inland or deep ocean regions. To derive the coastal zone mask, a zero thresholding was first applied to classify each single MNWDI images as land/water, from which the frequency of being classified as water was calculated. An initial coastal mask was derived as pixels with water frequency between 20 percent and 80 percent throughout the study period. A 5-pixel binary dilation of the initial mask was then applied to include adjacent land and water pixels. Finally, connected regions smaller than three pixels were removed. 500 samples per class were randomly extracted for each location.

2.2.3 Sentinel-1 Feature Extraction

Available Sentinel-1 bands include vh and vv polarisation bands, local incidence angle, normalised scattering area and nodata mask bands. In order to explore features that could potentially capture nonlinear and complex patterns of the data, ten features in total were calculated, consisting of:

- *median vh*: annual median of vh band
- *std vh*: annual standard deviation of vh band
- *median vv*: annual median of vv band
- *std vv*: annual standard deviation of vv band
- *median vv+vh*: annual median of the sum of vh and vv bands
- *std vv+vh*: annual standard deviation of the sum of vh and vv bands
- *median vv-vh*: annual median of the difference between vv and vh bands
- *std vv-vh*: annual standard deviation of the difference between vv and vh bands
- *median area*: annual median of normalised scattering area
- *median angle*: annual median of local incidence angle

2.2.4 Model Training and Prediction

A Random Forest model was trained for each location using the extracted training data. The number of trees for the model was set as 200. The fraction of samples used to build each individual tree was set as 0.5. Each trained model was then applied to predict classes and probabilities for all pixels within the location extent.

2.2.5 Post-processing and Assessment

To reduce noise observed from the predicted results, a median filter with 3 by 3 pixels window size was applied to the probability results. Final land/water classification results were derived by thresholding the filtered probability results using a 50 percent threshold. The results were assessed using all pixels within the coastal zone mask. In addition, classification results derived through thresholding the annual median vh band were generated for comparison, using minimum threshold determination method (Glasbey, 1993; Prewitt and Mendelsohn, 1966). The method smooths the histogram until there are only two maxima; the minimum in between is then derived as the threshold value.

3. Results and Findings

3.1 Feature Importance

79 locations were excluded from the analysis due to significant tidal height difference between Sentinel-1 and -2 data or insufficient training samples. Figure 3 shows the overall ordered importance of all features used for the model training across all locations. *Median vh*, *median vh+vv* and *median vv* were found to be the most important features on average. This indicated that other than the medians of the two polarisation bands, the combination of the two bands and standard deviations are also helpful in building the models. *Median angle* is the least important feature.

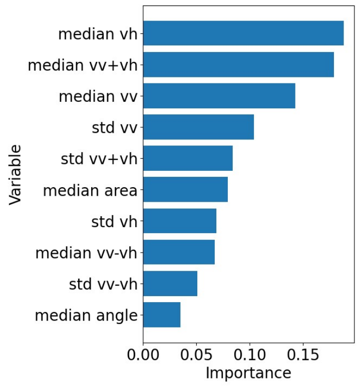
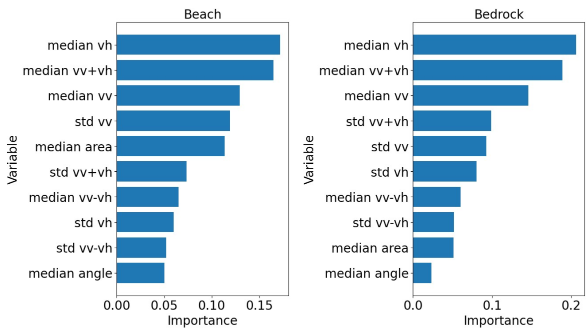


Figure 3: Overall feature importance.

Figure 4 shows feature importance for each of the coastal geomorphology types. Consistent across all geomorphology types, *median vh*, *median vh+vv* and *median vv* were found to be the most important features. Meanwhile, *std vv* is consistently amongst the top four or five important features. In addition, *median angle* is consistently least important to the models, confirming that the incidence angle effect is minimised in the Sentinel-1 RTC backscatter product. Median area was amongst the least important features for Bedrock and Wetland locations, while relatively more important for Beach locations.



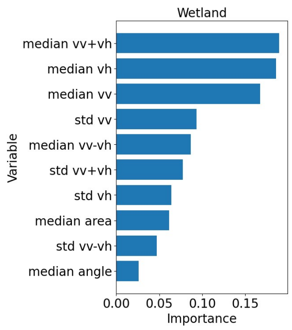


Figure 4. Feature importance by coastal geomorphology type.

3.2 Model Performance

Confusion matrices of the overall classification results are shown as Figure 5, where normalisation was applied to each row. The Overall Accuracy (OA) for the supervised results (0.92) is significantly higher than the thresholding results (0.78). Judging from the recall scores of land (0.82) and water (0.74), the minimum thresholding method tend to misclassify more water as land pixels.

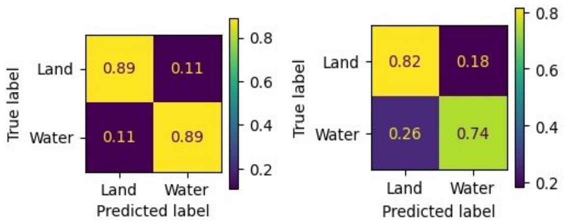


Figure 5. Confusion matrices (normalised along each row) for the supervised results (left) and minimum thresholding results (right).

Improvement of accuracies was also observed for all three coastal geomorphology types (see confusion matrices in Figure 6). OAs for all three types are presented in Table1. The supervised workflow has the best performance for Wetland type (OA=0.92) and worst for Bedrock type (OA=0.87). In contrast, the thresholding method has relatively the best performance for Bedrock type (OA=0.83) and worst performance for Beach type (OA=0.71). This indicates that distribution of pixel values is more consistent and easier to classify through image thresholding for Bedrock coastal regions. The most significant improvement was on Beach locations (from 0.71 to 0.90) compared to Wetland (0.76 to 0.92) and Bedrock (0.83 to 0.87) locations.

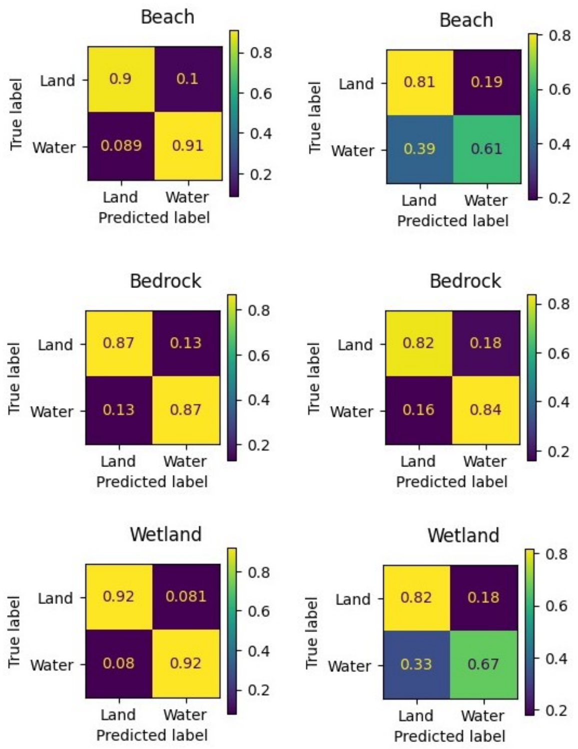


Figure 6. Confusion matrices (normalised along each row) by geomorphology type. Left column: proposed supervised results. Right column: minimum thresholding results.

	Beach	Bedrock	Wetland
Supervised	0.90	0.87	0.92
Thresholding	0.71	0.83	0.76

Table 1. OAs of the proposed supervised workflow and thresholding methods by coastal geomorphology type.

3.3 Classification Results

Figure 7 shows example classification results of year 2021 for the three coastal geomorphology types. Annual median RGB composites of the Sentinel-2 images are included in the figure (top row) for visual reference. Overall, it can be observed that the supervised workflow produced more consistent results with the Sentinel-2 classification than the thresholding method. Note that for some locations (not displayed here) the thresholding method was not able to identify a suitable threshold, resulting in all pixels classified as land. This demonstrates the advantage of the supervised workflow which is more robust.

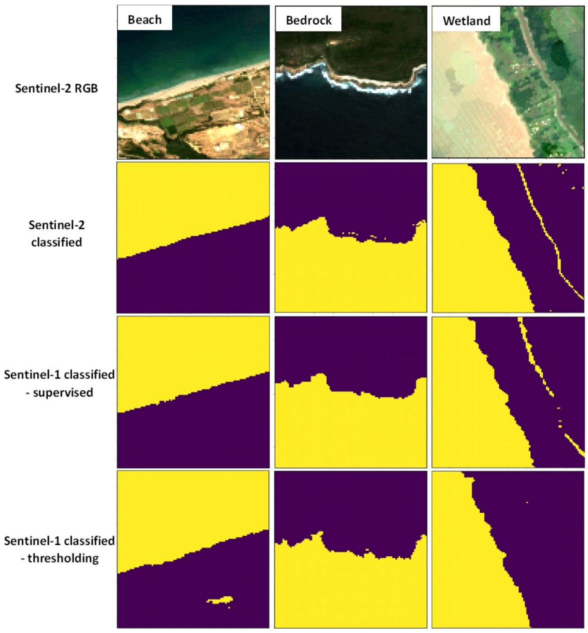


Figure 7. Example classification results. Top row: Annual median RGB composites of Sentinel-2 data. Second row: classified results from Sentinel-2 data. Third and fourth rows: Sentinel-1 classification results using the proposed supervised workflow and minimum thresholding respectively. Dark purple and yellow on the results represent land and water classes respectively. Each column represents an example location for one geomorphology type.

Figure 8 shows a challenging sandy beach example where significant confusion between flat and smooth sandy beaches and water presented on the Sentinel-1 results. The land-water boundaries on the Sentinel-1 results were generally more landward compared to Sentinel-2 image and result, as some land pixels were misclassified as water, e.g. the sandbar as marked in the green circle. This was due to the smooth and flat sandy surface which reflected very low backscattering signals. On the other hand, small regions of water pixels were misclassified as land class even with post-processing applied. This was likely due to the breaking waves of the near shore water.

While the result from thresholding method appears less noisy, more sandy beach pixels were missed in land class prediction, such as the sandbar (marked in green circle) and the landward beach areas (marked in red circle). In comparison, the supervised workflow was able to better classify these areas.

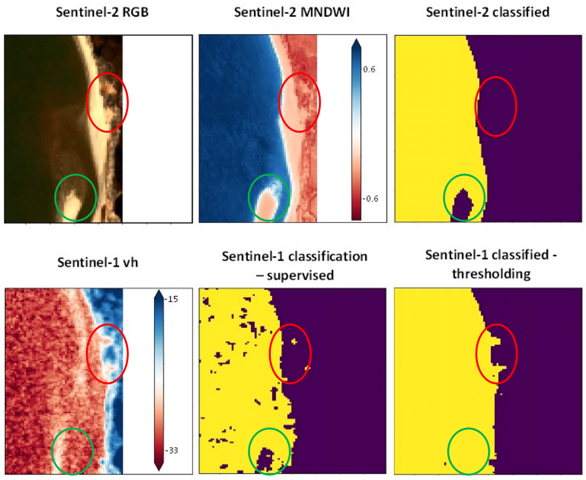


Figure 8. Example classification results of a challenging location for Sentinel-1 data. Red and green circles mark typical areas where misclassification presents on Sentinel-1 results. Dark purple and yellow on the results represent land and water classes respectively.

3.4 Discussions

A global classification model was also explored, with the expectation that a single model could be trained and applied to all locations, which would ensure consistent mapping results regardless of study area. However, significantly poorer results were derived. This was likely due to the complexity of feature distribution across locations with inconsistent feature distributions. Nevertheless, there is a possibility that a more advanced model, or a workflow with other datasets integrated might help improve the performance of a global model.

Spatial filtering was widely applied as one of the pre-processing steps to remove speckling noise on the Sentinel-1 data. In this work no spatial filtering was applied to avoid degradation of the resolution and potentially losing small features in the classification results. Meanwhile, it was observed that the annual aggregation helped to reduce speckle noise. Nevertheless, in the case when the proposed workflow is applied or adapted to seasonal or even finer temporal scale analysis, spatial filtering may be needed.

The workflow extracts training samples from Sentinel-2 data, which avoids the efforts to source or create manually training data. However, it should be noted that misclassification exists in the Sentinel-2 data, which could also affect the reliability of the Sentinel-1 results. Validation of the results using other reference data is encouraged.

Due to the difference in satellite orbits and subsequently different overpassing times, tidal height distributions are different between Sentinel-1 and -2 observations. In this work a significant number of initially randomly selected experiment locations were excluded from the analysis due to significant tidal difference. This indicates that in the annual coastal or shoreline change analysis, tidal bias needs to be considered if comparison of the changes is made between the two datasets.

Our workflow provides an alternative method for coastal change mapping where optical satellites provide insufficient observations free from cloud, e.g. over small tropical islands. While the shorelines to be extracted from Sentinel-1 may be different from Sentinel-2, it is expected that long-term changes can also be identified from the Sentinel-1 timeseries data.

4. Conclusion and Future Work

A supervised workflow was proposed to improve coastal land-water classification on Sentinel-1 data. The supervised workflow had both higher OAs and produced results that were visually more consistent with Sentinel-2 data. As the training labels are extracted from Sentinel-2 classifications, the workflow is fully automated. Nevertheless, it is still challenging to combine or compare Sentinel-1 with Sentinel-2 results, as SAR backscattering signals are affected by various factors including incidence angle, surface roughness and moisture. Consequently, distinguishing between land and water can be problematic over smooth sandy beaches or breaking waves.

Our workflow provides an alternative method for coastal change mapping where optical satellites provide insufficient observations free from cloud, e.g. over small tropical islands.

Future work will include developing a continentally applicable workflow to detect coastal changes and coastlines mapping using Sentinel-1 data. It is expected that spatially adaptive thresholding and multi-band thresholding methods may also be helpful.

Acknowledgements

We thank the European Space Agency for the provision of Sentinel-1 and -2 data. We also appreciate the two anonymous reviewers for providing valuable feedback on the extended abstract.

References

Bishop-Taylor, R., Nanson, R., Sagar, S., Lymburner, L., 2021. Mapping Australia's dynamic coastline at mean sea level using three decades of Landsat imagery. *Remote Sensing of Environment*, 267, 112734. doi.org/10.1016/j.rse.2021.112734

Bishop-Taylor, R., Sagar, S., Lymburner, L. and Beaman, R.J., 2019. Between the tides: Modelling the elevation of Australia's exposed intertidal zone at continental scale. *Estuarine, Coastal and Shelf Science*, *223*, pp.115-128.

Carrere, L., Lyard, F., Cancet, M. and Guillot, A., 2015. FES 2014, a new tidal model on the global ocean with enhanced accuracy in shallow seas and in the Arctic region. In *EGU general assembly conference abstracts* (p. 5481).

Chen, Z. and Zhao, S., 2022. Automatic monitoring of surface water dynamics using Sentinel-1 and Sentinel-2 data with Google Earth Engine. *International Journal of Applied Earth Observation and Geoinformation*, *113*, p.103010.

Digital Earth Africa. 2021a. DE Africa Coastlines: A Coastline Monitoring Service. URL: https://docs.digitalearthafrica.org/en/latest/data_specs/Coastlines_specs.html. Accessed in Mar 2024.

Digital Earth Africa. 2021b. Sentinel-1 Normalised Radar Backscatter. URL: https://docs.digitalearthafrica.org/en/latest/data_specs/Sentinel-1_specs.html. Accessed in Mar 2024.

Glasbey, C.A., 1993. An analysis of histogram-based thresholding algorithms. *CVGIP: Graphical models and image processing*, *55*(6), pp.532-537.

Li, Y., Niu, Z., Xu, Z. and Yan, X., 2020. Construction of high spatial-temporal water body dataset in China based on Sentinel-1 archives and GEE. *Remote Sensing*, *12*(15), p.2413.

Liang, J. and Liu, D., 2020. A local thresholding approach to flood water delineation using Sentinel-1 SAR imagery. *ISPRS journal of photogrammetry and remote sensing*, *159*, pp.53-62.

Mao, Y., Harris, D., & Phinn, S., 2022. Global coastal geomorphology dataset based on machine learning methods. *The University of Queensland*. Data Collection. doi.org/10.48610/f60606a.

Pelich, R., Chini, M., Hostache, R., Matgen, P. and López-Martínez, C., 2020. Coastline detection based on Sentinel-1 time series for ship-and flood-monitoring applications. *IEEE Geoscience and Remote Sensing Letters*, *18*(10), pp.1771-1775.

Philipp, M., Dietz, A., Ullmann, T. and Kuenzer, C., 2022. Automated extraction of annual erosion rates for Arctic permafrost coasts using Sentinel-1, deep learning, and change vector analysis. *Remote Sensing*, *14*(15), p.3656.

Prewitt, J.M. and Mendelsohn, M.L., 1966. The analysis of cell images. *Annals of the New York Academy of Sciences*, *128*(3), pp.1035-1053.

Tan, S., 2023. A New ground open water detection scheme using Sentinel-1 SAR images. *European Journal of Remote Sensing*, p.2278743.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*, *27*(14), pp.3025-3033.

Yuan, F., Repse, M., Leith, A., Rosenqvist, A., Milcinski, G., Moghaddam, N.F., Dhar, T., Burton, C., Hall, L., Jorand, C. and Lewis, A., 2022. An operational analysis ready radar backscatter dataset for the African continent. *Remote Sensing*, *14*(2), p.351.

Zhang, S., Xu, Q., Wang, H., Kang, Y. and Li, X., 2022. Automatic waterline extraction and topographic mapping of tidal flats from SAR images based on deep learning. *Geophysical Research Letters*, *49*(2), p.e2021GL096007.