Title: Seabed Classification in Northwest Fernandina Island using Integrated Acoustic Data and Machine Learning: Comparing Random Forest Performance at 5m vs. 10m Resolution

Authors: Esraa E. Abouelmaaty, Obed Omane Okyere, José Manuel Echevarría Rubio

Date: April 8th, 2025

Abstract:

Accurate seabed characterization is crucial for understanding deep-sea ecosystems, resource management, and habitat mapping. This study focuses on the largely unexplored volcanic terrain of the northwest flank of Fernandina Island, Galápagos Islands, aiming to develop and evaluate machine learning approaches for full-coverage seabed classification. We integrated multibeam echosounder (MBES) bathymetry, backscatter, and derived terrain features (Slope, Aspect, TRI, TPI, Roughness) at two spatial resolutions: 5m and 10m. Supervised Random Forest (RF) classifiers, tuned via GridSearch Cross-Validation, were compared alongside an unsupervised K-Means clustering approach (k=7) for exploratory analysis. Performance was evaluated using ground-truth samples derived from ROV imagery annotation (Overall Accuracy, Kappa, per-class metrics) and feature importance analysis. The 10m resolution RF model yielded slightly higher accuracy (Overall Accuracy: 77.3%, Kappa: 0.73) compared to the 5m model (OA: 74.1%, Kappa: 0.68). For both resolutions, Depth emerged as the most influential feature (>28%), followed by terrain morphology features (TPI notably significant at 10m, Aspect at 5m) and Backscatter, highlighting the combined importance of large-scale morphology, local relief, and sediment properties. K-Means clustering offered complementary insights into natural data groupings. The study demonstrates the utility of integrating multiple data streams within a machine learning framework for robust seabed interpretation in a complex deep-sea environment and provides insights into the impact of data resolution, while identifying areas for refinement concerning ground-truth data and feature optimization.

Keywords: Seabed Classification, Machine Learning, Terrain Analysis, Feature Importance.

1. Introduction

Understanding the composition and structure of the seabed is fundamental for a wide range of marine applications, particularly in largely unexplored deep-sea environments which remain poorly characterized. Benthic habitat mapping, geological resource assessment, marine protected area management, and planning for offshore infrastructure all rely on accurate seabed information (Brown et al., 2011; Harris & Baker, 2012). Acoustic remote sensing, particularly using Multibeam Echosounder Systems (MBES), has become the primary tool for acquiring high-resolution, full-coverage data on seabed depth (bathymetry) and acoustic reflectivity (backscatter intensity) (Lurton & Lamarche, 2015).

This study focuses on the northwest flank of the submarine rift system of Fernandina Island, Galápagos Islands, an area characterized by a variety of volcanic terrains (pillow lavas, collapse pits, volcanic mounds) and distinct biological communities across a depth range of approximately 350m to 1620m. Traditionally, seabed interpretation in such complex areas relied heavily on manual analysis of bathymetry and backscatter mosaics, supplemented by ground-truthing (e.g., ROV video). However, manual interpretation can be subjective, time-consuming, and difficult to apply consistently over large areas (Diesing et al., 2014).

Consequently, there has been a shift towards automated classification techniques. Machine learning (ML) algorithms are increasingly prominent due to their ability to handle complex, high-dimensional datasets and identify non-linear relationships between acoustic features and seabed properties (Li et al., 2020). While bathymetry and backscatter provide primary information, seabed morphology, captured through terrain derivatives (e.g., slope, roughness, TPI), offers crucial contextual information related to physical processes, depositional environments, and habitat structure (Wilson et al., 2007; Lecours et al., 2016). Integrating these morphological features with bathymetry and backscatter within an ML framework holds promise for improving classification accuracy.

However, the impact of input data resolution on classification performance and feature importance in complex deep-sea volcanic settings is less understood. This study aims to fill this gap by addressing the following objectives:

1. Characterize the seabed using multibeam sonar-derived bathymetry and backscatter data.
2. Analyze topographic terrain variables (Slope, Aspect, TRI, TPI, Roughness) to understand microhabitat structure.

3. Utilize annotated ROV still images as ground-truth data.
4. Evaluate the performance of a tuned Random Forest classifier for automated seabed mapping.
5. Compare the impact of input data resolution (5m vs. 10m) on classification accuracy and feature importance.

We apply and compare a supervised Random Forest (RF) classifier, optimized via GridSearchCV, and an unsupervised K-Means clustering algorithm, using integrated datasets at both 5m and 10m resolutions. This comparison evaluates the effectiveness of these approaches and the influence of resolution on capturing seabed heterogeneity within this specific deep-sea volcanic study area.

2. Methods

2.1 Data Acquisition and Study Area

The study area is located on the north flank of the submarine rift system northwest of Fernandina Island, Galápagos Islands. Acoustic data were acquired using hull-mounted shipboard MBES systems (EM714), implying potentially others for deeper waters). The primary datasets included:

- Bathymetry: Provided as GeoTIFF files at 5m and 10m resolution. The bathymetry grids served as the reference spatial framework.
- Backscatter: Provided as GeoTIFF files corresponding to the bathymetry resolutions.
- Ground Truth: Provided as a CSV file containing point coordinates (Longitude, Latitude) derived from ROV dive imagery annotated using BIIGLE software. A total of 292 initial points were provided, representing seven seabed classes: Biogenic mat, Coarse sediment, Coral reef, Coral rubble, Lava flows, Mixed, and Soft sediment.

The study area exhibits diverse volcanic terrains, including pillow lavas, collapse pits, and volcanic mounds, with varying sediment cover and biological features like bamboo corals and Madrepora reefs observed between ~350m and ~1620m depth.

2.2 Data Preprocessing and Feature Engineering

A standardized workflow was implemented using Python (3.12) libraries including GDAL/OGR (via osgeo), Rasterio, GeoPandas, Pandas, and Scikit-learn. The process was executed separately for the 5m and 10m datasets.

1. **Coordinate System and Alignment:** The ground truth points, initially in geographic coordinates (WGS 84, EPSG:4326), were reprojected to the projected coordinate system of the reference bathymetry grids (WGS 84 / UTM Zone 15S, EPSG:32715) using GeoPandas. The backscatter raster for each resolution was spatially aligned (warped) to precisely match the corresponding bathymetry grid (extent, resolution, CRS) using gdal.Warp with bilinear resampling. This ensured pixel-to-pixel correspondence.
2. **Terrain Feature Calculation:** Standard terrain derivatives were calculated from the reference bathymetry raster for each resolution using gdal.DEMProcessing. The following features, known for their relevance to seabed morphology and habitat structure (Wilson et al., 2007; Lecours et al., 2016), were generated:
    - Slope
    - Aspect
    - Terrain Ruggedness Index (TRI)
    - Topographic Position Index (TPI)
    - Roughness
    - Hillshade (for visualization only)

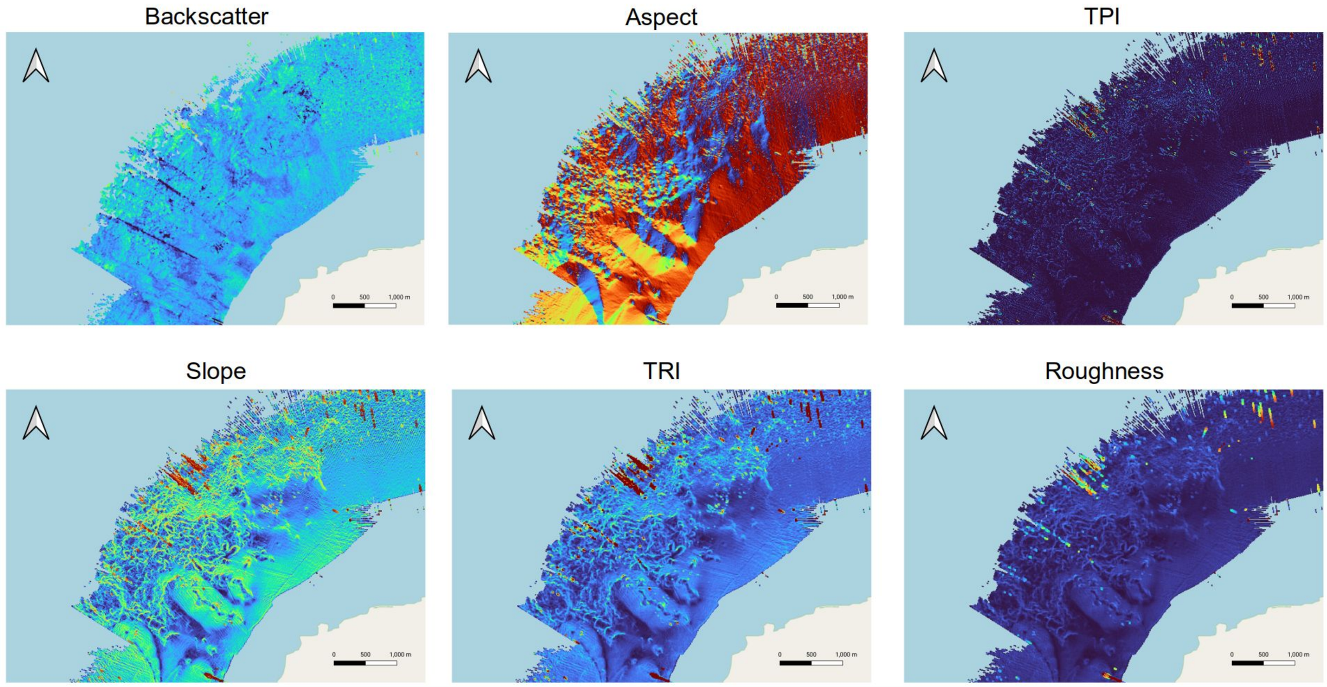

Figure 1. Derived geospatial layers for the study area used as inputs for classification: Backscatter, Aspect, Topographic Position Index (TPI), Slope, Terrain Ruggedness Index (TRI), and Roughness.

3. **Ground Truth Data Extraction:** Feature values from the respective stacked raster (5m or 10m) were extracted at the reprojected ground truth point locations. Points falling on NoData pixels (including the raster's defined NoData value and any NaNs) in *any* feature band were removed. This resulted in 282 valid points for the 5m dataset and 292 valid points for the 10m dataset.

2.3 Classification Approaches

Two distinct machine learning approaches were implemented and compared for both the 5m and 10m datasets:

1. **Supervised Classification: Random Forest (RF)**

    - RF was chosen for its robustness to noise, ability to handle high-dimensional data, inherent feature importance evaluation, and demonstrated high accuracy in remote sensing (Breiman, 2001; Belgiu & Drăguț, 2016).

      Implementation: The valid ground truth dataset (feature values and class labels) for each resolution was split into training (70%) and testing (30%) sets using stratified sampling (random_state=42). String class labels were mapped to integers. A GridSearchCV approach with 5-fold cross-validation was employed on the training set to optimize hyperparameters: n_estimators ([50, 100, 200, 300]), max_depth ([None, 10, 20]), min_samples_split ([2, 5]), and min_samples_leaf ([1, 3]). The class_weight='balanced' option was used to address class imbalance. The best estimator identified by GridSearchCV was trained on the full training set and used to predict classes for the test set and all valid pixels in the corresponding stacked feature raster.

2. **Unsupervised Classification: K-Means Clustering**

    - K-Means was selected as a computationally efficient unsupervised algorithm (MacQueen, 1967) to explore natural groupings within the feature data, independent of ground truth labels, providing a comparative perspective.

Implementation: The feature data for all valid pixels (extracted for RF prediction) was scaled using StandardScaler. K-Means was applied with n_clusters=7. Each valid pixel was assigned to one of the 7 clusters.

2.4 Performance Evaluation

- **Random Forest:** Performance was assessed using the independent test set for each resolution. Metrics included: Overall Accuracy (OA), Cohen's Kappa coefficient, Confusion Matrix, and a Classification Report (Precision, Recall, F1-score per class). Feature importance scores were derived from the trained RF models (Code Output Section 9).
- **K-Means:** Evaluation was primarily qualitative via visual inspection of the resulting cluster maps and their spatial patterns compared to RF maps and input features.

3. Results

3.1 Seabed Classification Maps

Full-coverage classification maps were generated for the study area using the tuned Random Forest model for both 5m and 10m resolutions.

- The Random Forest maps (Figure 2a - 5m RF Map, Figure 2b - 10m RF Map) depict the spatial distribution of the 7 predicted seabed classes based on the ground-truth labels.

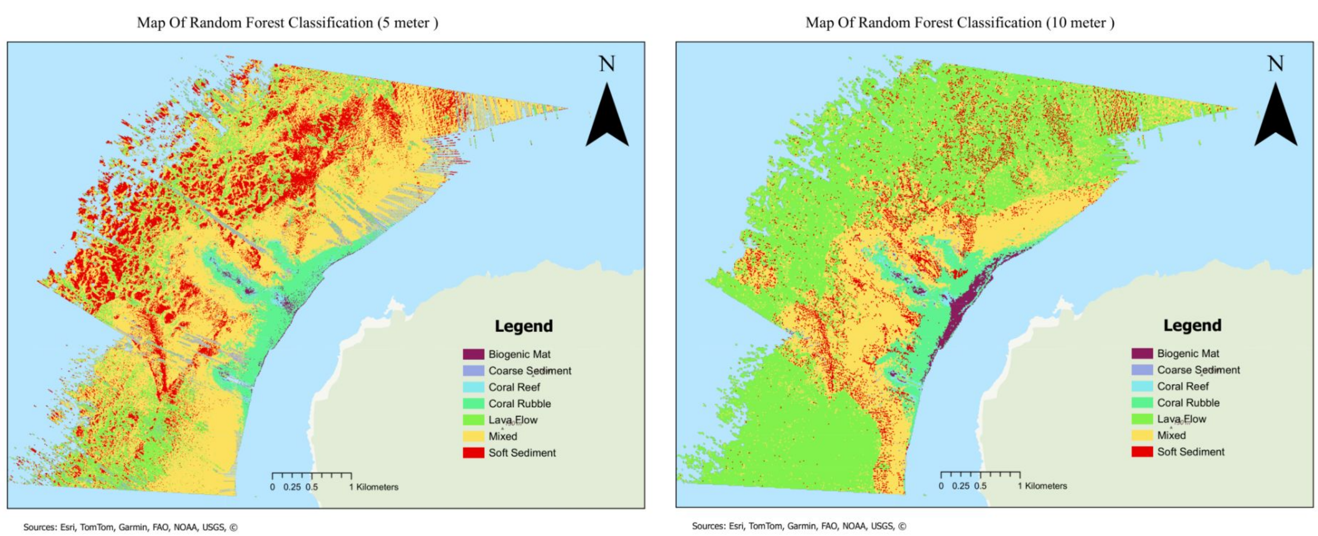

Figure 2. Random Forest classification maps of seabed substrate at 5m (2a) and 10m (2b) resolution for the same coastal area. Note the increased detail and different dominant class distributions (e.g., more Soft Sediment at 5m, larger Lava Flow extent at 10m) resulting from the change in processing resolution.

3.2 Random Forest Performance Metrics

The tuned Random Forest classifiers, evaluated on the respective hold-out test sets, achieved the following performance:

Metric	5m Resolution	10m Resolution
Overall Accuracy	74.12%	77.27%
Cohen's Kappa	0.683	0.727
Training Set Size	197	204
Test Set Size	85	88

Confusion Matrix (5m Resolution)

```
[ 7  0  0  2  0  0  0]
[ 0  4  0  0  0  1  0]
[ 1  0  9  2  0  0  0]
[ 0  0  3 11  0  1  0]
[ 0  0  0  0  8  0  0]
[ 1  0  0  1  4 19  2]
[ 0  0  0  0  0  4  5]
```

Confusion Matrix (10m Resolution)

```
[ 6  0  1  2  0  0  0]
[ 0  5  0  0  0  0  0]
[ 0  0 10  2  0  0  0]
[ 0  0  2 13  0  0  0]
[ 0  0  0  0 11  0  0]
[ 0  0  0  0  6 17  4]
[ 0  0  0  0  1  2  6]
```

The 10m resolution model showed slightly higher overall accuracy and kappa. Class performance varied, with 'Lava flows' and 'Coarse sediment' generally well-classified, while 'Mixed', 'Biogenic mat', and 'Soft sediment' showed more confusion, particularly in the 5m results.

3.3 Feature Importance

Feature importance analysis from the Random Forest models highlights the relative contribution of each variable.

Table 5: Feature Importance Comparison

Feature	Importance (5m)	Rank (5m)	Importance (10m)	Rank (10m)
Depth	28.43%	1	29.22%	1
Aspect	14.87%	2	10.05%	7
TRI	12.68%	3	11.96%	4
Roughness	12.10%	4	10.98%	6
TPI	11.73%	5	14.32%	2
Backscatter	10.10%	6	12.46%	3
Slope	10.09%	7	11.01%	5

Depth was consistently the most influential feature (>28%) for both resolutions. Notably, **TPI** became significantly more important at 10m (Rank 2) compared to 5m (Rank 5), suggesting larger-scale topographic context (ridges/valleys) was more informative at the coarser resolution. Conversely, **Aspect** was more important at 5m (Rank 2) than 10m (Rank 7), possibly indicating that finer-scale slope orientation details were more variable or relevant at 5m. Backscatter held moderate importance (Rank 3-6), confirming its value alongside morphology. TRI, Roughness, and Slope showed moderate and relatively consistent importance across resolutions.

4. Discussion

4.1 Seabed Heterogeneity Capture and Resolution Impact

The supervised Random Forest approach, integrating bathymetry, backscatter, and terrain derivatives, successfully captured significant seabed heterogeneity, achieving good overall accuracy (>74% for 5m, >77% for 10m). The finding that the **10m resolution data yielded slightly higher classification accuracy** metrics in this specific study is noteworthy. Potential reasons include:

- **Feature Scale:** Terrain derivatives calculated at 10m might capture broader morphological patterns (ridges, banks, larger textures) more relevant to the scale of the ground truth classes or dominant ecological zones than the finer details resolved at 5m.
- **Noise Reduction:** The coarser 10m resolution might implicitly smooth some finer-scale noise present in the 5m bathymetry or backscatter, potentially leading to more robust feature extraction for the given classes.
- **Ground Truth Scale:** The ground truth, derived from ROV imagery, might represent features or habitats whose characteristic scale aligns better with the 10m grid resolution.

The feature importance results reinforce the value of data integration. While **Depth** was paramount, the substantial contributions from **TPI** (especially at 10m), **Backscatter**, **TRI**, **Roughness**, **Slope**, and **Aspect** (especially at 5m) demonstrate that morphology and acoustic reflectance were crucial for differentiating classes within broad depth zones. The shift in importance of TPI and Aspect between resolutions further underscores how feature relevance can be scale-dependent.

4.2 Potential Sources of Error and Uncertainty

Several factors could influence the results:

1. **Data Quality:** Noise or artifacts in the MBES bathymetry (e.g., CUBE algorithm effects) or backscatter data (e.g., nadir saturation, angular dependence) can propagate into derived features.
2. **Ground Truth Data:** The accuracy, quantity (282-292 points for a large area), spatial distribution, and positional accuracy (ROV navigation) of ground truth samples heavily influence supervised model performance and evaluation. Subjectivity in visual annotation (BIIGLE) also introduces uncertainty. The observed confusion between certain classes (e.g., Mixed, Biogenic Mat, Coral Rubble/Reef, classification reports) might reflect genuine spectral/morphological overlap or limitations in the ground truth scheme.

Areas of higher uncertainty likely exist at class boundaries, in regions of complex small-scale morphology unresolved by the grid, and in areas far from ground-truth samples.

4.3 Context with Primary Literature

The successful application of RF for integrating diverse geospatial datasets aligns with findings in marine remote sensing (Belgiu & Drăguț, 2016). The results strongly support the incorporation of terrain derivatives for improved seabed characterization (Wilson et al., 2007; Lecours et al., 2016). The dominance of depth is common (Brown et al., 2011), but the significant role of morphology (TPI, TRI, Roughness, Aspect, Slope) and backscatter underscores the multi-faceted nature of seabed classification. The slight advantage of the 10m resolution contrasts with some studies where higher resolution always yields better results, highlighting that optimal resolution can be task- and environment-dependent, influenced by feature scale and noise characteristics. Comparing supervised and unsupervised results provides practical insights relevant to discussions on classification confidence (Diesing et al., 2014).

5. Conclusion

This project successfully applied and compared machine learning techniques for seabed classification in a complex deep-sea volcanic environment off Fernandina Island, integrating MBES bathymetry, backscatter, and derived terrain features at 5m and 10m resolutions. The supervised Random Forest classifier, tuned using GridSearchCV, achieved good performance (OA: 74.1%-77.3%, Kappa: 0.68-0.73), demonstrating the effectiveness of the integrated multi-feature approach. Feature importance analysis confirmed Depth as the primary driver but highlighted the significant, scale-dependent contributions of morphological features (TPI, Aspect, TRI, Roughness, Slope) and Backscatter. **Critically, the 10m resolution dataset yielded slightly higher classification accuracy than the 5m dataset in this study**, suggesting that for this environment and classification task, the broader patterns captured at 10m may be more informative or less affected by noise than finer details. Comparison with unsupervised K-Means clustering provided a complementary perspective on data structure. While the RF maps provide robust interpretations, uncertainties related to ground-truth limitations and potential feature overlaps were identified, suggesting pathways for future improvements, including targeted ground-truthing and further feature exploration.

6. Acknowledgements

The project team thanks the providers of the MBES and ROV datasets for the Northwest Fernandina Island study area.

7. References

- Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31.
- Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.
- Brown, C. J., Smith, S. J., Lawton, P., & Anderson, J. T. (2011). Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, *92*(3), 502–520.
- Diesing, Markus & Green, S. & Stephens, David & Lark, R. & Stewart, Heather & Dove, Dayton. (2014). Mapping seabed sediments: Comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Continental Shelf Research. 84. 107-119. 10.1016/j.csr.2014.05.004.
- Fonseca, Luciano & Brown, Craig & Calder, Brian & Mayer, Larry & Rzhanov, Yuri. (2009). Angular range analysis of acoustic themes from Stanton Banks Ireland: A link between visual interpretation and multibeam echosounder angular signatures. Applied Acoustics - APPL ACOUST. 70. 10.1016/j.apacoust.2008.09.008.
- Harris, P. T., & Baker, E. K. (Eds.). (2012). *Seafloor Geomorphology as Benthic Habitat: GeoHAB Atlas of Seafloor Geomorphic Features and Benthic Habitats*. Elsevier.
- Lecours, Vincent & Dolan, Margaret & Micallef, Aaron & Lucieer, Vanessa. (2016). A review of marine geomorphometry, the quantitative study of the seafloor. Hydrology and Earth System Sciences. 20. 3207-3244. 10.5194/hess-20-3207-2016.
- Lurton, X., & Lamarche, G. (Eds.). (2015). *Backscatter Measurements by Seafloor-Mapping Sonars: Guidelines and Recommendations*. GEBCO/Seafloor Mapping.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, *1*(14), 281–297.
- Wilson, M. F. J., O'Connell, B., Brown, C., Guinan, J. C., & Grehan, A. J. (2007). Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. *Marine Geodesy*, *30*(1–2), 3–35.