3rd Edition

Data Research meetup by MagIC

Improving nearshore bathymetry mapping through spatially adaptive machine learning

Telmo Geraldes Dias

NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal.

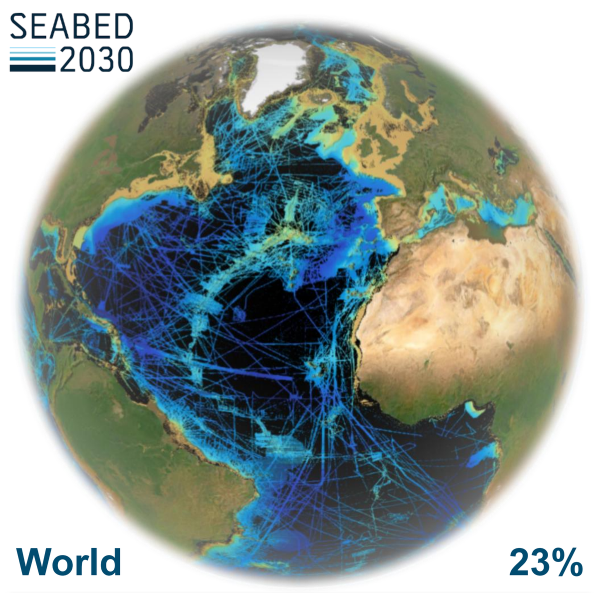
Instituto Hidrográfico (IH), Rua das Trinas 49, 1249-093 Lisboa, Portugal.

1 INTRODUCTION

➥ **Bathymetry** plays a pivotal role in **ocean science** and **blue economy** applications.

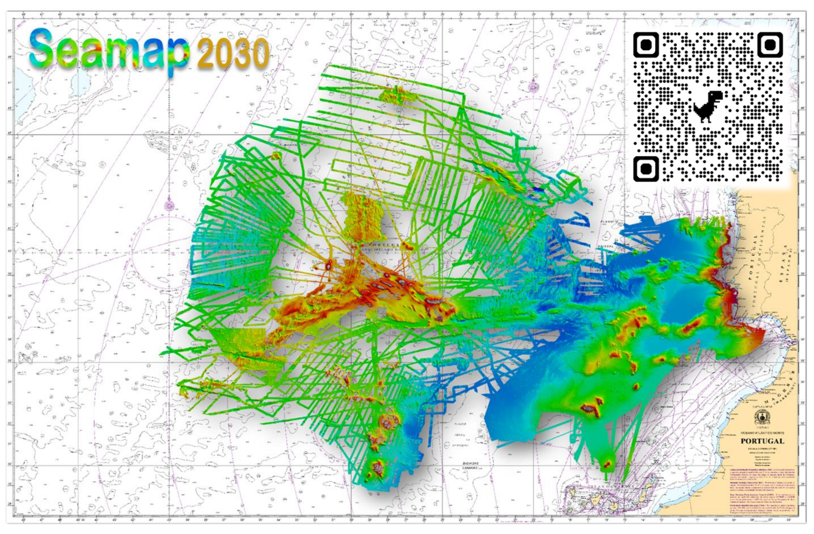
- 40% **World population** lives near the **coast**
- 80% **Global trade** transported by **sea**
- 99% **Communications** by **submarine** cables

➥ What is the current state of **seafloor mapping**?



World 23%

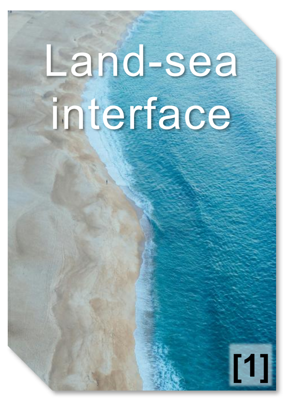
https://seabed2030.org/



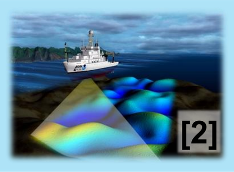
Portugal 66%

https://geomar.hidrografico.pt/

➥ The **land-sea interface** remains significantly **under-surveyed**.



In situ observations:
- Adverse sea conditions;
- Low efficiency.

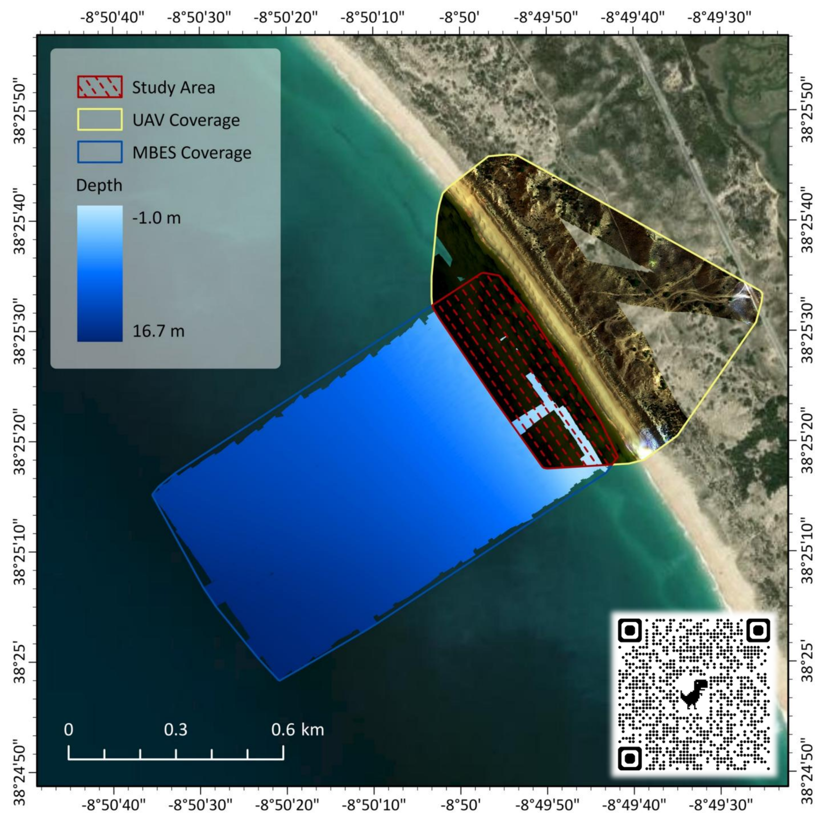


Remote sensing:
- LiDAR (expensive);
- Satellite (low resolution).

Unmanned Aerial Vehicle (UAV):
- Efficient and flexible;
- High spatial resolution;
- Optically derived bathymetry;
- Statistical/empirical approach;
- Machine learning (ML) techniques: GRF [4] and RF.

2 MATERIALS & METHODS

➥ **Study area:**



➥ **Data collection:**

Surface Survey Vessel
Multibeam Echosounder (MBES)

Unmanned Aerial Vehicle
Multispectral Camera

➥ **Methods:**

Data Acquisition → **Data Processing** → **ESDA** → **Model Development** → **Model Evaluation**

Data Acquisition
- *In situ* – MBES (0.25 m)
  Depth (target variable)
- Remote sensing – UAV (0.15 m)
  Multispectral bands (input features)
  (blue, green, red, red edge, near-infrared)

Data Processing
- Resampling
  1 660 412 data points (0.25 m)
  33 883 data points (1.75 m)

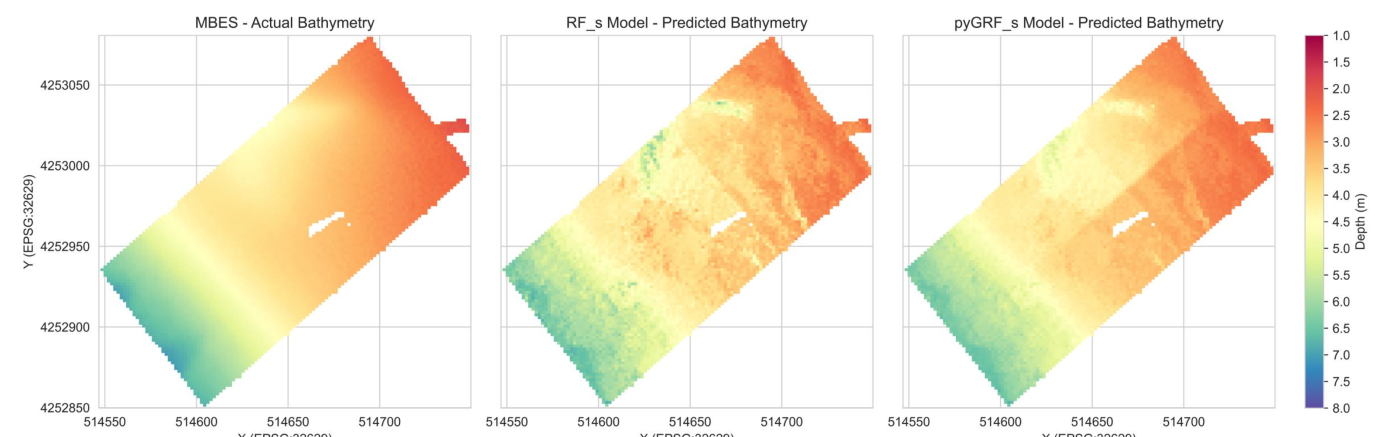
ESDA
- Global and spatial distribution
- Global correlation
  Pearson's (r) / Chatterjee's (ξ)
- Local bivariate relationships

Model Development
- RF / pyGRF (random split) – Baseline
- RF / pyGRF (geospatial split) – Evaluation
- Incremental spatial autocorrelation
- Geographically weighted regression
- Grid search

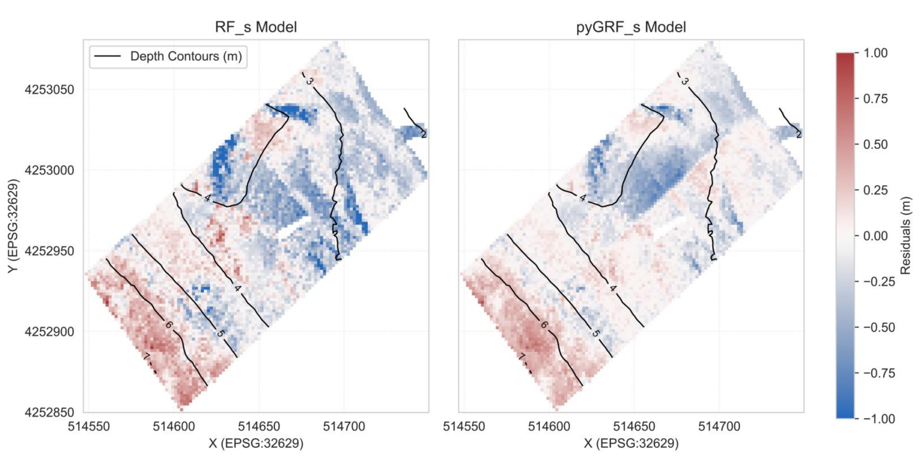
Model Evaluation
- Global metrics (MAE, MAPE, RMSE, R^2)
- Scatter plots (actual/predicted depth, residuals)
- Global and spatial distribution (residuals)
- Global and local feature importance
- Local R^2

3 RESULTS & DISCUSSION

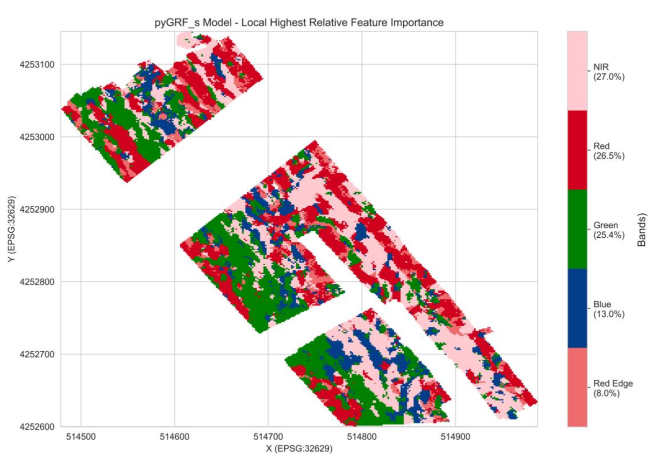
➥ **pyGRF** [5] improved all global metrics (MAE, MAPE, RMSE, R^2).



➥ Change of **residuals** sign (≈ 5 m depth).



➥ **Local** feature importance.



4 CONCLUSION

➥ **Contributions:**

- **Demonstrating** that Geographical Random Forest (GRF) improves **accuracy and interpretability** over a conventional Random Forest (RF) by capturing **spatially varying reflectance-depth relationships**;
- **Mapping** spatial variations in **feature importance** that remain hidden in global models:
  - **Red band**, globally the most influential predictor;
  - **Near-infrared band (NIR)**, locally dominant in shallow areas;
  - **Green band**, most relevant in the depth range with highest accuracy;
- **Identifying** a practical **optical depth limit** of approximately **5 m** for the study area, supported by multiple spatial diagnostics;
- **Providing** a **reproducible**, **spatially explicit** modelling **framework** suitable for broader application in coastal bathymetry.

➥ **Limitations:**

- Higher **computational cost**;
- Possible **prediction discontinuities** from spatial weighting.

➥ **Future work:**

- Extend the study to **other coastal regions** (e.g., cross-site testing, regional variability);
- Compare other **geospatial machine learning** models (e.g., XGBoost).

5 REFERENCES

[1] Photo by Sebastian Jacobsen on Unsplash (2024).

[2] Collecting Multibeam Sonar Data, NOAA's National Ocean Service (2009).

[3] Photo by SpaceX on Unsplash (2016).

[4] Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., & Kalogirou, S. (2021). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, *36*(2), 121–136. https://doi.org/10.1080/10106049.2019.1595177.

[5] Sun, K., Zhou, R. Z., Kim, J., & Hu, Y. (2024). PyGRF: An Improved Python Geographical Random Forest Model and Case Studies in Public Health and Natural Disasters. Transactions in GIS. https://doi.org/10.1111/tgis.13248.

6 ACKNOWLEDGEMENTS

This work was developed by the co-authors **Paulo Carvalho** (IH), **Ana Cristina Costa** (NOVA IMS) and **Márcia Lourenço Baptista** (NOVA IMS).

This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the project - UID/04152/2025 - Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS - https://doi.org/10.54499/UID/04152/2025 (2025-01-01/2028-12-31) and UID/PRR/04152/2025 https://doi.org/10.54499/UID/PRR/04152/2025 (2025-01-01/2026-06-30).

Funded by:

NOVA IMS MagIC · NOVA IMS · fct Fundação para a Ciência e a Tecnologia · PRR · República Portuguesa · Financiado pela União Europeia NextGenerationEU

NOVA Information Management School, 18 December 2025

ONLINE VERSION