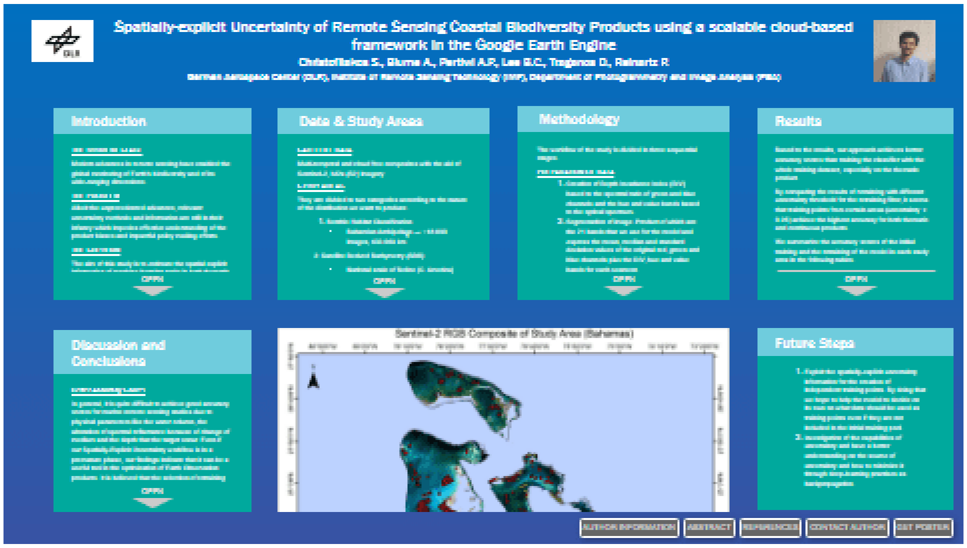
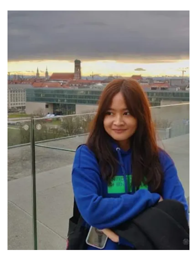
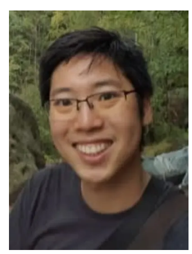
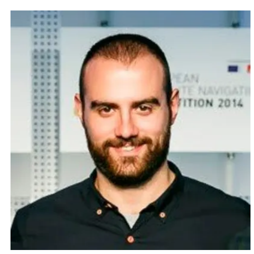
Spatially-explicit Uncertainty of Remote Sensing Coastal Biodiversity Products using a scalable cloud-based framework in the Google Earth Engine



Christofilakos S., Blume A., Pertiwi A.P., Lee B.C., Traganos D., Reinartz P.

German Aerospace Center (DLR), Institute of Remote Sensing Technology (IMF), Department of Photogrammetry and Image Analysis (PBA)



GLOBAL SEAGRASS WATCH
serverless is more

PRESENTED AT:

INTRODUCTION

THE CURRENT STAGE

Modern advances in remote sensing have enabled the global monitoring of Earth's biodiversity and of its wide-ranging dimensions.

THE PROBLEM

Albeit the unprecedented advances, relevant uncertainty methods and information are still in their infancy which impedes effective understanding of the product biases and impactful policy making efforts.

THE SOLUTION

The aim of this study is to estimate the spatial explicit information of machine learning tasks in both thematic and continuous products entirely through the cloud-based platform Google Earth Engine (GEE) and use it to optimize the accuracy of the model and moreover produce a map regarding the uncertainty level of the procedure.

DATA & STUDY AREAS

SATELLITE DATA

Multi-temporal and cloud free composites with the aid of Sentinel-2, lvl2a (S2) imagery.

STUDY AREAS

They are divided to two categories according to the nature of the distribution we want to produce.

1. Benthic Habitat Classification

- Bahamian Archipelago → ~18.800 images, 633.063 km^2

2. Satellite Derived Bathymetry (SDB)

- National scale of Belize (C. America)
- Quirimbas Archipelago (Mozambique)

TRAINING DATA

On the one hand we use the Allen Coral Atlas product in order to export 1000 training points for the 4-class classification and on the other hand, we ingest Ice-Sat2 depth values to our model for the Satellite-derived bathymetry workflow.

VALIDATION DATA

In both cases field data assist as validation dataset. In the case of the habitat mapping the size of the validation data are 300 points per class while for the bathymetry they are 636 and 777 points for Quirimbas and Belize respectively

METHODOLOGY

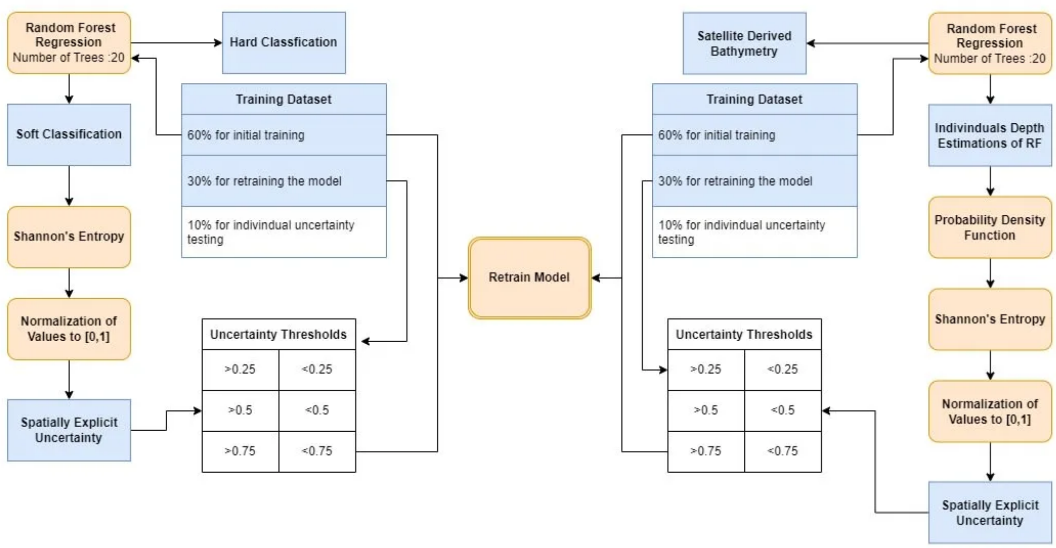
The workflow of the study is divided in three sequential stages.

PREPARATION OF DATA

1. Creation of Depth Invariance Index (DIV) based to the spectral ratio of green and blue channels and the hue and value bands based to the optical spectrum.
2. Segmenation of image. Product of which are the 21 bands that we use for the model and express the mean, median and standard deviation values of the original red, green and blue channels plus the DIV, hue and value bands for each segment.
3. Portional division of the training data to Training, Retraining and Testing.

- 60% of them remain as training data and help with the initial classification.
- 30 % of them are set as re-training data which are the keystone of the next stage.
- The remaining 10 % serves as testing points to measures the alteration of uncertainty due the process

MAIN PROCESSING STAGE



With the information of the classification and the probabilites of succesful classification per class we are able to estimate the spatial uncertainty per pixel. In context of a continuous distribution as a SDB, we estimate the spatially explicit uncertainty with the aid of the Probabily Density function based to the twenty different bathymetry estimations of the twenty trees of the Random Forest algorithm of GEE. Both regression and probabilities values are accessable through the platform. The Retraining data serves as a pool of data that we use for the second classification. By knowing the uncertainty value per pixel, we filter the retraining pool and accept new training data only from areas with low uncertainty.

RETRAINING OF THE MODEL

Last stage of the workflow is the retraining of the model where we use the initial training dataset merged with the retraining data that come from low uncertainty areas.

RESULTS

Based to the results, our approach achieves better accuracy scores than training the classifier with the whole training dataset, especially on the thematic product.

By comparing the results of retraining with different uncertainty threshold for the retraining filter, it seems that training points from certain areas (uncertainty < 0.25) achieve the highest accurasy for both thematic and continuous products.

We summarize the accuracy scores of the initial training and the retraining of the model in each study area in the following tables.

	Initial Classification	Retrained from Uncertain Areas lt(0.25)	Accuracy Gain
Overall Accuracy	57.83%	62.08%	4.25%
User's Accuracy	53.82%	60.30%	6.48%
Producer's Accuracy	54.00%	67.33%	13.33%

Benthic Habitat Classification Accuracy Matrix, Bahamas

Quirimbas

model	Initial Regression	Retrained from Uncertain Areas lt(0.25)	Accuracy Gain
MeanSqrError	2.6328	2.1955	0.4373
r_sqr	0.6289	0.6162	0.0127

Belize

model	Initial Regression	Retrained from Uncertain Areas lt(0.25)	Accuracy Gain
MeanSqrError	1.2306	1.1479	0.0827
r_sqr	0.6104	0.6026	0.0078

Satellite Derived Bathymetry Accuracy Matrix, Quirimbas & Belize

DISCUSSION AND CONCLUSIONS

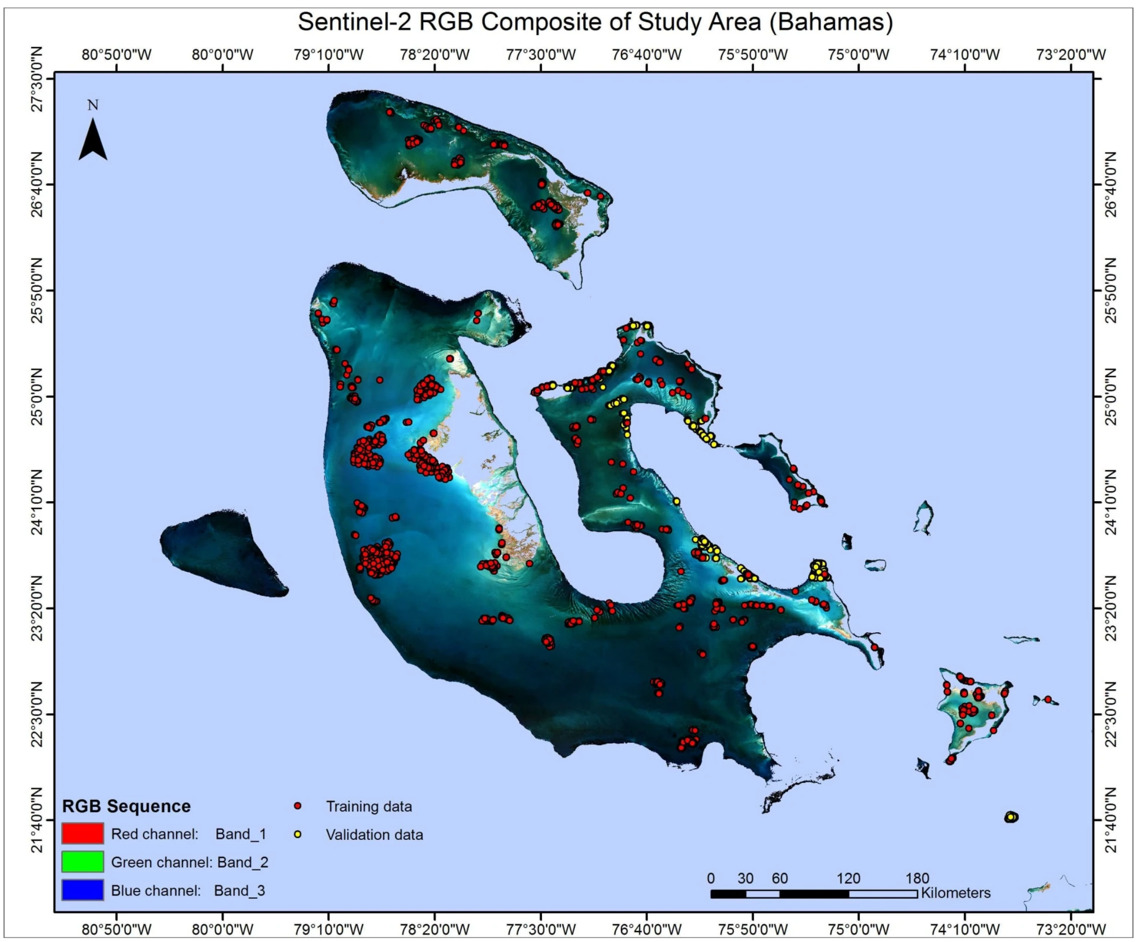
Better Accuracy Scores

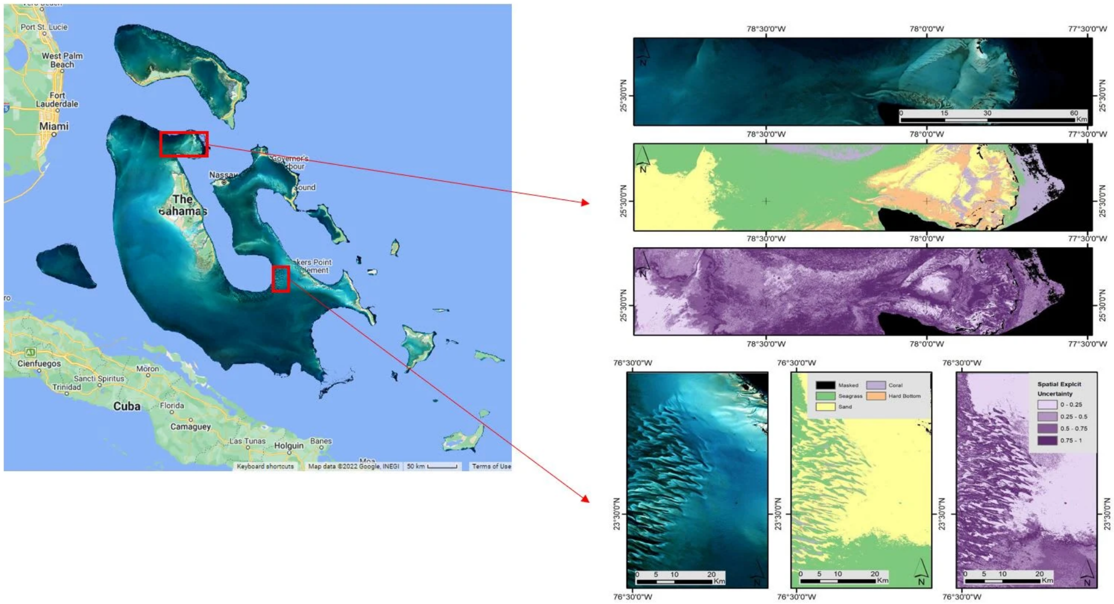
In general, it is quite difficult to achieve good accuracy scores for marine remote sensing studies due to physical parameters like the water column, the alteration of spectral reflactance because of change of medium and the depth that the target occur. Even if our Spatially-Explicit Uncertainty workflow is in a premature phase, our findings indicate that it can be a useful tool in the optimization of Earth Observation products. It is believed that the selection of retraining points with low uncertainty values helps the model to better perform due to the minimized noise that they introduce to the model.

Uncertainty Map

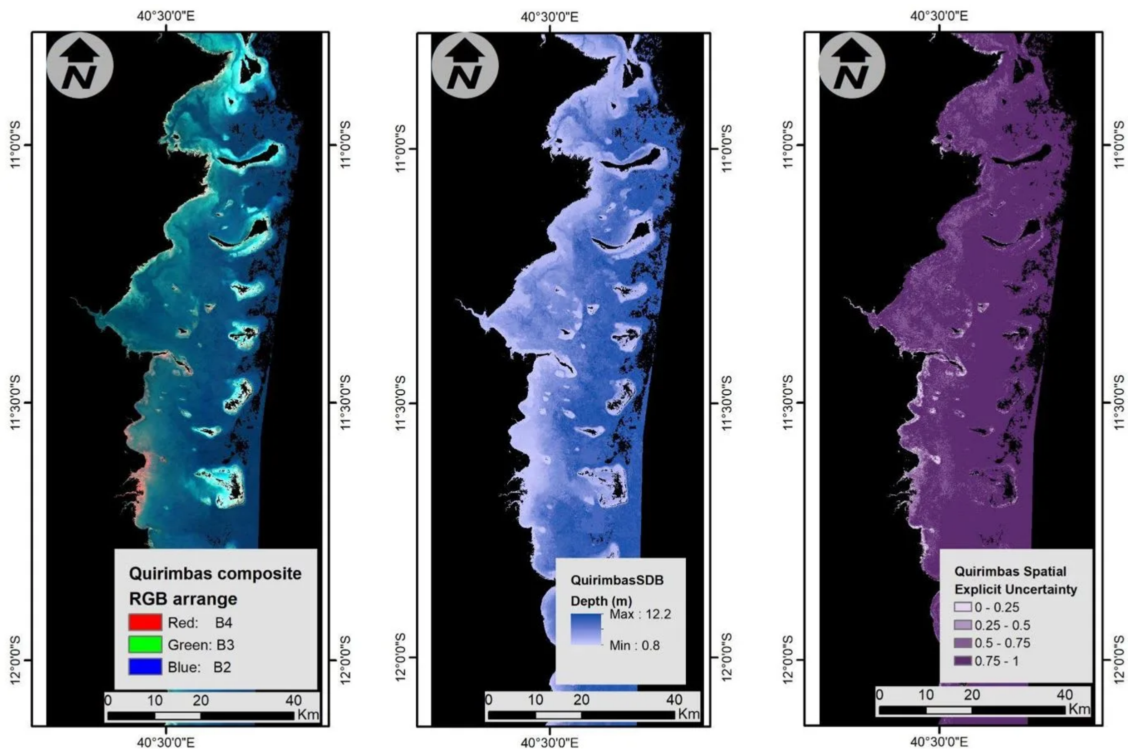
The spatially explicit uncertainty maps seem to agree with anticipated high uncertainty in areas as deep waters, delta of rivers and it is worth to mention that they highlight the borders of satellite imagery tiles.

In the same time, our approach seems to be the only by now which is able to visualize the level of uncertainty in marine remote sensing studies in a way that policy makers can take advantage of for better decision making. Albeit the scalable potentials through GEE platform, the size and the geo-spatial arrangment of the initial training pool appears to be the only limiting factor.

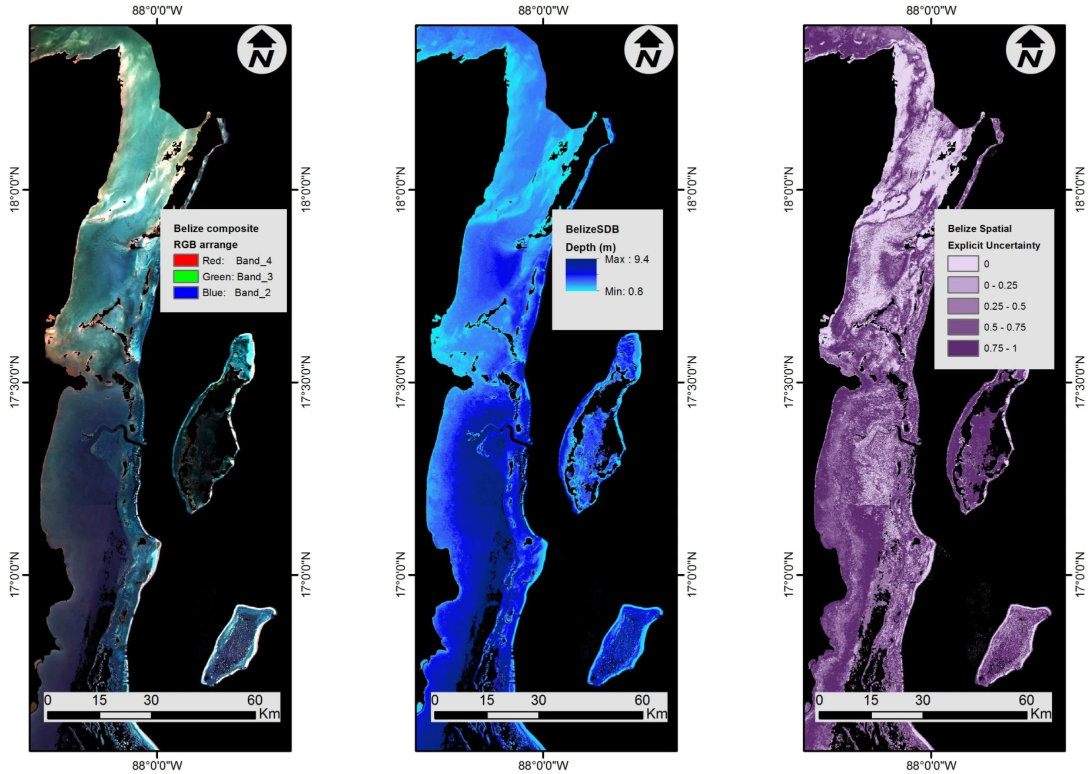




Port St. Lucie
West Palm Beach
Fort Lauderdale
Miami
The Bahamas
Nassau
Cienfuegos
Trinidad
Cuba
Florida
Camaguey
Las Tunas
Holguin
Banes
78°30'0"W
78°0'0"W
77°30'0"W
25°30'0"N
76°30'0"W
23°30'0"N
Masked
Seagrass
Sand
Coral
Hard Bottom
Spatial Explicit Uncertainty
0 - 0.25
0.25 - 0.5
0.5 - 0.75
0.75 - 1
Km



40°30'0"E
11°0'0"S
11°30'0"S
12°0'0"S
N
Quirimbas composite RGB arrange
Red: B4
Green: B3
Blue: B2
QuirimbasSDB
Depth (m)
Max : 12.2
Min : 0.8
Quirimbas Spatial Explicit Uncertainty
0 - 0.25
0.25 - 0.5
0.5 - 0.75
0.75 - 1
0 10 20 40
Km


88°0'0"W
18°0'0"N
17°30'0"N
17°0'0"N
N
Belize composite
RGB arrange
Red: Band_4
Green: Band_3
Blue: Band_2
BelizeSDB
Depth (m)
Max : 9.4
Min: 0.8
Belize Spatial
Explicit Uncertainty
0
0 - 0.25
0.25 - 0.5
0.5 - 0.75
0.75 - 1
0 15 30 60
Km

FUTURE STEPS

1. Exploit the spatially-explicit uncertainty information for the creation of independent training points. By doing that we hope to help the model to decide on its own on what data should be used as training points even if they are not included in the initial training pool.
2. Investigation of the capabilities of uncertainty and have a better understanding on the source of uncertainty and how to minimize it through deep-learning practises as backpropagation.

AUTHOR INFORMATION

Contact info of main author

email : spyridon.christofilakos@dlr.de

Linkdn: https://www.linkedin.com/in/spyros-christofilakos-1912aa22b/ (https://www.linkedin.com/in/spyros-christofilakos-1912aa22b/)

Linkdn of our project : https://www.linkedin.com/company/global-seagrass-watch/ (https://www.linkedin.com/company/global-seagrass-watch/)

Fellow co-author Alina Blume is currently working at ESA Centre for Earth Observation, Frascati, Italy

ABSTRACT

Modern advances in remote sensing have enabled the global monitoring of Earth's biodiversity and of its wide-ranging dimensions. These scalable advances are offering global information on the extent, structure, function and services of different ecosystem types, and their benefits to the environment and humans, in general. Albeit the unprecedented advances, relevant uncertainty methods and information are still in their infancy which impedes effective understanding of the product biases and impactful policy making efforts. In our study, we present a novel scalable uncertainty quantification framework, developed entirely within the Google Earth Engine, which assesses both thematic (e.g., ecosystem presence/absence, extent and distribution) and continuous products (e.g., satellite-derived bathymetry) related to coastal biodiversity using multi-temporal and cloud-free 10-m Sentinel-2, field data collections, and human-annotated data points. By exploiting the cloud-native machine learning classifier and its outputs, we estimate the uncertainty of the procedure per pixel. With that information, our semi-automated model is able to re-train itself in a data driven way and produce better results. There are three areas of interest in this study. The first is the Archipelago of Bahamas island complex, where we assess a four-class benthic habitat classification product per pixel. Our second and third study area is the national scale of Belize and the Quirimbas Archipelago (Mozambique), respectively, in which a satellite-derived bathymetry map. In the case of classification, our model achieved a better overall accuracy of ~4% in comparison with the initial classification while the producer and user accuracy of the habitat class that we are interested in, seagrass, rose by 13% and 7% respectively. On the regression results, our framework accurately highlights the areas with most uncertainty given the byproducts of the maximum likelihood regression that took place. While still in its alpha version, we envisage that further developments of the framework could allow better quantification of the data and model uncertainty, reducing the uncertainties in the coastal biodiversity monitoring and enabling more effective policy making efforts towards safeguarding global biodiversity health.

REFERENCES

B. Lyons, M., M. Roelfsema, C., V. Kennedy, E., M. Kovacs, E., Borrego-Acevedo, R., Markey, K., Roe, M., M. Yuwono, D., L. Harris, D., R. Phinn, S., Asner, G.P., Li, J., E. Knapp, D., S. Fabina, N., Larsen, K., Traganos, D., J. Murray, N., 2020. Mapping the world's coral reefs using a global multiscale earth observation framework. Remote Sens. Ecol. Conserv. 1–12. https://doi.org/10.1002/rse2.157

A. Blume, 2021. composition and related carbon stocks in support of the Nationally Eidesstattliche Versicherung Statutory Declaration in Lieu of an Oath

C.B. Lee, D. Traganos, P. Reinartz, 2022. A Simple Cloud-Native Spectral Transformation Method to Disentangle Optically Shallow and Deep Waters in Sentinel-2 Images. https://doi.org/10.3390/rs14030590

D. Traganos, B. Aggarwal, D. Poursanidis, K. Topouzelis, N. Chrysoulakis, P. Reinartz, 2018. Towards Global-Scale Seagrass Mapping and Monitoring Using Sentinel-2 on Google Earth Engine: The Case Study of the Aegean and Ionian Seas. Remote Sens. 10, 1227. https://doi.org/10.3390/rs10081227

N. Lang, N. Kalischek, J. Armston, K. Schindler, R. Dubayah, J.D. Wegner, 2022. Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles. Remote Sens. Environ. 268, 112760. https://doi.org/10.1016/j.rse.2021.112760

N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, R. Moore, 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.06.031

N. Marc Thomas et all., (2020).SPACE-BORNE CLOUD-NATIVE SATELLITE-DERIVED BATHYMETRY (SDB) MODELS USING ICESat-2 and SENTINEL-2
https://doi.org/10.1002/essoar.10504452.2

Malinin, A. ,2019. Uncertainty Estimation in Deep Learning with application to Spoken Language Assessment (Doctoral thesis). https://doi.org/10.17863/CAM.45912