Roinn Cumarsáide, Gníomhaithe ar son na hAeráide & Comhshaoil
Department of Communications, Climate Action & Environment

Geological Survey
Suirbhéireacht Gheolaíochta
Ireland | Éireann

Research Programme
- Short Call 2020, Final Report -

Lead Applicant Name	Vivienne Kelly
Host Organisation name	Compass Informatics Limited
Project Title	Using Earth Observation (Satellite) imagery for offshore renewable energy site selection and constraint analysis.
Contract Number	2020-sc-28

This report covers the period 1st December 2020 to project end date 31st August 2021

2. (a) Project information:

Final project report (*max 2000 words, excluding figures and headings*):

Abstract

This report outlines a methodology that can use machine learning to derive bathymetry data from satellite imagery. In particular, this report describes the development of Python Jupyter Notebooks to automatically create Satellite Derived Bathymetry products from Sentinel 1 Synthetic Aperture Radar data using Sentinel Hub API on the Euro Data Cube platform. Depth values are estimated using a machine learning approach. The Nearest Neighbour method is used to estimate the bathymetry to a root mean square deviation of 11.76m. The process produces a 17.89 km×17.89 km raster image for an area of interest with pixels of 1.28 km×1.28 km. Output SDB products can be visually compared with INFOMAR and EMODnet bathymetry using the generated ArcGIS Online web application.

(i) Objectives and scientific/engineering targets reached beyond the state of the art

The overall project objective was to implement a process to generate Satellite Derived Bathymetry for four areas of interest using Sentinel 1 SAR and to validate the output against INFOMAR bathymetry product.

This SDB output could be used to support initial "Go/No Go" areas for subsequent in-situ surveys and can be used as an input to initial OWF, Aquaculture or other site selection processes. Furthermore, there is potential for it to be used to track movement of dynamic topological features, e.g., sandbanks, in pre, during and post installation environmental impact assessments. The range of applications can be applicable for all coastal marine industries requiring information of water depths.

To demonstrate the process, we have identified the below areas of interests:

- Dublin Bay – this is the site of previous SDB products using Sentinel 2 optical sensors, which the GSI have been a part of[1][2][3]. The use of Sentinel 1 SAR for this site could allow for future projects to perform studies comparing these processes. Furthermore, Dublin Bay has 5 years of ground truthing data which can be used to validate against.
- Clew Bay, Mayo - chosen for the high production of aquaculture in the region and its relative calm waters in the protective bay, which is part of the Clew Bay Complex Special Area of Conservation (SAC) site - https://www.npws.ie/protected-sites/sac/001482.
- Off coast of Shannon Estuary, Atlantic - chosen for the proposed OWF development sites contained within it – e.g., Moneypoint Offshore, https://www.4coffshore.com/offshorewind/.
- Off the coast of Wexford, Irish Sea - contains the recently approved Shelmalere OWF feasibility study, https://dpenergy.info/shelmalere/.

Methodological objectives included:

- Assess the use of Sentinel Hub API, https://www.sentinel-hub.com/, in accessing Sentinel 1 SAR data.
- Assess the use of Euro Data Cube (EDC), https://eurodatacube.com/, platform in creating automated Python based Jupyter Notebooks to the generate SDB products.
- Assess the implementation of Principal Component Analysis (PCA) machine learning technique to assist in the process of generating SDB.
- Create a web application viewer to visualise and interact with data generated and pre-existing bathymetry products.
- Provide an online documentation site that describes the processes, methodologies, data, and final outputs.
- Provide an inventory of satellites for which the automated SDB scripts could be generated from.

The project has demonstrated **a process that can ingest Sentinel 1 SAR data, via Sentinel Hub API, and produce SDB product utilising PCA machine learning** in automated Jupyter Notebooks on Euro Data Cube platform, with the validation of outputs through comparisons with INFOMAR and EMODnet bathymetry.

[1] Casal, G., Hedley, J.D., Monteys, X., Harris, P., Cahalane, C., McCarthy, T. (2020). Satellite-derived bathymetry in optically complex waters using a model inversion approach and Sentinel-2 data. Estuarine, Coastal and Shelf Science, 241, 106814.
[2] Casal, G., Harris, P., Monteys, X., Hedley, J., Cahalane, C., McCarthy, T. (2020). Understanding satellite-derived bathymetry using Sentinel 2 imagery and spatial prediction models. GIScience and remote sensing, 57:3, 271-286
[3] Casal, G., Monteys, X., Hedley, J., Harris, P., Cahalane, C., McCarthy, T. (2019). Assessment of empirical algorithms for bathymetry extraction using Sentinel-2 data. International Journal of Remote Sensing, 40 (8): 2855-2879; DOI: 10.1080/01431161.2018.1533660

(ii) Implementation **(milestones, deliverables, project management)**

WP1. Project Management

Weekly steering group meetings and bi-weekly scrum type meetings were performed to ensure project remains on track as regards resources, timescales, and particularly objectives. Steering group members include:

- Vivienne Kelly – Compass Informatics COO / Project Manager
- Rob O'Loughlin – EO Data Analyst
- Dave Dallaghan – Project Manager (on ECHOES[4] project)
- John Lavelle – EO Data Analyst

WP2. Technical Research & Scope Definition

Scope amendment following discussions with stakeholders and the GSI was confirmed. Project scope was narrowed to focus on Bathymetry – a critical variable in ORE, Aquaculture and other nearshore economic activities. Bathymetry is of extreme importance to coastal and offshore stakeholders. Investigation into deriving bathymetry using satellite imagery to compliment exiting bathymetry products while also creating outputs in areas where no pre-existing bathymetry product occurs. The project scope was redefined to focus on EO data, using Sentinel 1 SAR, in the generation of a SDB product for pre-defined AOIs.

WP3. Earth Observation Data inventory and review

Sentinel 1 SAR was chosen as it covers both historical analysis and allows for on-going monitoring. In theory the approach taken should be reproducible for other SAR satellites, with higher degree of probability for other C-band satellites. The below image shows SAR missions through time with their band specified.

[4] https://echoesproj.eu/ - Compass producing web platform with EO based web tools for ECHOES project.

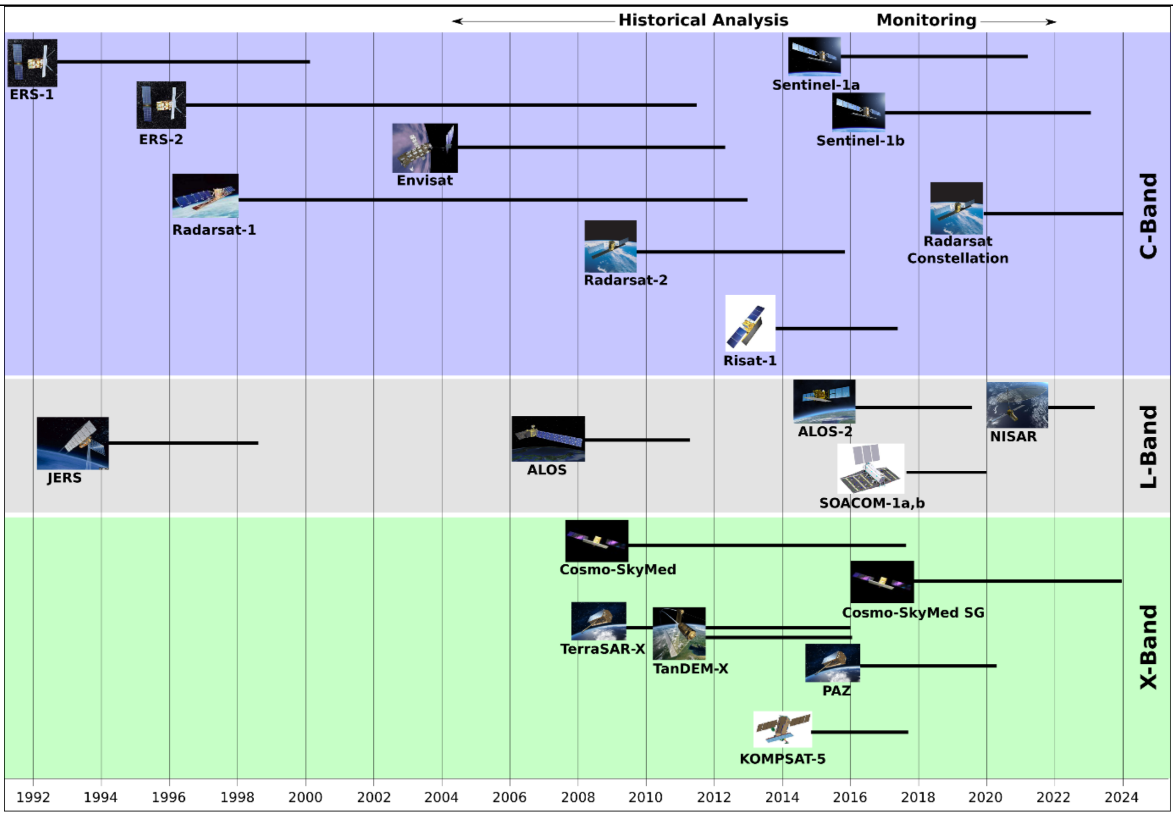


Figure 1: SAR missions accessed from https://www.unavco.org/instrumentation/geophysical/imaging/sar-satellites/sar-satellites.html

WP4. Stakeholder Input

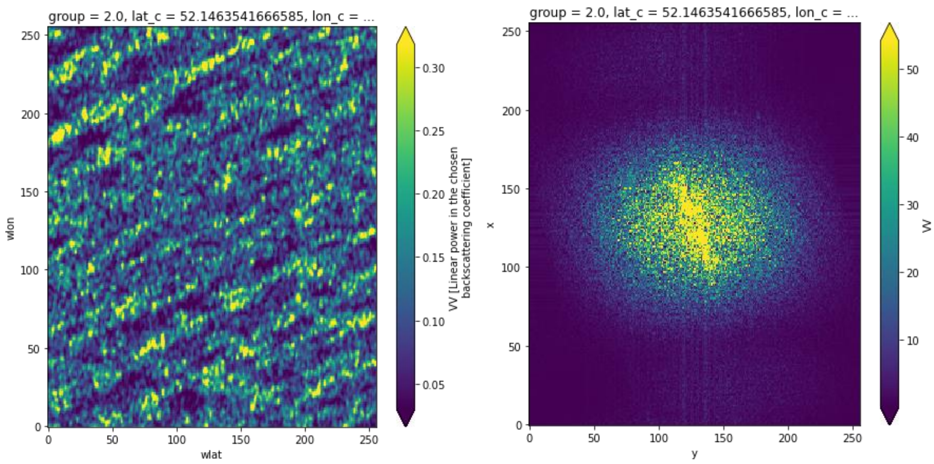
Compass has been in direct contact with several key stakeholders with experience in site selection for nearshore economic activity. Stakeholders that provided input to the project included: GDG Solutions http://www.gdgeo.com/ (via Geoscience Ireland), Simple Blue Energy https://simplyblueenergy.com/, BlueWise Marine https://bluewisemarine.ie/, Green Rebel Marine https://www.greenrebel.ie/, and Marine Ireland Industry Network (MIIN) https://marine-ireland.ie/. Subsequent contact with the GSI has led to the recommended redefining of project scope to align with market needs/gaps while continuing to utilise the vast untapped benefits of EO data.

WP5. Algorithm Development

An automated methodology to generate SDB product using Sentinel 1 SAR data was developed.

Euro Data Cube is the platform used to develop Python based Jupyter Notebooks. This platform allows for easy integration of Sentinel Hub API completely, which removed satellite data management and pre-processing overheads.

A gridded bathymetric approach is used to ensure maximum automation; this essentially the generation of a depth value for each grid (subscene) of an area of interest (scene). Two methods to estimate the bathymetry using Sentinel 1 SAR data have been implemented. In both methods, the SAR scene is partitioned into subscenes and the spectra of each are computed using fast Fourier transformations (FFTs); the change in the spectra, due to interaction with the seabed, is used to estimate the bathymetry. The methods are performed in the same process with the machine learning method aiming to improve the estimated bathymetry results.



(a) (b)

Figure 2: (a) The backscatter coefficients (VV) for a selected 2.56×2.56 km sea wave subscene. (b) The FFT of the subscene.

The first method is based on paper by Wiehle et al. (2019)[5], which involves computing the peak wavelength for each of the subscenes. The depth can then be estimated by solving the dispersion relation:

$$\omega_p^2 = \frac{2\pi g}{L_p}\tanh\left(\frac{2\pi d}{L_p}\right)$$

where $\omega_p = 2\pi/T_p$ is the peak wave frequency, T_p is the peak wave period, L_p is the peak wavelength and g is the gravitational acceleration. This can be solved for the depth d to

$$d = \frac{L_p}{2\pi}\tanh^{-1}\left(\frac{2\pi L_p}{g T_p^2}\right)$$

[5] Wiehle, S., Pleskachevsky, A. & Gebhardt, C. Automatic bathymetry retrieval from SAR images. CEAS Space J 11, 105–114 (2019). https://doi.org/10.1007/s12567-018-0234-4

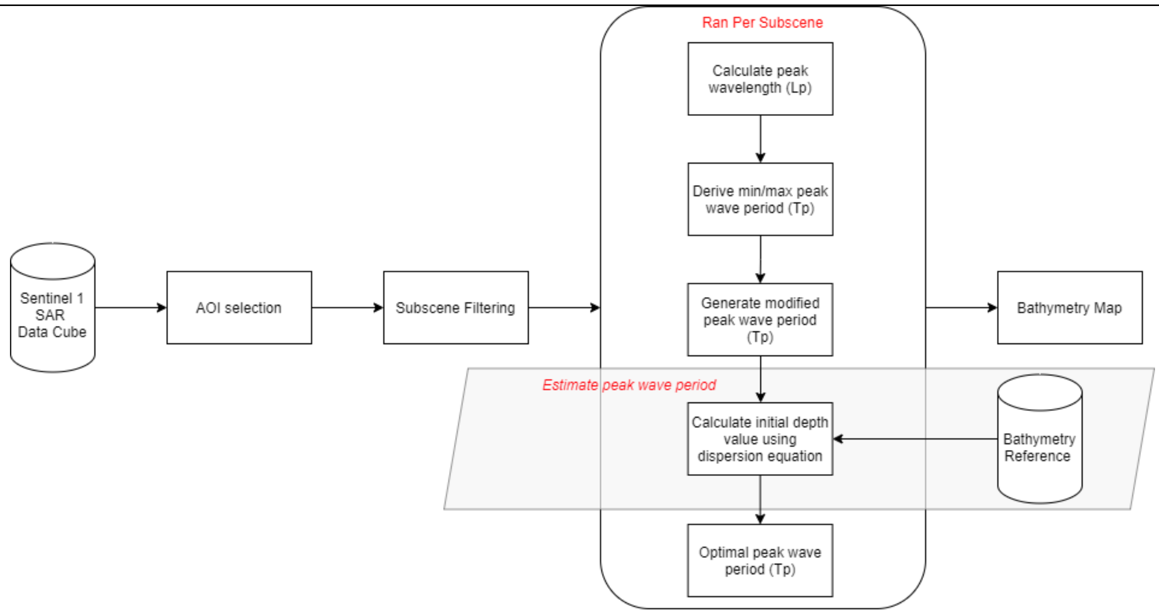


Figure 3: Process workflow diagram for the peak wavelength based method

Scene (satellite image) sizes of 17.89×17.89 km to maximise area being processed while ensuring processing limits of EDC are not exceeded. Large scenes should be used as according to Weihle et al. (2019), incoming swells in deep water may typically be considered constant in the wave period and wavelength for large scenes.

A rolling window was performed on subscenes of 2.56×2.56 km with an overlap of 50% between subscenes (1.28 km), across the 17.89×17.89 km scenes. This overlap was set to increase the resolution of the retrieved bathymetry, while having a subscene size sufficiently large for wavelength retrieval.

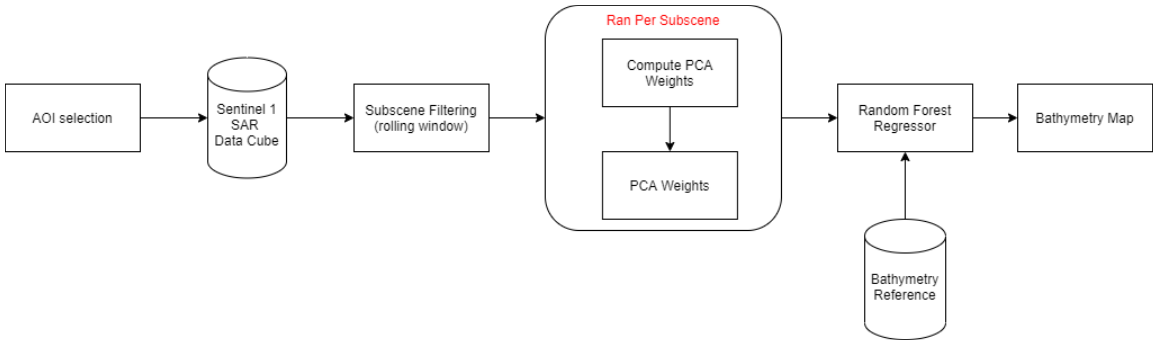
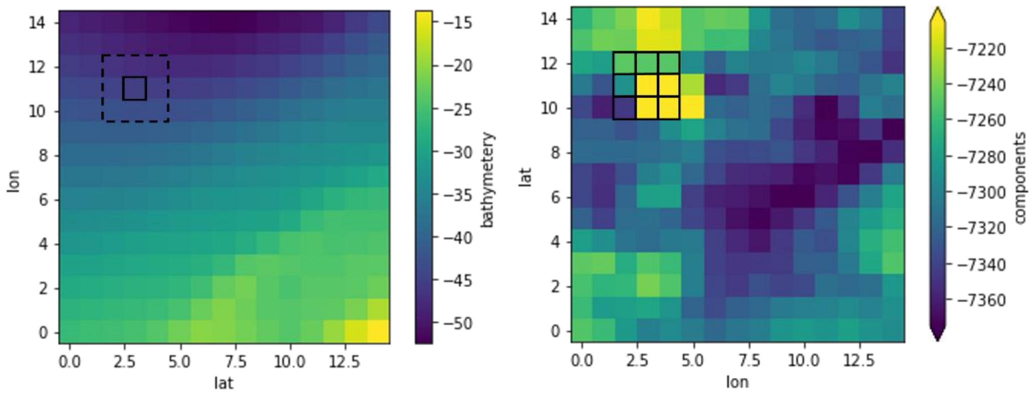


Figure 4: Process workflow diagram for the PCA approach.

The spectra are computed for each subscene. The spectra are decomposed into the principle components (see Figure 5). The Random Forest method is applied to predict the bathymetry using these components.

The second method is a novel approach, which involves applying machine learning to estimate the bathymetry from the spectra. This approach has the potential to be more accurate than the peak wavelength approach, as the components account for more of the variance in the spectra than the peak wavelength alone. This is a new process that Compass has implemented to improve the bathymetry estimate.

The workflow for this process is shown in Figure 4. The method involves applying PCA to the backscatter spectra (see Figure 5). Like the peak wavelength method, the spectra are computed on a rolling window across the 20.48×20.48 km (2048×2048 pixel) scene.

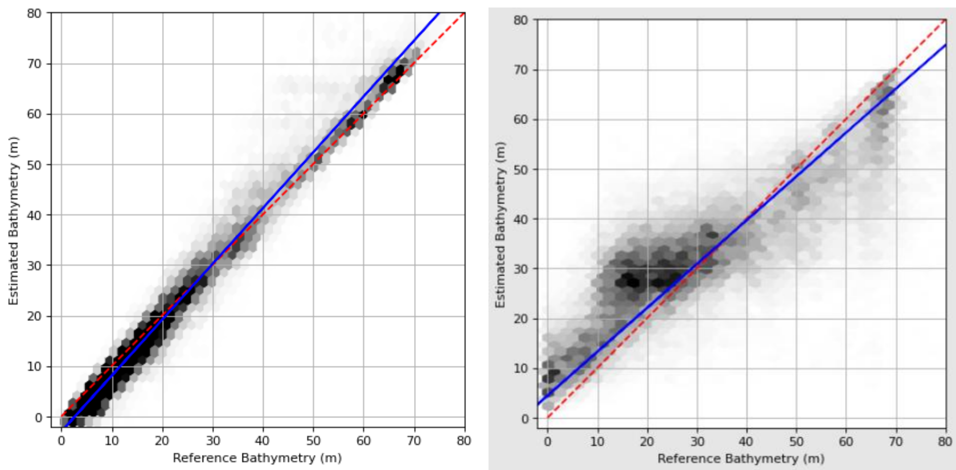


(a) (b)

Figure 5: 2.56×2.56 km sliding windows with a 50 % overlap are applied to the 20.48×20.48 km (2048×2048 pixel) Sentinel 1 scenes, producing arrays (with a pixel size of 1.28×1.28 km) of (a) the mean bathymetry and (b) the principle component weights from the PCA (only the first of a total of 6 is shown). The dimension of the resulting arrays are 16.64×16.64 km (13×13 pixels).

In Figure 5, the black grids show the 3.84×3.84 km (3×3 pixel) sliding windows (with a step size of 1) that are applied to arrays resulting from the first sliding window operation, to obtain the training datasets for the machine learning algorithm. Each sample in the machine learning algorithm consists of a 3×3 pixel sub-array of the PCA components in the window (i.e., an array of dimensions 3×3×6) and the central pixel of the corresponding 3×3 pixel bathymetry sub-array.

A scatter plot comparing the SDB to the EMODnet bathymetry is shown in Figure 6.



(a) (b)

Figure 6: Binned scatter plot of the estimated vs reference bathymetry from the machine learning approach. The Nearest Neighbour method was used to estimate the bathymetry.

The RMSD is 11.76m. Figure (a) displays the machine learning training data and figure (b) displays the bathymetry estimate using the test data.

.

Main Python packages used:

- The xcube_sh, https://xcube.readthedocs.io/en/latest/plugins.html, Python plugin is used to easily access the Sentinel Hub API on-the-fly. This plugin is based on xcube, which is a Python toolkit developed to provide EO data in an analysis-ready form to users, with data split into cubes or dimensions, e.g., time, latitude, longitude, etc.
- Xarray, http://xarray.pydata.org/en/stable/, is used to conveniently work with multi-dimensional arrays, e.g., data cubes.
- The inclusion of Dask, https://dask.org/, Python library allowed for parallel computing, which helped in the scaling of the machine learning workflow to improve processing times.

WP6. Climate Change Predictions

Contract amended to exclude this work package.

WP7. Compilation of Findings and Deliverables

The full set of WP findings and deliverables is described in full under Section (iii) Outputs to date.

(iii) Outputs to date

The following outputs have been produced:

- Online User Documentation

  Online user documentation was created using Python Sphinx package. This website contains all the information relating the project including links to reports generated throughout the project with detailed methodology and extracted Jupyter Notebooks.

  This is a detailed resource that can be used by GSI and other researchers in the future. https://w12-iis.compass.ie/GSI/SatelliteDerivedBathymetry/

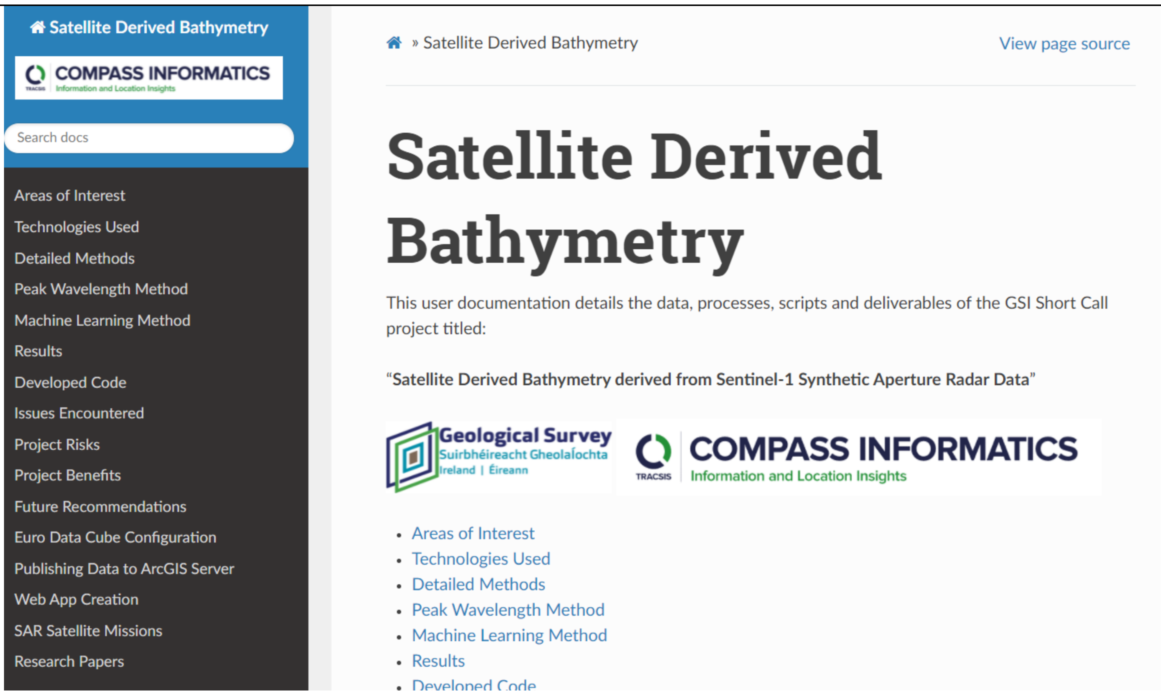


Figure 7: Online User Documentation

- <u>Jupyter Notebooks</u>
  Jupyter Notebooks consisting of Python processes to create and validate SDB outputs for predefined AOIs has been produced. The notebooks are stored in a private GitHub repository, https://github.com/johnlavelle/SDB_SAR, which Compass can provide access to on request. Description of code can be accessed via user documentation (discussed below) - https://w12-iis.compass.ie/GSI/SatelliteDerivedBathymetry/.

- <u>Validation of SDB</u>
  Included in our SDB code is an automated validation process that compares the output depth values against EMODnet bathymetry data. This is done by splitting the bathymetry data into test and training datasets, with 33 % of the data used for testing. The RMSD, of 11.76m, is computed on the test dataset.

- <u>Web Services</u>
  The following data is now available via an ArcGIS REST web service and available in the web application viewer:

Table 1: web services created

Name	Type	Description
Areas of Interest	Vector	Polygon dataset displaying areas of interest for the generation of Satellite Derived Bathymetry
Satellite Derived Bathymetry	Raster	Satellite Derived Bathymetry raster outputs for predefined areas of interests at a predefined date and time.

Any SDB imagery derived will be provided as a REST Service.
To view the web application viewer, click here.

- <u>Web Application Viewer</u>
  An Esri ArcGIS Online Web AppBuilder application was created to allow users to interact and visually compare SDB outputs with that of INFOMAR and EMODnet. Currently only INFOMAR and EMODnet are available. Any SDE outputs will be included in this viewer.
  To view the web application viewer, click here.

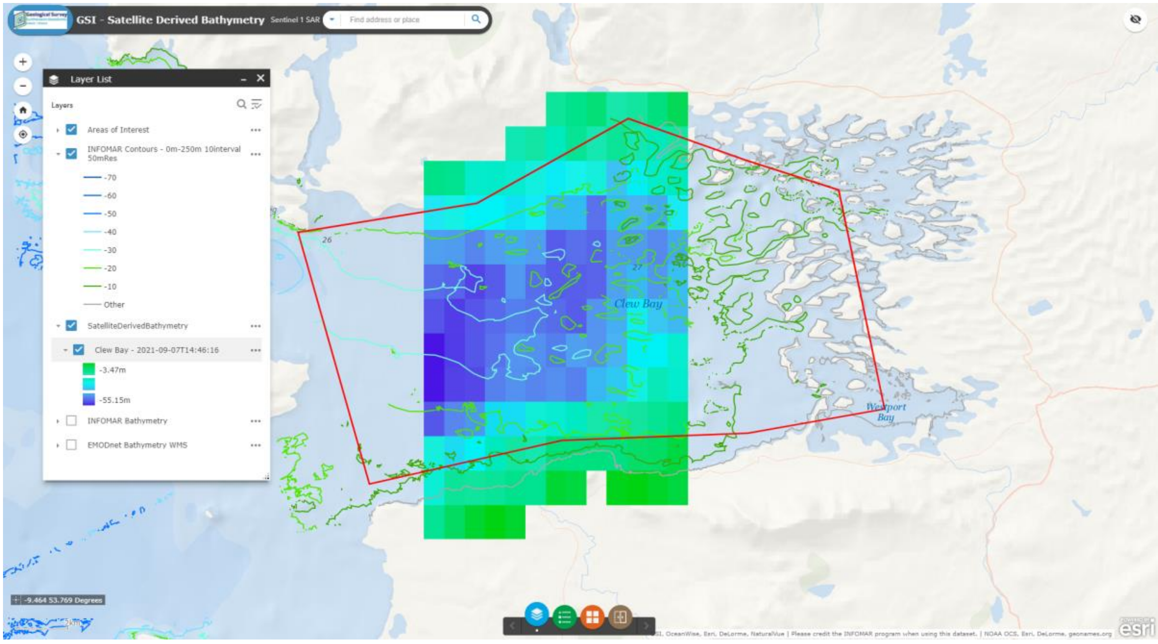


Figure 8: ArcGIS Online Web AppBuilder viewer showing Clew Bay SDB output for 07/09/2021 with INFOMAR Contours displayed in a similar symbology.

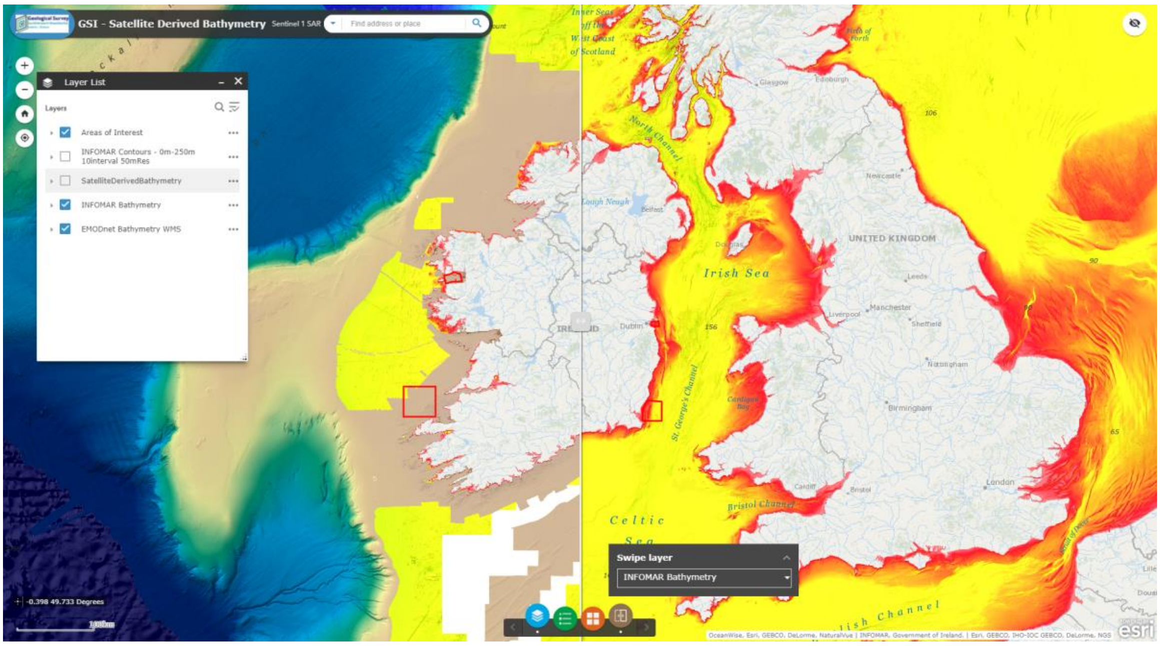


Figure 9: ArcGIS Online Web AppBuilder viewer demonstrating the use of the Swipe widget.

(iv) Impact

- Relevance/application to GSI programmes, where applicable

The GSI have been involved in SDB projects using Sentinel 2 optical analysis[6][7][8]. However, the application of SAR in SDB generation is not known to have been performed in the Irish marine environment and as such is the first of its kind here.

As SAR is not impacted by cloud or turbid waters, analysis can theoretically be done year round producing large historical data. It has the potential to be used to estimate bathymetry in dynamic sites, and those difficult to access by boat.

As SAR can provide depth estimates from 10m to 70m of water depths and optical methods up to 20 m, synergies between these sources are possible. The results of both optical and SAR can complement each other[9][10] and so the work conducted by the GSI on Sentinel 2 could be improved by integrating the Sentinel 1 SAR process from this project.

Sentinel Hub is an online cloud-based repository for satellite imagery. Instant access to Sentinel, and other EO imagery – archives of more than 5 PB of data, both historic and the latest acquisitions – are available through a subscription based licence. A free 1 month trial is available.

EDC is a newly developed cloud based European Space Agency (ESA) application. It is a one-stop-shop for global archive of pre-processed EO data that allows direct query access in the form of online Jupyter Notebook platform. A three month free trial was available to use with the ability to extend this for research purposes, also a marketplace for previously developed Jupyter Notebooks is available to increase uptake speed.

While the EDC platform was used to create and run the Jupyter Notebooks, it is not a requirement in the long term. The code essentially only requires access to Sentinel Hub and as result it is not a requirement to use Euro Data Cube. Another platform or hosting environment can be used to run this code, or it can be run locally.

[6] Casal, G., Hedley, J.D., Monteys, X., Harris, P., Cahalane, C., McCarthy, T. (2020). Satellite-derived bathymetry in optically complex waters using a model inversion approach and Sentinel-2 data. Estuarine, Coastal and Shelf Science, 241, 106814.
[7] Casal, G., Harris, P., Monteys, X., Hedley, J., Cahalane, C., McCarthy, T. (2020). Understanding satellite-derived bathymetry using Sentinel 2 imagery and spatial prediction models. GIScience and remote sensing, 57:3, 271-286
[8] Casal, G., Monteys, X., Hedley, J., Harris, P., Cahalane, C., McCarthy, T. (2019). Assessment of empirical algorithms for bathymetry extraction using Sentinel-2 data. International Journal of Remote Sensing, 40 (8): 2855-2879; DOI: 10.1080/01431161.2018.1533660
[9] Pleskachevsky, A., Lehner, S., Heege, T. et al. Synergy and fusion of optical and synthetic aperture radar satellite data for underwater topography estimation in coastal areas. Ocean Dynamics 61, 2099–2120 (2011). https://doi.org/10.1007/s10236-011-0460-1
[10] Pleskachevsky, A., Lehner, S. Synergy and fusion of optical and synthetic aperture radar satellite data for underwater topography estimation in coastal areas. SEASAR. ESA. Norway (2012). https://earth.esa.int/c/document_library/get_file?folderId=233742&name=DLFE-2153.pdf

Through the generation of Python Jupyter Notebooks, a reproducible methodology (in terms of code, informative text, and intermediary outputs) is provided to enable further development.

PCA was used for dimensionality reduction of the backscatter coefficient spectra. The first 64 components account for over 99 % of the variance, with a corresponding file size reduction of PCA weights by a factor of ~1000, allowing the PCA weights to be stored in a file of several megabytes compared to several gigabytes for the backscatter coefficients. This enables the PCA weights to be provided for download, for further development of the method.

- Other impacts

Benefits of the use of machine learning over the standard approach, based on computing the peak wavelength of subscenes:

- does not require knowledge or estimation of the peak wave period
- can potentially capture physical effects (in addition to the dispersion relation) to estimate the bathymetry, such as wave shoaling, refraction, and non-linear interactions at shallow depths.

Future Improvements:

- Use of Neural Networks in place of the Random Forest method. As wave conditions of each scene impacts the accuracy of the depth values produced using multiple scenes (e.g., multiple time periods) for an AOI by selecting the which scenes can be used to estimate the bathymetry. Also, a neural network approach could potentially capture more of the physics involved in the interaction of the waves and the seabed, enabling a more accurate bathymetry estimate.
- Inclusion of tidal data may result in improved accuracy of outputs.
- Run the code of a machine with more cores and memory
- Use the CreoDIAS, https://creodias.eu/, cloud computing platform (removes need for EDC and Sentinel Hub payments. Single payment for server with sliding costs)
- Further investigate PCA use in determining peak wavelength
- Improve PCA by running on more data

3. Supplemental information/datasets

All methodological and code details are provided in full via the online user documentation and GitHub.