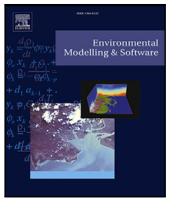
Environmental Modelling and Software 171 (2024) 105861

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Position Paper

Groundtruther: A QGIS plug-in for seafloor characterization

M. Di Stefano [a,b,*], G. Gonzalez Mirelis [c], L. Mayer [a,b]

[a] Department of Earth Science, University of New Hampshire, 214 James Hall, 56 College Road, Durham, 03824, NH, USA
[b] Center for Coastal and Ocean Mapping/Joint Hydrographic Center, Chase Ocean Engineering Lab, 24 Colovos Road, Durham, 03824, NH, USA
[c] Institute of Marine Research, Nordnesgaten 50, Bergen, 5005, Norway

ARTICLE INFO

Dataset link: https://github.com/epifanio/groundtruther/, https://github.com/epifanio/grass_api/, https://zenodo.org/records/7995674/

Keywords:
Remote sensing
Ground truth
Multibeam
Seafloor images
GIS

ABSTRACT

This work focuses on developing a software system for concurrently analyzing co-located multibeam echo sounder (MBES) datasets and seafloor imagery, fulfilling the need for a unified seafloor data exploration and analysis platform. The system comprises a graphical user interface where image browsing and geospatial data are linked and several toolboxes for extracting backscatter distribution, its angular response, and bathymetric derivatives to ultimately build detailed quantitative reports. The overall objective is to provide an efficient means of understanding the relationships between morphology, backscatter, and the observed biota and, thus, understanding the relationship between the physical and ecological elements of the seafloor. In addition, Groundtruther provides new ways to interpret remotely sensed information derived from MBES and aid the development of spatial distribution models. Furthermore, it could lead to the enrichment of ground-truth databases used to develop formal geophysical models that link acoustic backscatter observations to intrinsic properties of the seafloor.

Software availability

Groundtruther

- Name of the software: Groundtruther
- Developer: Massimo Di Stefano
- Contact information: massimods@met.no
- Year first available: 2023
- Program language: Python
- Cost: Free
- Software availability:
  - source code development : https://github.com/epifanio/groundtruther
  - latest release: https://github.com/epifanio/groundtruther/releases/tag/v02
  - QGIS plugin: https://plugins.qgis.org/plugins/groundtruther/
- Program size: 3.5 MB
- License: GPL V3

Grass API

- Name of the software: GRASS API
- Developer: Massimo Di Stefano
- Contact information: massimods@met.no
- Year first available: 2023
- Program language: Python
- Cost: Free
- Software availability:
  - source code development : https://github.com/epifanio/GRASS_API
  - latest release: https://github.com/epifanio/grass_api/releases/tag/0.1
- Program size: 41 kB
- License: GPL V3

The *Groundtruther* code is also available for installation from the *QGIS* plugin repository. The software has been tested on Linux Operating System. Although very flexible, the configuration settings are tailor-made for HabCam V4 data, a complete dataset to test the application has been made available and linked at the following DOI:10.5281/zenodo.7995674

1. Introduction

Seafloor characterization requires the collection and concurrent analysis of many kinds of datasets which may differ vastly in nature

* Corresponding author at: Department of Earth Science, University of New Hampshire, 214 James Hall, 56 College Road, Durham, 03824, NH, USA.
E-mail address: distefan@ccom.unh.edu (M. Di Stefano).

https://doi.org/10.1016/j.envsoft.2023.105861
Received 15 July 2023; Received in revised form 3 October 2023; Accepted 19 October 2023
Available online 25 October 2023

and spatial scale, including bathymetry, acoustic backscatter, oceanographic information, seafloor images, physical samples, etc. Bathymetric measurements and acoustic backscatter data are collected employing multibeam echo sounders (MBES) and together form a robust basis for seafloor mapping, covering large areas of the seafloor quickly and efficiently (Parnum, 2007; Hasan et al., 2014; Brown et al., 2011). Bathymetry and its derivatives, including geomorphology and terrain curvatures, are widely recognized as important predictors of habitat distribution (Misiuk et al., 2021; Lecours et al., 2017). Seafloor acoustic backscatter, acquired simultaneously with bathymetric data, refers to the measurement of the ratio of scattered versus incident acoustic intensity of the sound waves emitted by the sonar system. Assuming the use of calibrated hardware and that the data has been processed following best practice approaches (Lurton and Lamarche, 2005), the acoustic backscatter signal provides information about seafloor roughness and hardness, giving insight into the type and distribution of sediments, and can be converted into continuous mosaics to support habitat mapping (Porskamp et al., 2022).

While remotely-sensed data can significantly increase our understanding of seafloor processes, e.g., through providing insight into seabed geomorphology, substrate characteristics, or the distribution of discrete scatterers such as shell, cobble, gas bubbles, or vegetation over large areas, it is by coupling such data with direct "first-hand" observations of the seafloor, also known as ground-truth data, that their full potential for habitat mapping and seafloor characterization is unlocked. Seafloor optical imagery, whether still photographs, video footage, or stereo images, are widely used as ground truth (Ierodiaconou et al., 2007; Ilich et al., 2021; Che Hasan et al., 2012), as is sediment sampling (e.g., McGonigle and Collier (2014)) depending on the purpose for the mapping. Ground-truth data, if accurately georeferenced, helps both to interpret and to validate remote sensing data. The ability to link acoustic measurements to seafloor samples is instrumental for the development of physical models that can predict the properties of the seafloor through geoacoustic modeling (Jackson and Richardson, 2010). Also, spatially nesting seafloor images within MBES and other data has boosted species distribution modeling (SDM), which links species distributions to seafloor environmental characteristics (Brown et al., 2011) and is a technique widely applied in marine management.

The relationship between backscatter measurements and seafloor properties is notoriously complex and depends on various factors, including the frequency and angle of incidence of the sound waves and the nature of the seafloor materials (Porskamp et al., 2022; Misiuk and Brown, 2022). In particular, the acoustic intensity registered at different incidence angles may provide complementary, non-redundant information describing substrate characteristics (Clarke, 1994). For example, the scattering of inner acoustic beams is dominated by specular scattering, which is primarily a function of the acoustic impedance of the seafloor, defined as the ratio between acoustic pressure and soundwave velocity (Fonseca et al., 2009), and it can be related to the "hardness" of the seafloor (X. and G., 2015; Hou et al., 2018). At the outer beams, scattering is increasingly sensitive to sediment roughness.

Angular range analysis (ARA) is an approach for calculating the entire angular response curve (ARC) across each half of the MBES swath (i.e., from the nadir to the outer beam on both the port and starboard sides of the swath), which can be used to estimate substrate properties (Fonseca and Mayer, 2007). So as not to lose horizontal spatial resolution, it has been suggested to aggregate soundings into homogeneous patches of seabed identified from the backscatter mosaic (Fonseca et al., 2009). Porskamp et al. (2022) have recently demonstrated the importance of integrating ARA and backscatter angular response analysis into seafloor habitat mapping, thus highlighting a need for software tools for the quantitative analysis of backscatter data in relation to ground truth.

We have developed *Groundtruther*, a platform for exploring and analyzing co-located, geospatial marine datasets including accurately georeferenced, point-locality, ground-truth data. Among other things *Groundtruther* offers new ways to interact with and interpret seafloor imagery in relation to remotely sensed information derived from MBES. It allows the user to implement easily a few of the most common workflows in seafloor characterization, e.g., analyzing seafloor digital terrain models, querying acoustic point cloud data, and browsing seafloor images, where, in contrast to a geographic information system (GIS), everything is interlinked in a graphical user interface (GUI). *Groundtruther* is a system that enables the use of point-locality data (such as biotic data obtained from annotated seafloor images, environmental data from in-situ sensors, sediment data from grab samples, etc.) for the interpretation and validation of full-coverage data, whether obtained from remote sensors or from models (e.g., oceanographic). It can also be used to generate inter-dataset relationships suitable to develop species distribution modeling and to support geoacoustic modeling.

In this paper we demonstrate:

- A GUI to synchronize image browsing and map canvas
- A fast and efficient point location method for the regional sampling of backscatter point clouds
- A routine for the quantitative analysis of geomorphological features

We illustrate the use of the software through two example applications: a quantitative analysis of bedforms found in the Great South Channel (GSC) in the North East Atlantic, also described in Di Stefano and Mayer (2018), and a method to document seafloor variability through the visualization of seafloor images and their co-located MBES data.

2. Test data

This first release of *Groundtruther* was optimized to accommodate ground truth data (optical and sensor) obtained through the HabCam V4 (Howland et al., 2006; Taylor et al., 2008), a towed camera system that has a long history of use in the North East Atlantic, mainly by the National Oceanic and Atmospheric Administration (NOAA) - Northeast Fisheries Science Center (NEFSC).

In 2015 the University of New Hampshire helped fund a small (4-hour) HabCam survey, which was part of a larger NOAA survey, at the Great South Channel (GSC). This sub-survey uniquely used Ultra Short Base Line (USBL) positioning to track the vehicle's position. During this sub-survey continuous, high-resolution, still photographic imagery of the seafloor was collected with a Prosilica Giga Ethernet (stereo) camera. The vehicle was towed at an average speed of 6 kn and kept at a constant altitude of ~1.7 m above the seafloor. The camera system was set to collect six frames per second to ensure overlap between frames of ~ %40. Each frame covered an area of approximately one square meter. The entire 4-hour survey amounted to 59 000 accurately geolocated images. From this set, 5 900 were randomly selected and have been used to test *Groundtruther*. Regarding sensor data, this set comprises water depth, temperature, salinity, turbidity, dissolved oxygen, dissolved organic matter, and chlorophyll values for the same set of points.

All images were annotated both automatically and manually. Automated annotations were done using a "single class" feature detector available through the Video and Image Analytics for Multiple Environments (VIAME) (VIAME Development Team, 2023) framework. This detector attempted to find all the seabed-dwelling fauna in the images. Manual annotations were done at the species level by marine biologists at the University of New Hampshire for the purpose of training feature detection algorithms. Both annotation datasets were stored in a CSV format consistent with those produced by an automated object detection pipeline in use by NEFSC for the same towed camera system.

Uncalibrated MBES data was collected simultaneously with HabCam images with system settings optimized for collection of high-resolution bathymetry. Although the backscatter data is negatively affected by this

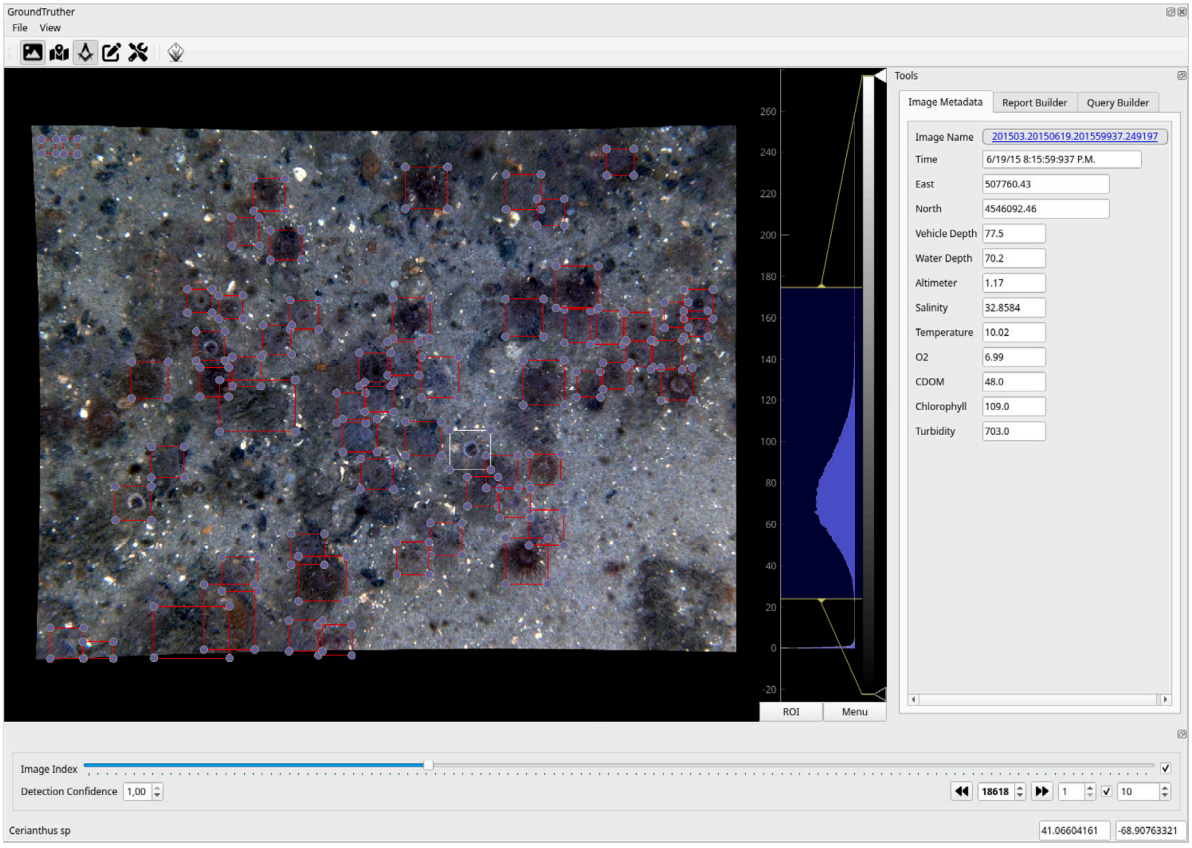


Fig. 1. Image browser. The image rendering is based on *OpenGL* for fast rendering of the images. The image browsing tools are available as a detachable widget on the bottom. The image metadata is embedded in a second detachable panel on the right side of the main window. The *Groundtruther* toolbar includes tools to change the application settings and to enable/disable its several components.

it can still be used as a relative measure of seafloor properties which can also be informative. Two data products derived from these data have been included in this test, namely: a 1 x 1 m bathymetric grid and a point cloud holding the backscatter output of the system (refer to Table 1, for a description of the point cloud data structure).

3. Description of the software

Groundtruther is entirely written in Python and is licensed as open-source software under the GPLv3 license. The system consists of a main GUI written in *PyQt* (PyQt Development Team, 2023), which hosts the following components:

- Image browser
- Map canvas
- Backscatter Query Builder
- Bathymetryc Processing Engine
- Report builder

All the components of the *Groundtruther* user interface are QT-Widgets connected through the QT signals and slots mechanism for inter-object communication (signaling).

3.1. Image browser

Browsing images is a fundamental part of the seafloor characterization process, but scanning over hundreds of thousands of photos and associated metadata, applying essential image enhancement, and overlaying additional information (i.e., image annotations), require a fast and dedicated graphical user interface. In *Groundtruther*, the GUI for image visualization, the Image Browser, is based on an *OpenGL* viewer provided by the *pyqtgraph* library (PyQtGraph Development Team, 2023) (Fig. 1). For browsing, the images are indexed and sorted by time of acquisition, and the scanning can be performed by means of browsing buttons, a slider, and a *spinbox* (a GUI element that combines numerical values with up-down controls). These widgets let the user track the image index and change the number of images to skip for each scanning step.

As the user browses the photos, a signal is sent to the image metadata GUI component which will update its contents according to the image index. The image browser has access to image location

Table 1
Point cloud file structure. "Backscatter Value" refers to the raw value, whereas "Corrected Backscatter Value" stores processed values, including signal level adjustment due to range and transmission loss, beam incidence angle and beam footprint area adjustments, and other corrections relating to the specific sonar. "True angle" is the 3D grazing angle of the beam; in the test dataset, it was corrected to account for seafloor slope. The data structure used here follows the same structure as the output provided by the QPS FMGT software.

Field	Type
Ping Time	Integer
Ping Number	Integer
Beam Number	Integer
Easting	float
Northing	float
Depth	float
Longitude	Double
Latitude	Double
Backscatter Value	float
Corrected Backscatter Value	float
True Angle	float
Datetime	string
Line	string

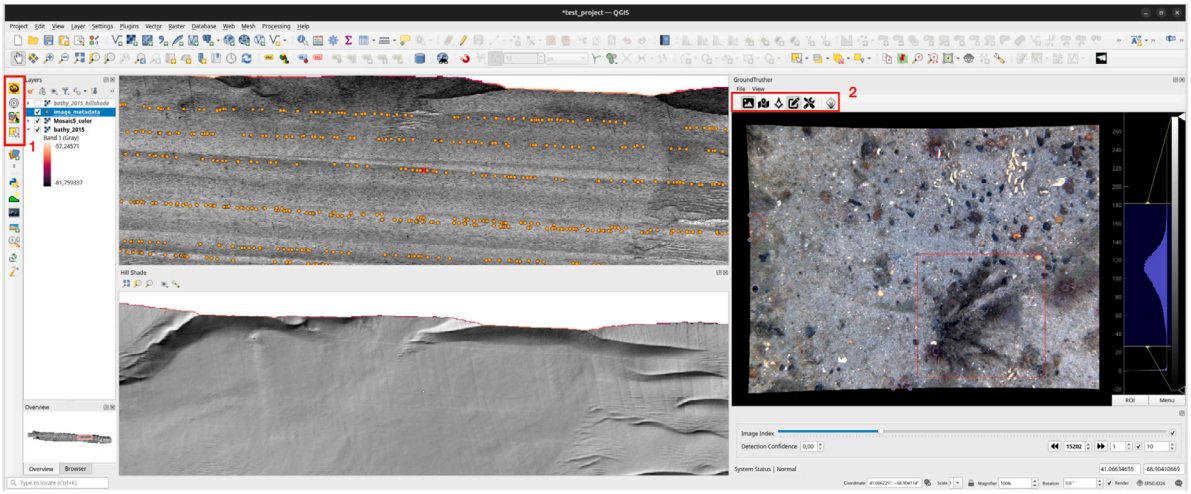


Fig. 2. Map canvas. Integrated into *QGIS*, *Groundtruther* includes two toolbars (highlighted), one (1) for activating the main window and to access the tools to interact with the map canvas (zoom to the closest image, GRASS GIS Query, and set the GRASS GIS computational region) and a second toolbar (2) to hide/show the main *Groundtruther* functionalities. The location of the selected image from the *image browser* is rendered in the QGIS map canvas as red marker, in this image overlaid on the acoustic backscatter mosaic.

information through an index built on the image metadata. Image metadata may also contain other information such as filename, and values for the several sensors available on the HabCam V4. This index is used as a reference to synchronize the action assigned to the image browsing widgets. Any activity on one of the image browsing widgets will trigger an update of the image metadata view as well as all other widgets. For example, when a checkbox is enabled, any action on the image browser section will send the image location and a zoom value, read from a *spinbox*, to the *Groundtruther* map canvas, which will then pan and zoom to the correct location. For signaling actions in the GUI, we use the *pyqt* signaling system, while for the image metadata indexing and query, we use the data structures and data analysis tools available in the *pandas* (Pandas Development Team, 2023) software library. The viewer can perform basic image enhancement (histogram stretching) and overlay image annotations as bounding boxes. Each annotation can be selected by means of pointing and clicking on the image, and if found, the associated label is rendered in the Image Browser status bar.

3.2. Map canvas

Linking images to geospatial information (including both raster and vector data) has been a critical driver for the development of this application. The ability to synchronize the location of geospatial datasets with the browsing of in-situ observations (images) can lead to new insights into seafloor variability (e.g., easily identify patches, boundaries, etc.) and the correlation of this variability with seafloor features, such as its morphology. Therefore, the *Groundtruther* software has been fully integrated into the open-source Geographic Information System (GIS) *QGIS* via its plug-in system and uses the *PyQGIS* API (Application Programming Interface) (QGIS Development Team, 2023) to interact with the *QGIS* map canvas (Fig. 2). The linkage between the Image Browser and map canvas is based on a KD-Tree spatial index built from the image metadata. The tool automatically builds the KD-Tree index once the image metadata is loaded. The synchronization is then enabled via a map-tool button on the map canvas toolbar, which, when enabled, will find and display the closest image in the image browser via point-and-click on the map canvas. The location of the selected image from the *image browser* is rendered in the QGIS map canvas as a red marker (Fig. 2).

Through the *Groundtruther* toolbar the user has access to some of the surface analysis tools available in the GRASS-GIS (Geographic Resources Analysis Support Software) software (GRASS Development Team, 2023a; Neteler et al., 2012). It also includes tools to set the GRASS GIS computational region and query the GRASS database at a given location. Only a limited set of geoprocessing tools is enabled and is available via a dedicated panel in the user interface. This is discussed further in the Bathymetric Processing section below.

3.3. Backscatter query builder

The input used by the Backscatter Query Builder is a point cloud of spatially georeferenced points where each location has associated attributes including backscatter strength and grazing angle calculated using the actual seafloor slope. For the backscatter values we included two fields namely “backscatter” and “corrected backscatter”, as derived from *Qimera-QPS* (QPS, 2023). Further details on how these data were acquired and processed can be found in Di Stefano and Mayer (2018). The file structure (1) is consistent with the output from the *Qimera-QPS* per-beam-details survey line export. The point cloud is created by merging all the available survey lines. The tool accepts inputs as a text file with a comma-separated value (CSV) or as a parquet binary file (which is preferred for performance reasons). The Backscatter Query Builder allows the user to spatially query the acoustic data with a user-defined geometric shape. The user can choose between a rectangular or an elliptical shape. The central position for the sampling shape is derived from the location of the selected image in the image browser; then, the user can set its size and orientation from the Backscatter Query Builder GUI (Fig. 3). For the selected shape, we also query the image metadata to select the available images within the sampling area, which are then rendered in the Image Selection tab of the Backscatter Query Builder toolbox.

Three types of widgets are included. The first is an angular range analysis (ARA) interactive widget to explore the variability of backscatter intensity (y-axis) as a function of the incidence angle (x-axis). In the ARA widget, all the selected beams are rendered as points in a scatterplot where the sign for the incidence angle represents the side of the swath, with positive values for beams coming from the starboard side and negative values for the port side. The user has the option to “fold” the data points by means of a checkbox. When checked, the app will consider the absolute value for the angles and will plot all the points on the positive side of the X-axis. In addition, a nth-degree polynomial model is fit to the data points; the user can choose the degree for the fitting curve as well as apply some axis and data transformation (i.e., down-sampling). This functionality works as a visual aid to detect trends in the data.

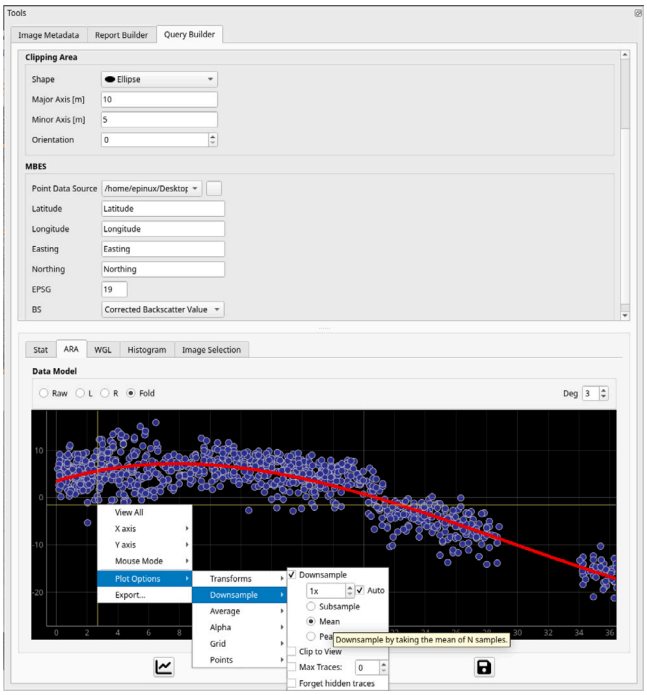


Fig. 3. The Backscatter Query Builder, available as a detachable panel from the main window, comprises two main sections, each of which can be hidden to increase the widget's usability. The first section on top is used to define the query parameters, like query shape size and orientation. Among other options, we include the possibility of selecting the field used to query the backscatter values, which can be configured to have multiple choices. The second section includes several widgets, starting from the left: a table view with a statistical description of the selected points; the ARA widget (selected in the figure, x-axis: angle [°], y-axis: backscatter [dB]); a 3D viewer; and a widget to render the selected images.

The second widget is an *OpenGL* scene containing a 3D surface generated using the depth information contained in the point cloud. The user can navigate the scene and save the current frame as an image to embed in a report (see below). The third widget is a plotting tool to explore the distribution of the selected backscatter, which by default, uses a Gaussian kernel for density estimation. The widget includes three types of density plots. The first option uses the whole dataset to draw a single density curve, and the other two offer the possibility to group the data by a categorical attribute; in this case, the plot consists of as many density curves as the number of available categories. The amplitude of the grouped density curves is normalized or scaled by the count computed by the density function based on the selected option. By default, we use the survey line as a grouping category, information available from the point cloud.

3.4. Bathymetric processing engine

Groundtruther grants access to geoprocessing routines to extract derivatives from digital bathymetry. As the main processing core, we developed the GRASS Representational State Transfer (REST) API, henceforth referred to as GRASS API. The *GRASS GIS* gives access to hundreds of geoprocessing routines, is fully scriptable, and has a 'multi-user' design (Neteler and Mitasova, 2008). For the REST interface of the *GRASS API*, we base the development on the *FastAPI* (FastAPI Development Team, 2023) library, which is used to develop the web APIs.

To deploy the GRASS API service, a *docker* container is made available, which can be deployed via *docker-compose*. The *docker-compose* also includes a *Jupyterlab* service which has access to the same libraries used by the *GRASS API* and by the *Groundtruther* application. This approach allows the user to access a full suite of development tools to consume the *GRASS API* and to prototype and test new geoprocessing routines that can then be integrated into the system. For the *GRASS API* and *Jupyterlab* services, we utilize *GRASS GIS* version 8.3 with the full suite of GRASS extension modules (GRASS Development Team, 2023b) preinstalled.

We access the GRASS API directly from the *Groundtruther* via the map canvas toolbar, where the user can enable the grass tools, define the *GRASS GIS* computational region, query the available *GRASS GIS* database, or run one of the available geoprocessing routines. The *GRASS API* developed here implements a subset of the geoprocessing tools available in the *GRASS GIS* software via its *Python* bindings, *PyGRASS* (Zambelli et al., 2013). Specifically, we exposed the basic tools to handle the import/export of datasets into the *GRASS GIS* database, tools to set the geo-computational region, and the main tools to extract bathymetric derivatives (including slope, aspect, and various types of terrain curvature) (Wood, 1996), and geomorphological features, and various descriptors of the form such as intensity, exposition, range, variance, elongation, azimuth, extent, and width (Jasiewicz and Stepinski, 2013).

The list of implemented modules includes:

- Set the GRASS session environment. It includes creations of new GRASS LOCATION and MAPSET via georeferenced files or EPSG codes (*g.gisenv*)
- Set the GRASS computational region (*g.region*)
- Get a list of raster and vector layers for a given LOCATION/MAPSET (*g.list*)
- Query a given list of Raster layer for a given location (*r.what*)
- Convert to and from map canvas coordinates and the projection system in use in a given GRASS Location (*m.proj*)
- Extract surface derivatives and morphological features (*r.geomorphon*, *r.param.scale*)
- Manage access to GRASS raster and vector layer (*g.remove*)
- Access to georeferenced file metadata (*gdalinfo*)

Due to the scriptable nature of *GRASS GIS* it is also possible to execute a complex workflow. As an example, we included a routine to compute a geospatial rule-based model (GRM) analysis (Di Stefano and Mayer, 2018) for the quantitative analysis of a bedform (Fig. 4).

3.5. Report builder

Narratives are effective tools for communicating scientific data and findings. In *Groundtruther*, the Report Builder makes it possible for various findings to be connected, illustrated, explained, and interpreted in a flexible platform, much like in a narrative. The Report Builder (Fig. 5) comprises two main components. The first part includes a number of widgets to set the geometry and rendering options for the vector output, which is in Keyhole Markup Language (KML) format. The included widgets are used to set some KML options like image icon, line width and color, options to extrude the marker from the ground, and parameters to specify the look-at camera settings. The second section of the Report Builder provides the basic features found in a rich text editor and, in addition, includes tools to embed images from other components of the *Groundtruther* system (e.g., plots saved from the Backscatter Query Builder or images and associated metadata from the Image Browser). The produced KML file includes rich text description, embedding in a single document geographic information, images, and quantitative information. The Report Builder also can export the rich text description as portable document format (PDF). Associated with the KML output, the tool also produces a machine-readable output (JSON format) suitable for passing information to other software/analytical workflow (e.g., ingesting the data in a routine for habitat modeling).

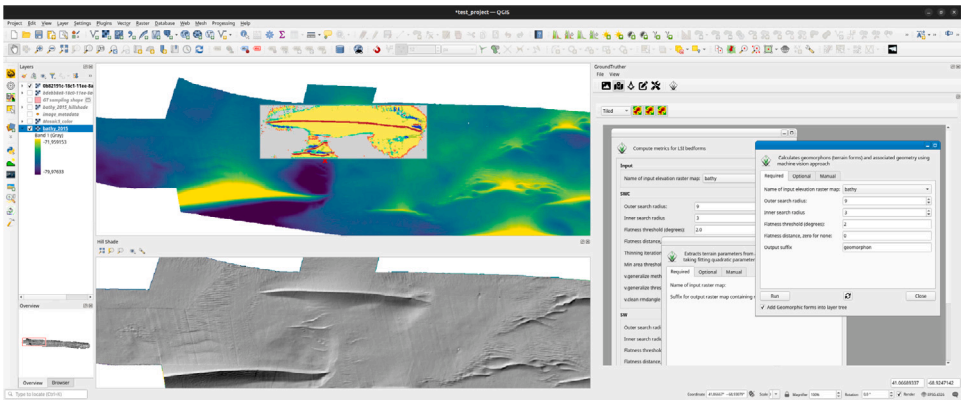


Fig. 4. The bathymetric processing engine is a front end to the *GRASS API* and provides a user interface to send HTTP requests to the API and retrieve the resulting objects from the API response. The returned objects are then decoded and embedded in the output widgets.

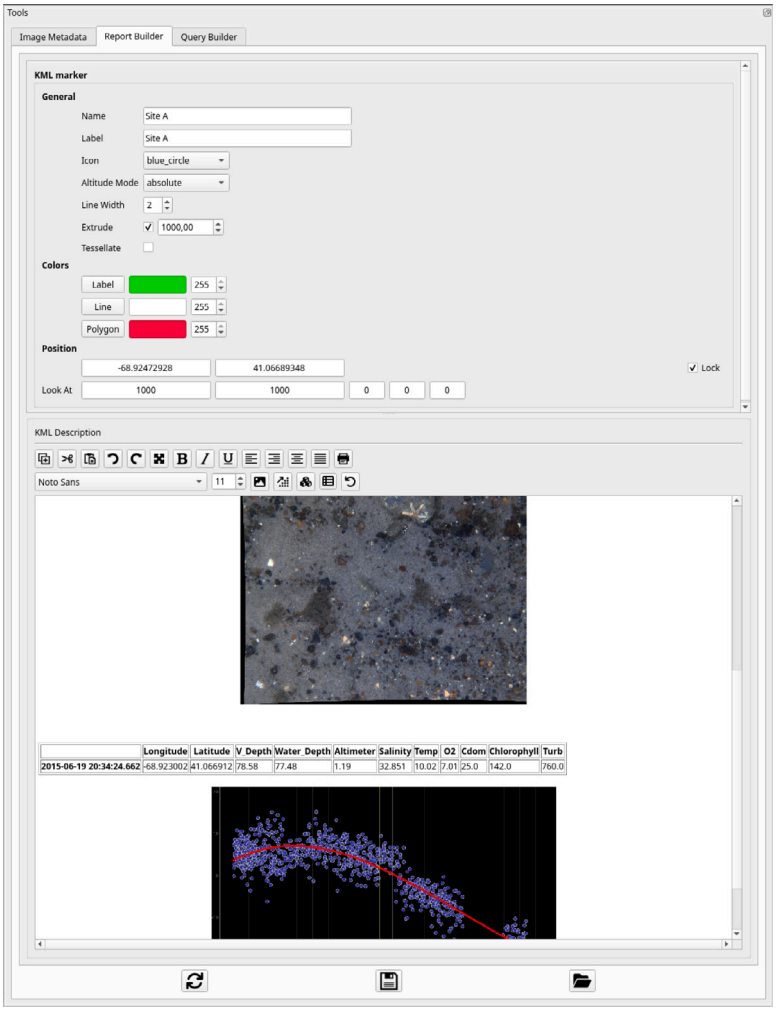


Fig. 5. The report builder comprises two parts, each of which can be hidden to increase tool usability. The first part sets the KML output option. The second part is a rich text editor and includes tool buttons to capture the active image from the Image Browser and the output generated by the Backscatter Query Builder.

4. Applications

4.1. Bathymetric analysis

4.1.1. Extraction of bathymetric derivatives

To demonstrate the usability of *Groundtruther* for bathymetric analysis, we started by setting the computational region to the area of the GSC surveyed in 2015, henceforth referred to as the study area. We then used the *Groundtruther* Bathymetric Processing Engine to run two *GRASS* commands that produce bathymetric derivatives, namely *r.param.scale*, and *r.geomorphon*. For this region, we generated the following derivatives: slope, aspect, profile curvature, plan curvature, longitudinal curvature, cross-sectional curvature, maximum curvature, and minimum curvature from *r.param.scale* (see Wood (1996) for definitions), as well as intensity, range, variance, extent, azimuth,

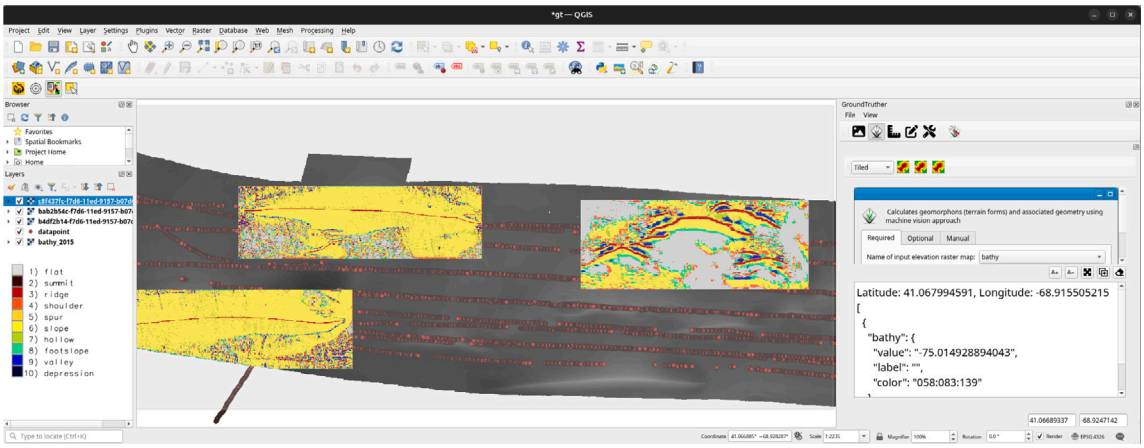


Fig. 6. Bathymetric analysis output and *GRASS* module dialog - *r.geomorphon*. A point-and-click query on the *GRASS* database results is shown in the panel on the right side of the map canvas.

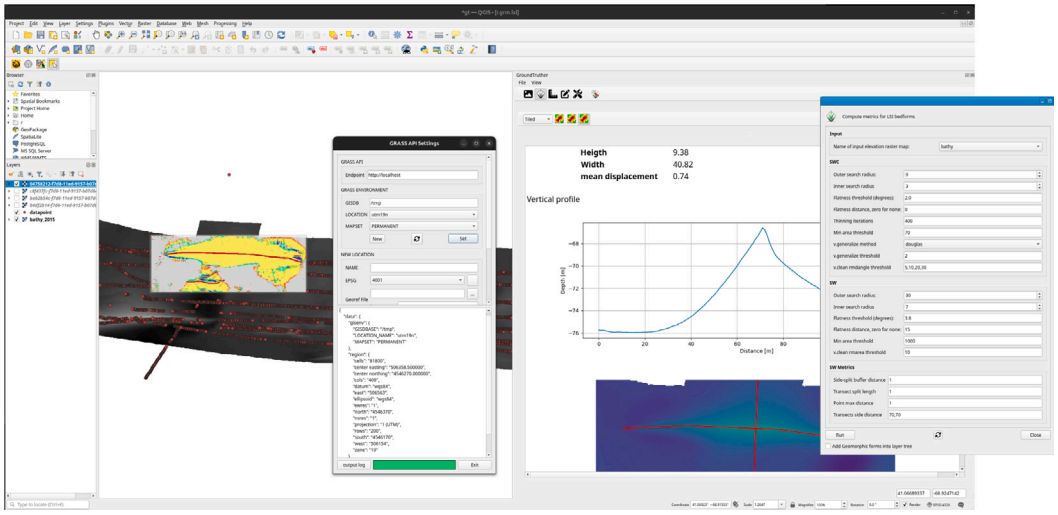


Fig. 7. *Groundtruther* map canvas and *GRASS GIS* tools. Here, we show a bathymetric grid overlaid with the bounding box of the selected *GRASS GIS* computational region, and the output of *r.geomorphon*. The *GRASS GIS* computational region is set by dragging a box on the map canvas when the relative map tool is checked from the map canvas toolbar. A *GRASS LOCATION* needs to be active for the action to work. The Location setting dialog is shown in the middle. On the right, we show the dialog to run the GRM LSI analysis, and its output is shown embedded in the query output panel.

elongation, and width, from *r.geomorphon* (see Jasiewicz and Stepinski (2013) for definitions). This stack of raster layers can be queried via the *GRASS* maptool query (by means of point and click on the map). The *GRASS* query output can then be sent to the Report Builder or used to compile a feature vector matrix (see Fig. 6).

We used *Groundtruther* to extract geomorphological features (namely, flat, peak, ridge, shoulder, spur, slope, hollow, footslope, valley, and pit) as generated by the *r.geomorphon* command. We found this result to be more robust (i.e., more scale-independent and less sensitive to terrain artifacts) relative to the features generated using the *r.param.scale* method, as shown in Di Stefano and Mayer (2018), and therefore we only present the output from *r.geomorphon*. The resulting categorical map is shown in Fig. 7.

4.1.2. Quantitative characterization of bedforms

We then set a smaller computational region around a single bedform clearly visible from the bathymetry. At this scale, geomorphological features can be used to isolate and delineate bedforms. Once a bedform was isolated, we performed a quantitative analysis thereof by applying the geospatial rule-based model (GRM) proposed by Di Stefano and Mayer (2018), which has been integrated into *Groundtruther* and is run through a dialogue interface (Fig. 8). The GRM tool returned the bedform's height, width, and length, identified its lee side and stoss side, detected its peak, and generated a maximum vertical profile in one single step. Raster and vector results are then overlaid and compiled in a map using the *GRASS GIS* rendering driver, which is included in the output.

4.2. Seafloor variability

We used three locations (hereafter referred to as A, B, and C) to document seafloor variability from relative backscatter point cloud data patterns. These locations were selected on the basis of what was visible from the bathymetry and images, as well as with the additional criterion of being in areas with a high density of survey lines so that the corresponding point clouds would also be relatively dense. Locations A and B were located in dynamic areas dominated by fine, mobile sediment. Location C was located in a more stable area of consolidated seafloor with sessile fauna present. We used the Backscatter Query Builder to specify the dimensions of the shape to query (i.e. extract) the data, which in this case it was an ellipse. We set major and minor axis length to 10 m and 5 m, and oriented the minor axis parallel to the MBES survey track lines. Within the extracted point clouds it is possible to analyze the distribution of backscatter values (by means of density

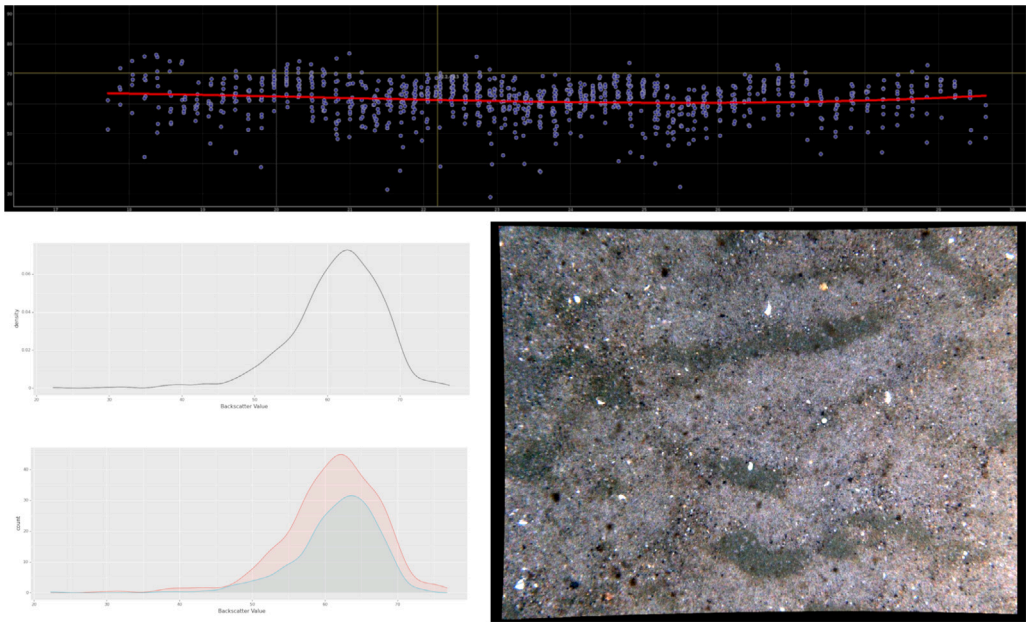


Fig. 8. Sampling location "A", Latitude: 41.0676, Longitude: −68.9275. The query shape was an ellipse with major axis direction: North-South, and length: 10 m, minor axis direction: East-West, and length: 5 m. Top: ARA plot, Backscatter amplitude (dB) on the *y*-axis, Incidence angle (deg) on the *x*-axis; Lower left: density plots (x-axis: Backscatter amplitude (dB), y-axis: sample density), whole dataset (Top), and grouped by survey lines (bottom); Lower right: HabCam V4 Image for the selected location. Notice the oscillatory pattern in the backscatter angular response and its unimodal distribution. The selected images showed an accumulation of gravel and coarse sand in the ripple's trough.

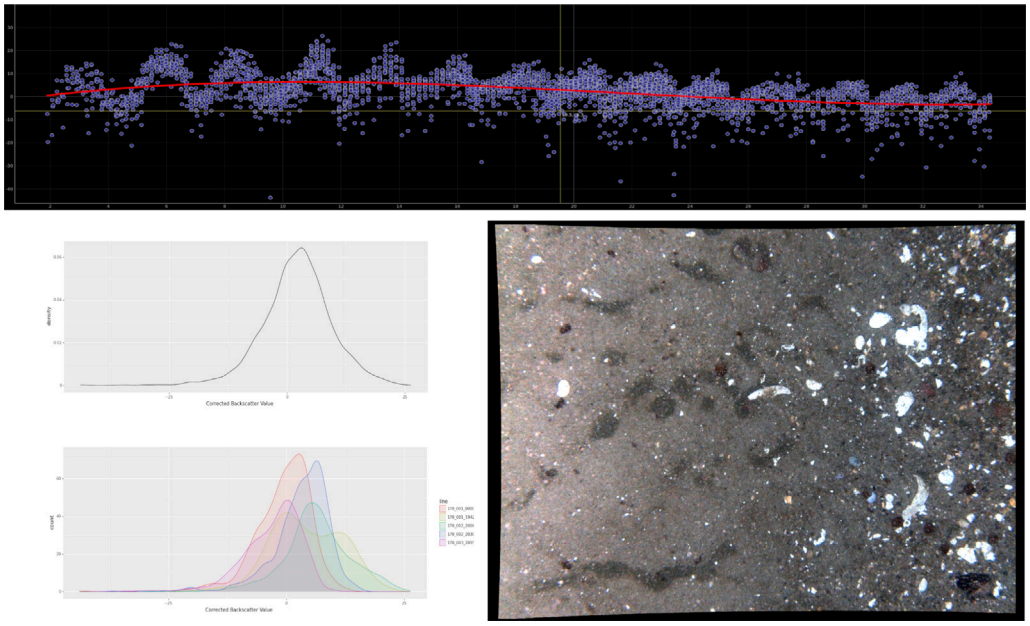


Fig. 9. Sampling location "B", Latitude: 41.0660 Longitude: −68.898. The query shape was an ellipse with major axis direction: North-South, and length: 10 m, minor axis direction: East-West, and length: 5 m. Top: ARA plot, Backscatter amplitude (dB) on the *y*-axis, Incidence angle (deg) on the *x*-axis; Lower left: density plots (x-axis: Backscatter amplitude (dB), y-axis: sample density), whole dataset (Top), and grouped by survey lines (bottom); Lower right: HabCam V4 Image for the selected location. Notice the oscillatory pattern in the backscatter angular response and its unimodal distribution. The selected images showed an accumulation of shell hash in the ripple's trough.

plots), as well as the relationship between backscatter values and angle of incidence (by means of ARA plots).

In locations A (Fig. 8), and B (Fig. 9), we observed a similar oscillatory trend in the ARA plots, though this pattern was slightly more clear in B. In addition, the density plots are both unimodal. Both locations were situated in the middle of two different ripple fields, which could explain the oscillatory pattern observed in the ARA plot. The main difference from locations A and B emerged when looking at the image selection, where images from location A show a high content of gravel in the troughs of the sand ripples, and images from location B contain more shell hash.

In contrast, the ARA plots for location C do not show any clear pattern and have a relatively clear bi-modal distribution (shown on the corresponding density plot). This pattern (bi-modality) is typically associated with the presence of two types of substrate, but before making that inference, a more careful inspection is required. After

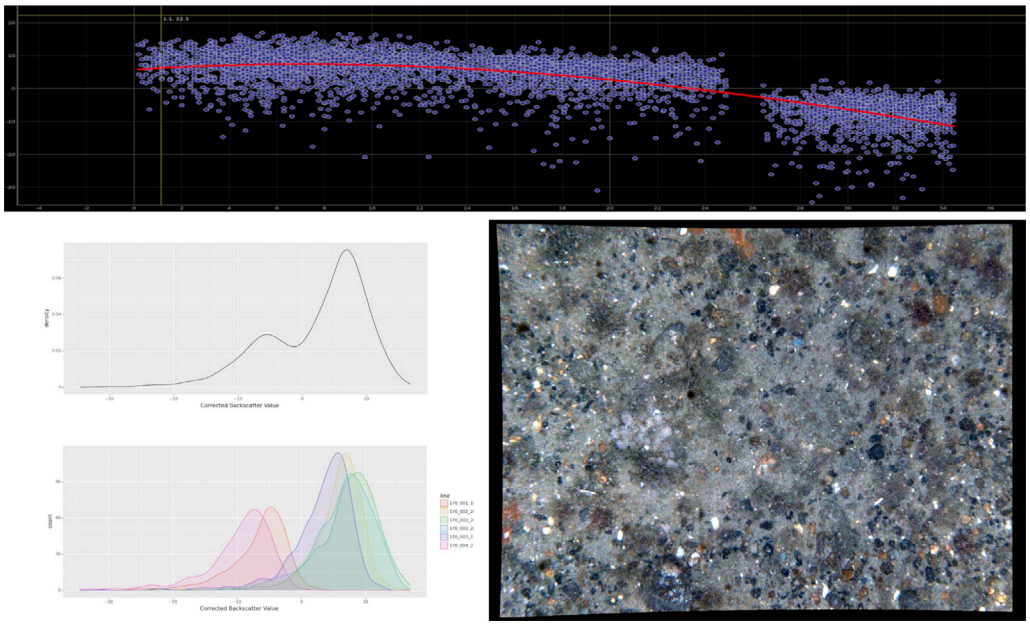


Fig. 10. Sampling location "C", Latitude: 41.0658 Longitude: −68.9037. The query shape was an ellipse with major axis direction: North-South, and length: 10 m, minor axis direction: East-West, and length: 5 m. Top: ARA plot, Backscatter amplitude (dB) on the *y*-axis, Incidence angle (deg) on the *x*-axis; Lower left: density plots (x-axis: Backscatter amplitude (dB), y-axis: sample density), whole dataset (Top), and grouped by survey lines (bottom); Lower right: HabCam V4 Image for the selected location. Notice the density plot distribution, which shows the high dependency on the survey lines.

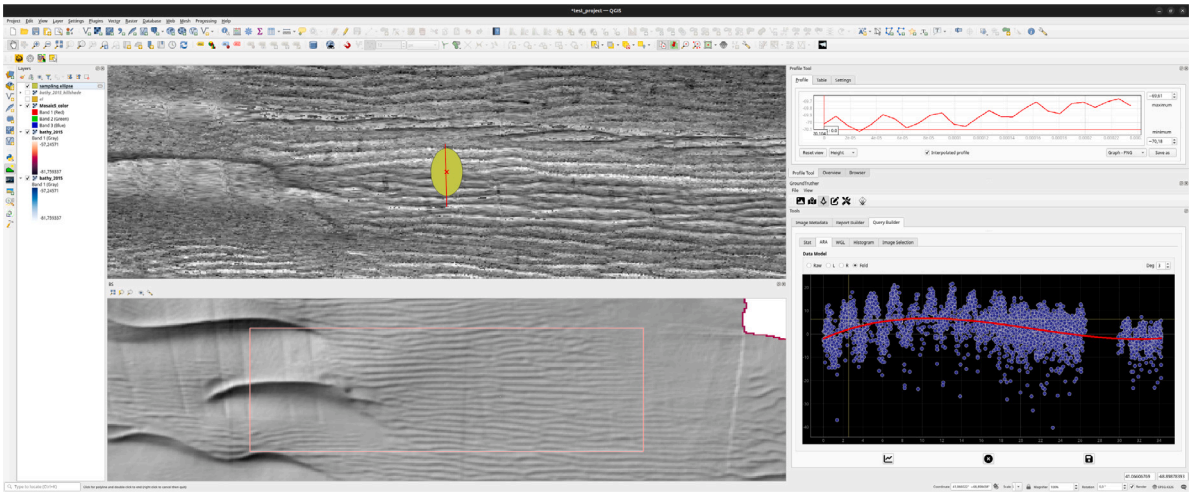


Fig. 11. QGIS and Groundtruther: example view displaying integrated tools for the exploration in proximity of location "B". On the left: QGIS map-canvases: top panel Backscatter mosaic, overlay: sampling shape generated from the Backscatter query builder (green); Vertical profile path (red); bottom panel: bathymetric shaded relief; On the right: top panel vertical profile tool; on the bottom ARA scatterplot for the selected sampling shape.

splitting the density plot by survey line, it becomes apparent that at location C (Fig. 10) the separate distributions had very different means, confirming the dependency of the backscatter values on the survey line azimuth (Lurton et al., 2018). Indeed, the (over 30) images show a constant type of sediment within the queried ellipse.

5. Discussion

Despite some data limitations (see below), we have been able to generate new knowledge about the seafloor of the GSC using *Groundtruther*, even in its current prototype state. For example, by browsing images known to be on ripple fields, we determined two types of sand ripples based on the material accumulated in their troughs, which can lead to the generation of new hypotheses on sediment dynamics. We have also successfully reproduced previously published analyses and results by Di Stefano and Mayer (2018) in a much more automated and repeatable way. We have found additional evidence for the strong dependence of backscatter strength on the angle of incidence, but, more importantly, we offer tools to more easily and precisely address this well-known problem. By being able to quickly co-register seafloor imagery with the acoustic response of the seafloor, the Backscatter Query Builder can help distinguish between situations where the effects from the incidence angle are due to the slope of the terrain, its roughness, or the azimuth of the survey line (Lurton et al., 2018). While analyzing these data sets independently may well lead to similar conclusions as was the case for the results published in Di Stefano and Mayer (2018), *Groundtruther* has made the process faster, less subjective, and more repeatable. An example of Groundtruther in action is shown in Fig. 11.

The main strength of *Groundtruther* lies in its ability to browse images while keeping track of image location, such that it is possible to quickly scan multiple images that have a common spatial context. Alternatively, a series of contiguous images can be used to define zones

by finding the exact location where a change is detected. These two alternative approaches have been described as a top-down and bottom-up approach in habitat mapping (Eastwood et al., 2006; LaFrance et al., 2014). The latter approach was used in Di Stefano and Mayer (2020), where microhabitats found around bedforms were described using the same datasets and for the same area used here, albeit lacking the ability to spatially browse images (it was indeed this constraint that prompted the development of *Groundtruther*). *Groundtruther* has the potential to significantly facilitate habitat mapping by helping (1) identify patches and/or features, and (2) characterize these, whether quantitatively (e.g., seafloor composition, bedform attributes) or qualitatively (e.g., images, reports, etc.). In addition, *Groundtruther* can boost benthic SDM. Firstly, it can import image annotations, such as those detected through Artificial Intelligence (AI) feature detection algorithms, e.g., FathomNet (Katija et al., 2022). Secondly, it has the ability to quickly generate feature vector matrices for sets of predefined spatial points, ready for ingestion into predictive models of species distributions. We believe that *Groundtruther* would also be an ideal front-end for AI training or other AI applications like data labeling. The image browser allows the user to browse images by time as well as space. This can simulate the process of image acquisition in the field, which can help interpret the data. However, the functionality of *Groundtruther* is fully exploited when images are accurately georeferenced, namely, when position error is, at most, less than the size of the mapping unit. Indeed, in order to use the HabCam V4 dataset, the use of a USBL has been crucial. In addition to increased navigational safety, the image quality is also improved as this setup facilitates more intelligent navigation of the camera over the seafloor, for example, by enabling it to maintain a constant altitude.

For testing the application, we have used datasets, particularly the acoustic point cloud, which are affected by issues related to the fact that the hardware was not calibrated and that the acquisition settings were optimized for collecting high-resolution bathymetry. By choosing the highest available frequency in the MBES settings (400 kHz), we achieved a 1 m resolution in the bathymetry at 90 m depth. At such depth and high frequency, the signal attenuation negatively affected the quality of the backscatter. Assuming the availability of high-quality, calibrated backscatter data, *Groundtruther* would allow users to create a database of backscatter strength and its angular variability, that could be associated to typical bottom types, leading to the enrichment of ground-truth databases used to develop formal geophysical models that link acoustic backscatter observations to intrinsic properties of the seafloor.

The choice of software when it comes to analyzing backscatter and its angular response curve is quite limited. The widely used Geocoder (Fonseca and Calder, 2005) which is integrated into a number of commercial software packages allows the user to extract ARA information, but there is a lack of tools to use and correlate such information with physical measurements or visual observations. We consider *Groundtruther* to be a step in this direction, adding context into data exploration (e.g., offering tools to explore acoustic backscatter in relation to images or to surface properties). *Groundtruther* puts the ground-truth data into context, and allows the user to interact with it, validate hypothesis and make new discoveries.

CRediT authorship contribution statement

M. Di Stefano: Conceptualization of this study, Methodology, data preparation, and Software development. **G. Gonzalez Mirelis:** Review of this study, provided feedback on software functionalities, and Software testing. **L. Mayer:** Review of this study, and provided feedback on software functionalities.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Source code is released as open source software and made publicly available at - https://github.com/epifanio/groundtruther/, and - https://github.com/epifanio/grass_api/.A sample dataset was also produced to test the application and has been made available under a ccby4 open license - https://zenodo.org/records/7995674/.

Acknowledgments

This work was supported by NOAA, United States GRANT NA15NOS 4000200.

References

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. Estuar. Coast. Shelf Sci. 92 (3), 502–520. http://dx.doi.org/10.1016/j.ecss.2011.02.007.

Che Hasan, R., Ierodiaconou, D., Laurenson, L., 2012. Combining angular response classification and backscatter imagery segmentation for benthic biological habitat mapping. Estuar. Coast. Shelf Sci. 97, 1–9.

Clarke, J.H., 1994. Toward remote seafloor classification using the angular response of acoustic backscattering: a case study from multiple overlapping GLORIA data. IEEE J. Ocean. Eng. 19 (1), 112–127. http://dx.doi.org/10.1109/48.289456.

Di Stefano, M., Mayer, L., 2018. An automatic procedure for the quantitative characterization of submarine bedforms. Geosciences 8 (1), 28. http://dx.doi.org/10.3390/geosciences8010028.

Di Stefano, M., Mayer, L., 2020. Geomorphology and microhabitats of large, isolated, immobile bedforms in the great south channel, northwest atlantic ocean. In: Seafloor Geomorphology as Benthic Habitat. Elsevier, pp. 503–518. http://dx.doi.org/10.1016/b978-0-12-814960-7.00029-4.

Eastwood, P.D., Souissi, S., Rogers, S.I., Coggan, R.A., Brown, C.J., 2006. Mapping seabed assemblages using comparative top-down and bottom-up classification approaches. Can. J. Fish. Aquat. Sci. 63 (7), 1536–1548.

FastAPI Development Team, 2023. Fastapi framework. URL: https://fastapi.tiangolo.com/.

Fonseca, L., Brown, C., Calder, B., Mayer, L., Rzhanov, Y., 2009. Angular range analysis of acoustic themes from stanton banks Ireland: A link between visual interpretation and multibeam echosounder angular signatures. Appl. Acoust. 70 (10), 1298–1304. http://dx.doi.org/10.1016/j.apacoust.2008.09.008.

Fonseca, L., Calder, B.R., 2005. Geocoder: An efficient backscatter map constructor. In: U.S. Hydrographic Conference. URL: http://ushydro.thsoa.org/hy05/08_3.pdf.

Fonseca, L., Mayer, L., 2007. Remote estimation of surficial seafloor properties through the application angular range analysis to multibeam sonar data. Mar. Geophys. Res. 28 (2), 119–126. http://dx.doi.org/10.1007/s11001-007-9019-4.

GRASS Development Team, 2023a. GRASS - geographic resources analysis support system. URL: https://grass.osgeo.org/.

GRASS Development Team, 2023b. GRASS - geographic resources analysis support system - extension modules. URL: https://github.com/OSGeo/grass-addons.

Hasan, R.C., Ierodiaconou, D., Laurenson, L., Schimel, A., 2014. Integrating multibeam backscatter angular response, mosaic and bathymetry data for benthic habitat mapping. In: Thrush, S. (Ed.), PLoS ONE 9 (5), e97339. http://dx.doi.org/10.1371/journal.pone.0097339.

Hou, Z., Chen, Z., Wang, J., Zheng, X., Yan, W., Tian, Y., Luo, Y., 2018. Acoustic impedance properties of seafloor sediments off the coast of southeastern Hainan, south China sea. J. Asian Earth Sci. 154, 1–7.

Howland, J., Gallager, S., Singh, H., Girard, A., Abrams, L., Griner, C., Taylor, R., Vine, N., 2006. Development of a towed survey system for deployment by the fishing industry. In: OCEANS 2006. IEEE, http://dx.doi.org/10.1109/oceans.2006.307098.

Ierodiaconou, D., Laurenson, L., Burq, S., Reston, M., 2007. Marine benthic habitat mapping using multibeam data, georeferencedvideo and image classification techniques in victoria, Australia. J. Spat. Sci. 52 (1), 93–104.

Ilich, A.R., Brizzolara, J.L., Grasty, S.E., Gray, J.W., Hommeyer, M., Lembke, C., Locker, S.D., Silverman, A., Switzer, T.S., Vivlamore, A., Murawski, S.A., 2021. Integrating towed underwater video and multibeam acoustics for marine benthic habitat mapping and fish population estimation. Geosciences 11 (4), http://dx.doi.org/10.3390/geosciences11040176, URL: https://www.mdpi.com/2076-3263/11/4/176.

Jackson, D., Richardson, M., 2010. High-Frequency Seafloor Acoustics. Springer, New York, NY.

Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology 182, 147–156. http://dx.doi.org/10.1016/j.geomorph.2012.11.005.

Katija, K., Orenstein, E., Schlining, B., Lundsten, L., Barnard, K., Sainz, G., Boulais, O., Cromwell, M., Butler, E., Woodward, B., Bell, K.L.C., 2022. FathomNet: A global image database for enabling artificial intelligence in the ocean. Sci. Rep. 12 (1), http://dx.doi.org/10.1038/s41598-022-19939-2.

LaFrance, M., King, J.W., Oakley, B.A., Pratt, S., 2014. A comparison of top-down and bottom-up approaches to benthic habitat mapping to inform offshore wind energy development. Cont. Shelf Res. 83, 24–44.

Lecours, V., Devillers, R., Simms, A.E., Lucieer, V.L., Brown, C.J., 2017. Towards a framework for terrain attribute selection in environmental studies. Environ. Model. Softw. 89, 19–30. http://dx.doi.org/10.1016/j.envsoft.2016.11.027.

Lurton, X., Eleftherakis, D., Augustin, J.-M., 2018. Analysis of seafloor backscatter strength dependence on the survey azimuth using multibeam echosounder data. Mar. Geophys. Res. http://dx.doi.org/10.1007/s11001-017-9318-3.

Lurton, X., Lamarche, G., 2005. Backscatter measurements by seafloor-mapping sonars guidelines and recommendations.

McGonigle, C., Collier, J.S., 2014. Interlinking backscatter, grain size and benthic community structure. Estuar. Coast. Shelf Sci. 147, 123–136. http://dx.doi.org/10.1016/j.ecss.2014.05.025.

Misiuk, B., Brown, C.J., 2022. Multiple imputation of multibeam angular response data for high resolution full coverage seabed mapping. Mar. Geophys. Res. 43 (1), http://dx.doi.org/10.1007/s11001-022-09471-3.

Misiuk, B., Lecours, V., Dolan, M.F.J., Robert, K., 2021. Evaluating the suitability of multi-scale terrain attribute calculation approaches for seabed mapping applications. Mar. Geod. 44 (4), 327–385. http://dx.doi.org/10.1080/01490419.2021.1925789.

Neteler, M., Bowman, M.H., Landa, M., Metz, M., 2012. GRASS GIS: a multi-purpose open source GIS. Environ. Model. Softw. 31, 124–130. http://dx.doi.org/10.1016/j.envsoft.2011.11.014.

Neteler, M., Mitasova, H. (Eds.), 2008. Open Source GIS. Springer US, http://dx.doi.org/10.1007/978-0-387-68574-8.

Pandas Development Team, 2023. Pandas: python data analysis software. URL: https://pandas.pydata.org/.

Parnum, I.M., 2007. Benthic habitat mapping using multibeam sonar systems. URL: https://espace.curtin.edu.au/handle/20.500.11937/1131. doi:http://hdl.handle.net/20.500.11937/1131.

Porskamp, P., Young, M., Rattray, A., Brown, C.J., Hasan, R.C., Ierodiaconou, D., 2022. Integrating angular backscatter response analysis derivatives into a hierarchical classification for habitat mapping. Front. Remote Sens. 3, http://dx.doi.org/10.3389/frsen.2022.903133.

PyQt Development Team, 2023. Pyqt. URL: https://www.riverbankcomputing.com/static/Docs/PyQt5/index.html.

PyQtGraph Development Team, 2023. PyQtGraph: graphics library for PyQt5/PyQt6/PySide2/pyside6. URL: www.pyqtgraph.org.

QGIS Development Team, 2023. QGIS python API. URL: https://qgis.org/pyqgis/3.0/.

QPS, 2023. QPS - qimera. URL: https://qps.nl/qimera/.

Taylor, R., Vine, N., York, A., Lerner, S., Hart, D., Howland, J., Prasad, L., Mayer, L., Gallager, S., 2008. Evolution of a benthic imaging system from a towed camera to an automated habitat characterization system. In: OCEANS 2008. IEEE, http://dx.doi.org/10.1109/oceans.2008.5152032.

VIAME Development Team, 2023. VIAME - video and image analytics for multiple environments. URL: https://github.com/VIAME/VIAME.

Wood, J., 1996. The geomorphological characterisation of digital elevation models. URL: http://hdl.handle.net/2381/34503.

X., L., G., L., 2015. Backscatter Measurements by Seafloor-Mapping Sonars Guidelines and Recommendations. Technical Report, URL: https://niwa.co.nz/static/BWSG_REPORT_MAY2015_web.pdf.

Zambelli, P., Gebbert, S., Ciolli, M., 2013. Pygrass: An object oriented python application programming interface (API) for geographic resources analysis support system (GRASS) geographic information system (GIS). ISPRS Int. J. Geo-Inf. 2 (1), 201–219. http://dx.doi.org/10.3390/ijgi2010201.