Megafauna community assessment of polymetallic nodule fields with cameras: Platform and methodology comparison

Timm Schoening[1], Autun Purser[2], Daniel Langenkämper[3], Inken Suck[1], James Taylor[4],
Daphne Cuvelier[5,6], Lidia Lins[7], Erik Simon-Lledó[8], Yann Marcon[9,10], Daniel O.B. Jones[8],
Tim Nattkemper[3], Kevin Köser[1], Martin Zurowietz[3], Jose Gomes-Pereira[11,6], and Jens Greinert[1]

[1]GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany
[2]Alfred Wegener Helmholtz Institute for Polar and Marine Studies, Bremerhaven, Germany
[3]Bielefeld University, Bielefeld, Germany
[4]Senckenberg, Wilhelmshaven, Germany
[5]MARE - Marine and environmental sciences centre/IMAR - Instituto do Mar
[6]Centro OKEANOS - Universidade dos Açores, Horta, Portugal
[7]Ghent University, Ghent, Belgium
[8]National Oceanography Centre, University of Southampton Waterfront Campus, Southampton, UK
[9]MARUM Center for Marine Environmental Sciences, Bremen, Germany
[10]Department of Geosciences, University of Bremen, Bremen, Germany
[11]Naturalist, Lda. and Atlantic Naturalist Association, Horta, Portugal

Correspondence: Timm Schoening (tschoening@geomar.de)

Abstract. With the mining of polymetallic nodules from the deep sea seafloor once more evoking commercial interest, decisions must be taken on how to most efficiently regulate and monitor physical and community disturbance in these remote ecosystems. Image-based approaches allow non-destructive assessment of the abundance of larger fauna to be derived from survey data, with repeat surveys of areas possible to allow time series data collection. At time of writing, key underwater imaging platforms commonly used to map seafloor fauna abundances are Autonomous Underwater Vehicles (AUVs), Remotely Operated Vehicles (ROVs) and towed camera "Ocean Floor Observation Systems" (OFOSs). These systems are highly customisable, with cameras, illumination sources and deployment protocols changing rapidly, even during a survey cruise. In this study, 8 image datasets were collected from a discrete area of polymetallic nodule rich seafloor by an AUV and several OFOSs deployed at various altitudes above the seafloor. A fauna identification catalogue was used by 5 annotators to estimate the abundances of 20 fauna categories from the different data sets. Results show that, for many categories of megafauna, differences in image resolution greatly influenced the estimations of fauna abundance determined by the annotators. This is an important finding for the development of future monitoring legislation for these areas. When and if commercial exploitation of these marine resources commences, robust and verifiable standards which incorporate developing technological advances in camera-based monitoring surveys should be key to developing appropriate management regulations for these regions.

15 *Copyright statement.* TEXT

1 Introduction

The increasing demand for tech metals for consumer and industrial high technology devices has again stoked interest in the potential use of global deep-sea polymetallic nodule fields as exploitable sources of these materials in the near future (**???**). This increasing interest, simultaneously driving the technological development of marine mining equipment and the granting of exploration contracts within the Clarion Clipperton Fracture Zone (CCFZ) (**?**), has stimulated several recent European research projects (e. g. JPI Oceans MiningImpact 1&2 and MIDAS). These projects focused on the study of these remote ecosystems to better understand the nodule distribution (**?**), as well as the community structure of macro-fauna (**?**) and mega-fauna (**?**), ecosystem functioning and susceptibility to damage following anthropogenic perturbation and / or resource removal (**??**). Despite the occurrence of nodule fields in the Atlantic, Pacific and Indian Oceans, the majority of research efforts have been focused on the CCFZ, located in the North Central Pacific, as it has the highest known density of nodules (**???**), and the Peru Basin (South Central Pacific) (**???**). Both regions have been considered as potentially hosting commercial abundances of nodules at some point in history. Focused scientific study commenced in the 1980s, with simulated mining studies conducted in both areas, to assess the response of fauna to mining activities (**?**). These studies are summarised in **?**, with the "DISturbance and COLonization" (DISCOL) long-term study in the Peru Basin being the most extensively perturbated region of seafloor studied to date (**?**). Prior to the 1980s, only occasional opportunistic fauna collection records had been published from these areas. Since the 1980s, regular biological box core sampling has been conducted in the CCFZ, whereas the majority of fauna sampling at the DISCOL area has been image based, augmenting some initial trawl sampling deployments. The DISCOL experiment was designed to simulate what effects physical disturbances, such as those caused by future commercial deep-sea mining, might have on the seafloor and its inhabitants. In 1989, a plough-harrow was used to create a large-scale disturbance on the seafloor in the DISCOL Experimental Area (DEA). The plough harrow was deployed 78 times in 1989 with the aim of driving all polymetallic nodules from the sediment surface into the underlying soft sediments (Figure **??**) (**?**). This ploughing action destroyed the majority of surface megafauna and drove manganese nodules within 8 m diameter swathes down into the sediments. As a result, fauna that lived attached to the nodules was removed and thus destroyed. The soft-bottom community, however, did show signs of recovery 7 years after the plough disturbance. Several monitoring cruises of the impacted areas commenced in the following years and decades. The repopulation of the disturbed areas by highly motile and scavenging animals started shortly after the area was ploughed (**?**). Seven years later hemi-sessile animals had returned to the disturbed areas, but the total abundance of soft-bottom taxa was still low compared to the pre-impact study. However, nearby reference areas not impacted by the experiment indicated pronounced temporal variability in megafauna communities in the region (**?**). The ploughing activities also created a sediment plume that resettled in the surrounding areas. In these indirectly impacted areas, animal densities declined immediately after the ploughing event and although densities later (i.e. 3+ years) appeared to be greater than in the pre-impact study reference areas (**?**), megafaunal community composition in these areas remains significantly different than that found within plough tracks and reference areas (**?**). As has been reported from many ecosystems, the methodologies used to quantify fauna abundances and species diversity can greatly influence assessments. This challenges the direct comparison of regions sampled differently (**????**). Further, small variations in deployment techniques or sampling

set-ups (e.g variables such as mesh size or trawl speed for direct sampling, illumination, camera and lenses for remote sampling) can also influence the quality of the collected data (**?**), hampering comparison within the same study site. In this study, a range of commonly used imaging platforms were deployed at varying altitudes above the seafloor to survey megafauna across a defined region of the DEA which is a region of the Peru Basin with abundant seafloor nodule coverage. These collected images were then placed into the online image annotation system BIIGLE (**?**) and the fauna was identified in the different image sets by 5 annotators using a predetermined taxon catalogue. The hypothesis tested was that both composition and abundance of fauna differ between different imaging methodologies in polymetallic nodule fields. This study aims to provide useful information and guidance on how future optical monitoring of these and other remote ecosystems should most effectively and efficiently be conducted, should commercial exploitation of these remote resource fields commence.

1.1 Polymetallic nodules and associated fauna

Polymetallic nodules, as well as representing a potential commercial resource (**???**), are a key hard substratum that, in combination with the background soft sediment, act to increase habitat complexity and promote the occurrence of some of the most biologically diverse seafloor assemblages in the abyss (**??**). Nodule fields at the abyssal Pacific can comprise of nodules of up to 25 cm diameter (**?**), and at a range of abundance densities (e. g. 0-30 kg / m^2 (**?**). Processes of nodule formation are uncertain, though each individual nodule tends to form around a small shell fragment, shark tooth or equivalent small hard foci. With growth, individual nodules become heavier and capable of supporting, as an anchor or hard substrate, a range of larger filter feeding organisms (**??**), such as sponges (stalked (**?**) and encrusting (**?**)), stalked and non-stalked crinoids, soft and hard corals (**?**), xenophyophores (**?**), and sabellid worms etc. (**?**). Sessile organisms in turn support a diverse array of mobile and sessile epibenthic organisms, including further sponges, corals and worms, as well as mobile and semi mobile fauna such as amphipods, isopods, anemones, brooding octopodes (**??**) and many others (**?**). Although soft sediment stalked sponge fauna are found in nodule abundant regions, the nodule-based epifauna supports increased local biodiversity and abundance of species. In addition to providing a hard substrate for living attachment, nodules also increase the range of hydrodynamic niches available to the local ecosystem fauna (**?**), as well as adding complexity to food fall transport pathways. Recent cruise observations from the DISCOL region showed rapid transport of dead pyrosomes, following a surface bloom, to the seafloor (**?**). These dead pyrosomes were then hydro-dynamically trapped by benthic currents alongside nodules, providing a local food supply to the nodule community which might otherwise have been transported from the region by the ambient benthic flow conditions. This flow dynamic variability also impacts on the habitat niches available for infauna (across all infauna size classes) below and surrounding the nodules, with their presence influencing local biogeochemical activity and oxygen penetration pathways. At this crucial time point in research into polymetallic nodules and associated fauna, it is important to highlight also the gaps in current knowledge, and that any management plans developed take these shortfalls into consideration. At time of writing it is clear even from the sparsity of published megafauna papers from nodule regions that these ecosystems are not all alike. The Peru Basin region of the south Pacific seems to support a generally higher abundance of stalked fauna than the Clarion Clipperton Fracture Zone (CCFZ) nodule domains (**??**). Some large megafauna, such as benthic octopodes have thus far only been observed within these nodule ecosystems in the south Pacific (**?**), as have some fish species (**?**) despite the recent increased

sampling effort across the CCFZ. Conversely, the abundant sessile sponges recently characterised from the CCFZ, Plenaster craigi, (**?**) is not apparent in images or analysed samples collected from south of the equator. Whether these discrepancies are due to oceanographic, nutrient or habitat niche differences is not yet known. It may be considered that the larger nodule sizes found in the Peru Basin region are more suitable as anchors of sufficient stability for stalked fauna to allow brooding by octopodes for the hypothesised years required by deep sea incirrates (**?**). Another major absence in the scientific data set is sampled voucher specimens from nodule provinces. Opportunistic direct sampling by Remove Operated Vehicle (ROV) has taken place on a limited scale, though the ground-truthing of image and video data collected by ROV and Automated Underwater Vehicle (AUV) to species level is at present, not possible. Though this is an obvious disadvantage over direct sampling of the seafloor by trawl etc. to determine the present fauna mix, this is perhaps to some extent countered by the far larger areas which may be surveyed rapidly by towed and remote camera systems - an important point given the extremely sparse distribution of many fauna individuals of morphospecies in nodule ecosystems (**????**). These sparse distributions make impact assessments more problematic than for denser fauna categories which have historically been subject to direct impact by the offshore fishery or petrochemical industries, such as coral and sponge reefs, where atolls and accumulations can be directly surveyed prior and post-cruise, either via imaging or direct sampling (**???**). Whether future management plans favour a direct or an image-based monitoring approach to megafauna diversity and stock assessment, the requirement to fill these holes in extant voucher specimen collections from these regions is equally prescient.

1.2 Potential impacts associated with nodule extraction

Nodule collection will locally remove the major source of hard substrate in nodule field areas, rendering the remaining habitat unsuitable for some fauna (i. e. suspension feeders), as observed in experimental mining studies in the CCFZ (**??**) and DISCOL areas (**?**). Further, depending on the removal technique, the seafloor will likely be perturbed, with compaction tracks potentially formed, and all overcast by plume deposits (**??**). These features will increase the complexity of biogeochemical activity in the region (**?**), and influence local hydrodynamic conditions. Experimental tracks made with both epibenthic sled (**?**) and plough harrow (**?**) have created seafloor micro-topography which focused deposition of salps following a surface bloom event which occurred during SO242-2 (**?**). Such localised food input variability in the deep sea will likely result in a further modification of the fauna communities found in these exploited regions.

1.3 Methodologies for fauna abundance assessment

Box coring or multi coring are common survey methods in impact assessments and monitoring programmes, conducted to assess impacts on small fauna (e. g. less than 1 cm) following an anthropogenic impact event (**?**). For larger fauna, image-based surveys usually provide much more accurate estimations of benthic taxa richness and numerical density than traditional trawling techniques (**??**), and have no direct physical impact on the ecosystem being investigated. When planning to assess polymetallic nodule fauna abundance following commercial exploitation of these remote resource fields, the associated human impacts of monitoring programmes should be as little as possible. We therefore focus within this paper on the contrasting suitability of various image-based approaches to assess fauna abundance in polymetallic nodule ecosystems. Furthermore, image

data can be made publicly available to regulators, interested NGOs and other players easily via online platforms (**?**) allowing these stakeholders to conduct their own studies or analyses with the same primary data. To assure reliable monitoring, contractors need to publish data including uncorrupted location and timing metadata. The acquisition technology of that metadata needs to be fraud-proof (e.g. by incorporating navigation data into the imagery). In case of monitoring activities utilising directly collected fauna from box core, multicore or Remotely Operated Vehicle (ROV) collection, much of the material will be processed once, by one lab, and can degrade during the processing steps, preventing further studies. Image data also facilitate the straightforward archiving of collected data (**?**) for later comparison with subsequent images potentially collected up to decades after experimental or industrial disturbance, to assess long-term recovery rates. Given the extremely long lifespans of many deep-sea organisms (**??**), this is an important consideration when developing monitoring strategies for efficient and useful impact assessment within these ecosystems.

1.4 Factors determining the quality of deep sea image data

Samples collected by box cores, multicores or trawl are directly related to the surface area sampled. In this case, the type of trawl or corer may influence the comparability of the results to some extent (i.e. net size and tow speed important for trawls, closing mechanism for box corers). For image-base derived data, there are possibly a greater number of factors affecting the estimations of fauna abundance. The most significant of those are introduced below:

1.4.1 Camera optics

The area of seafloor which may be imaged by an optical platform is determined by the lens parameters used in the camera system, distance and orientation to the seafloor, sensitivity of the system to motion and illumination, and a range of other factors (**?**). Larger areas of the seafloor can be imaged with wide angle or 'fish eye' camera systems (**?**), though there is an associated vignetting effect rendering the details collected from the extremities of an image less rich than areas of seafloor more directly located below the lens centre (**??**). The raw images collected by those camera systems can appear quite distorted and manual labelling of fauna within these images is more difficult towards the edges of each image. Digital post-processing of these distorted images can be reasonably straightforward when the arrangement of optics for an imaging platform is known, and for larger fauna these processed images can be suitable for subsequent analysis (**??**). However, image processing cannot create 'newly improved' data and therefore there will always be a loss of information at the image extremities after lens correction. Image analysis could therefore focus on central parts of the image and the boundary area of images be used to display e.g. navigation metadata. Lenses of a more 'telephoto' or narrower angle will allow collection of less distorted images, though these collected images will capture a significantly smaller area of seafloor than may be achieved with wider angle systems.

1.4.2 Illumination and power provision

The deep sea is a dark environment with no sunlight penetration. It is therefore essential that camera systems are supplemented by artificial illumination. To provide sufficient illumination for video and still camera systems, abundant power reserves must

either be mounted on the platform or delivered via a cable from the support vessel. The amount of power which can be provided to a platform is determined by a range of design and operational parameters. Autonomous Underwater Vehicles (AUVs) for example must remain reasonably light-weight and must carry sufficient power to provide mobility and to take images at depth. Towed camera systems in contrast are always attached to a cable (e.g. coaxial, fibre-optic) which may provide sufficient power for continuous seafloor illumination. Positioning of the lights on an imaging platform can be difficult, and optimising the spread of light, i.e. maintaining an equal light balance across the imaged area, challenging. Illumination vignetting can be partially addressed prior to analysis by excluding the image edges from analysis (**??**). Given that AUVs must carry all required power (for mobility and imaging) with them can result in a less than optimal illumination of the seafloor (see **??**). No doubt light emitting diode (LED) technology will become more efficient, but at present these prevalent lower light condition data sets constrain the seafloor resolution which may be achieved during imaging surveys. Additionally, when lights and camera are mounted close to each other, a significant amount of light might be scattered by the water column into the camera, leading to a degraded "foggy" image, which is an issue for small platforms and/or high-altitude photography. Finally, the colour spectrum of the light also needs to be considered, as for instance the returned yellow, orange and red components of the signal may be too weak to support taxonomic identification, depending on the type of light source. The illumination system needs to be setup to accommodate the target altitude of the camera platform above the seafloor, as well as the expected altitude variation.

1.4.3   Platform altitude

The distance to an object can greatly alter the quality of an image. Although this may sound a straightforward parameter, it may play a hugely important role when analysing fauna abundances in an area. Maintaining a uniform altitude throughout and between survey deployments is highly desirable (i. e. to standardise the object/fauna detectability rates), but may be difficult. In regions of the World Ocean where the seafloor is highly complex, such as at deep water coral reefs (**?**) or within canyon systems (**?**), it can be a struggle to maintain an equal distance from camera optics from towed, autonomous, remote and submersible-based imaging platforms to the seafloor. For polymetallic nodule fields however, the seafloor is generally fairly uniform in depth, with very gentle slopes more the norm than occasional sudden slopes or cliff walls. Even so, towed platform altitude stability can be greatly influenced by operator skill, experience, environmental conditions (i.e. wave conditions at surface) or ship infrastructure (winch operational parameters / presence or absence of heave compensators). AUV imaging platforms are improving in stability and mission planning at a rapid rate (**??**), and maintaining flight altitudes is now a standard surveying procedure. Operations with these expensive devices tend to err on the side of caution; ground tracking often set with a conservative 5 - 10 m flight altitude. At these higher flight altitudes, more light is required to illuminate the seafloor than when a comparable AUV is deployed close to the seafloor (see **??**).

1.4.4   Data volume

Pioneer image-based studies in polymetallic nodule fields were conducted with analogue film-based camera systems (although live, black and white seafloor views were provided to towed systems via a basic TV camera setup) (**?**). This limitation constrained deployments to the collection of a few 100s of images. At present, camera systems can deliver many images per

second, even under low light conditions. This potentially high flow of image data however requires either an adequate digital storage space on the imaging platform (**?**), or the facility to be transferred directly to a shipboard storage system (**?**). This increased data flux allows for more complete spatial studies of the seafloor to be made with an imaging platform, but to get this additional information from the data set, increased processing time is required.

1.4.5 Dataset resolution

Image resolution is derived from a combination of the camera optics and the deployment altitude, and allows to compare image datasets numerically. The camera optics determine the pixel resolution (usually in the tens of megapixels for state of the art camera systems). The field of view of the camera objective lens and the deployment altitude determine the image footprint i.e. the area in square meters that is covered by a single image acquisition. These two values can be combined to a measure of Megapixels per square meter (MPix / m^2) or the numerically identical pixels per square millimeter to analyse the annotator performance and fauna density estimates consistently.

1.4.6 Time series studies

To determine the level of impact an event has had on a specific region of seafloor, repeated visits to a locale are required. It is important to conduct baseline and impact monitoring surveys in a region-specific manner to accommodate differences in faunal composition. Baseline information acquired in one nodule area (e.g. the CCZ) cannot directly be transferred to another (e.g. the Peru basin). Ideally, a number of surveys at differing times of the year would be conducted before an impacting event, to gauge the background fauna community of a region, and to identify natural variation and seasonality in community patterns. These baseline studies would be followed by repeated surveys at different time points during and after the impacting event. These repeated visits should allow quantification of the duration and recovery of impacts. Planning such a study may sound straightforward, but given the remoteness of many deep-sea regions, getting the same equipment and survey crew together may be difficult. One such study, aimed at gauging the impact of oil and gas exploration drilling on cold-water coral reefs on the Norwegian Margin visually surveyed a number of reefs on 5 occasions (**?**). Despite these 5 survey cruises taking place within a 3 year period in a relatively accessible area of Norwegian shelf, each cruise used different ROV systems and survey protocols. Analysis of collected data was further complicated by the mounting of different camera and illumination systems on each ROV, and contrasting flight altitudes and dive plans being used for each deployment.

2 Methodology

For this comparative study of the effectiveness of various imaging platforms for assessing megafauna abundances in polymetallic nodule ecosystems, 8 distinct image data sets DS_A to DS_H (see Table **??**) were collected. All data sets were acquired in a discrete area of seafloor of ca. 600 x 150 m^2. These 8 datasets were collected by 3 different towed camera platforms (one of which was deployed at several altitudes above seafloor) and an AUV (deployed at two different altitudes above seafloor) during 3 research cruises. One dataset (DS_C) was acquired during RV Sonne cruise SO106, the other seven during RV Sonne cruises

SO242/1 (DS_A, DS_B, DS_D) and SO242/2 (DS_E-DS_H) in 2015. DS_H was created by producing a mosaic of the seafloor from overlapping AUV imagery, then dividing the mosaic into smaller image tiles for fauna analysis. All image sets were analysed by 5 annotators a_1 - a_5, using a predesigned fauna catalogue to label a selected group of 20 fauna categories ω_a - ω_t within each discrete image (see Figure **??**). The term category refers to an arbitrary object type, across various taxonomic levels and also includes the category litter. The group of annotators selected the 20 categories by including fauna that is frequent enough for statistical interpretation. The 20 categories do not cover all objects visible in the images nor do they represent all the fauna known to occur in the area. The majority of categories represented morphotypes and could thus potentially include different cryptic species. Numbers of annotations per category and per data set vary. No organism size cut-off was defined for annotation, rather the image resolution determines which size of objects are still discernible. From this labelling effort, the densities of the various identified fauna categories in each data set were statistically compared.

2.1 Imaging platforms, resolutions and deployment altitudes

2.1.1 DS_A (4.49 MPix/m^2) and DS_B (3.89 MPix/m^2): Low altitude imagery from AWI OFOS camera sled

Towed still image and video sleds are equipment often used for gleaning some information on seafloor physical and megafauna community structure (examples can be found in Figures **??** a, b, d). These platforms consist of a solid frame which is connected to a survey vessel by an umbilical cable, in most cases capable of supplying power and data transfer between the ship and the platform. To operate, an altitude above the seafloor is set by the users, as a function of seafloor topographical structure, items of interest, vessel speed and weather conditions. A winch operator maintains the appropriate flight altitude above seafloor as the survey vessel tows the device over the requested course. These systems can utilise reasonably simple cable systems to allow live TV signals from the seafloor to reach a towing support vessel, or modern fibre-optic cables through which high data loads can be transmitted in real-time. The simplicity and relatively low costs of these towed systems, coupled with their moderate personnel requirements have made them an attractive choice to use in scientific expeditions, and particularly in time series studies where the same equipment is required for each revisit to a location. For this current study, the Alfred Wegener Helmholtz Institute for Polar and Marine Studies (AWI) Ocean Floor Observation System (OFOS) system was used for collection of several data sets (see Table **??**). Developed for time series analysis of the HAUSGARTEN marine time series station, the system has seen 15 years of regular use and numerous megafauna fauna papers have been published based on collected data (**?????**). The AWI OFOS consists of a solid frame containing vertically downward facing still image and video cameras (Figure **??**). Additionally, the system mounts LED lights to supply light for the video camera, as well as powerful flash units to allow 26 megapixel still images to be taken from an optimal altitude of 1.5 m above the seafloor. The AWI OFOS also incorporates 3 parallel lasers, to allow seafloor coverage (and fauna sizes) to be quantified in the images and video data collected. Figures **??** (a) and (b) show typical images collected from the DISCOL area from an operational altitude of 1.6 m (DS_A) and 1.7 m (DS_B).

2.1.2 DS_C (1.05 MPix/m^2): High altitude, digitised analogue imagery from EXPLOS camera sled

Prior to the equipping of research vessels with fibre-optic cables allowing HD video to be transmitted directly to the support vessel during a dive, it was common practice to set up a low quality video link to the seafloor to allow the operators of a towed device to maintain an appropriate flight altitude above the seafloor during a deployment. The scientific data collected were still images manually triggered from the ship but recorded onto analogue photographic film using a PHOTOSEA 5000 camera mounted on the "Exploration System" (EXPLOS) towed device. This required the mounting of actual film canisters on the towed platforms, resulting in deployments with less than 400 images collected (the capacity of standard, extended 35mm magazines of the era). In 1989, after the seafloor ploughing, such an analogue towed camera rig was used to image in the DISCOL area (Figure **??** (a)). The 1989 data set was recently digitized by the Mining Impact project of the Joint Programming Initiative Ocean (JPIO) and made available for this study. An example image is given in Figure **??** (c).

2.1.3 DS_D (0.98 MPix/m^2): High altitude imagery from AWI OFOS camera sled

With increasing distance from the seafloor, a particular optical system can image a greater area for a given set of optics, assuming correct focusing etc. can be achieved. With a doubling of distance however, effectiveness of illumination is reduced by 75%. For towed systems this may be compensated for by additional supply of power / a greater number of lights. For the current study however, the same AWI OFOS system introduced in section **??** was redeployed with the same standard lighting configuration at a flight altitude of 3.3 m. Figure **??** (d) shows a typical seafloor image taken from this altitude.

2.1.4 DS_E (0.24 MPix/m^2): Low altitude imagery from AUV Abyss

During SO242-1, GEOMAR's AUV Abyss (**?**) was deployed for several photographic mapping missions (see Figure **??** (c)). The vehicle's original camera had been replaced by a Canon-6D DSLR camera and the Xenon strobe by an LED flash system (**?**), placed 2 m apart from one another. The low altitude vertical imagery of DS_E was captured from a target altitude of 4.5 m, at a speed of 1.5 m/s and at a frame rate of 1 Hz. The system was equipped with a Canon 8-15mm fisheye lens (fixed to 15mm) centred in a dome port. Owing to weak illumination in the outer image regions, only the central 90° (across track) resp. 74° (along track) of the fisheye images were used and tri-linearly resampled to a picture that an ideal rectilinear 18mm lens would have taken. An example picture is shown in Figure **??** (a).

2.1.5 DS_F (0.16 MPix/m^2): Low altitude imagery from custom OFOS camera sled

During SO242-1 the area of interest was surveyed with a colour video camera (Oktopus GmbH) in conjunction with one Oktopus HID 50 light mounted vertically on a towed frame (see Figure **??** (b)). The signal was transmitted to a deck unit (Oktopus GmbH VDT 3) and recorded using an external video converter (Hauppauge - HD PVR), which converted the signal to .mp4 files and was then recorded in a PC using ArcSoft Total Media Extream software. For this study, frames were extracted from these video files at a rate of 0.1 Hz. The custom OFOS was put together in an 'ad-hoc' fashion, from a range of off the

shelf components, to mimic "pioneer" image-based methodology, rather than as a fully designed and integrated device. An example image is given in Figure **??** (b). Further details of the custom OFOS and its deployments can be found in (**?**).

2.1.6 DS_G (0.07 MPix/m^2): High altitude imagery from AUV Abyss

As a result of the fixed distance of roughly 2 m between camera and light source on AUV Abyss, images taken by the above system at higher altitudes were increasingly suffering from very strong backscatter, additionally to the loss of colour resulting from the large distance from the light source to the seafloor and back into the camera. Although the AUV imaged at altitudes above 10 m, those images were deemed of a quality unsuited for fauna analysis. Consequently, besides the 4.2 m "low altitude" AUV imagery in DS_G, AUV imagery acquired at 7.5 m altitude represents the dataset of maximum altitude in this contribution. Apart from the different altitude, all capture parameters in DS_G remained the same as in DS_E. An example image for this dataset is shown in Figure **??** (c).

2.1.7 DS_H (0.04 MPix/m^2): Low altitude imagery AUV Abyss and extracted from a photo mosaic

AUV images of station SO242-1_102 were collected at ca. 4.5 m above the seabed with 80% along track and 50% across track overlap in order to build one large photo mosaic out of the images. In order to mitigate water and illumination effects otherwise dominant in the final mosaic, a robust statistical estimate of the illumination component has been performed. For this, each image was robustly averaged with the 7 images taken before and after, producing an image without nodules that represents the illumination effects. The raw image was then - pixel-wise - divided by the illumination image and multiplied by the expected seafloor colour, which was obtained from box core photographs of the same cruise. For each track of a multi-track AUV mission, the images were registered against each other, leading to relative AUV localisation information with sub-cm accuracy. Afterwards, the photos were projected to the seafloor and rendered into a virtual ortho-photo with 5 mm/pixel resolution (reflecting the best resolution in the fisheye images) of roughly 7 hectares size. The photo mosaic was then subdivided into ca. 11,000 tiles and uploaded to BIIGLE for megafaunal assessment. An example tile is shown in Figure **??** (d). A similar mosaic of the same area was used in **?**.

2.2 Image annotation methodology

Within the study, 1340 seafloor images (or mosaic tiles) were analysed for megafauna abundance and community structure estimation (see Table **??**). All images used in the study were imported into the BIIGLE online annotation system (**?**). Once imported, 5 annotators inspected the images independently and annotated objects by placing a circle around each instance using the BIIGLE annotation interface (see Figure **??**). To assist in this, an identification guide with 20 categories was produced (see Figure **??**), from which the annotators could work.

2.3 Observer Agreement

Manual annotation was conducted independently. To compare results from the 5 annotators a_1 to a_5 inter-observer agreement was computed (**?**). First, the individual annotations of each pair of annotators were compared regarding the annotation location (i. e. the detection step) and annotation label (i.e. the classification step). Annotations of individual experts were then grouped to gold standard annotations to increase the robustness of the dataset comparison. A gold standard is the best-possible ground truth information if no actual ground truth is available **?**. Grouping of annotations was conducted by fusing annotations which overlap within one image and are of similar size to one grouped annotation. The position and radius of a grouped annotation represent the mean of the positions and radii of the single, overlapping annotations. The support of one annotation quantifies how many experts found this individual and thus ranks between 1 and 5. The label of the grouped annotation was selected as the most frequent label within the grouped annotations. Annotations that were supported by only one annotator were discarded. Also, if no two annotators assigned the same label to an annotation it was discarded. As a further measure of observer agreement, Cohen's kappa was computed (**?**).

2.4 Fauna-specific statistical analysis

The average abundance estimations of each individual fauna category computed for each of the 8 image sets was derived from the annotations made by each independent annotator. The 5 density estimates obtained for each fauna category, as generated from the labels made by the individual image annotators across the 8 imaging platform data sets were compared using nonparametric Kruskal-Wallis tests. These tests were conducted using the software package SPSS 17.0. Significant differences were considered when $p < 0.05$.

3 Results

3.1 Aggregated results for datasets

Aggregated results for various characteristics of the 8 datasets and annotations were computed by averaging across all fauna categories (see Figure **??** and Table **??**). All figures except **??** (g) further visualise the results of the grouped annotations. Most obvious is the increase of fauna density with imaging resolution (see **??** (a)). This trend is mirrored in the observation that the median size of the annotated fauna decreases with increasing resolution (see Figure **??** (b)). Together it can be reasoned, that the increased resolution allows to annotate smaller objects, increasing the total amount of individuals annotated. Nevertheless, it is also obvious that the increased resolution comes with an increase in observer disagreement. Figure **??** (c) shows that the standard deviation of fauna densities created by the 5 experts increases with increasing resolution. Figures **??** (d) - (f) highlight the tradeoff between resolution and seafloor inspection effort. In Figure **??** (d) it can be seen that the increase in resolution comes with a decrease in acquisition efficiency in terms of the area per hour (m^2 / h) that can be imaged. This negative correlation exists also when removing data set DS_G. Figure **??** (f) shows that, although higher densities of fauna are detected for high resolution datasets, it still requires to manually inspect more megapixels per annotation compared to lower resolution

datasets. The annotation effort for such high resolution data sets is thus over-proportionally large. Removing single points that appear as outliers in the different data dimensions (Figures **??** (a) - (l)) does not change the general trends of the correlation lines.

3.2 Observer agreement

Figure **??** (g) outlines the importance in image-based studies of incorporating annotations created by more than one annotator. It shows the generally poor observer agreement in this study when considering the single expert annotations (see also Table **??**). It further highlights that the observer agreement drops with increasing image resolution, echoing the results in Figure **??** (c). When grouping the single observer annotation to form the gold standard annotations, the observer agreement increases significantly (see Figure **??** (h)). This increase is similarly reflected by the Cohen's kappa values; all but one above 0.7, which is deemed as "substantial agreement" (0.6-0.8).

3.3 Fauna-specific statistical Analysis

The seafloor densities of the 20 categories of fauna and seafloor features, as quantified by the 5 independent annotators are given in Figure **??** (mobile fauna) and Figure **??** (sessile fauna). Kruskal-Wallis tests indicated that for all fauna categories (with the exception of 'molluscs') observed, individual densities differed by imaging platform at the 95% threshold ('Small Encrusting', 'Starfish') or <99% threshold (all other fauna categories). For sessile fauna, average individual densities observed were highest across fauna categories in DS_A. Generally, the averaged densities for this dataset acquired at 1.6 m altitude were roughly double to triple those observed in DS_B which was collected in the same year from a slightly higher median altitude of 1.7 m. Densities of sessile fauna derived from AUV data were generally lower than those derived from OFOS data. Sessile fauna densities derived from AUV data acquired at 4.2 m altitude (DS_E) were invariably higher than those derived from 7.5 m AUV data (DS_G). Sessile fauna densities determined from the mosaicked images were roughly equivalent or a little lower than the densities determined from both uncombined AUV data sets (see Figure **??**). For mobile fauna, trends in densities of fauna categories were less dependent on the observing platform. Even though differences were indicated as significant for many fauna categories (see Table **??**), these differences were not clearly relatable to either imaging platform deployment altitude or methodology and observers (see Figure **??**).

4 Discussion

4.1 Spatial and Temporal factors

The current study attempts to estimate the effectivity of a range of imaging devices across an overlapping area of seafloor based on experts' manual annotations. Given the inaccuracies of about 1% achievable with the POSIDONIA underwater positioning system used for the majority of imaging deployments (**?**) and the lack of distinct seafloor features in the DISCOL polymetallic nodule province, sampling exactly the same areas of seafloor was not possible. Nevertheless, due to the reason-

ably homogenous nature of the seafloor (from meters to hundreds of meters scale) in the survey region, it seems likely that comparable organisms were present across areas. Temporal differences in community structure, particularly between years, cannot be wholly discounted as explanatory factors of differences between data sets **??**. Highly mobile fauna, such as fish and jellyfish, can vary in local abundances on temporal scales of minutes, and even the less mobile ophiuroids and holothurians

365 can respond relatively swiftly to changes in seafloor conditions, such as a food fall or hydrodynamic conditions. Even so, we assume that temporal and spatial differences between the collected data are of minor significance in explaining the differences in densities observed.

4.2   Deployment altitude and image resolution

Even though it was not possible to deploy all platforms at different altitudes within the same cruise, it was feasible to collect

370 material altogether from both the AUV (two altitudes) and the AWI OFOS (three altitudes). For virtually all fauna categories used, the highest density estimates were made from data collected at the lowest deployment altitude and highest pixel resolution. At these altitudes, less water is present between the camera and the target, reducing distortion and light attenuation effects. The only exceptions to this trend were the highly mobile, water column dwelling fauna, such as jellyfish and fish. Given the three dimensionality of the habitat utilised by these organisms, observation from a greater altitude is beneficial, and it is

375 thus more likely to image such fauna. This is potentially coupled with avoidance mechanisms triggered by the lights on the imaging platform or the sound of thrusters (in the case of the AUV deployments). The way in which fauna density estimations are subject to the deployment altitude does not appear to be linear or comparable across fauna categories. Larger fauna, such as 'stalked sponges' (see Figure **??** (d)) and 'starfish' (see Figure **??** (j)) were spotted with equivalent ease across all data sets, whereas smaller fauna, such as 'sessile polychaetes' and 'sponges' (see Figures **??** (b) and (i)) were annotated more frequently

380 in data collected from lower altitudes. These altitude-based trends in density estimation were observed in both AUV and OFOS data sets. Interestingly, an average deployment altitude difference of just 10 cm, from 1.7 to 1.6 m average altitude between SO242-2 OFOS deployments corresponded to a much greater difference in fauna density estimations than the 1.6 m difference in deployment altitudes between the 3.3 m and 1.7 m data sets. Both, the attenuation of light in water and the variable impact of this reduction on the wavelengths of reflected light, as well as the size of the fauna image received by the camera likely play a

385 role in determining the fauna abundance accuracy achievable from a data set. This extreme subjectivity to deployment altitude of derived density estimations is an important consideration when comparing results from different deployments.

4.3   Annotator skill / Observer effect

To label fauna to species level from imagery requires a certain amount of skill and awareness of the fauna likely to occur in a particular survey region. In many cases, annotation categories will only refer to morphotypes. This is due to the fact

390 that most fauna in the areas is either still unknown or impossible to identify from images alone. To properly assess fauna occurring in a habitat, especially when addressing human impacts, requires not only ecological expertise but also support from taxonomists. Nevertheless, even when specialists are analysing the same data set, inter-observer differences in annotations can be significant (**??**). Here however, differences between platform altitude proved to be more significant than observer effect for

all faunal categories. Therefore, given the sparsity of many deep-sea taxa in nodule provinces (**?**), the use of key species is
of more applicability when determining monitoring strategies for impact assessment, where statistically significant differences
in abundances may reflect differences in populations of pre-impacted or control areas and those within impacted areas. These
key fauna are likely to differ between different locations and ecosystems. For deep sea manganese nodule provinces, the level
of understanding of ecosystem functioning is probably insufficient to select species and/or taxa of major importance for the
ecosystem. Certainly, some easily annotated fauna play important roles as habitat engineer species, such as the stalked fauna,
which add the vertical axis to increase habitat niche availability (**??**). Biogeochemical processes within and at the sediment
/ seawater interface may well be influenced by mega-, macro- and meiofauna not visible even in high-resolution imagery.
Some large fauna spend variable amounts of time within the sediments, and smaller fauna may be below the resolution limit
of the imagery. Though densities of these less-visible organism categories may be measured with a range of methodologies
(**?**), the number of samples required, coupled with the remoteness of resource sites, renders these as probably inappropriate
for cost effective monitoring. By providing a clear identification catalogue, ideally with a limited number of categories (as
used in the current study) annotators with little or no experience will be able to identify fauna within an image set with an
ample degree of confidence. For complex studies of detailed community change, trained scientific personnel would be required
in order to have more accurate annotations. In either case, manual annotations need to be quality controlled, e.g. by creating
a gold standard, to produce more reliable data. Moreover, employing several experts for the image annotation would add a
considerable financial cost to any monitoring program. In the future, it is probable that the ongoing developments of computer
algorithms for resource quantification (**??**) and fauna identification (**?????**) will allow a near real-time assessment of fauna
abundances in a surveyed region for a given platform and deployment strategy. At present, however, as commercial nodule
mining approaches viability, traditional monitoring approaches like manual image annotation or physical sampling are the
only ones available for integration into regulatory frameworks and work plans. Nevertheless, expected technological advances
should be incorporated into the regulations.

5   Conclusions

The results from the current study highlight how tightly fauna abundance estimations in manganese nodule ecosystems may
be related to the investigative methodology used. Small differences in imaging platform operational altitude, illumination and
lens type analysed by a particular annotator can alter estimations of community structure. The results obtained by this study
are similar to other studies conducted in shallow reef environments (**?**), though they are highly prescient given the commercial
interest in these nodule resources and the current lack in background knowledge to estimate the impact of mining activities on
ecosystem function. For the first time, quantitative information was provided on the effect of using different platform altitudes
and the resulting imagery resolution. The authors of the current study do not intend to recommend a 'perfect' imaging platform
for megafauna abundance monitoring in manganese nodule ecosystems, as more work is still needed to determine whether
there are megafauna species that are of particular significance in maintaining current community structures and biodiversity in
the nodule regions, and because the commercial viability of the various platforms available for study will surely change during

the forthcoming years. With this study, we intend to give some general guidelines on how longterm monitoring studies in these regions should be planned to allow the collection of good-quality data which can be further used in time series analyses of larger fauna community composition:

430   1. For a given study location, a comparable survey deployment plan should be used at each time step of analysis: same sensor payload, instrument platform altitude, deployment speed, seafloor area imaged, sample unit size.

2. A well documented camera system should be used: aperture, sensitivity, lens arrangement, mounting angle.

3. Illumination should be maintained across deployments: intensity, wavelength, mounting angle.

4. Annotations by several observers need to be collected and thoroughly merged to create robust data for interpretation

435   5. The lowest feasible altitude above seabed using a given platform will always provide higher-resolution data and higher taxonomical resolution in the faunal identification.

Although many of these points may seem obvious requirements for a monitoring campaign or study, the extended duration of deep-sea surveys may lead to technological changes taking place between survey visits, or changes in personnel involved in conducting the work. Even during a recent 3 year study conducted within medium depths offshore Norway, the majority of
440   these points were missed (**?**). We highly recommend that in the developing industry of polymetallic nodule extraction, such guidelines be integrated into licensing agreements, with appropriate commitments made by companies to ensure longterm adherence (commitments such as maintaining appropriate equipment for the duration of the monitoring campaign, providing accurate blueprints/design specification of platforms used at each monitoring stage etc). We also recommend an increase in vigour of studies focusing on the biogeochemical processes at work in these remote ecosystems. Hence, relevance of any
445   observation on the short or the long term regarding changes in fauna density or communities associated with the exploitation of these resources and their possible impacts can be evaluated with greater confidence.

Author contributions.   Designed the study: JG, ESL, KK, TS, JNGP; Provided data: TS, AP, ESL, JNGP, KK, JG; Provided infrastructure: DL, TN, MZ, DOBJ, Annotated the images: AP, IS, JT, DC, LL; Analysed the data: TS, AP, DL, JT, DC, LL, ESL, JNGP, YM, MZ; Wrote the manuscript: TS, AP, DC, ESL, KK

450   *Competing interests.*   The authors report no competing interests.

Disclaimer.   TEXT

Acknowledgements. We thank the crew and scientific parties of cruises SO106, SO242/1 and SO242/2 for their indispensable support in making this study possible as well as the participants of the open discussion reviewing process of this manuscript.

References

455 Aguzzi, J., Costa, C., Fujiwara, Y., Iwase, R., Ramirez-Llorda, E., and Menesatti, P.: A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna, Sensors, 9, 8438–8455, 2009.

Amante, C. and Eakins, B. W.: ETOPO1 arc-minute global relief model: procedures, data sources and analysis, 2009.

Ayma, A., Aguzzi, J., Canals, M., Lastras, G., Bahamon, N., Mechó, A., and Company, J.: Comparison between ROV video and Agassiz trawl methods for sampling deep water fauna of submarine canyons in the Northwestern Mediterranean Sea with observations on behavioural

460 reactions of target species, Deep Sea Research Part I: Oceanographic Research Papers, 114, 149–159, 2016.

Beaulieu, S.: Life on glass houses: sponge stalk communities in the deep sea, Marine Biology, 138, 803–817, 2001.

Bergmann, M., Soltwedel, T., and Klages, M.: The interannual variability of megafaunal assemblages in the Arctic deep sea: Preliminary results from the HAUSGARTEN observatory (79 N), Deep Sea Research Part I: Oceanographic Research Papers, 58, 711–723, 2011.

Bluhm, H.: Re-establishment of an abyssal megabenthic community after experimental physical disturbance of the seafloor, Deep Sea Re-

465 search Part II: Topical Studies in Oceanography, 48, 3841–3868, 2001.

Boetius, A.: RV Sonne Fahrtbericht/Cruise Report SO242-2: JPI OCEANS Ecological Aspects of Deep-Sea Mining, DISCOL Revisited, Guayaquil - Guayaquil (Equador), 28.08.-01.10. 2015, Kiel: Helmholtz-Zentrum fü r Ozeanforschung, 2015.

Borowski, C.: Physically disturbed deep-sea macrofauna in the Peru Basin, southeast Pacific, revisited 7 years after the experimental impact, Deep Sea Research Part II: Topical Studies in Oceanography, 48, 3809–3839, 2001.

470 Burns, R. and Burns, V. M.: The mineralogy and crystal chemistry of deep-sea manganese nodules, a polymetallic resource of the twenty-first century, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 286, 283–301, 1977.

Cairns, S. D.: New abyssal Primnoidae (Anthozoa: Octocorallia) from the Clarion-Clipperton Fracture Zone, equatorial northeastern Pacific, Marine Biodiversity, 46, 141–150, 2016.

Cauwerts, C., PhD, M. B., and Deneyer, A.: Comparison of the vignetting effects of two identical fisheye lenses, Leukos, 8, 181–203, 2012.

475 De Smet, B., Pape, E., Riehl, T., Bonifácio, P., Colson, L., and Vanreusel, A.: The community structure of deep-sea macrofauna associated with polymetallic nodules in the eastern part of the Clarion-Clipperton Fracture Zone, Frontiers in Marine Science, 4, 103, 2017.

Drazen, J. C., Leitner, A. B., Morningstar, S., Marcon, Y., Greinert, J., and Purser, A.: Observations of deep-sea fishes and mobile scavengers from the abyssal DISCOL experimental mining area, Biogeosciences, 16, 3133–3146, 2019.

Durden, J. M., Schoening, T., Althaus, F., Friedman, A., Garcia, R., Glover, A. G., Greinert, J., Stout, N. J., Jones, D. O., Jordt, A., et al.:

480 Perspectives in visual imaging for marine biology and ecology: from acquisition to understanding, in: Oceanography and Marine Biology, pp. 9–80, CRC Press, 2016.

Gage, J. and Bett, B.: Deep-sea benthic sampling, Methods for the Study of the Marine Benthos. Oxford, UK: Blackwell, pp. 273–325, 2005.

Gardner, J. and Struthers, C.: Comparisons among survey methodologies to test for abundance and size of a highly targeted fish species, Journal of Fish Biology, 82, 242–262, 2013.

485 Gollner, S., Kaiser, S., Menzel, L., Jones, D. O., Brown, A., Mestre, N. C., Van Oevelen, D., Menot, L., Colaço, A., Canals, M., et al.: Resilience of benthic deep-sea fauna to mining activities, Marine Environmental Research, 129, 76–101, 2017.

Gooday, A. J., Holzmann, M., Caulle, C., Goineau, A., Kamenskaya, O., Weber, A. A.-T., and Pawlowski, J.: Giant protists (xenophyophores, Foraminifera) are exceptionally diverse in parts of the abyssal eastern Pacific licensed for polymetallic nodule exploration, Biological conservation, 207, 106–116, 2017.

490   Greinert, J.: RV SONNE Fahrtbericht/cruise report SO242-1 [SO242/1]: JPI OCEANS ecological aspects of deep-sea mining, DISCOL revisited, Guayaquil - Guayaquil (Equador), 28.07.-25.08. 2015, 2015.

Howell, K.-L., Piechaud, N., Downie, A.-L., and Kenny, A.: The distribution of deep-sea sponge aggregations in the North Atlantic and implications for their effective spatial management, Deep Sea Research Part I: Oceanographic Research Papers, 115, 309–320, 2016.

Huvenne, V., Bett, B., Masson, D., Le Bas, T., and Wheeler, A. J.: Effectiveness of a deep-sea cold-water coral Marine Protected Area,
495   following eight years of fisheries closure, Biological Conservation, 200, 60–69, 2016.

Jaffe, J. S.: Underwater optical imaging: the past, the present, and the prospects, IEEE Journal of Oceanic Engineering, 40, 683–700, 2014.

Jones, D. O., Kaiser, S., Sweetman, A. K., Smith, C. R., Menot, L., Vink, A., Trueblood, D., Greinert, J., Billett, D. S., Arbizu, P. M., et al.: Biological responses to disturbance from simulated deep-sea polymetallic nodule mining, PLoS One, 12, e0171 750, 2017.

Kersken, D., Kocot, K., Janussen, D., Schell, T., Pfenninger, M., and Arbizu, P. M.: First insights into the phylogeny of deep-sea glass sponges
500   (Hexactinellida) from polymetallic nodule fields in the Clarion-Clipperton Fracture Zone (CCFZ), northeastern Pacific, Hydrobiologia, 811, 283–293, 2018.

Kwasnitschka, T., Köser, K., Sticklus, J., Rothenbeck, M., Weiß, T., Wenzlaff, E., Schoening, T., Triebe, L., Steinführer, A., Devey, C., et al.: DeepSurveyCam - A deep ocean optical mapping system, Sensors, 16, 164, 2016.

Lam, K., Shin, P. K., Bradbeer, R., Randall, D., Ku, K. K., Hodgson, P., and Cheung, S. G.: A comparison of video and point intercept transect
505   methods for monitoring subtropical coral communities, Journal of Experimental Marine Biology and Ecology, 333, 115–128, 2006.

Langenkämper, D., Zurowietz, M., Schoening, T., and Nattkemper, T. W.: Biigle 2.0-browsing and annotating large marine image collections, Frontiers in Marine Science, 4, 83, 2017.

Lim, S.-C., Wiklund, H., Glover, A. G., Dahlgren, T. G., and Tan, K.-S.: A new genus and species of abyssal sponge commonly encrusting polymetallic nodules in the Clarion-Clipperton Zone, East Pacific Ocean, Systematics and biodiversity, 15, 507–519, 2017.

510   Linke, P. and Lackschewitz, K.: Autonomous Underwater Vehicle ABYSS, Journal of large-scale research facilities JLSRF, 2, 79, 2016.

Lodge, M., Johnson, D., Le Gurun, G., Wengler, M., Weaver, P., and Gunn, V.: Seabed mining: International Seabed Authority environmental management plan for the Clarion–Clipperton Zone. A partnership approach, Marine Policy, 49, 66–72, 2014.

Marcon, Y. and Purser, A.: PAPARA (ZZ) I: An open-source software interface for annotating photographs of the deep-sea, SoftwareX, 6, 69–80, 2017.

515   McHugh, M. L.: Interrater reliability: the kappa statistic, Biochemia medica: Biochemia medica, 22, 276–282, 2012.

McPhail, S., Furlong, M., and Pebody, M.: Low-altitude terrain following and collision avoidance in a flight-class autonomous underwater vehicle, Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, 224, 279–292, 2010.

Mewes, K., Mogollón, J. M., Picard, A., Rühlemann, C., Kuhn, T., Nöthen, K., and Kasten, S.: Impact of depositional and biogeochemical
520   processes on small scale variations in nodule abundance in the Clarion-Clipperton Fracture Zone, Deep Sea Research Part I: Oceanographic Research Papers, 91, 125–141, 2014.

Morris, K. J., Bett, B. J., Durden, J. M., Huvenne, V. A., Milligan, R., Jones, D. O., McPhail, S., Robert, K., Bailey, D. M., and Ruhl, H. A.: A new method for ecological surveying of the abyss using autonomous underwater vehicle photography, Limnology and Oceanography: Methods, 12, 795–809, 2014.

525   Mullineaux, L. S.: Organisms living on manganese nodules and crusts: distribution and abundance at three North Pacific sites, Deep Sea Research Part A. Oceanographic Research Papers, 34, 165–184, 1987.

Mullineaux, L. S.: Vertical distributions of the epifauna on manganese nodules: implications for settlement and feeding, Limnology and Oceanography, 34, 1247–1262, 1989.

Murphy, H. M. and Jenkins, G. P.: Observational methods used in marine spatial monitoring of fishes and associated habitats: a review, Marine and Freshwater Research, 61, 236–252, 2010.

Norse, E. A., Brooke, S., Cheung, W. W., Clark, M. R., Ekeland, I., Froese, R., Gjerde, K. M., Haedrich, R. L., Heppell, S. S., Morato, T., et al.: Sustainability of deep-sea fisheries, Marine policy, 36, 307–320, 2012.

Orejas, C., Gori, A., Iacono, C. L., Puig, P., Gili, J.-M., and Dale, M. R.: Cold-water corals in the Cap de Creus canyon, northwestern Mediterranean: spatial distribution, density and anthropogenic impact, Marine Ecology Progress Series, 397, 37–51, 2009.

Paul, S. A., Gaye, B., Haeckel, M., Kasten, S., and Koschinsky, A.: Biogeochemical regeneration of a nodule mining disturbance site: trace metals, DOC and amino acids in deep-sea sediments and pore waters, Frontiers in Marine Science, 5, 117, 2018.

Petersen, S., Hannington, M., and Krätschell, A.: Technology developments in the exploration and evaluation of deep-sea mineral resources, in: Annales des Mines-Responsabilite et environnement, 1, pp. 14–18, FFE, 2017.

Peukert, A., Petersen, S., Greinert, J., and Charlot, F.: Seabed mining, in: Submarine Geomorphology, pp. 481–502, Springer, 2018a.

Peukert, A., Schoening, T., Alevizos, E., Köser, K., Kwasnitschka, T., and Greinert, J.: Understanding Mn-nodule distribution and evaluation of related deep-sea mining impacts using AUV-based hydroacoustic and optical data, Biogeosciences (BG), 15, 2525–2549, 2018b.

Peyronnet, J.-P., Person, R., and Rybicki, F.: Posidonia 6000: a new long range highly accurate ultra short base line positioning system, in: IEEE Oceanic Engineering Society. OCEANS'98. Conference Proceedings (Cat. No. 98CH36259), vol. 3, pp. 1721–1727, IEEE, 1998.

Pham, C. K., Ramirez-Llodra, E., Alt, C. H., Amaro, T., Bergmann, M., Canals, M., Davies, J., Duineveld, G., Galgani, F., Howell, K. L., et al.: Marine litter distribution and density in European seas, from the shelves to deep basins, PloS one, 9, e95 839, 2014.

Purser, A.: A time series study of Lophelia pertusa and reef megafauna responses to drill cuttings exposure on the Norwegian margin, PloS one, 10, e0134 076, 2015.

Purser, A., Bergmann, M., Lundälv, T., Ontrup, J., and Nattkemper, T. W.: Use of machine-learning algorithms for the automated detection of cold-water coral habitats: a pilot study, Marine Ecology Progress Series, 397, 241–251, 2009.

Purser, A., Marcon, Y., Hoving, H.-J. T., Vecchione, M., Piatkowski, U., Eason, D., Bluhm, H., and Boetius, A.: Association of deep-sea incirrate octopods with manganese crusts and nodule fields in the Pacific Ocean, Current Biology, 26, R1268–R1269, 2016.

Purser, A., Marcon, Y., Dreutter, S., Hoge, U., Sablotny, B., Hehemann, L., Lemburg, J., Dorschel, B., Biebow, H., and Boetius, A.: Ocean Floor Observation and Bathymetry System (OFOBS): a new towed camera/sonar system for deep-sea habitat surveys, IEEE Journal of Oceanic Engineering, 44, 87–99, 2018.

Roark, E. B., Guilderson, T. P., Dunbar, R. B., Fallon, S. J., and Mucciarone, D. A.: Extreme longevity in proteinaceous deep-sea corals, Proceedings of the National Academy of Sciences, 106, 5204–5208, 2009.

Schoening, T., Bergmann, M., Ontrup, J., Taylor, J., Dannheim, J., Gutt, J., Purser, A., and Nattkemper, T. W.: Semi-automated image analysis for the assessment of megafaunal densities at the Arctic deep-sea observatory HAUSGARTEN, PloS one, 7, e38 179, 2012.

Schoening, T., Kuhn, T., Jones, D. O., Simon-Lledo, E., and Nattkemper, T. W.: Fully automated image segmentation for benthic resource assessment of poly-metallic nodules, Methods in Oceanography, 15, 78–89, 2016a.

Schoening, T., Osterloff, J., and Nattkemper, T. W.: RecoMIA - Recommendations for marine image annotation: Lessons learned and future directions, Frontiers in Marine Science, 3, 59, 2016b.

Schoening, T., Jones, D. O., and Greinert, J.: Compact-morphology-based poly-metallic nodule delineation, Scientific reports, 7, 13 338, 2017.

565 Schoening, T., Köser, K., and Greinert, J.: An acquisition, curation and management workflow for sustainable, terabyte-scale marine image analysis, Scientific data, 5, 180 181, 2018.

Schriever, G., Thiel, H., Beckmann, W., Bluhm, H., Boetius, A., Borowski, C., Bussau, C., Foell, E., Hoffmann, G., Maybury, C., et al.: Cruise report DISCOL 3, SONNE cruise 77: Cruise report DISCOL 3, Sonne cruise 77: January 26-February 27, 1992, Balboa/Panama-Balboa/Panama, 1992.

570 Sharma, R.: Deep-Sea Mining, Springer, 2017.

Sharma, R.: Environmental Issues of Deep-Sea Mining: Impacts, Consequences and Policy Perspectives, Springer, 2019.

Siddiqui, S. A., Salman, A., Malik, M. I., Shafait, F., Mian, A., Shortis, M. R., and Harvey, E. S.: Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data, ICES Journal of Marine Science, 75, 374–389, 2017.

575 Simon-Lledó, E., Bett, B. J., Huvenne, V. A., Köser, K., Schoening, T., Greinert, J., and Jones, D. O.: Biological effects 26 years after simulated deep-sea mining, Scientific reports, 9, 8040, 2019a.

Simon-Lledó, E., Bett, B. J., Huvenne, V. A., Schoening, T., Benoist, N. M., Jeffreys, R. M., Durden, J. M., and Jones, D. O.: Megafaunal variation in the abyssal landscape of the Clarion Clipperton Zone, Progress in oceanography, 170, 119–133, 2019b.

Simon-Lledó, E., Bett, B. J., Huvenne, V. A., Schoening, T., Benoist, N. M., and Jones, D. O.: Ecology of a polymetallic nodule occurrence
580 gradient: Implications for deep-sea mining, Limnology and Oceanography, 2019c.

Taylor, J., Krumpen, T., Soltwedel, T., Gutt, J., and Bergmann, M.: Regional-and local-scale variations in benthic megafaunal composition at the Arctic deep-sea observatory HAUSGARTEN, Deep Sea Research Part I: Oceanographic Research Papers, 108, 58–72, 2016.

Taylor, J., Krumpen, T., Soltwedel, T., Gutt, J., and Bergmann, M.: Dynamic benthic megafaunal communities: Assessing temporal variations in structure, composition and diversity at the Arctic deep-sea observatory HAUSGARTEN between 2004 and 2015, Deep Sea Research
585 Part I: Oceanographic Research Papers, 122, 81–94, 2017.

Thiel, H.: Evaluation of the environmental consequences of polymetallic nodule mining based on the results of the TUSCH Research Association, Deep Sea Research Part II: Topical Studies in Oceanography, 48, 3433–3452, 2001.

Tilot, V., Ormond, R., Moreno Navas, J., and Catalá, T. S.: The benthic megafaunal assemblages of the CCZ (Eastern Pacific) and an approach to their management in the face of threatened anthropogenic impacts, Frontiers in Marine Science, 5, 7, 2018.

590 Vanreusel, A., Hilario, A., Ribeiro, P. A., Menot, L., and Arbizu, P. M.: Threatened by mining, polymetallic nodules are required to preserve abyssal epifauna, Scientific reports, 6, 26 808, 2016.

Volkmann, S. E. and Lehnen, F.: Production key figures for planning the mining of manganese nodules, Marine Georesources & Geotechnology, 36, 360–375, 2018.

Watling, H.: Review of biohydrometallurgical metals extraction from polymetallic mineral resources, Minerals, 5, 1–60, 2015.

595 Wilson, S., Graham, N., and Polunin, N.: Appraisal of visual assessments of habitat complexity and benthic composition on coral reefs, Marine Biology, 151, 1069–1076, 2007.

Yamazaki, T. and Brockett, F. H.: History of Deep-Ocean Mining, Encyclopedia of Maritime and Offshore Engineering, pp. 1–9, 2017.

Yu, C., Xiang, X., Wilson, P., and Zhang, Q.: Guidance-error-based robust fuzzy adaptive control for bottom following of a flight-style AUV with delayed and saturated control surfaces, IEEE Transactions on Cybernetics, 2018.

600 Zurowietz, M., Langenkämper, D., Hosking, B., Ruhl, H. A., and Nattkemper, T. W.: MAIA - A machine learning assisted image annotation method for environmental monitoring and exploration, PloS one, 13, 2018.
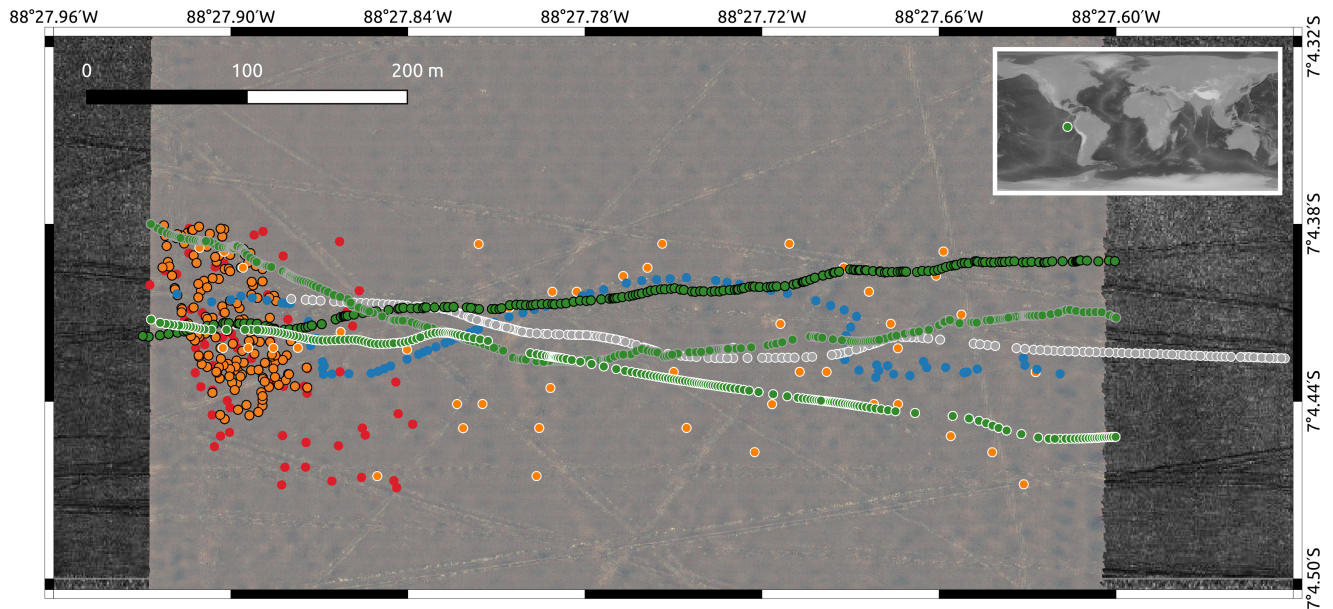
Figure 1. Overview map of imaging locations of the 8 different datasets. DS_A (green dots, grey border), DS_B (green dots, black border), DS_C (blue dots), DS_D (green dots, white border), DS_E (orange dots, black border), DS_F (grey dots), DS_G (orange dots, white border), DS_H (red dots). The world map in the top right corner shows the geographical location of the DISCOL area in the Eastern South Pacific (green dot, ©NOAA (**?**)). The study area covers ca. 600x150 m^2. The background map shows another photo mosaic, created from the full image set of which DS_G is a subset. Criss-crossing lines are plough tracks by the mining simulation in 1989.

Table 1. Summary of image data collected for each dataset considered in this study. Columns marked by (*) represent median values across the dataset.

Dataset	Station	Date [dd/mm/yyyy]	Platform	Resolution* [MPix / m^2]	Altitude* [m]	Footprint* [m^2 / image]	Number of images
DS_A	SO242-2_171	25/09/2015	AWI OFOS	4.49	1.6	4.9	311
DS_B	SO242-2_155	25/09/2015	AWI OFOS	3.89	1.7	5.7	206
DS_C	SO106_OFOS35	1997	EXPLOS OFOS	1.05	3.4	12.5	80
DS_D	SO242-2_233	25/09/2015	AWI OFOS	0.98	3.2	22.5	209
DS_E	SO242-1_107	17/08/2015	AUV Abyss	0.24	4.2	52.9	154
DS_F	SO242-1_111	18/08/2015	Custom OFOS	0.16	2.0	2.6	272
DS_G	SO242-1_083	13/08/2015	AUV Abyss	0.07	7.5	169.1	46
DS_H	SO242-1_102 (Mosaic)	16/08/2015	AUV Abyss	0.04	4.5	32.8	62
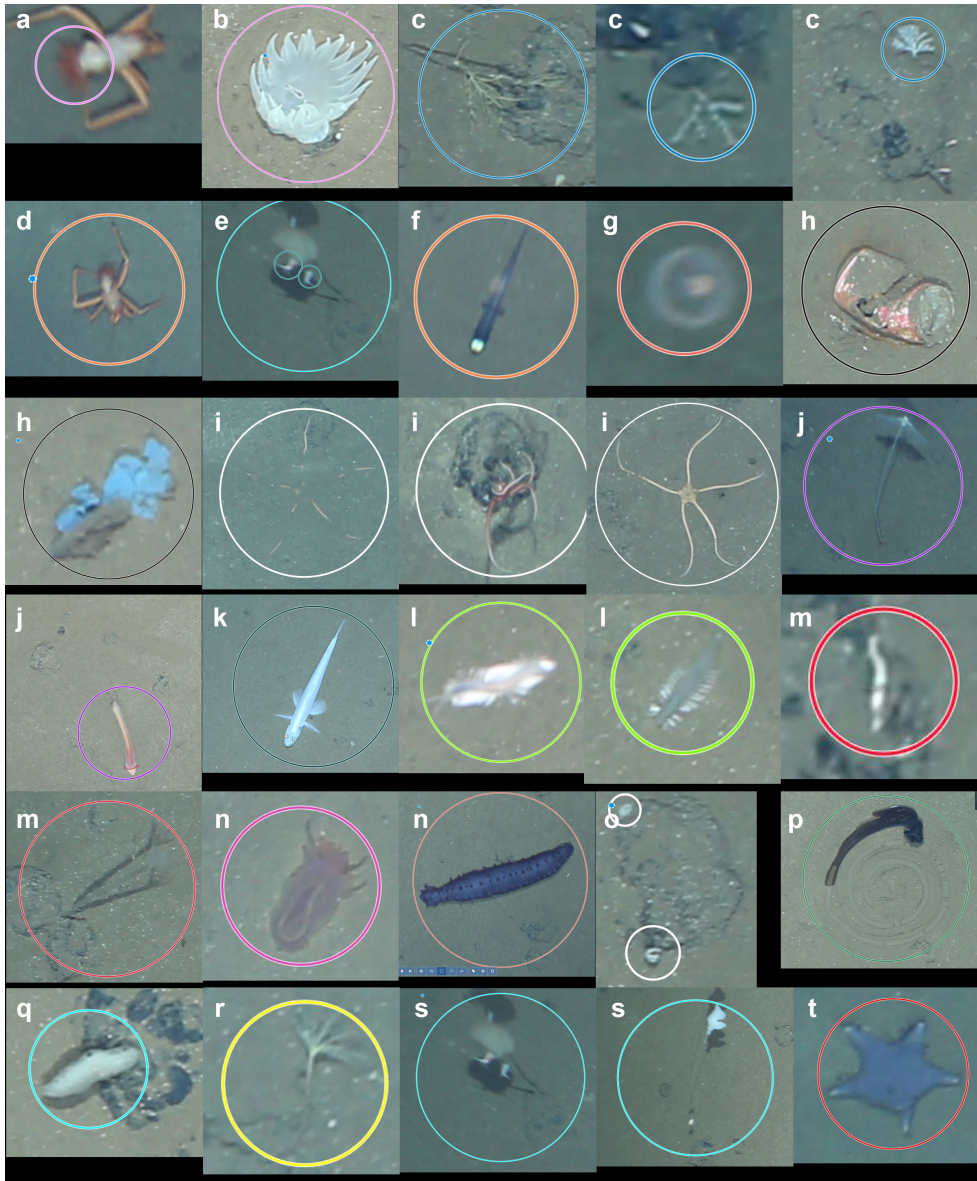
Figure 2. Fauna categories used in the current study for the DISCOL area. Circles correspond to annotations in BIIGLE. Colours of annotations visualise the category type. a,b: anemones; c: corals; d: crustacea; e: epifauna; f: Ipnops fish; g: jellyfish; h: litter; i: ophiuroidea; j: cladorhizidae; l,p: enteropneusta; k: fish; l: polychaeta worms; m: tubeworms polychaeta; n: holothuroidea; o: small encrusting; q: porifera; r: stalked crinoid; s: stalked porifera; t: asteroidea. All examples have been scaled for visualisation purpose, some like l and m are small and close to the resolution limit.
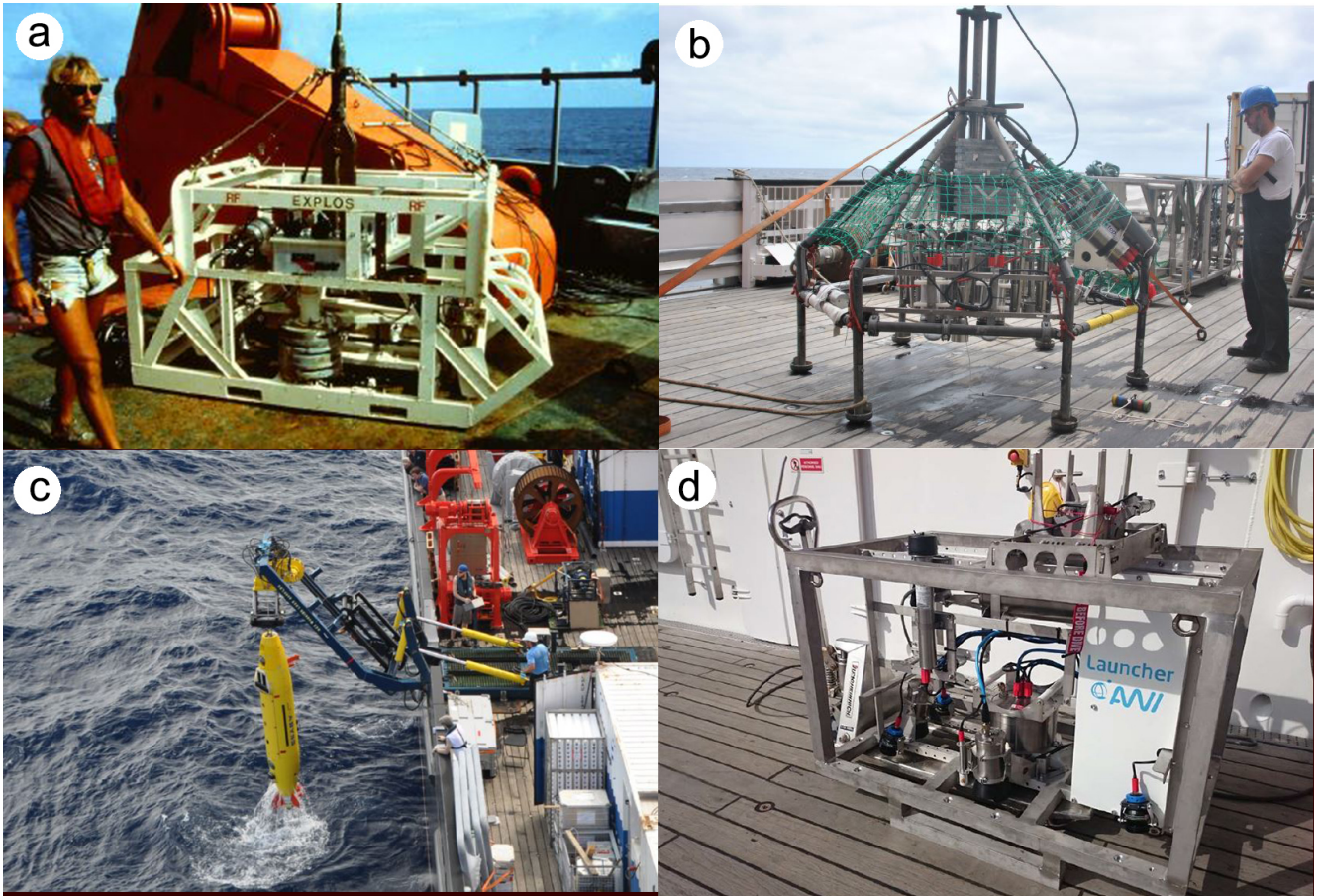
Figure 3. Imaging platforms used in the current study. a) The EXPLOS OFOS analog camera sled from 1997 (**?**) b) A custom OFOS used during SO241/1 c) GEOMAR AUV Abyss d) AWI OFOS

Table 2. Annotation results for the 8 different datasets considered in this study.

Dataset	No. annotations (grouped)	No. Categories found	Observer Agreement (single annotators)	Observer Agreement (grouped)	Cohens' Kappa (grouped)	Fauna Density [ind. / m^2]
DS$_A$	741	22	0.06	0.65	0.75	0.0194
DS$_B$	264	22	0.11	0.66	0.76	0.0092
DS$_C$	78	18	0.12	0.71	0.82	0.0085
DS$_D$	1077	22	0.14	0.66	0.81	0.0065
DS$_E$	231	22	0.20	0.69	0.82	0.0009
DS$_F$	70	15	0.24	0.40	0.49	0.0029
DS$_G$	61	13	0.16	0.65	0.74	0.0007
DS$_H$	202	22	0.23	0.66	0.77	0.0030
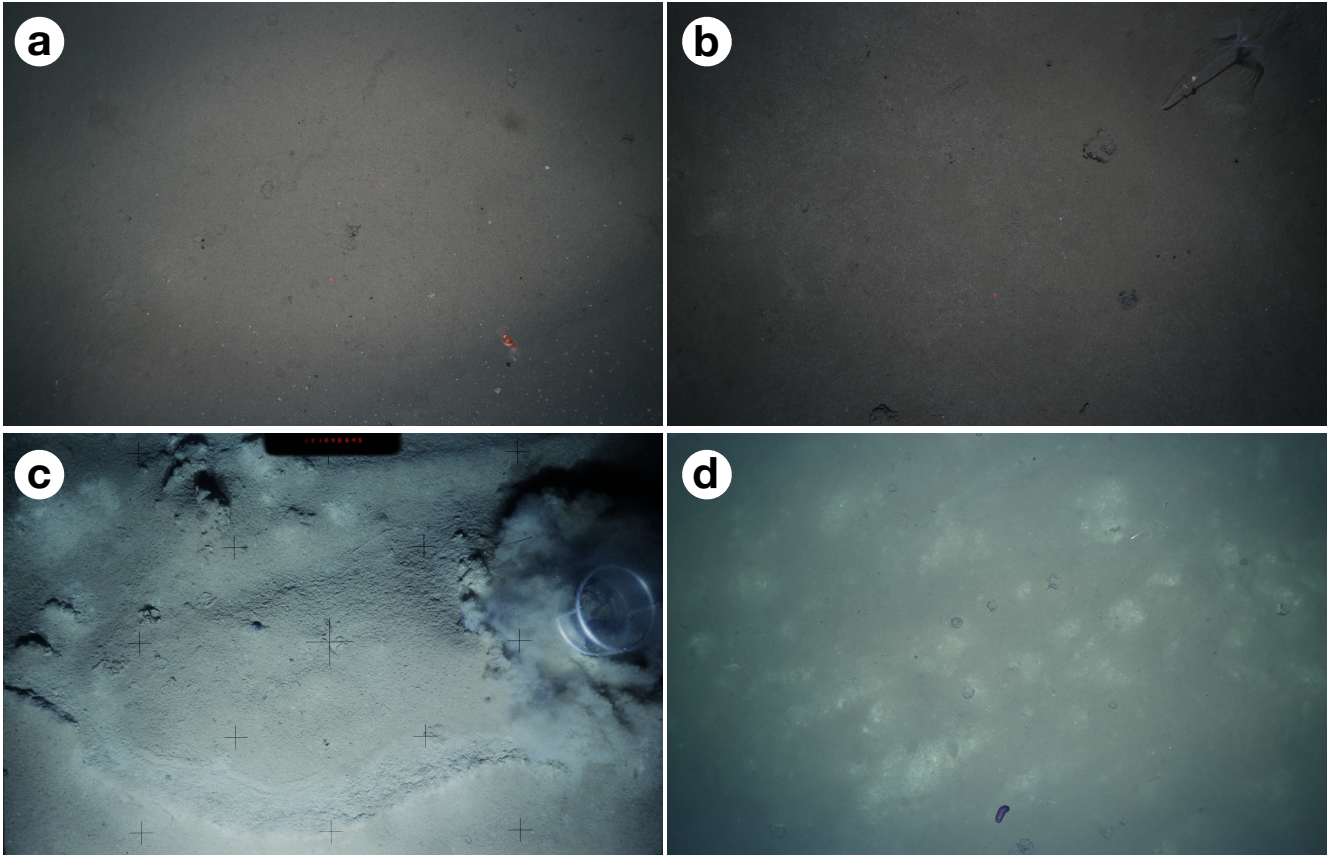
Figure 4. Example images of datasets DS_A - DS_D, with platform information and mean image footprints of: a) DS_A - OFOS - 4.9 m^2 b) DS_B - OFOS - 5.7 m^2 c) DS_C - OFOS - 12.5 m^2 d) DS_D - OFOS - 22.5 m^2
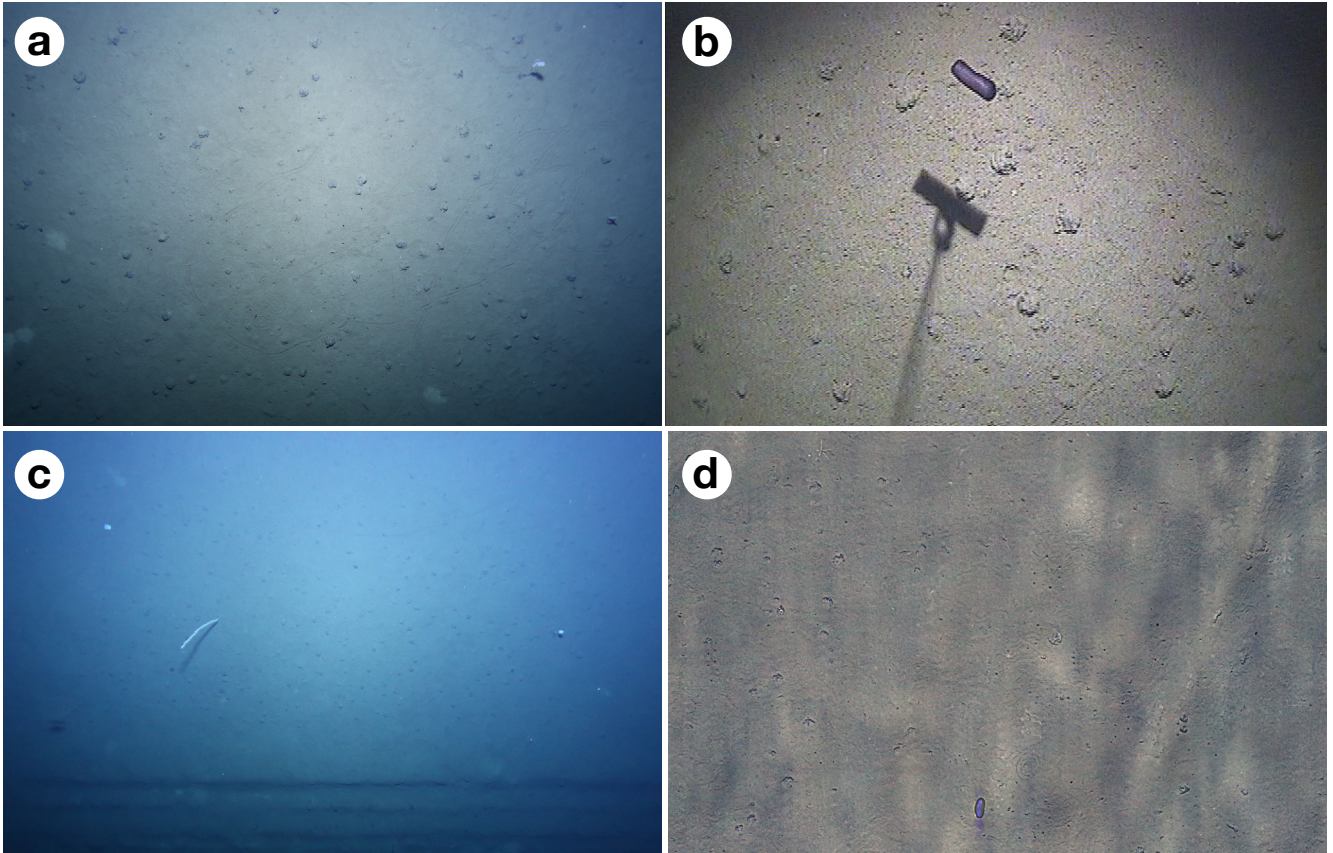
Figure 5. Example images of datasets DS_E - DS_H, with platform information and mean image footprints of: a) DS_E - AUV - 52.9 m^2, b) DS_F - OFOS - 2.6 m^2, c) DS_G - AUV - 169.1 m^2 d) DS_H - AUV - 32.8 m^2.

Figure 6. Circular fauna identifications made by several operators using the BIIGLE software application. Each circle corresponds to one annotation by one annotator. Colours of circles correspond to categories.
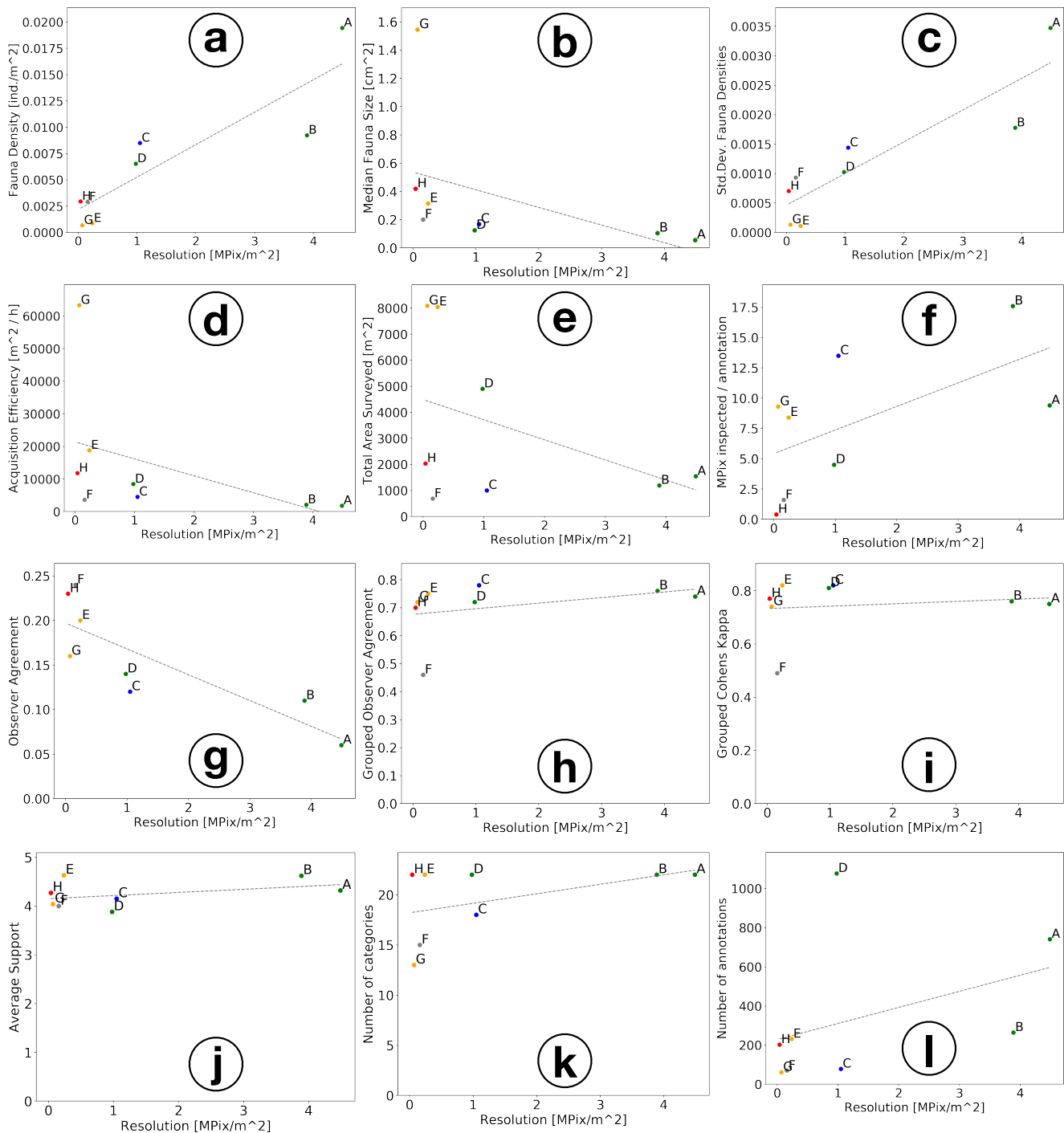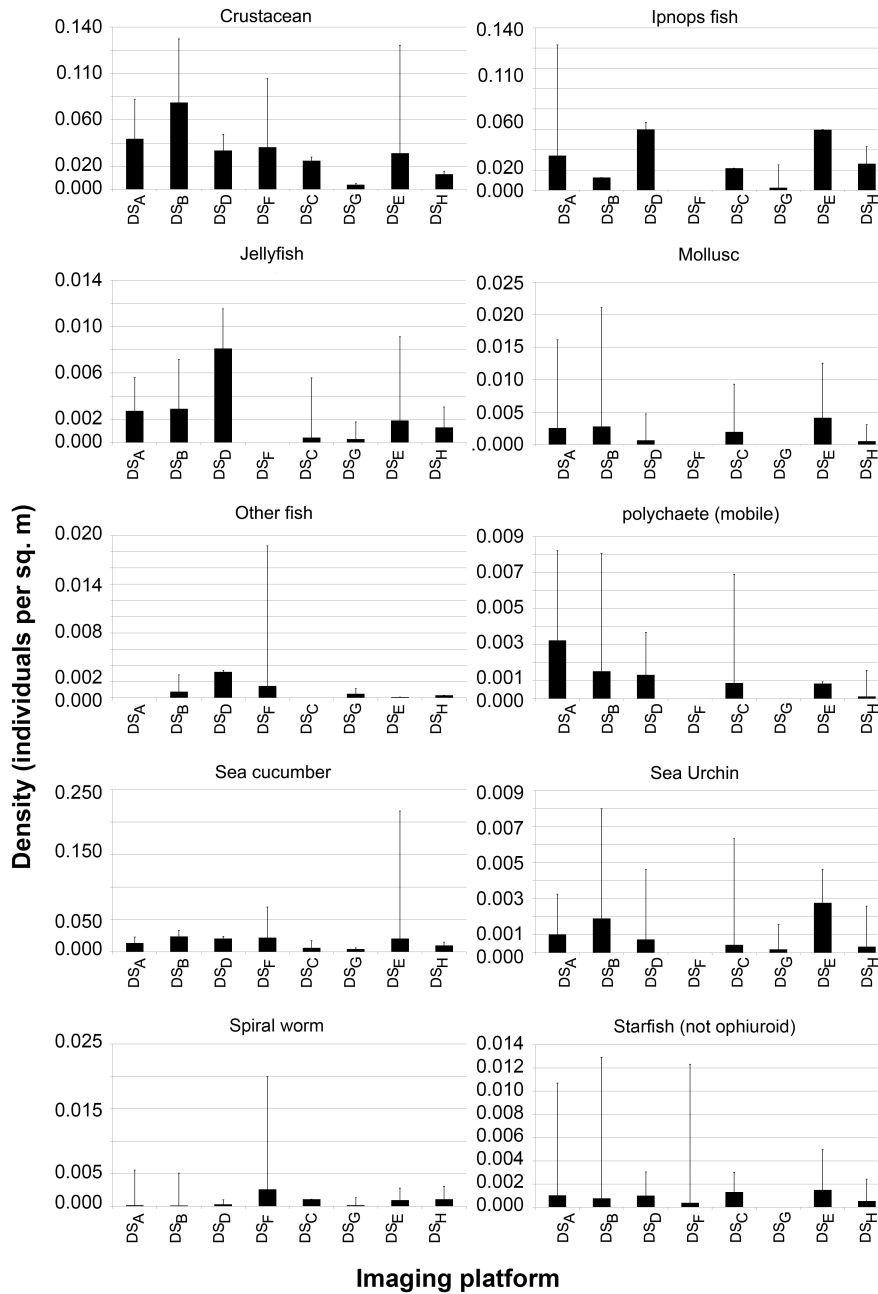
Figure 8. Mobile fauna abundances averaged across 5 annotators independently annotating image data collected during the 8 survey deployments.
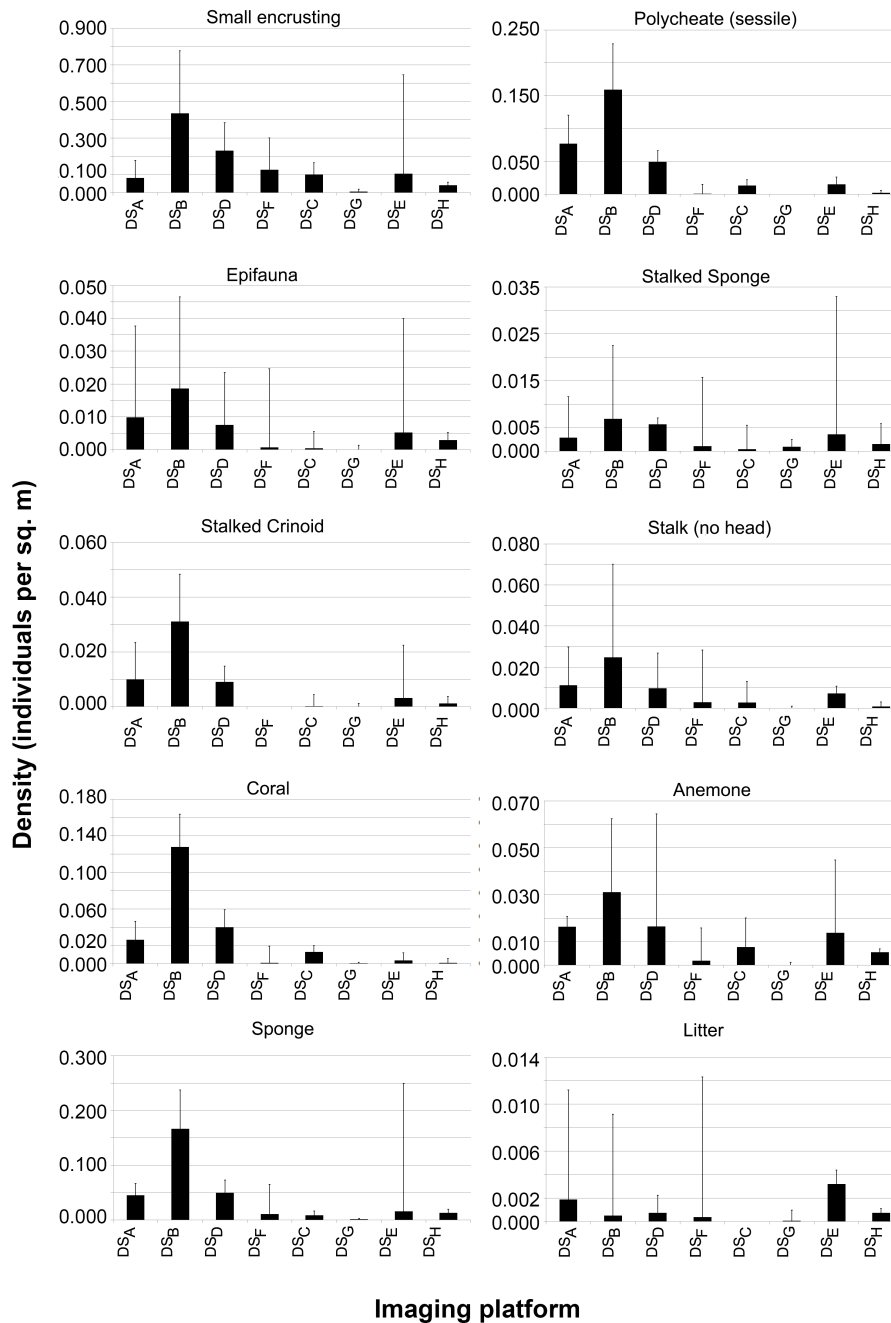
Figure 9. Sessile fauna abundances averaged across 5 annotators independently annotating image data collected during the 8 survey deployments.

Table 3. Kruskal-Wallis test assessment of whether differences in fauna abundance derived from the DISCOL seafloor data are significant for each fauna category used in the current study. H = Test statistic, N = number of observers, df = degrees of freedom , i.e. number of data types compared - 1, p = significance. P values of less than 0.05 indicate significance at the 95% percentile.

ω	Fauna	H	N	df	p
ω_a	Anemone	34.09	5	7	<0.001
ω_c	Coral	34.63	5	7	<0.001
ω_d	Crustacea	24.20	5	7	<0.001
ω_e	Epifauna	33.61	5	7	<0.001
ω_f	Ipnops fish	36.92	5	7	<0.001
ω_g	Jellyfish	32.86	5	7	<0.001
ω_h	Litter	25.68	5	7	<0.001
	Mollusc	13.65	5	7	0.46
ω_k	Other fish	29.09	5	7	<0.001
ω_l	Polychaete mobile	27.14	5	7	<0.001
ω_m	Polychaete sessile	35.16	5	7	<0.001
ω_n	Sea Cucumber	23.73	5	7	<0.001
	Sea Urchin	25.22	5	7	<0.001
ω_o	Small encrusting	16.56	5	7	0.013
ω_p	Spiral worm	25.37	5	7	<0.001
ω_q	Sponge	32.011	5	7	<0.001
ω_r	Stalked crinoid	35.54	5	7	<0.001
ω_s	Stalked sponge	23.99	5	7	<0.001
	Stalk no head	25.82	5	7	<0.001
ω_t	Starfish	16.93	5	7	0.011