Waikato Regional Council Technical Report 2025/08

Biodiversity data extraction for the Waikato area of the Hauraki Gulf Marine Park.

ISSN 2230-4363 (Online)
waikatoregion.govt.nz

Prepared by	Matt Bennion Carolyn Lundquist National Institute of Water & Atmospheric Research Ltd
For	Waikato Regional Council Private Bag 3038 Waikato Mail Centre HAMILTON 3240
Publication date	June 2022
Document ID	24899155

	Name	Date
Peer Reviewer	Michael Townsend	25th February 2025
Approving Manager	Mike Scarsbrook	26th February 2025

Biodiversity data extraction for the Waikato area of the Hauraki Gulf Marine Park

Prepared for Waikato Regional Council

June 2022

Climate, Freshwater & Ocean Science

Prepared by:
Matt Bennion
Carolyn Lundquist

For any information regarding this report please contact:

Matt Bennion
Quantitative Marine Ecologist
Marine Ecology
+64 7 838 8362
matt.bennion@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 11115
Hamilton 3251

Phone +64 7 856 7026

NIWA CLIENT REPORT No: 2022192HN
Report date: June 2022
NIWA Project: EVW22213

Quality Assurance Statement		
	Reviewed by:	Tom Brough
	Formatting checked by:	Carole Evans
	Approved for release by:	Michael Bruce

Contents

Tables

Figures

Executive summary

- Biodiversity data compiled for the project titled "Recovery of biogenic habitats: assessing the recovery potential offered by spatial planning scenarios proposed in the Sea Change Plan – ZBD2020-06" has been extracted for the Waikato region of the Hauraki Gulf Marine Park.
- Layers produced using species distribution modelling for the ZBD2020-06 project are included in the geodatabase created for this project. The modelled layers and methods used to produce them are briefly described herein.
- Point records of species presence (and in some cases absence) are used to provide a broad overview of the biodiversity in the Waikato area of the Hauraki Gulf Marine Park. Patterns of species assemblages are visualised and briefly discussed, and a focus is placed on the Mercury Islands group and Matarangi sub-region.
- Limitations of the extracted data are discussed. Principally, these limitations include missing information (e.g., sampling gear type), year of collection and low taxonomic resolution of some of the species records. Limitations of the modelled layers are also discussed, with reference to the visual expert evaluation of the layers and the data that was used to model biogenic habitat distributions in the Hauraki Gulf Marine Park.

1 Introduction

In 2021 NIWA was contracted by Fisheries New Zealand (FNZ) to provide scientific support for a project titled "Recovery of biogenic habitats: assessing the recovery potential offered by spatial planning scenarios proposed in the Sea Change Plan" (project no. ZBD2020-06, hereafter referred to as 'the FNZ Project'). To assess the protection afforded to current habitats and the recovery potential of spatial planning scenarios, information on biogenic habitats within the bounds of the Hauraki Gulf Marine Park (HGMP) was compiled. The types of information assembled included species location data from online data repositories, museums, and research institutes, as well as video and photo surveys from previous NIWA investigations (internal and external contracts) of biodiversity within the HGMP. Data was compiled with a focus on biogenic habitats, given the aims of the FNZ Project. While compiling data of biogenic habitats, data of 'non-target' species were also obtained (i.e., data of species that are not considered biogenic habitat formers), and while not used in the aforementioned project, were retained. To assess the recovery potential offered by spatial planning scenarios proposed in the Sea Change Plan, modelled layers of biogenic habitats were produced using species distribution models (SDMs). In total, 20 modelled layers of biogenic habitat were produced for the FNZ Project. The taxonomic resolution of data used to model biogenic habitats for the FNZ Project varied, thus available modelled layers represent broad biogenic habitat types for instance, 'encrusting sponges' or are of coarse taxonomic resolution i.e., 'miscellaneous Anthozoa'.

The data compiled for the FNZ Project comprise two high level taxonomic groups: macroalgae and invertebrates. Given the focus of the FNZ Project was on the recovery potential of biogenic habitats, data on fish or birds, and pelagic species (e.g., crustacea) were not compiled to inform spatial planning scenarios. Most of the species data compiled was sourced from research institutes (NIWA) and museums (Te Papa and Auckland Museum) that hold national invertebrate collections. Video/image and trawl surveys also provided a source of species occurrences, and in some cases, also provided point records of species absences. These absences (referred to as 'true absences') are important for the species distribution modelling process. Similarly, for spatial planning, goal setting and prioritisation, knowledge of where species are, and are not, found is crucial for informed decision making. Successful spatial planning relies on biodiversity information, either up to date (current distribution) or historical (conservation/recovery potential) depending on spatial planning goals. For the FNZ Project, both recent and historical data was of interest as 'recovery potential' was a significant assessment criterion. Recovery potential in this case could be any location within the bounds of the study area where recovery could potentially occur, if protected from the impacts of bottom-contact commercial fishing. For this reason, the data compiled for the FNZ Project covers several decades. Therefore, in some cases species occurrence records may represent historical locations where a given species may no longer occur.

The data compiled for the FNZ Project was entirely within the bounds of the HGMP, which contains the administrative coastal and marine boundaries of both Auckland Council and Waikato Regional Council. While the FNZ Project was in process, a request was made by Waikato Regional Council for the data compiled for the FNZ Project to be shared. Regional authorities have a responsibility to manage and guide sustainable development in their respective administrative regions, and effective management relies on best available data to inform decision making. However, there is a paucity of species distribution and abundance data for the oceans, an issue that is often mediated using species distribution models (Robinson et al. 2017). Here, data compiled for the FNZ Project, in addition to data sourced outside of the timeline of the Project (e.g., scallop by-catch information), were compiled for the Waikato Region of the HGMP. In addition, modelled layers produced to inform the spatial management scenarios within the Project, are described and provided in a geodatabase. Broad descriptions of the biodiversity within the Waikato Region of the HGMP are given, with specific focus placed on the Mercury Islands group (east of the Coromandel Peninsula) and Matarangi sub-region.

2 Methods

Most of the data within this report is a subset of the data compiled for the FNZ Project. However, the data compiled for the FNZ Project built on several previous contracts to NIWA (e.g. Lundquist et al. 2014; Lundquist et al. 2020a) that involved data compilations. The sources of data are therefore diverse, as well as the permissions for use and attributions (to the data collector or source, but also the data compiler) (Table 2-1). For some data sources, data are shared with a confidentiality clause, such that data are available for use by Waikato Regional Council in informing coastal management and planning in the Waikato region, but permission is not granted for any further sharing of this information (permissions information is included in the extract database). In this section, the sources which represent the data extract are outlined, as well as the types of data. In addition, the data extract includes modelled biogenic habitat layers created for the FNZ Project (ZBD2020-06). The methods used to create these layers are briefly discussed here; for more information on the methodology used to create these models see Stephenson et al. (2021b), Stephenson et al. (2020a) and Stephenson et al. (submitted).

2.1 Data extraction

2.1.1 Types of data

The primary data extracted from the database are point (location) records of species presence (occurrence) and/or absence (Figure 2-1). Where available, abundance data has been included. There is a total of 7,625 species occurrence records in the Waikato regional extract and 16,351 species absence records, across 1,783 unique locations (e.g., unique combination of latitude and longitude in the database). For 2,444 species records, abundance information was also available. All records from the source: The Ministry of Business, Innovation and Employment (MBIE) funded Juvenile Bottlenecks Programme (source: Mark Morrison, NIWA) are accompanied by abundance information. The abundance data in this instance is volume (ml), derived from graduated bins following research beam trawls. This MBIE project is still in process, and this dataset has not yet been finalised or published; thus, it is requested that all data from this source is kept confidential by Waikato Regional Council, and not shared externally until it has been made publicly available. Many of the records in the scallop database (Williams and Parkinson 2010; Williams et al. 2010; Williams et al. 2013) are accompanied by abundance information. Abundance information from the scallop database is either count (number of individuals) or volume (litres) of individuals caught (dredge) and is present for the 2010, 2012 and 2021 (J. Williams, unpublished data) survey years. That said, abundance data does exist for the 2009 survey year (J. Williams, pers. comm.). When using the abundance data from these sources, careful consideration should be placed on the standardisation of abundance information. That is, taking the trawl/dredge distance and area swept into account to determine an abundance estimate within a standardised area (e.g., km^2).

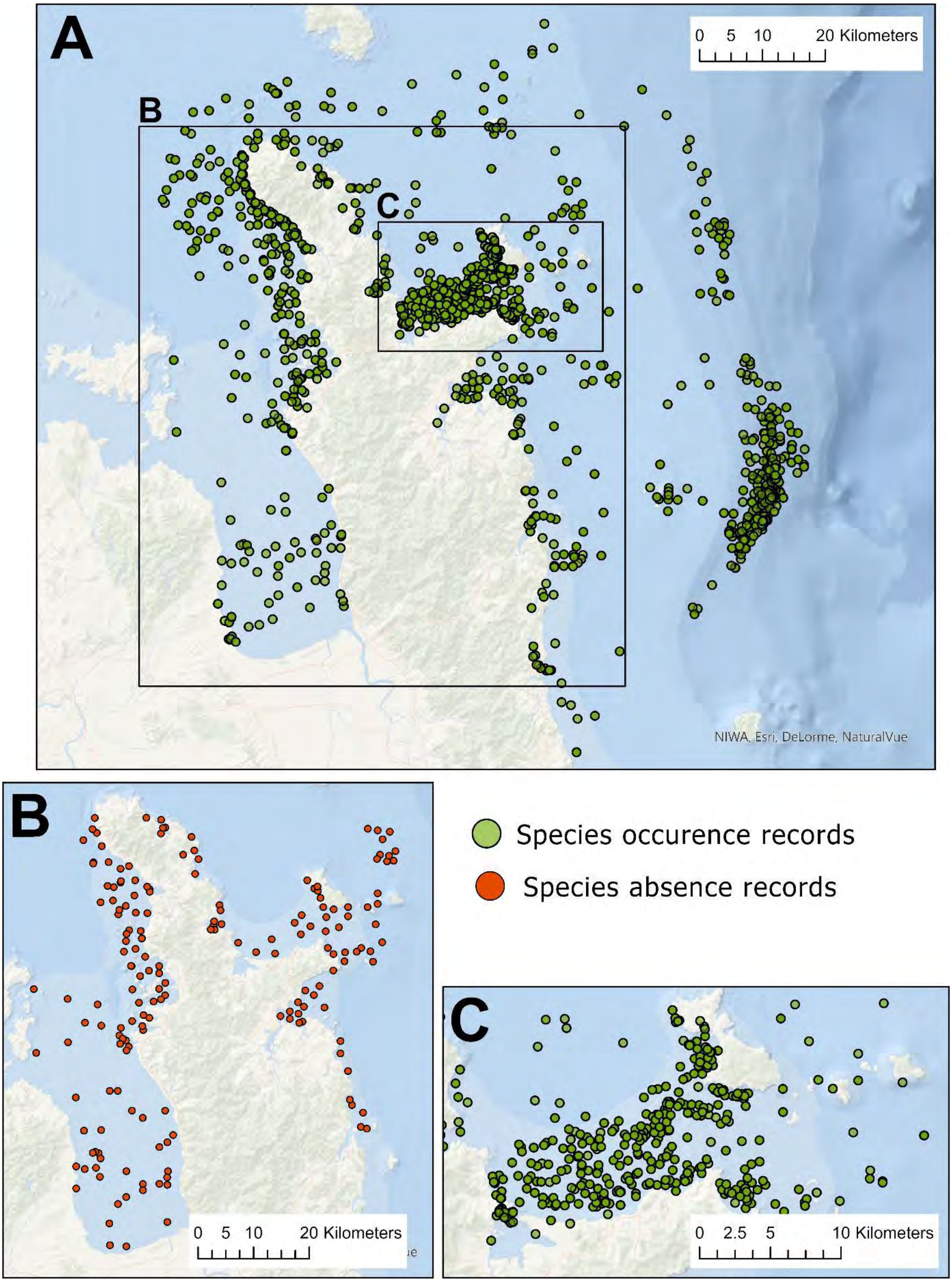

Figure 2-1: Species occurrence/absence. Maps showing the distribution of point records in the HGMP database. A) full HGMP study area of FNZ project with distribution of species presence records, B) distribution of species absence data in the Waikato regional extract, C) distribution of species presence records in the Mercury Islands group and Matarangi sub-region.

Table 2-1: Sources of species records in the database containing the Waikato regional data extract. 'Database' here indicates where data was sourced from (i.e., 'core data source'). The number of unique species record locations within each compiled database is shown, as well as the total number of species presence and absence data compiled from each respective database.

Database	Core data source(s)	No. unique locations [a]	No. species occurrence records	No. species absence records	No. species abundance records
Museum records	Database representing data on invertebrate collections maintained by Te Papa Museum, Auckland Museum, *niwainverts*	838	3,489	0	0
National macroalgae database	Compiled and maintained by NIWA on behalf of Te Papa Museum, Auckland Museum, and NIWA	85	434	0	0
Juvenile Bottlenecks	Unpublished data from MBIE Endeavour Programme, PI Mark Morrison. * please note confidentiality with respect to this dataset until it is published	126	501	15 875	501
Horse mussel records	Various sources collated by Clinton Duffy (all *Atrina zelandica*)	35	35	0	0
Firth of Thames acoustic habitat mapping	Video records as ground truth data for NIWA client report for the Department of Conservation (Morrison et al. 2002)	28	43	473	0
Tier 1 Marine Biodiversity reporting	OBIS, NIWA, TRAWL Invertebrate point records compiled for NIWA client report for Fisheries New Zealand (Lundquist et al. 2014)	338	620	0	0
Scallop	Scallop by-catch database (Fisheries New Zealand) (Williams and Parkinson 2010; Williams et al. 2010; Williams et al. 2013)	333	2,503	0	1,943

[a] Based on unique combinations of latitude and longitude

2.1.2 Data sources

A list of data sources for all information in the extract database is shown in Table 2-1. Sources of data are given in two levels. The upper level given is the compiler source, for instance "NIWA macroalgae database", whereas the lower-level data source gives attribution to the core data source(s) e.g., Te Papa Museum or Auckland Museum. For invertebrates, the sources of data vary greatly from research institutes (NIWA, Te Papa Museum and Auckland Museum) to online data repositories (OBIS: Ocean Biodiversity Information System) and NIWA contracts (Morrison et al. 2002). Three separate counts are given for each respective source, (1) number of unique locations per data source, (2) number of presences i.e., individual species records per source and (3) number of absences per species source (Table 2-1).

2.1.3 Missing metadata

For most of the species records, date/year and original source are included. There are, however, several records with either missing year and/or original source information. For year, source, and gear type, if no information was available to populate these fields, the respective field has been left blank. Information on sampling gear type is missing for approximately a third of the occurrence records in the data extract database.

For each species record in the extract database, as much taxonomic information as was available has been provided. Nevertheless, the resolution of taxonomic information varies considerably. For repository sources, i.e., OBIS, NIWA macroalgae database and records from Auckland Museum and Te Papa Museum, there is generally complete (or almost complete) taxonomic information from phylum to genus. Approximately half of the records in the database contain taxonomic information to 'species' level. For some data, primarily from research or NIWA contract sources (e.g., Juvenile Bottlenecks Programme, Table 2-1), the taxonomic resolution varies considerably, with some records identified to species level and others with little or no taxonomic information. For these records, the entry may be descriptive, e.g., "feathery branched red algae". In these cases, a common name has been given to the record, e.g., "red macroalgae", but no taxonomic information has been extrapolated for the Waikato regional extract, i.e., phylum has been left blank. For records where a common name was available, e.g., cushion star, high level taxonomic information has been generated, e.g., 'Echinodermata' for inclusion in the extract database.

Due to missing metadata, it is possible that there are duplicate records within the database in some instances. Potential duplicates within the database have been flagged using the 'duplicated' function in R (R Core Team, 2018). A record has been marked as a 'potential duplicate' with a corresponding TRUE/FALSE if the latitude and longitude, phyla, genera and year for a given row matched another row in the extract database.

2.2 Biodiversity in the Waikato Region of the Hauraki Gulf Marine Park

In addition to the data extraction and compilation, a secondary aim of this project involves an overview of the biodiversity in the Waikato region of the HGMP. A focus is placed on the Mercury Islands group and Matarangi sub-region (Figure 2-1) and a general comparison of the taxa present inside and outside of this sub-region is drawn. The Mercury Islands group consists of the large Great Mercury Island (Ahuahu), Red Mercury Island (Whakau) and five islets (Korapuki, Green, Atiu, Kawhitu and Moturehu).

For this overview, the proportion of species records within each given phylum has been visualised for the entire study area and within the sub-region (Mercury Islands group and Matarangi). All analyses and data visualisation were carried out in R (R-Core-Team 2018) using the 'dplyr' and 'ggplot2' packages (Wickham 2011) and ArcGIS Pro (Esri).

2.3 Species distribution models

Species distribution models (SDMs) were produced for 20 biogenic habitat groups for the FNZ Project. Three of these modelled layers were given low evaluation scores following expert evaluation, where marine ecologists/taxonomists with knowledge specific to each biogenic habitat group gave modelled layers (predicted spatial distributions) a score based on their knowledge of a biogenic groups or species' distribution. The three low scoring models were mussels, Brachiopoda and calcareous tubeworms. Given the poor performance of these models, they have also been excluded from the compiled geodatabase for this project. In total, 17 modelled layers of biogenic habitat distribution are provided with the Waikato region extract. These modelled layers were produced using similar methods described by Stephenson et al. (2020a), Stephenson et al. (2021a) and Stephenson et al. (2021b), with minor adjustments. Models were fit using many of the same gridded (250x250 m) environmental variables (layers) described in detail in Table 2-4 of Stephenson et al. (2020b).

Briefly, two types of binomial models were used to create modelled layers of biogenic habitat distribution. Boosted regression trees (BRTs) and Random Forests (RF) models are often used to model species distribution, and our approach used a combination of these two models, referred to as ensemble models (see Stephenson et al. 2021). BRT and RF models were fit with training data and evaluated using withheld 'evaluation' data repeatedly for 100 bootstraps. The BRT models were fit with a Bernoulli distribution, an out-of-bag fraction of 0.6 and 5 folds (cross validation). The RF models were fit with a step factor of 1.5. Model fit was assessed using the area under curve (AUC) and true skill statistic (TSS). For each biogenic group, a predicted spatial distribution was produced, as well as an uncertainty layer, in this case standard deviation (Table 3-1) based on the 100 bootstraps. The predicted distribution reflects the habitat suitability index (HSI) for each biogenic habitat type. HSI represents the relative suitability of the environmental conditions for taxa occurrence and ranges from 1 (highly suitable) to 0 (highly unsuitable) habitat (Stephenson et al. submitted).

3 Results & discussion

3.1 Biodiversity in the Hauraki Gulf Marine Park

3.1.1 Waikato region of the Hauraki Gulf Marine Park

The Waikato region of the Hauraki Gulf Marine Park contains a diversity of benthic habitats which support a wide range of flora and fauna. For instance, the Coromandel Peninsula contains forests of large brown macroalgae (evidenced by records of *Ecklonia radiata*, *Carpophyllum* spp. and *Cystophora* spp. in the species database) along parts of its coastline. In estuaries from Matarangi to Tairua, the intertidal sediments host numerous taxa typically expected in the intertidal estuaries of New Zealand. These include species critical for healthy ecosystem functioning in these soft-sediment habitats like polychaetes (e.g., order Maldanidae) and infaunal bivalves (e.g., *Austrovenus stutchburyi* and *Macomona liliana*).

Offshore from the Coromandel Peninsula in deeper waters of the Waikato region of the HGMP, records for sponges (Porifera), sea anemones (Actiniaria), sea pens (Pennatulacea), and other cnidarians (Anthozoa) and isopods and decapods (Arthropoda) are more common. This 'deep' region is characterised by depths >200 m and contains comparatively unique taxa (within the database), when considered next to the rest of the HGMP which includes predominantly shallower waters (<200 m). Porifera records within the Waikato regional extract include sponges of the order Tetractinellida (Class: Demospongiae). For Cnidaria, this area contains the only records for sea pens (order: Pennatulacea) in the HGMP, as well as many of the records for the orders Scleractinia (stony corals) and Actiniaria (sea anemones). The biodiversity in this area is so distinct compared to the 'inner gulf' that for previous spatial planning optimisation exercises (i.e. Lundquist et al. 2020b), this deep portion has been removed from the modelled area, to avoid skewing prioritization towards an area that while unique to the HGMP due to the ecologically arbitrary administrative boundaries, is not necessarily unique in broader New Zealand deep waters. A majority of the records in the Waikato regional extract in in this deep area are for Arthropoda, in particular *Metanephrops challengeri* (scampi). Proportionally, the large number of scampi records within this area is expected, as this area supports a high proportion of the fishing effort for the scampi commercial fishery.

Proportionally, Mollusca account for the highest number of occurrence records in the Waikato regional extract (46%), followed by Arthropoda (16%) and Echinodermata (12%) (Figure 3-1). Many species of Echinodermata, Mollusca, Arthropoda and Annelida are cosmopolitan in their distribution within the study region. This is expected given the diversity of taxa within these phyla and the subsequent broad ecological ranges/niches that taxa within these phyla can occupy (e.g., starfish, gastropods, bivalves and polychaetes). On the other hand, some taxa could be considered comparatively rare within the compiled database and by proxy, the study area. Yet, while the data compiled here could be used to identify areas that are species rich (with caveats placed on underlying spatial biases due to differences in sampling effort), identifying rare taxa and areas with comparatively low species richness or diversity should be done with caution. Underlying spatial biases may result in reduced sampling of habitats or areas that are difficult to access. Taxonomic biases may also exist as rare species may be less easily identified, or not yet taxonomically identified. For instance, Bryozoa can be found in an extremely wide range of habitats, given their epibiotic nature, i.e., co-occurring with many other species (Wood et al. 2012), yet the low proportion of Bryozoa records within the database (0.8%, Figure 3-1) is undoubtedly lower than 'real' proportions due to the high level of expertise required to accurately identify Bryozoa taxa. Smaller species such as Nematoda (where there is only one record in the database) are less commonly included in macroinvertebrate museums as they are considered meiofauna and therefore would often be intentionally excluded; these taxa often parasitise hosts and therefore may remain undetected, unless diagnostics like routine histopathology of hosts or bespoke studies of nematode life-histories are undertaken.

Many of the Bryozoa records that are contained within the database are concentrated in one area, the Colville Channel. In the database there are 62 occurrence records for Bryozoa, with 25 of these found at one unique location in the Colville Channel (location: -36.400, 175.380), and most records are from the class Gymnolaemata. For Porifera, many of the taxonomically identified sponge records within the database are found on the west coast of the Coromandel Peninsula from near Waiaro to Amodeo Bay (though most of the unidentified sponge records are situated elsewhere). This area also contains most of the Ascidian (Chordata) occurrence records with many records either unclassified taxonomically or described as *Styela* species.

3.1.2 Mercury Islands Group and Matarangi sub-region

Sixty percent of the Porifera (sponge) records in the database are contained within the Mercury Islands Group and Matarangi sub-region. Though almost of these sponge records are only identified to phylum level (source: scallop database). Of Porifera records that are taxonomically identified in the sub-region study area (Figure 3-2), two ‘clusters’ of sponge records are present. On the west of Great Mercury Island (Ahuahu), three records for sponges in the genus *Halicondria* are present, and off the point of Opito Bay, three records are present for upright, finger sponges including Ras*pailia topsenti, Iophon minor* and *Callyspongia ramosa*. In the sub-region, primarily in the channel between Opito Bay and the Mercury Islands group, many records for large brown macroalgae (~80) are present (phylum: Ochrophyta), the majority of which are for *Ecklonia* species. In the same area, several red algae (described but taxonomically unidentified) records (12) and green algae (mainly *Ulva* spp. and *Caulerpa* spp.) records (12) are present. In addition, three records for rhodoliths (calcareous, crustose algae) and 18 records for ‘coralline turf’ are present in this area.

Throughout the sub-region (primarily west of the Mercury Island group) there are hundreds of records for Echinodermata (434) Arthropoda (310) and Mollusca (1109) most of which are ‘by-catch’ from the scallop database (derived from scallop dredge surveys). Given the objective of these scallop surveys (target taxa: *Pecten novaezelandiae*), almost a quarter of the mollusc records in this area (259 of 1109) are for scallops (*Pecten* spp.). Notably, 55 of the mollusc records in this sub-region are for dog cockles (*Tucetona laticostata*). For Echinodermata, almost 85% of records in this area are comprised of two starfish genera *Astropecten* (232) and *Luidia* (132). Similarly, for Arthropoda 130 of 310 records are for unidentified crabs and 95 are for the genus *Pagurus* (hermit crabs). Therefore, >70% of species records for Athropoda in the sub-region are for crab species. On the west side of Greater Mercury Island (outside Hurihuri Harbour), most of species occurrence data in the extract is for macroalgae; brown, green and red. Though there are also dozens of records for unidentified Porifera (sponges) and Mollusca (class: Gastropoda), as well as scallops (*P. novaezelandiae*) and starfish (Astropecten). Within Huruhi Harbour, there are three records for horse mussels (*Atrina zelandica*).

To the east of Greater Mercury Island, dozens of records for Arthropoda are present within the database, all within the class Malacostraca (location: -36.595, 175.836). Drawing a line through the Mercury Islands Group from the southern coast of Great Mercury Island (Ahuahu) to Red Mercury Island (Whakau) (Figure 3-2), the majority of datapoints are for macroalgae (brown, red, and one green). Gastropod and asteroid (*Astropecten* spp.) records are also present in this area, as well as a single bryozoan record (morphologically described but taxonomically unidentified). Additionally, to the west of Korapuki Island, there is a single rhodolith occurrence record. South of this arbitrary line through the Mercury Islands group, between Opito Bay and Ohinau Island/Ohinauiti Island, there is a cluster of algae records (mostly *E. radiata* and other unidentified brown algae) and a cluster of sponge records (*n* = 11) all in the class Demospongiae (e.g., *Tethya*, *Raspailia*, *Aaptos* and *Polymastia* spp.).

Within the Whangapoua Harbour (Matarangi) there is a cluster of species records on the north of the harbour entrance (close to Opera Point) (Figure 3-2). Most of the data from this area are Mollusca records (bivalves and gastropods) with a wide range of genera present e.g., the bivalves *Dosinia*, *Paphies* and *Talochlamys* spp. and gastropods *Cominella* and *Nomoacmea* species. In this small area, there are 189 individual species records at 14 unique locations, indicating considerable sampling effort within this area. In contrast, the same sampling effort is not replicated to the north in Kennedy Bay, where 31 individual species records are present at 9 unique locations (taxa present: macroalgae and Echinodermata, i.e., starfish).

There is a single occurrence record for Brachiopoda (lamp shells) within this sub-region, at Whangapoua, for *Calloria inconspicua*, with only five brachiopod records in total in the entire Waikato region study area. There is a general paucity of Hydrozoa records, with only 14 records at 11 unique locations, widely spread within the Waikato region of the HGMP. However, in the sub-region, there are 5 records at three unique locations for the hydrozoans *Aglaophenia*, *Solanderia* and *Amphisbetia* spp. On visual assessment, the number of records within each phyla is similar between the sub-region and the entire study area (Figure 3-1). That said, there are proportionally higher numbers of occurrence records for macroalgae (Ochrophyta and Rhodophyta), Echinodermata and sponges (Porifera) in the Whangapoua/Matarangi/Mercury Islands sub-region compared to the entire study area.

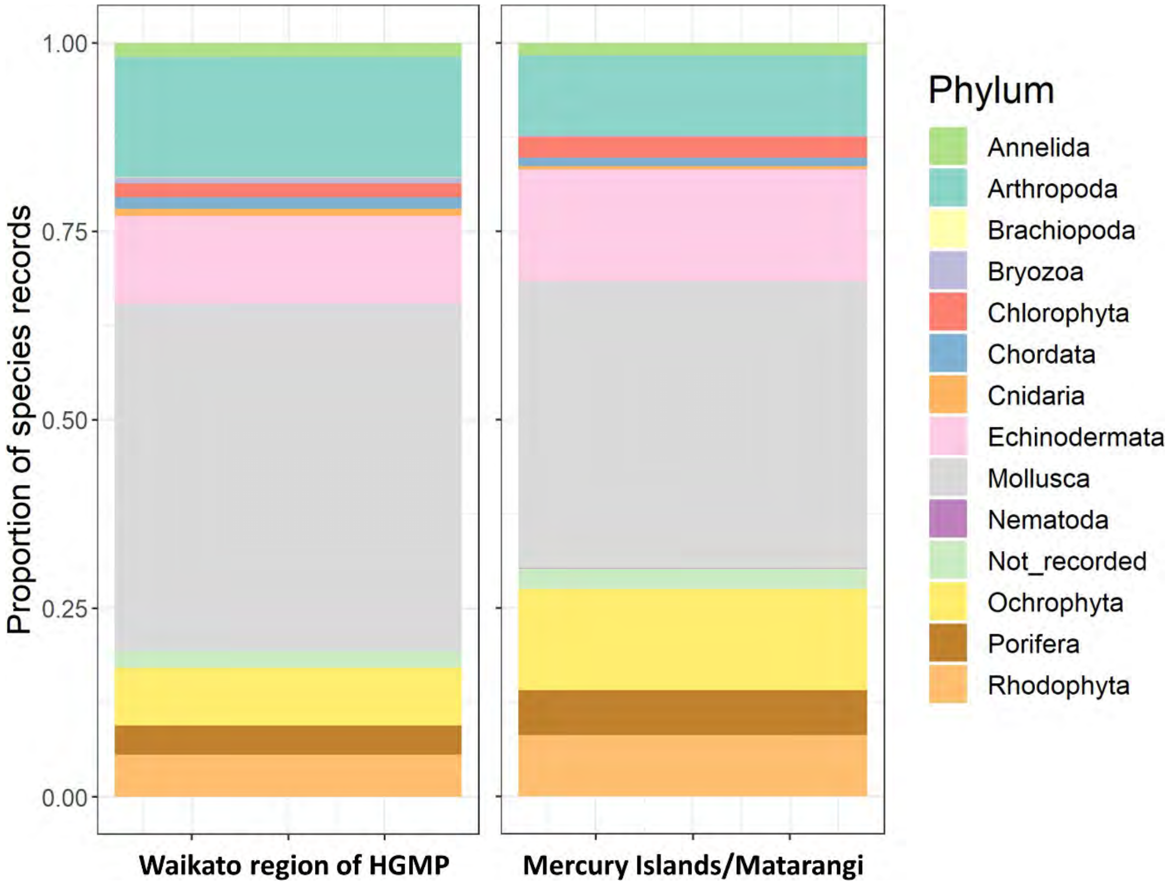

Figure 3-1: Proportion of species occurrence records in each phylum. Stacked bars show the proportion species occurrence records in each given phylum for the entire study area (Waikato Region of the HGMP) and the Mercury Islands Group and Matarangi sub-region.

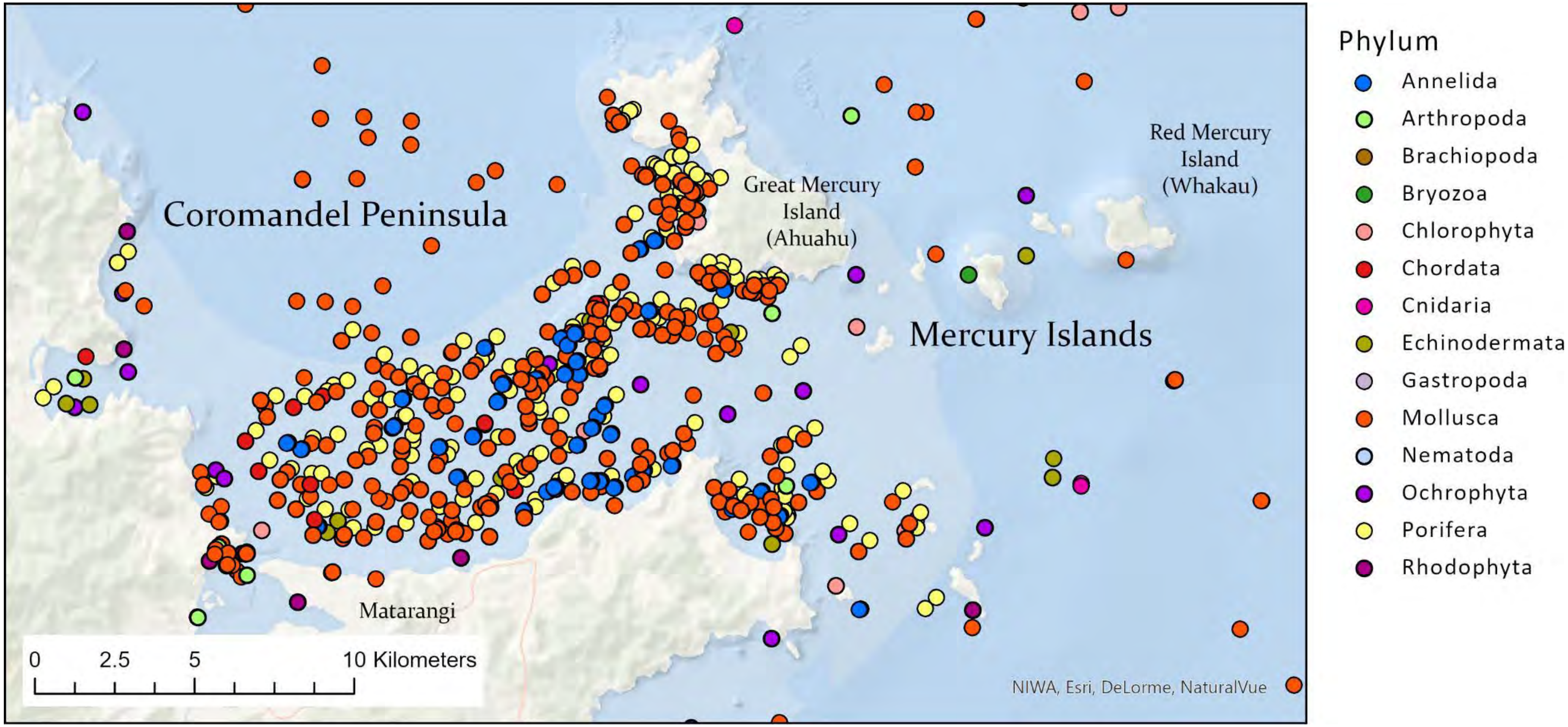

Figure 3-2: Phyla represented in the Mercury Islands group sub-region. Biodiversity in the Mercury Islands group and Matarangi sub-region. Different phyla are represented by different coloured points. Note: many locations contain multiple species records; biodiversity shown here is therefore not exhaustive. Additionally, jittering has been used for visualisation purposes.

3.2 Species distribution models

In the geodatabase accompanying this report, 17 modelled layers of biogenic habitats are included. A full list of modelled layers and a description of each modelled layer is provided in Table 3-1. The number of species occurrence records used to create each modelled layer is also shown in Table 3-1. For each layer, there is an accompanying uncertainty layer (standard deviation). The modelled layers ensemble predictions made from BRT and RF models; the spatial predictions are mean habitat suitability index (HSI) values based on 100 bootstraps. Therefore, the accompanying uncertainty layers are the standard deviation (SD) based on the same 100 bootstraps. The uncertainty layer is influenced in part by the spatial distribution of input species records, because of the random sampling of training and evaluation data that takes place in the modelling process for each bootstrap. Sometimes uncertainty layers can be also interpreted as a proxy of biodiversity data paucity, i.e., reflecting high uncertainty in areas of low sampling effort for a given taxa group or biogenic habitat. Though this should be done with caution as uncertainty is also influenced by the spread of data across environmental gradients (variables used in modelling). Environmental coverage is better explored by tailored modelling approaches (see Stephenson et al. 2021b). Each modelled layer and accompanying uncertainty layer have been provided in raster format at a resolution of 250 m x 250 m (Table 3-1).

Table 3-1: Modelled layers of biogenic habitat. A list of the modelled layers is provided in the table below, including descriptions of modelled layers, the number of unique species records (one per 250 m x 250 m grid) and a list of model outputs included in the accompanying geodatabase.

Layer name / habitat modelled	Description	No. of occurrences records use to create models	Model outputs (raster format)
Hydrozoa	All point records for taxa in the class Hydrozoa	30	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Sea anemones	All sea anemone taxa	137	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Cup corals	All cup coral records, mostly *Monomyces* and *Flabellum*	44	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Misc. Anthozoa	All Anthozoa records except for sea anemones, cup corals and Hydrozoa	67	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Biogenic patches	Multi-species aggregations that indicate biogenic habitat. Models created with records for 'biogenic lumps', dog cockles and dead shell debris	346	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Canopy-forming algae	Large brown macroalgae e.g., *Ecklonia radiata* & *Carpophyllum* spp.	308	HSI layer & uncertainty layer (SD) 250 m x 250 m grid

Layer name / habitat modelled	Description	No. of occurrences records use to create models	Model outputs (raster format)
Rhodoliths	All rhodolith point records; *Lithothamnion*, *Sporolithon* and observations (videos and M. Morrison, pers. obs.)	24	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Misc. macroalgae	All macroalgae taxa records remaining following the removal of canopy-forming macroalgae and rhodolith records	424	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Encrusting sponges	Encrusting sponge taxa, list created based on literature and expert advice (M. Kelly, NIWA)	248	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Erect/upright sponges	Encrusting sponge taxa, list created based on literature and expert advice (M. Kelly, NIWA)	531	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Encrusting Bryozoa	Encrusting Bryozoa taxa, list created based on literature and expert advice (D. Gordon, NIWA)	120	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Erect Bryozoa	Erect/frame building Bryozoa taxa, list created based on literature and expert advice (D. Gordon, NIWA)	48	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Erect and rooted Bryozoa	Erect and rooted Bryozoa taxa, list created based on literature and expert advice (D. Gordon, NIWA)	24	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Horse mussels	All point records for *Atrina* sp.	477	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Oysters	All point records for oyster spp. e.g., *Saccostrea* and *Ostrea*	42	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Non-calcareous tubeworms	Generally hard substrate associated taxa; mainly Sabellida	195	HSI layer & uncertainty layer (SD) 250 m x 250 m grid
Misc. Annelida assemblages	Generally soft sediment associated taxa, some tube forming; mostly taxa in the orders Terebellida, Eunicida and Spionida	209	HSI layer & uncertainty layer (SD) 250 m x 250 m grid

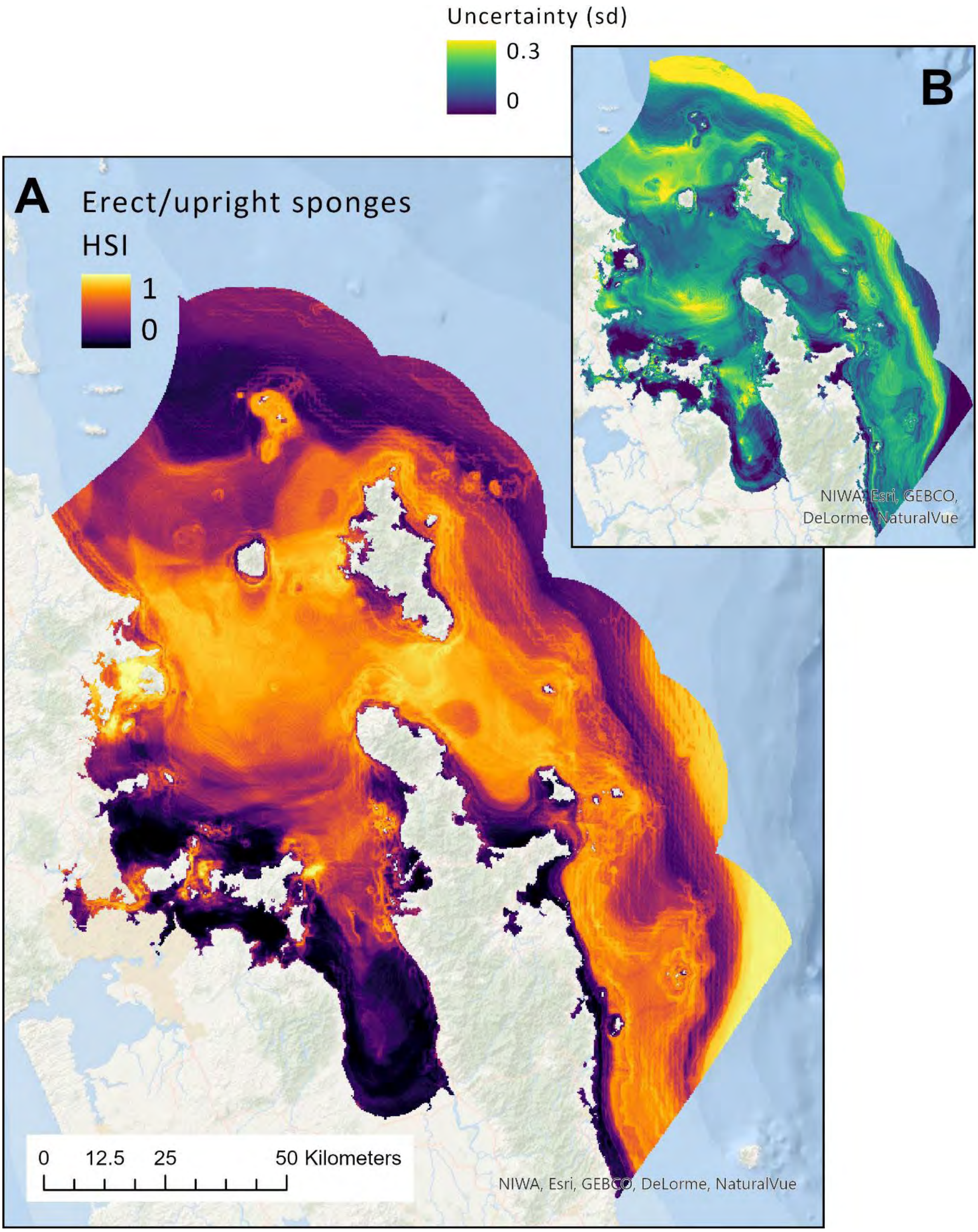

Figure 3-3: Species distribution model example with uncertainty. A) Habitat suitability index (HSI) modelled layer for the erect/upright sponge group, B) accompanying uncertainty layer (standard deviation).

3.3 Limitations of the compiled data

As with almost all spatial datasets, limitations exist which should be considered when using the data for research or to inform spatial management. In the compiled species occurrence database, there are several key limitations which should be considered: 1) missing sampling gear type; 2) missing collection year; 3) age of records; 4) low taxonomic resolution/depth; and 5) spatial bias in sampling effort. For many spatial analyses, gear type might not be of concern to the data user but can challenge comparability of point records based on likelihood of capture. Some gear types are either highly selective (cores or grabs) or indiscriminate (trawls or dredge). For absence records, if the gear type used is highly selective, any absences should only be used to denote absences of the specific target species, i.e., absence in this case is not indicative of absence of all biodiversity. Additionally, for gear types like trawl and dredge the presence record is given with a start, end and mean latitude and longitude. Often a single location is used, with consideration not given to the length of trawl, or where a record was found within a trawl. In this report, and in the extract database, start latitude and longitude for records from the 'scallop' database were used/provided in the 'latitude' and 'longitude' columns for consistency, as end points are not available for the 2009 scallop dredge survey. For a given use, it could be important that fine resolution spatial data is used, in these instances, records derived from trawls might not be applicable. If required, the user can derive uncertainty of the true spatial location of the species record using the start and end positions to elucidate the length of trawl. Where gear type information is missing, these potential limitations are amplified.

The date that a given record was collected may be of considerable importance. For instance, if the data is being used to inform current species or habitat distributions, it may be necessary to threshold data beyond a given age as spatial distributions may have changed since the record was created. Again, this limitation is amplified if the date/year information is missing from the database (this is the case for approximately a fifth of the database). On the other hand, if the desired use for the data is to inform on historical distributions, these historical records might be crucial for the use case. For instance, for the FNZ Project the goals included developing modelled layers of potential recovery distribution. In this case, historical records were equally important for modelling as records that could inform 'current distribution' as well as 'recovery distribution' should fishing impacts be removed. When using the modelled layers of biogenic habitat, it should be recognised that modelled layers are raw outputs that reflect environmental correlates between available presence and absence data and show suitable habitat for a taxa. The layers do not account for instances where historical stressors may have resulted in that taxa's removal. These spatial distributions should be used with caution to inform spatial management without the application of 'condition' layers representing current and historical stressors. Condition layers could reflect aggregated fishing impact (i.e., effort), sedimentation or other impacts which could have altered biogenic habitat spatial distribution.

While most of the records in the accompanying database contain taxonomic information of the respective species (especially those from data repositories), some records within the database have not been taxonomically identified and are therefore, without associated taxonomic data. For instance, some records gathered by trawl, dredge or video were not identified to genus level (1554 occurrence records in total). In these instances, taxonomic data have been added (depending on the description given) where possible. Nevertheless, for 161 records, a description (colour or shape) is provided in lieu of any taxonomic information.

A common issue in databases like the one compiled for this project is spatial bias of sampling effort. This bias is an inevitable facet of species data repositories that rely on observations from fishers (by-catch data), members of the public and researchers with limited systematic survey effort for the region as a whole. Essentially records are often concentrated where there is reasonable access, i.e., common fishing grounds or safe access from the shore. Spatial bias issues are typically only mitigated with inclusion of datasets compiled systematically with careful attention placed on the reduction of spatial bias, that is, bespoke research effort, with study designs created with spatial coverage in mind (e.g. Morrison et al. 2002).

Finally, all records extracted from the scallop database (2,503) are accompanied by a life status column (i.e., alive or unknown). Records do exist for dead by-catch too (J. Williams, pers. comm.) though depending on the desired use for the data, caution should be exercised when using the 'unknown' life status species records (466 of 2,503 records), considering they may represent deceased individuals.

4 Conclusion

There is considerable information on biodiversity captured in the accompanying database for the Waikato regional area, but also the Mercury Islands and Matarangi sub-region study area. Almost half of the occurrence records compiled are for marine molluscs (bivalves and gastropods). This is unsurprising given the research effort placed on these taxa compared to Brachiopoda, Nematoda and Bryozoa. Proportionally, many of the records within the database are for Arthropoda too (16%), which is attributed in part to the scampi fishery that takes place in the southeast of the HGMP.

When using the data compiled for this project, careful attention should be paid to the previous section 'Limitations of compiled data' of this report. Inevitably, any underlying biases in a compiled dataset are amplified if mixed with data with the same or new biases, like missing metadata or spatial variation in sampling effort. For instance, several datasets included in the compiled database are related to or derived from commercial fisheries, which result in spatial bias of species occurrence data toward areas of commercial fishing effort.

Finally, the species occurrence data described in this report and within the extract database were compiled for the Project (ZBD2020-06), which was possible due to previous efforts compiling data for past NIWA contracts such as Lundquist et al. (2014). All efforts have been made to correctly attribute data to its original source, while acknowledging its secondary source i.e., compiler(s).

In summary, the data compiled for this project includes:

1. A database of species occurrence records with taxonomic information, year, source (and secondary source where applicable), position (latitude and longitude) and gear.
2. A database of species absences with taxonomic information, year, source (and secondary source where applicable), position and gear.
3. A geodatabase containing 17 biogenic habitat modelled layers accompanied by uncertainty (standard deviation) for the HGMP area at 250 m x 250 m resolution.

5 Acknowledgements

We thank various scientists for contributions of data and advice from the National Institute of Water and Atmospheric Research (NIWA) including James Williams, Mark Morrison, Dennis Gordon, Di Tracy, Diana MacPherson, Judi Hewitt, Michelle Kelly, Kate Neill, Wendy Nelson. We also thank Clinton Duffy (Department of Conservation) for data included in the database. We thank Ian Tuck (Fisheries New Zealand) for support and various institutions for permission/authorisation and support including Fisheries New Zealand, Te Papa Museum and Auckland Museum. Finally, we thank Tom Brough and Michael Bruce for reviewing this report.

6 References

Lundquist, C.J., Julian, K., Costello, M., Gordon, D., Mackay, K., Mills, S., Neill, K., Nelson, W., Thompson, D. (2014) Development of a Tier 1 National Reporting Statistic for New Zealand's Marine Biodiversity. *New Zealand Aquatic Environment and Biodiversity Report*, 147: 61.

Lundquist, C.J., Stephenson, F., McCartain, L., Watson, S., Brough, T., Nelson, W., Neill, K., Anderson, T., Anderson, O., Bulmer, R., Gee, E., Pinkerton, M., Rowden, A., Thompson, D. (2020a) Evaluating Key Ecological Areas datasets for the New Zealand Marine Environment. *Prepared for Department of Conservation (project DOC19206)*, 2020244HN: 131.

Lundquist, C.J., Tablada, J., Watson, S. (2020b) Evaluation of Biodiversity Protected by Sea Change Tai Timu Tai Pari – Marine Protected Area Proposals. *Prepared for Department of Conservation (project DOC20206)*, 2020244HN: 154.

Morrison, M., Drury, J., Shankar, U., Hill, A. (2002) A broad scale seafloor habitat assessment of the Firth of Thames using acoustic mapping, with associated video and grab sample ground-truthing. *Prepared for the Department of Conservation*, AKL2002-014: 73.

R-Core-Team (2018) R: A language and environment for statistical computing. https://www.R-project.org/.

Robinson, N.M., Nelson, W.A., Costello, M.J., Sutherland, J.E., Lundquist, C.J. (2017) A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice. *Frontiers in Marine Science*, 4: 421.

Stephenson, F., Bowden, D.A., Finucci, B., Anderson, O.F., Rowden, A.A. (2021a) Developing updated predictive models for benthic taxa and communities across Chatham Rise and Campbell Plateau using photographic survey data. *New Zealand Aquatic Environment and Biodiversity*, 276: 82.

Stephenson, F., Brough, T., Lohrer D., Leduc, D., Geange, S., … Lundquist, C., et.al. (submitted) An atlas of seabed biodiversity for Aotearoa New Zealand. *Frontiers in Marine Science*.

Stephenson, F., Goetz, K., Sharp, B.R., Mouton, T.L., Beets, F.L., Roberts, J., MacDiarmid, A.B., Constantine, R., Lundquist, C.J. (2020a) Modelling the spatial distribution of cetaceans in New Zealand waters. *Diversity and Distributions*, 26(4): 495-516.

Stephenson, F., Rowden, A., Brough, T., Leathwick, J., Bulmer, R., Clark, D., Lundquist, C., Greenfield, B., Bowden, D., Tuck, I., Neill, K., Mackay, K., Pinkerton, M., Anderson, O., Gorman, R., Mills, S., Watson, S., Nelson, W., Hewitt, J. (2020b) Development of a New Zealand Seafloor Community Classification (SCC). *Prepared for Department of Conservation* 2020243HN: 84.

Stephenson, F., Rowden, A.A., Anderson, O.F., Pitcher, C.R., Pinkerton, M.H., Petersen, G., Bowden, D.A. (2021b) Presence-only habitat suitability models for vulnerable marine ecosystem indicator taxa in the South Pacific have reached their predictive limit. *ICES Journal of Marine Science*, 78(8): 2830-2843.

Wickham, H. (2011) ggplot2. *Wiley interdisciplinary reviews: computational statistics*, 3(2): 180-185.

Williams, J.R., Parkinson, D.M. (2010) Biomass survey and stock assessment for the Coromandel scallop fishery, 2010. *New Zealand Fisheries Assessment Report 2010/37*: 30.

Williams, J.R., Parkinson, D.M., Bian, R. (2013) Biomass survey and yield calculation for the Coromandel commercial scallop fishery, 2012. *New Zealand Fisheries Assessment Report 2013/18*: 57.

Williams, J.R., Parkinson, D.M., Tuck, I.D. (2010) Biomass survey and stock assessment for the Coromandel scallop fishery, 2009. *New Zealand Fisheries Assessment Report 2010/33*: 40.

Wood, A.L., Probert, P.K., Rowden, A.A., Smith, A.M. (2012) Complex habitat generated by marine bryozoans: a review of its distribution, structure, diversity, threats and conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 22(4): 547-563.