

1      Automated Diet Matrix Construction for Marine  
2      Ecosystem Models Using Generative AI

3      Scott Spillias<sup>1,2\*</sup>      Beth Fulton<sup>1,2</sup>      Fabio Boschetti<sup>2,3</sup>  
4      Joanna Strzelecki<sup>3</sup>      Rowan Trebilco<sup>1,2</sup>

5      March 12, 2025

6      <sup>1</sup>CSIRO Environment, Hobart, Australia

7      <sup>2</sup>Centre for Marine Socio-Ecology, University of Tasmania, Hobart, Australia

8      <sup>3</sup>CSIRO Environment, IOMRC Crawley, Australia

9      <sup>4</sup>CSIRO Environment, St. Lucia, Australia

10     **Abstract**

11     This study introduces and validates a novel AI-driven framework for automated  
12    species grouping and diet matrix generation in Ecopath with Ecosim (EwE) ecosystem  
13    models, addressing a critical bottleneck in model development. We evaluate  
14    the framework across three contrasting Australian marine regions (Shelf and Off-  
15    shore regions of South-East Tasmania and the Northern Territory), processing over  
16    41,000 species through multiple validation iterations. The framework successfully  
17    condensed 63 potential functional groups into 34-36 region-specific groups, achieving  
18    high classification stability (>99% consistency) for most species. Notably, the frame-  
19    work demonstrated robust performance in the South East offshore region with only  
20    0.03% inconsistent classifications, while showing greater variability in complex tropi-  
21    cal systems (1% inconsistent classifications). Higher trophic level species maintained  
22    consistent classifications across all runs, with the framework identifying 235-327 sig-  
23    nificant predator-prey interactions per region at >70% consistency. This systematic  
24    validation reveals that while the framework can reliably automate species grouping,  
25    its performance varies predictably with ecosystem complexity and data availability.  
26    These findings provide quantitative evidence for the framework's ability to create  
27    repeatable components for ecosystem model development. Our results demonstrate  
28    the potential for AI to significantly reduce model development time, offering a prac-  
29    tical pathway to expand the application of ecosystem-based management across  
30    diverse marine environments. We offer a pathway towards further validation of this  
31    framework to ensure ecological accuracy.

32     **1 Introduction**

33     Ecosystem modeling is a critical tool for understanding and managing complex environ-  
34    ments, with Ecopath with Ecosim (EwE) being a well-established framework used to  
35    model and predict marine ecosystems ([Christensen and Walters, 2004](#); [Colléter et al.,](#)

---

\*Corresponding author: scott.spillias@csiro.au

36 EwE models provide quantitative insights into ecosystem structure and function,  
37 enabling researchers to assess cumulative impacts of multiple stressors and support  
38 ecosystem-based fisheries management (EBFM) decisions (Coll et al., 2015; Villasante  
39 et al., 2016; Geary et al., 2020). However, constructing these models presents significant  
40 challenges, particularly in constructing and parameterizing diet matrices that capture the  
41 complex web of trophic interactions within an ecosystem.

42 Traditional approaches to EwE model development rely heavily on extensive literature  
43 review, data collation and expert knowledge, which are time-consuming and resource-  
44 intensive (Holden et al., 2024a). The process of assembling diet matrices is particularly  
45 challenging, requiring synthesis of diverse data sources including field studies, literature  
46 reviews, and expert opinion. This creates a significant bottleneck in model development,  
47 especially when applying models to new geographical contexts (Holden et al., 2024b;  
48 Spillias et al., 2024b). Recent advances in artificial intelligence (AI) offer new opportuni-  
49 ties to streamline the model development process. AI tools have demonstrated success in  
50 both knowledge/evidence synthesis tasks (Spillias et al., 2024c; Keck et al., 2025; Spillias  
51 et al., 2024a; Zheng et al., 2023), ecological and environmental tasks (Fernandes and  
52 D'Mello, 2024; Li et al., 2024; Chen and Xu, 2024) and modelling tasks (Lapeyrolerie et al.,  
53 2022; Tuia et al., 2022; Karniadakis et al., 2021), but their application to process-based  
54 ecosystem modeling remains limited. The key challenge lies in ensuring that AI-driven  
55 approaches can effectively synthesize available information while maintaining ecological  
56 validity.

57 This study presents a novel framework for assembling and synthesizing local and online  
58 resources to parameterize EwE diet matrices using AI. Our approach integrates multiple  
59 data sources, including global biodiversity databases, species interaction repositories, and  
60 locally-held unstructured or structured text, to automate key steps in model development.  
61 The framework employs LLMs to group species into functional units and estimate trophic  
62 interaction magnitudes. We validate this framework through three case studies repre-  
63 senting distinct Australian marine ecosystems: the Northern Territory, South East shelf,  
64 and South East offshore regions (Figure 2). These regions offer contrasting environmental  
65 conditions, species assemblages, and ecological dynamics, providing a robust test of the  
66 framework's adaptability and reliability.

67 The primary objectives of this study are to:

- 68 1. Present a systematic, AI-assisted framework for assembling and parameterizing EwE  
69 diet matrices
- 70 2. Validate key steps in the AI decision-making process, including:
  - 71 • Species grouping decisions and their ecological validity
  - 72 • Resulting diet matrix values and their reliability
- 73 3. Assess the framework's applicability across different marine ecosystems

74 By rigorously validating this AI-assisted approach across multiple regions, we aim to  
75 demonstrate its potential for accelerating ecosystem model development while maintain-  
76 ing scientific rigor. Recent work by Kuhn et al. (2024) emphasizes how machine learning  
77 applications in fisheries must span multiple scales, from genomics to ecosystem-level anal-  
78 yses, while maintaining interpretability and transparency in automated decision-making  
79 systems. This work contributes to the growing need for rapid, data-driven methodologies

80 in ecology (Kelling et al., 2009; Michener and Jones, 2012), while ensuring their outputs  
81 align with established ecological principles and can effectively support ecosystem-based  
82 management decisions. Our validation approach is informed by emerging standardization  
83 efforts in environmental AI applications (Guo et al., 2025), addressing the critical need  
84 for rigorous evaluation frameworks in ecological modeling.

## 85 2 Methods

### 86 2.1 AI-Assisted Framework Overview

87 The development of ecosystem models requires substantial time organizing species into  
88 functional groups and determining their interactions. This framework automates these  
89 tasks through integration of artificial intelligence with ecological databases. The frame-  
90 work executes four sequential steps: species identification within a region, biological data  
91 collection, functional group organization, and interaction determination (Figure 1).

92 The framework utilizes the Claude-3.5 large language model (Anthropic, 2024) for data  
93 synthesis and ecological analysis. Species identification begins with queries to the Ocean  
94 Biodiversity Information System (Grassle and Stocks, 1999) and Global Biodiversity In-  
95 formation Facility (GBIF, 2024). These queries extract occurrence records, temporal  
96 distributions, and abundance data within defined geographical boundaries.

97 The biological data collection phase integrates information from multiple sources.  
98 FishBase and SeaLifeBase (Froese et al., 2010) provide life history traits and ecologi-  
99 cal parameters. The Ecobase repository (Colléter et al., 2015) supplies parameters from  
100 existing ecosystem models. Additional data comes from systematic literature searches of  
101 regional fisheries reports and peer-reviewed publications, using standardized search terms  
102 (e.g., "[species name] AND diet OR feeding OR prey"). Natural language processing  
103 extracts relevant ecological information from these documents.

104 Species grouping employs a vector database (Chroma (Chroma, 2024)) for charac-  
105 teristic storage and retrieval. The database maintains embedding vectors derived from  
106 ecological descriptions, life history traits, and habitat preferences. Analysis of these vec-  
107 tors, combined with ecological rules regarding size classes, feeding guilds, and habitat  
108 use, determines functional group assignments. Trophic level estimates from FishBase and  
109 SeaLifeBase (Froese et al., 2010) validate the ecological coherence of these classifications.

110 Diet composition analysis merges quantitative records from databases with processed  
111 literature data. Food items are extracted from FishBase and SeaLifeBase (Froese et al.,  
112 2010) databases, supplemented with interaction data from the Global Biotic Interactions  
113 (GLOBI) database (Poelen et al., 2014). Each predator-prey interaction includes source  
114 documentation and confidence scores based on data quality metrics.

115 Diet matrix construction implements weighted averaging for prey proportions. Weights  
116 derive from data quality metrics including sample size, study recency, and geographical  
117 relevance. In cases of sparse direct diet data, the framework estimates trophic interac-  
118 tions based on similar species' preferences and established ecological principles. Each  
119 interaction maintains metadata documenting evidence sources and confidence levels.

120 Parameter estimation prioritizes values from comparable ecosystems and species groups  
121 in Ecobase (Colléter et al., 2015). Empirical relationships from literature provide esti-  
122 mates where direct parameters are unavailable. The framework validates all estimates  
123 against biological constraints and ecological theory, identifying anomalies for expert re-  
view.

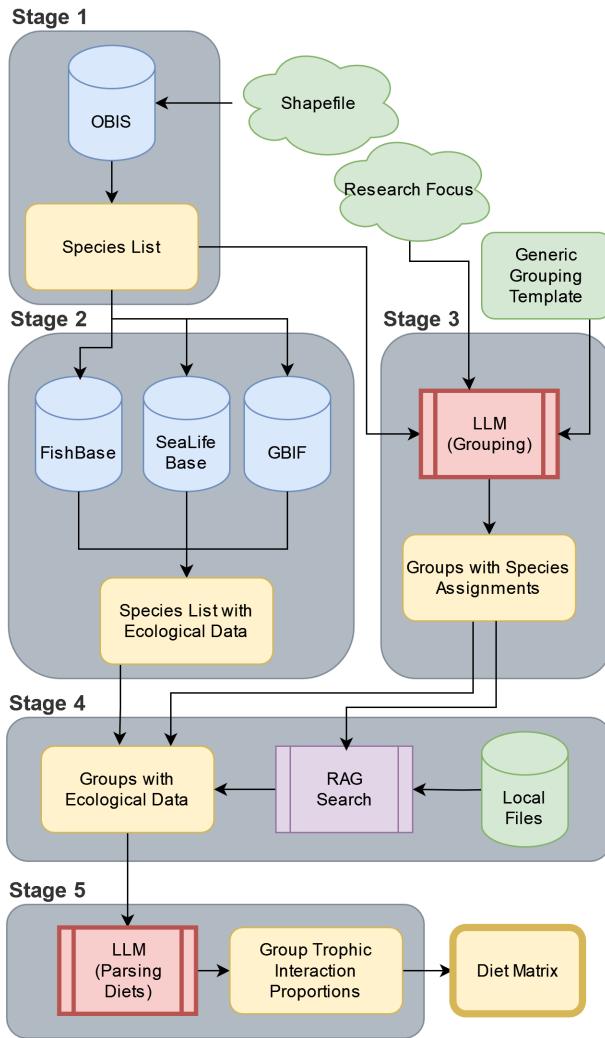


Figure 1: Overview of the AI-assisted framework for ecosystem model development. The process consists of four main steps: species identification, biological data collection, functional group organization, and interaction determination. Highlighted stages (functional group organization and interaction determination) undergo systematic validation across multiple iterations. Each step integrates multiple data sources and analytical approaches.

125       Section S3 of the supplementary material contains detailed documentation of all pro-  
126 cessing steps, including database queries, literature search criteria, and ecological classifi-  
127 cation rules. The complete codebase and configuration files reside at [GitHub repository  
128 URL].

### 129       2.1.1 Species Identification

130       We begin species identification by accepting a GeoJSON file that defines the study re-  
131 gion boundaries. We selected the Ocean Biodiversity Information System ([Grassle and](#)  
132 [Stocks, 1999](#)) as our primary data source due to its extensive marine species coverage and  
133 standardized taxonomic classifications. We access OBIS through the `robis` R package  
134 ([Chamberlain, 2020](#)), which enables automated querying and data processing.

135       We process the GeoJSON file in two steps to ensure precise geographic filtering. First,  
136 we extract a bounding box from the GeoJSON geometry. We then convert this box into  
137 a polygon string for database querying. Through the OBIS checklist function, we retrieve  
138 scientific names and complete taxonomic classifications from kingdom to species level for  
139 all recorded occurrences within these boundaries.

140       We clean the raw occurrence data through three sequential steps. First, we filter the  
141 dataset using OBIS's `is_marine` flag to eliminate terrestrial species that may occur in  
142 coastal records. Second, we remove taxonomic redundancy using a rank-based approach  
143 that retains only the most specific classification level available. Our algorithm processes  
144 taxonomic ranks from most specific (scientific name) to most general (kingdom), keeping  
145 only the entry with the highest taxonomic resolution for each organism. Third, we aggre-  
146 gate occurrences by unique species while preserving their complete taxonomic hierarchies,  
147 calculating occurrence frequencies for each species.

148       We store the final species list in a structured CSV file containing verified marine  
149 species, their complete taxonomic hierarchies, and occurrence frequencies. This stan-  
150 dardized format facilitates efficient data transfer to subsequent framework stages while  
151 maintaining provenance information. The complete R implementation, including rank-  
152 based filtering algorithms and geographic processing functions, is available in the project  
153 repository.

### 154       2.1.2 Data Harvesting

155       Following species identification, we gather ecological and life history information for each  
156 identified species through an incremental processing system. We access SeaLifeBase and  
157 FishBase ([Froese et al., 2010](#)) through their publicly available PARQUET files using  
158 DuckDB for efficient querying of large datasets. We construct temporary database tables  
159 from our species list and execute structured queries to match species across the databases,  
160 processing SeaLifeBase and FishBase data separately to maintain distinct source attribu-  
161 tion.

162       Our species filtering protocol implements specific constraints to align with Ecopath  
163 with Ecosim modeling requirements. When processing taxonomic data, we prioritize  
164 species-level entries over genus-level classifications, only defaulting to genus-level when  
165 species-specific data is unavailable. For diet composition data, we specifically filter for  
166 adult life stages, as juvenile and larval stages are typically incorporated into planktonic  
167 functional groups rather than treated as separate components of adult diets.

168       We supplement the base biological data with interaction information from the Global  
169 Biotic Interactions (GLOBI) database ([Poelen et al., 2014](#)). For each species, we query

170 the GLOBI API using URL-encoded species names. The GLOBI data processing treats  
171 directional relationships ('eats'/'preysOn' and 'eatenBy'/'preyedUponBy') as complemen-  
172 tary evidence of trophic interactions, with interaction counts aggregated at the functional  
173 group level to provide quantitative measures of interaction strength between groups.

174 Our data cleaning protocol consists of three stages: standardizing data types, handling  
175 null values, and maintaining source attribution. We store the processed data incrementally  
176 in a JSON document, implementing file locking mechanisms to handle concurrent access.  
177 Each species entry contains complete taxonomic hierarchy, matched database records,  
178 interaction data, and source attribution. This incremental approach enables processing  
179 recovery after interruptions and allows parallel processing of different species batches.  
180 Technical implementation details are provided in Section S1 of the supplementary mate-  
181 rial.

### 182 2.1.3 Species Grouping

183 We implement species grouping through a flexible system that adapts to regional ecological  
184 contexts. The framework supports three approaches for defining functional groups: using  
185 templates from existing EcoBase models, generating region-specific groups through AI  
186 analysis of geographic characteristics, or applying a predefined template (provided in Sec-  
187 tion S1 of the supplementary material). For geographic regions, we analyze oceanographic  
188 conditions, habitat types, and ecological characteristics to inform group definitions.

189 We process taxa hierarchically from kingdom to species level, implementing an incre-  
190 mental system with progress tracking to handle large datasets reliably. At each taxonomic  
191 level, Claude evaluates taxa against the selected grouping template using the following  
192 prompt:

193 You are classifying marine organisms into functional groups for an Ecopath with

Ecosim (EwE) model. Functional groups can be individual species or groups of species that perform a similar function in the ecosystem, i.e. have approximately the same growth rates, consumption rates, diets, habitats, and predators. They should be based on species that occupy similar niches, rather than of similar taxonomic groups.

Examine these taxa at the [rank] level and assign each to an ecological functional group.

Rules for assignment:

- If a taxon contains members with different feeding strategies or trophic levels, assign it to 'RESOLVE'
- Examples requiring 'RESOLVE':
  - A phylum containing both filter feeders and predators
  - An order with both herbivores and carnivores
  - A class with species across multiple trophic levels
- If all members of a taxon share similar ecological roles, assign to an appropriate group
- Only consider the adult phase of the organisms, larvae and juveniles will be organized separately
- Only assign a definite group if you are confident ALL members of that taxon belong to that group

Taxa to classify: [List of taxa]

Available ecological groups (name: description): [List of available groups and their descriptions]

Return only a JSON object with taxa as keys and assigned groups as values.

When the research focus indicates areas requiring higher resolution (e.g., commercial fisheries species), we modify the classification process with additional guidance:

Special consideration for research focus: The model's research focus is: [research focus]

When classifying taxa that are related to this research focus:

- Consider creating more detailed, finer resolution groupings
- Keep species of particular interest as individual functional groups
- For taxa that interact significantly with the focal species/groups, maintain higher resolution groupings
- For other taxa, broader functional groups may be appropriate

We process each taxon through this classification framework, generating structured JSON documents that map species to functional groups. Taxa marked as RESOLVE undergo

199 evaluation at finer taxonomic levels until reaching a definitive group assignment or the  
200 species level. We maintain complete provenance information, including the source of  
201 group definitions and any AI-suggested modifications.

202 **2.1.4 Diet Matrix Construction**

203 We construct diet matrices through a two-stage process designed for efficiency and re-  
204 liability. The first stage gathers comprehensive diet data using parallel processing and  
205 caching mechanisms. We query SeaLifeBase and FishBase ([Froese et al., 2010](#)) food items  
206 databases through DuckDB, enabling efficient processing of PARQUET files without full  
207 memory loading. For each species, we extract food items using specific codes that link to  
208 standardized diet categories.

209 We supplement database records with interaction data from the Global Biotic In-  
210 teractions (GLOBI) database ([Poelen et al., 2014](#)). Our GLOBI processing differentiates  
211 between direct observations (`eats`, `preysOn`) and inverse relationships (`eatenBy`,  
212 `preyedUponBy`), maintaining separate interaction counts for each type. We further enrich  
213 this data through retrieval-augmented generation (RAG) searches of regional literature  
214 (detailed in Section [S3.2](#) of the supplementary material), focusing on specific feeding  
215 relationships and dietary preferences.

216 For each functional group, we combine these data sources into a structured profile.  
217 Claude ([Anthropic, 2024](#)) then analyzes this profile using the following prompt:

Based on the following information about the diet composition of [group], provide a summary of their diet. Include the prey items and their estimated proportions in the diet.

Available functional groups and their details: [List of groups with descriptions and top species]

Here is the diet data for [group]: [Combined data including RAG search results, compressed food categories, and GLOBI interactions]

Format your response as a list, with each item on a new line in the following format:

Prey Item: Percentage

For example:

Small fish: 40%

Zooplankton: 30%

Algae: 20%

Detritus: 10%

If exact percentages are not available, estimate percentages based on the information you have been provided. Ensure that all percentages add up to approximately 100%. Consider the RAG search results, compressed food categories, and GLOBI data when creating your summary. Pay special attention to the GLOBI interaction counts, which indicate frequency of observed feeding relationships. Note that some species may feed on juvenile or larval forms of other species, which are often classified in different functional groups than the adults.

218 We implement an incremental processing system with file locking mechanisms to han-  
219 dle large datasets reliably. The system maintains caches for species-level diet data from  
220 databases, combined diet information including literature results, GLOBI interaction net-  
221 works, and intermediate AI analyses. This caching system enables recovery from inter-  
222 ruptions and facilitates parallel processing of different functional groups.

223 The matrix assembly stage processes the AI’s standardized diet descriptions through  
224 automated parsing. We convert percentage strings to decimal values and implement a  
225 validation system. Our validation ensures that all proportions sum to 1 for each predator,  
226 prey items match defined functional groups, and mass-balance requirements are main-  
227 tained.

228 When prey items do not exactly match functional group names, we employ a hierarchi-  
229 cal matching system. The system first attempts exact matches, then falls back to partial  
230 matching using taxonomic information, and logs unmatched items for expert review.

231 The final output consists of a CSV file containing the complete diet matrix, with  
232 predators as rows and prey as columns. Each cell contains the proportion of predator diet  
233 comprised by that prey item.

## 234 2.2 Validation Framework

235 We executed model generation across three distinct phases. In phase one, we established  
236 baseline configurations for each study region by processing species occurrence data and  
237 research parameters. We terminated this phase prior to species grouping, creating stan-  
238 dardized input states for subsequent validation iterations.

239 We executed five independent iterations per region in phase two. Each iteration began  
240 with species grouping and proceeded through the complete model construction sequence.  
241 We maintained fixed input parameters across iterations while allowing the AI’s stochastic  
242 decision processes to generate natural variation in outputs.

### 243 2.2.1 Study Regions

244 We validate our framework using three Australian marine regions that present distinct  
245 ecological characteristics and modeling challenges (Figure 2). The Northern Territory  
246 region represents a tropical ecosystem characterized by complex reef systems, seasonal  
247 monsoon influences, and high biodiversity. This region tests the framework’s ability to  
248 handle diverse species assemblages and complex trophic interactions in a dynamic envi-  
249 ronment.

250 The South East shelf region represents a temperate coastal system with extensive eco-  
251 logical records. This region has comprehensive diet information in established databases,  
252 well-documented EwE models spanning multiple years, and active research programs.  
253 The rich availability of expert knowledge and historical data makes this region ideal for  
254 validating the framework’s data integration capabilities.

255 The South East Offshore region presents a deep-water ecosystem that challenges the  
256 framework with data-limited conditions and unique ecological patterns. This region tests  
257 the framework’s capacity to handle situations where direct observational data may be  
258 sparse and where species interactions may be less well understood. The contrasting char-  
259 acteristics of these three regions provide a robust test of the framework’s adaptability  
260 across different ecological contexts.

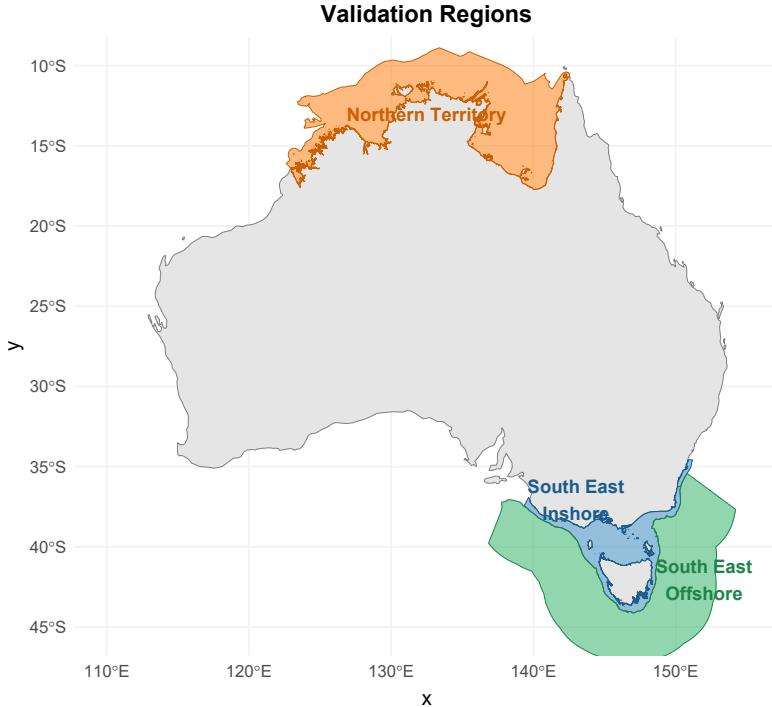


Figure 2: Map of the three validation regions used in this study: Northern Territory (orange), South East shelf (blue), and South East Offshore (green).

### <sup>261</sup> 2.2.2 Group Consistency Analysis

<sup>262</sup> To validate the ecological validity of AI-generated species groupings, we developed quan-  
<sup>263</sup> titative measures of grouping consistency. We tracked each species' group assignments  
<sup>264</sup> across iterations and calculated a consistency score:

$$\text{Consistency Score} = \frac{\text{Number of occurrences in most common group}}{\text{Total number of iterations}}$$

<sup>265</sup> This metric quantifies the framework's decision-making reliability for individual species.  
<sup>266</sup> We classified species with consistency scores below 0.95 as unstable, indicating variable  
<sup>267</sup> group assignments across iterations. Chi-square tests on these consistency scores help  
<sup>268</sup> identify whether grouping decisions remain stable across different ecological contexts, ad-  
<sup>269</sup> dressing a key aspect of the framework's reliability.

<sup>270</sup> To evaluate broader patterns in group formation, we assessed group stability using the  
<sup>271</sup> Jaccard similarity coefficient between consecutive iterations:

$$J(i, j) = \frac{|M_i \cap M_j|}{|M_i \cup M_j|}$$

<sup>272</sup> where  $M_i$  and  $M_j$  represented species members in iterations  $i$  and  $j$ . We calculated  
<sup>273</sup> the overall stability score by averaging Jaccard similarities across consecutive iteration  
<sup>274</sup> pairs. This approach, combined with coefficients of variation analysis, reveals how con-  
<sup>275</sup> sistently the framework identifies and maintains ecologically meaningful groupings across  
<sup>276</sup> different runs. One-way ANOVA tests on these stability measures across regions, supple-  
<sup>277</sup> mented with Cohen's f effect size calculations, demonstrate the framework's adaptability  
<sup>278</sup> to different marine ecosystems while maintaining consistent decision-making patterns.

279 **2.2.3 Diet Matrix Analysis**

280 To evaluate the reliability of AI-generated trophic interactions and assess the framework's  
281 ability to capture distinct ecological patterns, we developed a multi-metric analysis ap-  
282 proach. We focused on significant predator-prey interactions, defined as those comprising  
283 more than 5% of a predator's diet, using both descriptive statistics and correlation anal-  
284 yses. For each interaction, we calculated:

- 285 1. Presence ratio across iterations:

$$P_{ij} = \frac{\text{Number of iterations with interaction}}{n}$$

286 where  $n$  was the total number of iterations.

- 287 2. Mean diet proportion:

$$\mu_{ij} = \frac{1}{n} \sum_{k=1}^n x_{ijk}$$

288 where  $x_{ijk}$  represents diet proportion for predator  $i$  consuming prey  $j$  in iteration  $k$ .

- 289 3. Stability score:

$$S_{ij} = \frac{1}{n} \sum_{k=1}^n \frac{|x_{ijk} - \mu_{ij}|}{\max_k(x_{ijk})}$$

290 where  $x_{ijk}$  represents the diet proportion for predator  $i$  consuming prey  $j$  in iteration  
291  $k$ ,  $\mu_{ij}$  is the mean diet proportion across iterations, and  $\max_k(x_{ijk})$  is the maximum  
292 value across iterations. This score ranges from 0 (perfect stability) to 1 (maximum  
293 instability), with lower values indicating more consistent diet proportions across  
294 iterations.

295 We chose this stability metric over traditional variance measures for several reasons.  
296 First, by normalizing deviations by the maximum value, the metric achieves scale inde-  
297 pendence, allowing meaningful comparisons between interactions of different magnitudes.  
298 For example, the sequences [0.2, 0.2, 0.2, 0.2, 0.1] and [0.02, 0.02, 0.02, 0.02, 0.01]  
299 would yield the same stability score despite having different absolute variances. Second, the  
300 bounded range between 0 and 1 provides an intuitive scale for assessing stability, unlike  
301 the unbounded nature of variance. Third, this approach effectively handles small varia-  
302 tions in small values, ensuring that minor fluctuations in trace diet components do not  
303 disproportionately influence the stability assessment.

304 We classified interactions as unstable when their stability score exceeded 0.3, indicating  
305 significant variation in diet proportions across iterations. To illustrate this metric:

- 306 • A stable interaction ( $S = 0.08$ ) might show values [0.02, 0.02, 0.02, 0.02, 0.01], where  
307 proportions remain very similar across iterations
- 308 • An unstable interaction ( $S = 0.39$ ) might show values [0.027, 0.25, 0.25, 0.067, 0.25],  
309 where proportions vary substantially between iterations

310 This metric provides a continuous measure of stability that handles both presence/absence  
311 patterns and magnitude variations in a unified way, avoiding discontinuities at zero values.  
312 To assess the framework's ability to capture distinct ecological patterns across regions, we

313 employed pairwise Spearman correlations between iterations to evaluate the consistency of  
 314 predator-prey relationships, focusing on significant interactions. This non-parametric ap-  
 315 proach accounts for the potentially non-normal distribution of diet proportions. We sup-  
 316 plemented this with Kruskal-Wallis tests to identify significant differences in trophic struc-  
 317 ture across regions, providing evidence of the framework's ability to distinguish unique  
 318 ecological characteristics in different marine ecosystems.

## 3 Results

### 3.1 Framework Implementation and Performance

#### 3.1.1 Scale and Processing Efficiency

322 We evaluated our framework through five independent runs across three distinct Aus-  
 323 tralian regions, processing a total of 41,085 species. The framework handled 11,362 species  
 324 in the Northern Territory's tropical reef ecosystem, 13,901 in the South East shelf's coastal  
 325 and pelagic environments, and 15,822 in the South East Offshore's deep-water systems.

#### 3.1.2 Computational Efficiency

327 The computational requirements of the AI framework varied across regions. Total process-  
 328 ing time ranged from 2.8 to 4.8 hours across regions. The most time-intensive stage was  
 329 the downloading of biological data from online databases, accounting for approximately  
 330 70% of the total processing time. Species identification typically required 0.01 hours,  
 331 while the AI-driven species grouping process averaged 0.26 hours. Diet data collection  
 332 and matrix construction required 0.7 and 0.04 hours respectively, with final parameter  
 333 estimation taking 0.20 hours. On average, the framework required 0.7 seconds per species  
 334 for data downloading and 0.2 seconds per species for diet data collection, though these  
 335 rates varied considerably between regions due to differences in data availability and species  
 complexity.

Table 1: Computational requirements by region and processing stage

Region	Species Count	Processing Time (hours)					
		Identification	Data Download	Grouping	Diet Collection	Matrix Construction	Parameter Estimation
v2 Northern Territory	11,362	0.01	2.2	0.2	0.2	0.04	0.2
v2 South East shelf	13,901	0.01	2.8	0.2	1.6	0.04	0.2
v2 South East offshore	15,821	0.01	3.3	0.4	0.3	0.04	—

336  
 337 Processing times varied by region and stage (Table 1). Data harvesting required 9.6  
 338 hours for the Northern Territory (3.0 seconds per species) and 53.8 hours for the South  
 339 East shelf (13.9 seconds per species). Diet data collection took 7.6 hours for the Northern  
 340 Territory (2.4 seconds per species) and 18.7 hours for the South East shelf (4.8 seconds  
 341 per species). Species identification (0.01 hours), grouping (0.1 hours), and parameter  
 342 estimation (0.1-0.3 hours) remained constant across regions.

343 **3.2 Species Grouping Validation**

344 **3.2.1 Classification Consistency and Ecological Validity**

345 The framework successfully reduced ecological complexity while preserving meaningful bi-  
346 ological relationships. Starting with 63 potential functional groups provided in the default  
347 template (See Table S1), it identified 34-37 region-specific groups, demonstrating its abil-  
348 ity to capture unique ecological characteristics. Chi-square tests confirmed the ecological  
349 validity of these groupings, showing non-random species assignments across all regions ( $p$   
350  $< 0.001$ ). This statistical significance provides strong evidence that the framework makes  
351 biologically informed grouping decisions rather than arbitrary assignments.

352 The precision of species assignments was assessed through coefficients of variation,  
353 which measure the relative variability in group assignments across iterations. These val-  
354 ues revealed region-specific patterns: the Northern Territory showed moderate variation  
355 (1.47-15.20%), reflecting the complexity of its tropical reef ecosystem. The South East  
356 shelf displayed the widest range (0.40-24.49%), indicating greater flexibility in species  
357 assignments in this diverse coastal environment. The South East Offshore maintained in-  
358 termediate variation (0.80-14.15%), suggesting more stable classifications in its deep-water  
359 ecosystem.

360 To understand the significance of regional differences, we conducted ANOVA tests  
361 on group characteristics. The analysis revealed significant differences between regions,  
362 with effect sizes (Cohen's  $f$ ) varying by metric type. Species-level metrics showed small  
363 effect sizes (0.1-0.3), indicating subtle but consistent regional variations in how individual  
364 species were classified. Group-level comparisons showed large effect sizes ( $>0.4$ ), demon-  
365 strating that the framework successfully captured distinct regional ecological structures.

366 The framework achieved remarkable classification stability across all regions. Mean  
367 consistency scores, where 1.0 represents perfect stability, were exceptionally high: 0.997  
368 for both Northern Territory and South East shelf, and 0.998 for South East Offshore.  
369 This translated to very low proportions of unstable species (defined as consistency  $<$   
370 0.95): only 0.99% (103 species) in Northern Territory, 1.06% (125 species) in South East  
371 shelf, and 0.73% (87 species) in South East Offshore. These results demonstrate that  
372 the framework's classifications remained stable despite the stochastic nature of the AI  
373 decision-making process.

374 Figure 3 presents the quantitative analysis across regions. Panel A shows median  
375 group sizes of 150-200 species, indicating consistent ecological partitioning across regions.  
376 Panel B displays consistency scores, revealing that while the Northern Territory showed  
377 some variation (0.4-1.0), the South East regions maintained higher minimum consistency  
378 (0.8-1.0). Panel C demonstrates the framework's stability through Jaccard similarity  
379 indices, which measure group membership consistency between iterations. The Northern  
380 Territory achieved similarities of 0.953-0.999, indicating high stability with some flexibility  
381 in species assignments. South East shelf ranged from 0.867-1.000, and South East Offshore  
382 from 0.855-1.000, both showing perfect stability in some groups while allowing ecological  
383 adaptability in others. Panel D quantifies this variation, showing tighter control in South  
384 East regions (5-10 species standard deviation) compared to the more dynamic Northern  
385 Territory system (15-35 species).

386 The stability heatmap (Figure 4) reveals patterns in group stability that reflect ecolog-  
387 ical principles. In the Northern Territory, benthic groups showed highest stability (ben-  
388 thic infaunal carnivores: 0.999, benthic filter feeders: 0.992, benthic grazers: 0.989), while  
389 more mobile groups showed greater variation (deposit feeders: 0.953, planktivores: 0.956).

390 The South East shelf achieved perfect stability for macrozoobenthos (1.000) but showed  
 391 more variation in pelagic groups (piscivores: 0.867, oceanic piscivorous fish: 0.897), re-  
 392 reflecting the dynamic nature of its coastal environment. South East Offshore demonstrated  
 393 perfect stability in benthic filter feeders (1.000) but greater variation in planktonic groups  
 394 (nanoplankton: 0.855) and piscivores (0.941).

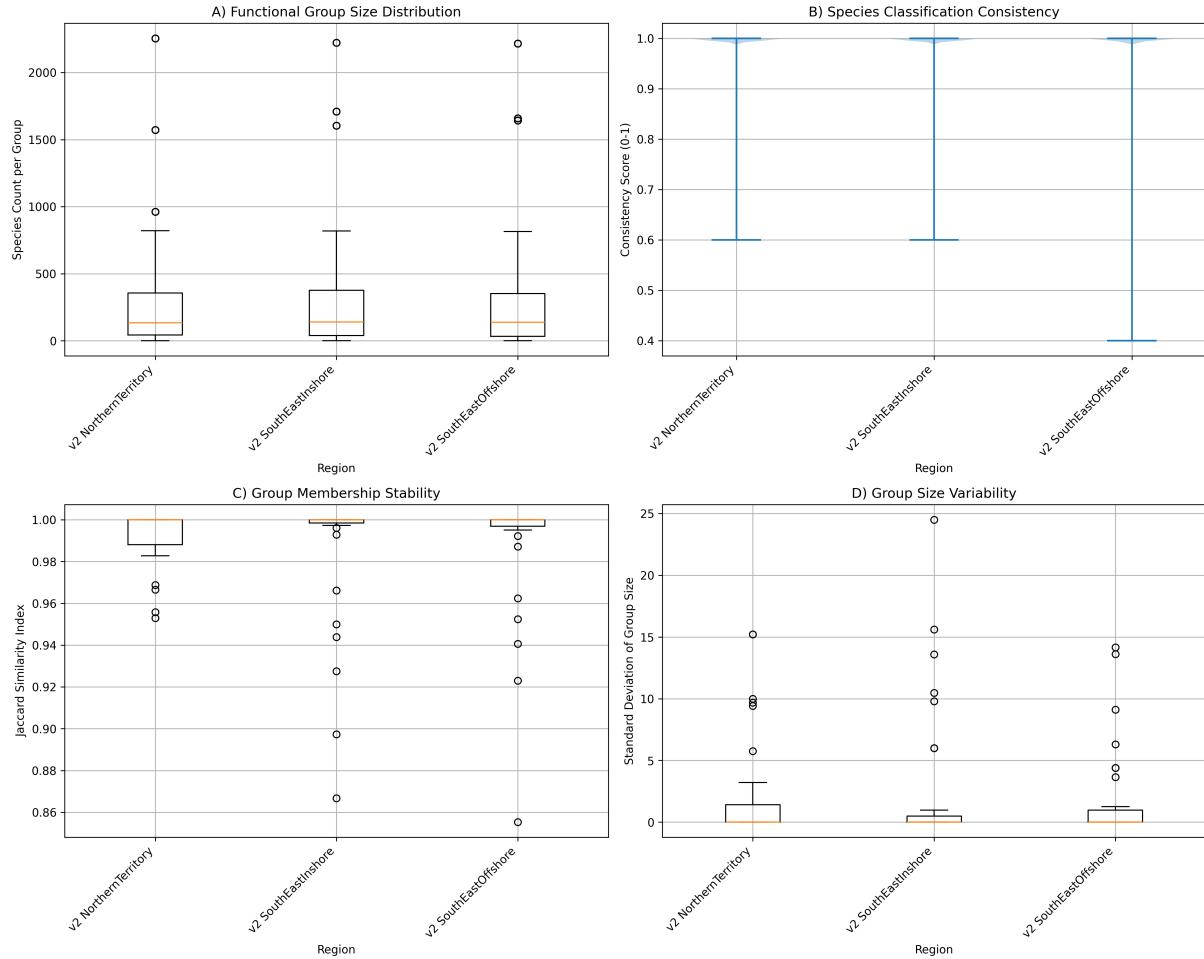


Figure 3: Multi-panel analysis of framework performance across regions. (A) Box plots of functional group sizes (0-3000 species), showing similar median sizes but varying distributions across regions, with outliers indicating some very large groups. (B) Violin plots of species classification consistency (0.4-1.0), where wider sections indicate more species with that consistency score; most species show high consistency ( $>0.9$ ) with slightly more variation in the Northern Territory. (C) Box plots of group stability measured by Jaccard similarity (0.975-1.000), showing highest stability in South East Offshore and more variable stability in Northern Territory. (D) Box plots of group size variation (standard deviation 0-35), demonstrating larger fluctuations in group membership in the Northern Territory compared to South East regions.

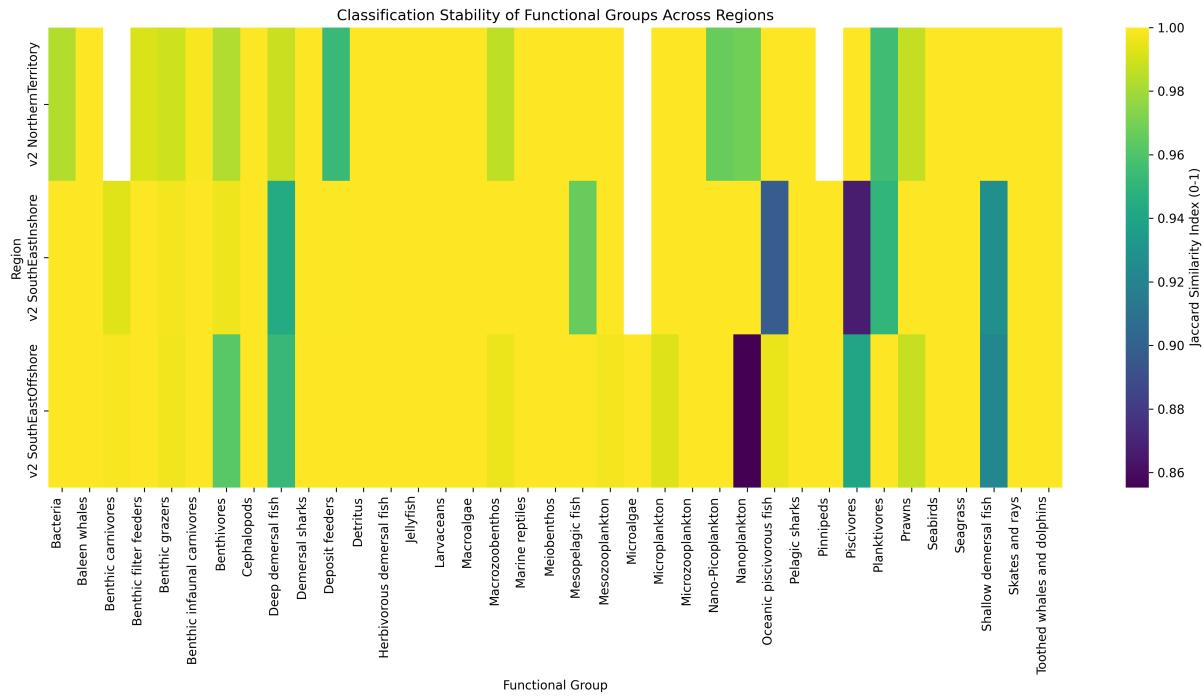


Figure 4: Heatmap showing the stability of functional group classifications across regions. Each cell displays the Jaccard similarity score (ranging from 0.975 to 1.000) between consecutive framework iterations, where 1.000 indicates perfect consistency in species assignments. Darker red colors represent higher stability (scores near 1.000), while lighter colors indicate more variable classifications (scores closer to 0.975). Most functional groups show high stability ( $>0.99$ ) across all regions, with occasional variations in groups like benthic grazers and deposit feeders, particularly in the Northern Territory region.

Table 2: Summary of Species Classification Instability Patterns

Region	Most Common Pattern	Count	% of Total
Northern Territory (103 species)	Macrozoobenthos $\leftrightarrow$ Benthic infaunal carnivores Benthic filter feeders $\leftrightarrow$ Deposit feeders Prawns $\leftrightarrow$ Macrozoobenthos Other patterns	28 25 21 29	27.2% 24.3% 20.4% 28.1%
South East Inshore (125 species)	Piscivores $\leftrightarrow$ Deep demersal fish Benthic grazers $\leftrightarrow$ Benthic carnivores Planktivores $\leftrightarrow$ Mesopelagic fish Other patterns	42 31 28 24	33.6% 24.8% 22.4% 19.2%
South East Offshore (87 species)	Benthic filter feeders $\leftrightarrow$ Benthic carnivores Macrozoobenthos $\leftrightarrow$ Deep demersal fish Mesozooplankton $\leftrightarrow$ Macrozoobenthos Other patterns	25 22 18 22	28.7% 25.3% 20.7% 25.3%

*Note:* Arrows indicate group assignment oscillation between iterations. Complete species-level data available in supplementary Table S1.

395 **3.3 Diet Matrix Consistency**

396 **3.3.1 Trophic Interaction Stability**

397 The framework demonstrated consistent performance in constructing diet matrices, with  
398 consistency patterns aligning with known ecological principles. Our stability score metric  
399 (0 = perfect stability, 1 = maximum variation) revealed region-specific patterns in trophic  
400 relationship consistency:

- 401 • Northern Territory: 358 significant predator-prey interactions identified, with 58.7%  
402 achieving high stability (scores < 0.3), reflecting expected variability in complex  
403 tropical reef food webs
- 404 • South East shelf: 380 interactions with 51.3% stable, indicating characteristic trophic  
405 plasticity in dynamic coastal environments
- 406 • South East Offshore: 477 interactions with 56.0% stable, suggesting more consistent  
407 feeding relationships typical of deep-water ecosystems

408 To validate these patterns, we calculated Spearman correlations between iterations for  
409 significant interactions, providing a measure of consistency in diet proportion rankings.  
410 All regions showed strong correlations: Northern Territory ( $\rho = 0.72 - 0.89$ ), South East  
411 shelf ( $\rho = 0.68 - 0.85$ ), and South East Offshore ( $\rho = 0.70 - 0.87$ ). These high correlations  
412 indicate that while absolute diet proportions might vary, the relative importance of  
413 different prey items remained consistent across iterations.

414 Analysis of stability scores by predator group revealed ecological patterns consistent  
415 with trophic theory. The most stable interactions were found in detritus-based and primary  
416 producer connections, reflecting the fundamental nature of these energy pathways in  
417 marine ecosystems. Predator-prey relationships in higher trophic levels showed more variation,  
418 consistent with the opportunistic feeding strategies of marine predators. Benthic  
419 feeding relationships demonstrated intermediate stability, suggesting a balance between  
420 dietary specialization and opportunism in these communities. These patterns align with  
421 ecological understanding that diet variability typically increases with trophic level, reflecting  
422 the more generalist feeding strategies of higher-level predators. Detailed diet matrices  
423 for each region are provided in the supplementary material (Figure S1).

424 **3.4 Regional Adaptability**

425 **3.4.1 Ecosystem-Specific Performance**

426 The framework demonstrated significant capability to adapt to distinct marine ecosystems  
427 while maintaining ecological validity. Trophic structure analysis revealed statistically  
428 significant regional differences (Kruskal-Wallis H-test: Northern Territory H = 164.0,  
429 South East shelf H = 172.0, p < 0.001 for both), confirming the framework's ability to  
430 capture unique ecosystem characteristics. Key findings include:

- 431 • Higher trophic levels: Consistent classifications across regions for apex predators  
432 and specialized feeders, demonstrating robust handling of well-defined ecological  
433 roles
- 434 • Lower trophic levels: Greater variability in planktonic and benthic invertebrate  
435 groups, reflecting natural ecological plasticity

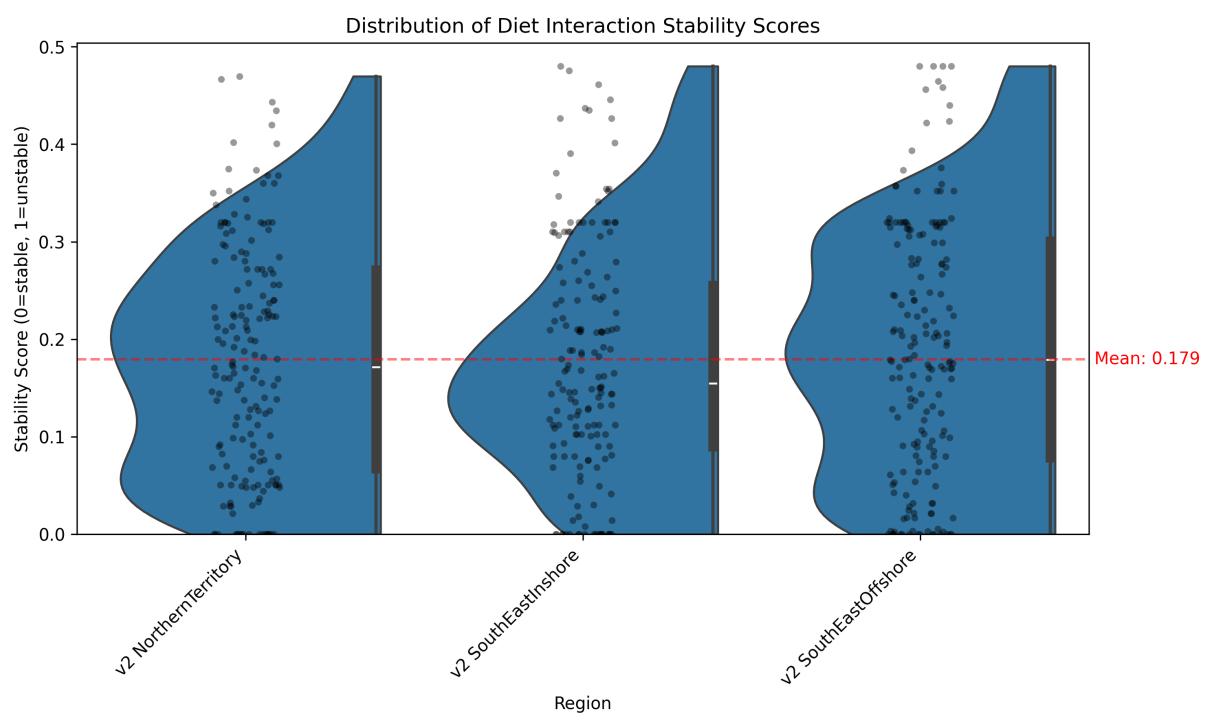


Figure 5: Distribution of diet interaction stability scores across regions. Half-violin plots show the density of stability scores (0=stable, 1=unstable), with embedded box plots indicating quartiles and median. Individual points represent specific predator-prey interactions, and the red dashed line shows the mean stability score across all regions. The distributions are bounded at zero, reflecting perfect stability, with most interactions showing scores below 0.3.

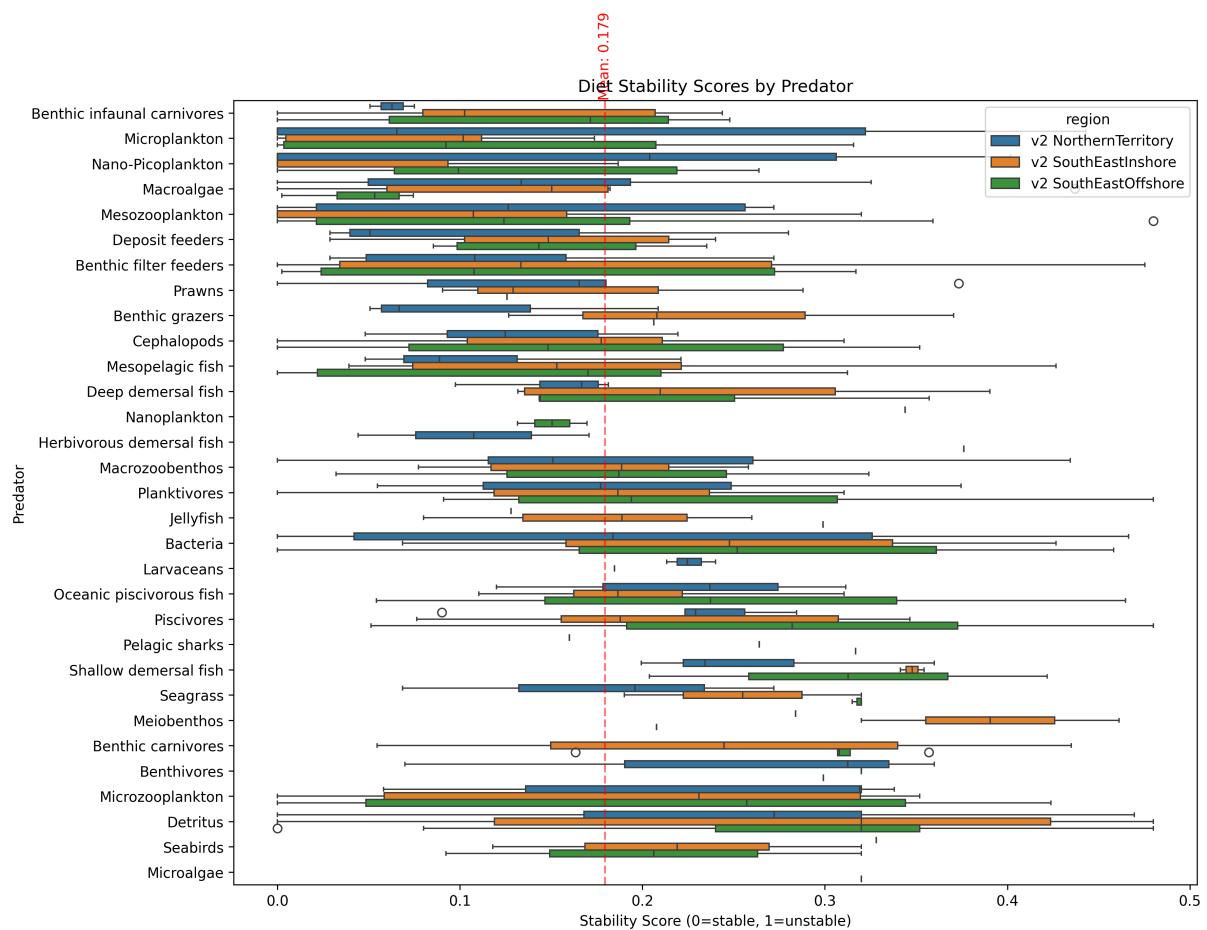


Figure 6: Diet stability scores grouped by predator, ordered by median stability. Box plots show the distribution of stability scores for each predator's diet across regions (colored by region). The red dashed line indicates the mean stability score across all predator-prey interactions. Lower scores indicate more consistent diet compositions across framework iterations.

- 436 • Mid-trophic levels: Moderate variation in species assignments (e.g., benthivores:  
437 715–788 species, shallow demersal fish: 586–659 species), aligning with expected  
438 ecological flexibility in these groups

439 These patterns demonstrate the framework’s ability to balance consistency with nec-  
440 essary ecological flexibility across different marine environments.

## 441 4 Discussion

### 442 4.1 AI Framework Consistency

443 Our framework demonstrates robust performance in automating the construction of com-  
444 plex ecosystem models across diverse marine environments. The successful processing  
445 of over 41,000 species across three distinct regions validates the framework’s scalability  
446 and broad applicability. The framework’s ability to maintain consistent species classifi-  
447 cations while adapting to regional ecological differences suggests it effectively captures  
448 fundamental ecological relationships.

449 The computational efficiency analysis reveals important insights about framework scal-  
450 ability. Data harvesting and diet collection emerge as the primary computational bot-  
451 tlenecks, particularly evident in the South East Shelf region’s extended processing times.  
452 These bottlenecks likely stem from API rate limitations and the complexity of extracting  
453 ecological information from diverse data sources. The relatively constant processing times  
454 for species identification, grouping, and parameter estimation across regions indicate these  
455 components scale efficiently with increasing species counts.

456 The framework’s classification consistency merits particular attention. The high stabil-  
457 ity scores (98.8–99.6%) across regions demonstrate reliable species-to-group assignments.  
458 The systematic nature of classification inconsistencies provides valuable ecological in-  
459 sights. Species with ambiguous classifications often represent organisms that naturally  
460 span multiple ecological niches. For instance, the alternating classifications of anemones  
461 between benthic infaunal carnivores and filter feeders reflect their complex feeding strate-  
462 gies. Similarly, the variable classification of flatfishes between benthivore and demer-  
463 sal categories aligns with their known ecological plasticity. These classification patterns  
464 suggest the framework captures meaningful ecological uncertainty rather than arbitrary  
465 assignment errors.

466 The diet matrix validation reveals a nuanced picture of trophic relationship stabili-  
467 ty. The moderate negative correlations between iterations initially appear concerning.  
468 However, the strong correlations in predator-prey rankings ( $\rho = 0.68 – 0.89$ ) suggest the  
469 framework maintains consistent broad-scale trophic structure while allowing flexibility  
470 in fine-scale interactions. This pattern aligns with ecological theory, where core trophic  
471 relationships remain stable while peripheral feeding interactions may vary with resource  
472 availability and environmental conditions.

473 The observed trophic level patterns provide compelling evidence for the framework’s  
474 ecological validity. The consistent classification of higher trophic level species reflects the  
475 relatively constrained niches of specialized predators. Conversely, the greater variability  
476 in lower trophic level classifications mirrors the natural complexity and adaptability of  
477 these groups. The framework’s ability to capture this fundamental ecological pattern  
478 suggests it successfully incorporates biological realism into its classification decisions.

479 The regional differences in group size variation and classification stability offer in-  
480 sights into ecosystem complexity. The Northern Territory's higher variation in group  
481 sizes and slightly lower classification stability likely reflect the increased ecological com-  
482 plexity of tropical reef systems. The more stable classifications in the South East regions  
483 may indicate more clearly defined ecological niches in temperate marine environments.  
484 These regional patterns demonstrate the framework's sensitivity to underlying ecological  
485 differences while maintaining consistent overall performance.

486 The reduction from 63 potential functional groups to 34-37 region-specific groups in-  
487 dicates the framework's ability to identify ecologically relevant groupings while avoiding  
488 artificial complexity. The statistical consistency across regions suggests these groupings  
489 represent fundamental ecological units rather than arbitrary divisions. This optimiza-  
490 tion of functional group complexity balances model detail with practical utility, a crucial  
491 consideration for ecosystem modeling applications.

## 492 4.2 Applications for EBFM

493 The framework offers significant practical value for ecosystem-based fisheries management  
494 through several key capabilities. Managers can now construct initial EwE models for new  
495 regions in days rather than months, enabling faster response to management needs. This  
496 addresses a key bottleneck identified by [Zheng et al. \(2023\)](#) in marine resource manage-  
497 ment. The framework's explicit quantification of uncertainty in species classifications and  
498 trophic relationships enables managers to identify areas requiring additional data collec-  
499 tion or careful monitoring, following approaches recommended by [Kuhn et al. \(2024\)](#).

500 The demonstrated adaptability across diverse ecosystems enables customization for  
501 specific regions while maintaining methodological consistency, supporting standardized  
502 approaches to ecosystem management across jurisdictions. Through systematic processing  
503 of available data, the framework also reveals specific areas where additional research or  
504 monitoring would most improve model reliability, aligning with recent work by [Chen and](#)  
505 [Xu \(2024\)](#) on marine protected area management.

## 506 4.3 Limitations and Uncertainties

507 Our framework faces several AI-specific limitations identified in recent ecological mod-  
508 eling research. The framework's performance depends on the quality and completeness  
509 of available training data. The framework's classification patterns showed regional varia-  
510 tions in stability, though further research is needed to determine the relationship between  
511 data availability and classification performance. This consideration aligns with findings  
512 from [Kuhn et al. \(2024\)](#) regarding machine learning applications in fisheries, where data  
513 quality significantly impacts model reliability.

514 An important limitation of our approach stems from its reliance on Claude 3.5 Sonnet,  
515 a closed-source large language model. The proprietary nature of this model introduces  
516 uncertainty regarding the training data used in its development and potential biases that  
517 may affect ecological interpretations. While our validation demonstrates consistent per-  
518 formance, the inability to examine the model's training data or internal decision-making  
519 processes raises important considerations for scientific reproducibility. Future iterations  
520 of the framework may benefit from exploring open-source alternatives or implementing  
521 multiple model approaches as demonstrated by [Kommineni et al. \(2024\)](#) in their work  
522 with various LLMs for biodiversity research.

523 The framework's interpretability presents another key challenge. While our validation  
524 demonstrates robust performance metrics, the underlying AI decision-making processes,  
525 particularly in parameter estimation, require careful scrutiny. This challenge mirrors  
526 concerns raised by [Fernandes and D'Mello \(2024\)](#) regarding the "black box" nature of AI  
527 systems in aquaculture applications. Our framework partially addresses these concerns  
528 through explicit uncertainty quantification and validation metrics, but further work is  
529 needed to enhance model transparency.

530 Technical limitations include computational resource requirements and processing time  
531 constraints, particularly evident in data harvesting operations. These limitations align  
532 with implementation barriers identified by [Fernandes and D'Mello \(2024\)](#), including ac-  
533 quisition costs and technical expertise requirements. The framework's sensitivity to data  
534 availability varies across ecological roles and regions, affecting both classification stability  
535 and diet matrix reliability.

## 536 4.4 Validation Roadmap

537 The framework's reliability requires validation through targeted experiments by the eco-  
538 logical community. Here we test consistency, or precision, of the framework to demon-  
539 strate that the process is repeatable. However, further work and testing by experienced  
540 ecologists and EwE modellers is needed to validate and confirm that the groupings and  
541 diet matrices that are produced are ecologically sound and useful for modelling. A fu-  
542 ture step would be to test the framework by providing it with published functional group  
543 classifications from their regions of expertise. This direct comparison would reveal the  
544 system's ability to sort species according to established ecological understanding. The  
545 framework's diet proportion estimates require validation against expert knowledge, focus-  
546 ing on both common and rare trophic relationships. These experiments would establish  
547 clear boundaries for the framework's operational use in ecosystem modeling.

548 We encourage ecologists to experiment with the framework as a collaborative tool  
549 rather than a replacement for expert judgment. The framework can rapidly generate ini-  
550 tial model structures for expert review, enabling efficient iteration between computational  
551 suggestions and ecological expertise. Systematic testing across different ecosystem types,  
552 from data-rich temperate systems to data-sparse tropical environments, would establish  
553 the framework's utility across diverse ecological contexts. This collaborative validation  
554 approach would strengthen the bridge between artificial intelligence and ecological exper-  
555 tise in ecosystem modeling.

## 556 4.5 Conclusion

557 Our validation analysis demonstrates both the capabilities and limitations of AI-assisted  
558 ecosystem modeling. The framework shows remarkable stability in many aspects while  
559 highlighting areas of ecological uncertainty that deserve attention. The clear regional  
560 patterns in performance suggest that the approach can adapt to different ecological con-  
561 texts while maintaining scientific rigor. These findings support the framework's utility for  
562 ecosystem-based management while providing clear directions for future improvement.

563 The observed trade-offs between consistency and complexity reflect fundamental chal-  
564 lenges in ecosystem modeling rather than simple methodological limitations. By quan-  
565 tifying these trade-offs and their regional variations, our analysis provides a foundation  
566 for more informed application of ecosystem models in fisheries management. As marine

567 ecosystems face increasing pressures from climate change and human activities, this un-  
568 derstanding of model behavior and limitations becomes increasingly crucial for effective  
569 ecosystem-based management (Geary et al., 2020).

## 570 Acknowledgements

571 [Add acknowledgements here]

## 572 Data Availability

573 The complete codebase, including all scripts, configuration files, and analysis tools, is  
574 available at [GitHub repository URL]. The validation framework, including reference  
575 group definitions and classification rules, is documented in the project repository to ensure  
576 reproducibility.

## 577 Author Contributions

578 SS: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation,  
579 Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visual-  
580 ization, Project administration. BF: Validation, Writing - Review & Editing, Supervision,  
581 Funding acquisition. FB: Methodology, Software, Validation, Writing - Review & Editing,  
582 Supervision. CB: Investigation, Validation, Writing - Review & Editing. RB: Investiga-  
583 tion, Validation, Writing - Review & Editing. JPG: Investigation, Validation, Writing -  
584 Review & Editing. JS: Conceptualization, Validation, Investigation, Writing - Review &  
585 Editing. RS: Investigation, Validation, Writing - Review & Editing. RT: Methodology,  
586 Software, Validation, Investigation, Writing - Review & Editing, Supervision, Funding  
587 acquisition.

## 588 Statement on the Use of Generative AI

589 Generative AI tools, specifically Claude Sonnet 3.5, were utilized in the preparation of this  
590 manuscript to assist with tasks such as language refinement, text structuring, and summa-  
591 rization. All scientific content, data interpretation, and conclusions were independently  
592 developed and verified by the authors to ensure accuracy and integrity.

## 593 References

- 594 Anthropic (2024). Claude 3 model card. Technical documentation, Anthropic.
- 595 Chamberlain, S. (2020). *robis: An R Client for the Ocean Biodiversity Information Sys-*  
596 *tem*. R package version 0.2.0.
- 597 Chen, M. and Xu, Z. (2024). A deep learning classification framework for research methods  
598 of marine protected area management. *Journal of Environmental Management*.
- 599 Christensen, V. and Walters, C. J. (2004). Ecopath with Ecosim: methods, capabilities  
600 and limitations. *Ecological Modelling*, 172(2-4):109–139.

- 601 Chroma (2024). Chroma - the ai-native open-source embedding database. Accessed: 2024.
- 602 Coll, M., Akoglu, E., Arreguín-Sánchez, F., Fulton, E. A., Gascuel, D., Heymans, J. J.,  
603 Libralato, S., Mackinson, S., Palomera, I., Piroddi, C., et al. (2015). Modelling dynamic  
604 ecosystems: venturing beyond boundaries with the Ecopath approach. *Reviews in Fish  
605 Biology and Fisheries*, 25(2):413–424.
- 606 Colléter, M., Valls, A., Guitton, J., Gascuel, D., Pauly, D., and Christensen, V. (2015).  
607 Global overview of the applications of the Ecopath with Ecosim modeling approach  
608 using the EcoBase models repository. *Ecological Modelling*, 302:42–53.
- 609 Fernandes, S. and D'Mello, A. (2024). Artificial intelligence in the aquaculture industry:  
610 Current state, challenges and future directions. *Aquaculture*, page 742048.
- 611 Froese, R., Pauly, D., et al. (2010). FishBase.
- 612 GBIF (2024). GBIF: The global biodiversity information facility. *Global Biodiversity  
613 Information Facility*. What is GBIF?
- 614 Geary, W. L., Bode, M., Doherty, T. S., Fulton, E. A., Nimmo, D. G., Tulloch, A. I.,  
615 Tulloch, V. J., and Ritchie, E. G. (2020). A guide to ecosystem models and their  
616 environmental applications. *Nature Ecology & Evolution*, 4:1459–1471.
- 617 Grassle, J. F. and Stocks, K. (1999). A global ocean biogeographic information system  
618 (OBIS) for the census of marine life. *Oceanography*, 12(3):12–14.
- 619 Guo, J., Li, N., and Xu, M. (2025). Environmental large language model evalua-  
620 tion (ELLE) dataset: A benchmark for evaluating generative AI applications in eco-  
621 environment domain. *arXiv preprint arXiv:2501.06277*.
- 622 Holden, M. H., Akinlotan, M. D., Binley, A. D., Cho, F. H., Helmstedt, K. J., and  
623 Chadès, I. (2024a). Why shouldn't I collect more data? Reconciling disagreements  
624 between intuition and value of information analyses. *Methods in Ecology and Evolution*,  
625 15:1580–1592.
- 626 Holden, M. H., Plagányi, É. E., Fulton, E. A., Campbell, A. B., Janes, R., Lovett, R. A.,  
627 Wickens, M., Adams, M. P., Botelho, L. L., Dichmont, C. M., et al. (2024b). Cost–  
628 benefit analysis of ecosystem modeling to support fisheries management. *Journal of  
629 Fish Biology*, 104:1667–1674.
- 630 Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L.  
631 (2021). Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440.
- 632 Keck, F., Broadbent, H., and Altermatt, F. (2025). Extracting massive ecological data on  
633 state and interactions of species using large language models. *bioRxiv*, pages 2025–01.
- 634 Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., and  
635 Hooker, G. (2009). Data-intensive science: a new paradigm for biodiversity studies.  
636 *BioScience*, 59(7):613–620.
- 637 Kommineni, V. K., König-Ries, B., and Samuel, S. (2024). Harnessing multiple LLMs  
638 for information retrieval: A case study on deep learning methodologies in biodiversity  
639 publications. *arXiv preprint arXiv:2411.09269*.

- 640 Kuhn, B., Cayetano, A., and Fincham, J. (2024). Machine learning applications for  
641 fisheries—at scales from genomics to ecosystems. *Reviews in Fisheries Science*.
- 642 Lapeyrolerie, M., Chapman, M. S., Norman, K. E., and Boettiger, C. (2022). Deep  
643 reinforcement learning for conservation decisions. *Methods in Ecology and Evolution*,  
644 13:2649–2662.
- 645 Li, H., Liu, J., Wang, Z., Luo, S., Jia, X., and Yao, H. (2024). Lite: Modeling environ-  
646 mental ecosystems with multimodal large language models. *Conference on Language  
647 Models (COLM)*.
- 648 Michener, W. K. and Jones, M. B. (2012). Ecoinformatics: supporting ecology as a  
649 data-intensive science. *Trends in Ecology & Evolution*, 27(2):85–93.
- 650 Poelen, J. H., Simons, J. D., and Mungall, C. J. (2014). Global biotic interactions:  
651 An open infrastructure to share and analyze species-interaction datasets. *Ecological  
652 Informatics*, 24:148–159.
- 653 Spillias, S., Ollerhead, K., Andreotta, M., Annand-Jones, R., Boschetti, F., Duggan, J.,  
654 Karcher, D., Paris, C., Shellock, R., and Trebilco, R. (2024a). Evaluating generative ai  
655 to extract qualitative data from peer-reviewed documents.
- 656 Spillias, S., Trebilco, R., Adams, M. P., Boschetti, F., Constable, A., Dunstan, P., Fer-  
657 rier, S., Porobic, J., Grimberg, E., Grigg, N., et al. (2024b). The future of artificial  
658 intelligence in ecosystem modelling. *SSRN preprint*.
- 659 Spillias, S., Tuohy, P., Andreotta, M., Annand-Jones, R., Boschetti, F., Cvitanovic, C.,  
660 Duggan, J., Fulton, E. A., Karcher, D. B., Paris, C., et al. (2024c). Human-ai collabora-  
661 tion to identify literature for evidence synthesis. *Cell Reports Sustainability*, 1(7).
- 662 Tuia, D., Kellenberger, B., Beery, S., Costeloe, B. R., Zuffi, S., Risso, B., Mathis, A.,  
663 Mathis, M. W., van Langevelde, F., Burghardt, T., et al. (2022). Perspectives in  
664 machine learning for wildlife conservation. *Nature Communications*, 13:1–15.
- 665 Villasante, S., Arreguín-Sánchez, F., Heymans, J. J., Libralato, S., Piroddi, C., Chris-  
666 tensen, V., and Coll, M. (2016). Modelling complex systems of multiple species for  
667 ecosystem based management. *Ecological Modelling*, 326:68–76.
- 668 Zheng, Z., Zhang, J., Vu, T. A., Diao, S., Tim, Y. H. W., and Yeung, S. K. (2023).  
669 MarineGPT: Unlocking secrets of ocean to the public. *arXiv preprint arXiv:2310.13596*.

670 **Supplementary Material**

671 **S1 Data Harvesting Implementation**

672 Our data harvesting system employs DuckDB for efficient querying of PARQUET files,  
673 enabling complex joins and aggregations without full memory loading. For species matching  
674 across databases, we use structured SQL queries that join on concatenated genus and  
675 species names:

```
676 SELECT * FROM sealifebase_df  
677 WHERE SpecCode = {spec_code}  
678 AND PreyStage LIKE '%adult%'  
679 AND PredatorStage LIKE '%adult%'
```

680 When combining interaction data from GLOBI with diet information, we implement  
681 a comprehensive interaction mapping system that creates bidirectional records:

```
682 interaction_data[source_group]['preys_on'][target_group] = count  
683 interaction_data[target_group]['is_preyed_on_by'][source_group] = count
```

684 Our data cleaning protocol standardizes types by converting numerical values to con-  
685 sistent formats and timestamps to ISO format. We handle null values by removing empty  
686 values, ‘NA’ strings, and null entries while preserving data structure. Source tracking  
687 maintains database origin information for all data points.

688 The system implements file locking mechanisms for concurrent access, with separate  
689 locks for species data and interaction networks. We use exponential backoff retry logic  
690 for API interactions, with configurable parameters including maximum retries (5), initial  
691 delay (1 second), and maximum delay (60 seconds).

692 The completion check system verifies the presence of required fields including:

- 693 • Complete taxonomic hierarchy
- 694 • Species-specific database records (when available)
- 695 • Interaction data
- 696 • Source attribution
- 697 • Data quality indicators

698 The final JSON output maintains a consistent structure across all species entries,  
699 facilitating automated processing in subsequent framework stages.

700 **S2 Diet Matrix Analysis**

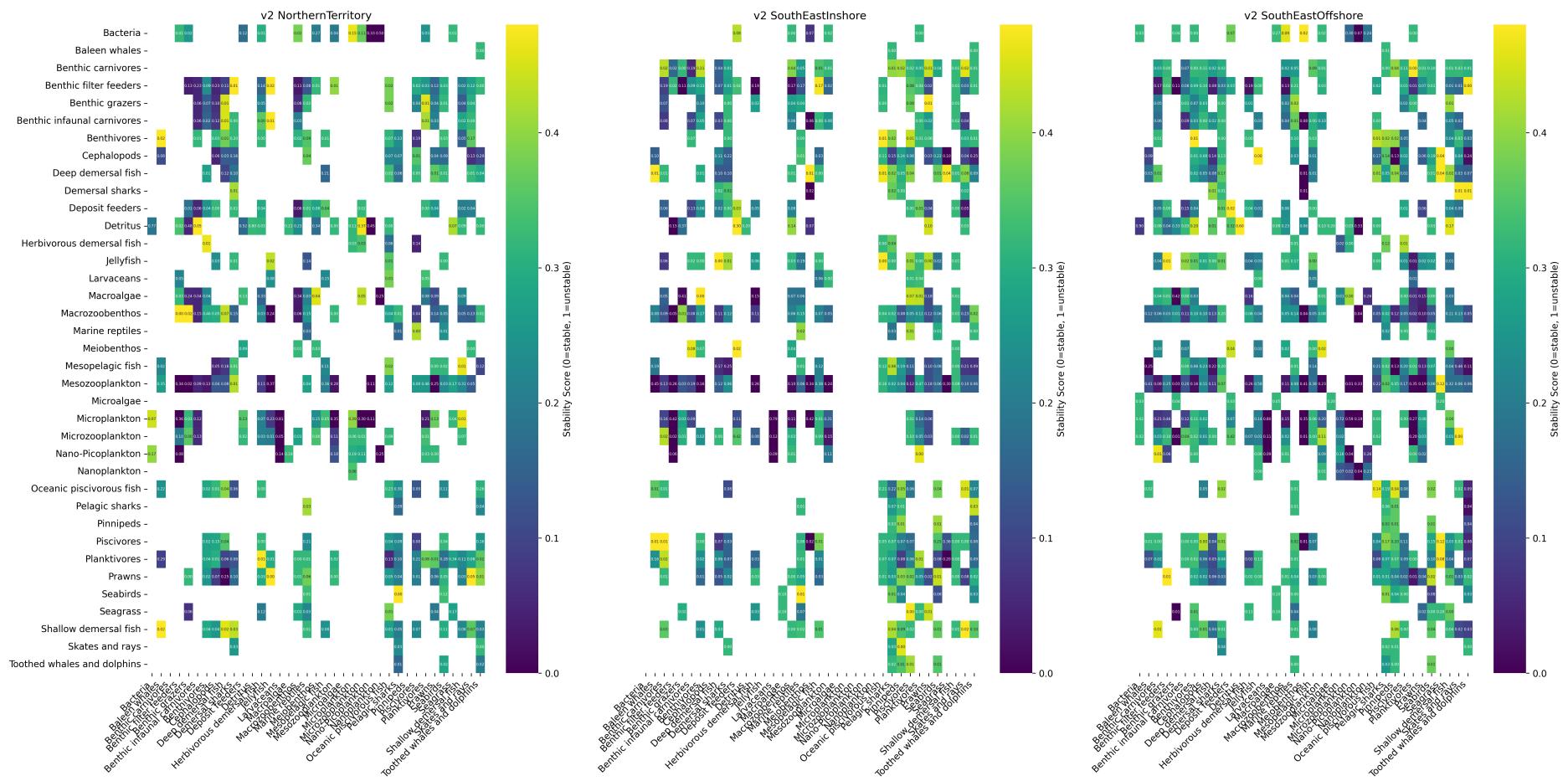


Figure S1: Detailed diet matrix consistency across five iterations for each geographic region. Column names represent predator groups and row names represent their prey groups. Numbers in each cell indicate the mean diet proportions across five iterations, while cell colors indicate the stability score (0-1, where 0 represents perfect stability and 1 represents maximum variation). White cells represent absent feeding relationships. This comprehensive visualization complements the stability score distributions and predator-specific analyses presented in the main text (Figures 4 and 5).

## 701 S3 Technical Implementation

### 702 S3.1 Default Grouping with Descriptions

703 Table S1 presents the complete template of potential functional groups used by the system.  
704 This template serves as a reference for group classification, though the system can create  
705 new groups or modify existing ones based on specific ecosystem characteristics.

Table S1: Complete Functional Group Template

Group Name	Description
Skates and rays	Bottom-dwelling cartilaginous fish that play a role in controlling benthic prey populations
Nearshore and smaller seabirds	Small gulls, terns etc that feed near shore (possibly include penguins here too) - avian predators that link marine and terrestrial ecosystems
Albatrosses	Large seabirds that forage exclusively at sea, feeding on marine prey (fishes, squids, gelatinous organisms)
Skuas and giant petrels	Large predatory seabirds that feed both at sea and on land, including predation on other birds
Fish-eating pinnipeds	Marine mammals (seals, sea lions) that primarily prey on fish in coastal and pelagic ecosystems
Invertebrate-eating pinnipeds	Marine mammals (particularly Antarctic seals) that primarily feed on krill and other invertebrates
Baleen whales	Large filter-feeding marine mammals that regulate zooplankton populations and contribute to nutrient cycling
Orcas	Apex predators that uniquely prey upon other top predators including marine mammals, sharks, and large fish
Sperm whales	Deep-diving cetaceans that primarily feed on deep-water squid and fish
Small toothed whales and dolphins	Smaller cetaceans that primarily feed on fish and squid in surface and mid-waters
Sea snakes	Marine reptiles that prey primarily on fish, particularly eels and fish eggs
Crocodiles	Large predatory reptiles in coastal and estuarine waters that prey on fish, birds, and mammals
Turtles	Herbivores and omnivores that breed on land
Planktivores	Small fishes that feed on plankton, crucial in transferring energy from plankton to larger predators
Flying fish	Epipelagic fish capable of gliding above the water surface, important prey for many predators
Remoras	Fish that form commensal relationships with larger marine animals, feeding on parasites and food scraps
Large oceanic piscivorous fish	Fish-eating predators in open ocean environments, mid-sized non-migratory species (e.g. barracuda)
Tuna and Billfish	Large oceanic predatory fish, highly mobile, often dive to feed deeper into the water column

Continued on next page

Table S1 – Continued

<b>Group Name</b>	<b>Description</b>
Shelf small benthivores	Small bodied fish that feed on benthic organisms, playing a key role in benthic-pelagic coupling, live in shelf waters
Shelf demersal omnivorous fish	Medium sized demersal fish that feed on invertebrates as well as smaller fish, live in shelf waters
Shelf medium demersal piscivores	Medium sized demersal fish living near the bottom in shallow waters, often important in benthic food webs, feed on other fish primarily, live in shelf waters
Shelf large piscivores	Fish-eating predatory fishes found in various marine habitats, important in controlling prey fish populations
Herbivorous demersal fish	Bottom-associated fish that primarily feed on plants, important in controlling algal growth
Slope/deep water benthivores	Small to mid sized fish that feed on benthic organisms and live on the shelf or seamounts
Slope/deep demersal omnivorous fish	Medium sized demersal fish that feed on invertebrates as well as smaller fish, live in slope or seamount waters
Slope/deep medium demersal piscivores	Medium sized demersal fish that feed on other fish primarily, live in slope or seamount waters
Slope/deep large piscivores	Fish-eating predatory fishes found in various marine habitats in deeper water, live in slope or seamount waters
Migratory mesopelagic fish	Fish living in the mesopelagic zone, undertake diel vertical migration, important in energy transfer between depths
Non-migratory mesopelagic fish	Fish living in the mesopelagic zone, non-migratory species, important in energy transfer between depths
Reef sharks	Top predators in coral reef ecosystems, controlling fish populations and maintaining reef health
Pelagic sharks	Open-ocean predators that help regulate populations of fishes and squids
Demersal sharks	Bottom-dwelling sharks, including dogfishes, that control populations of fishes and invertebrates on and near the seafloor
Cephalopods	Intelligent mollusks like squid and octopus, important predators in many marine ecosystems
Hard corals	Reef-building colonial animals that create complex habitat structure through calcium carbonate deposition
Soft corals	Colonial animals that contribute to reef habitat complexity without building calcium carbonate structures
Sea anemones	Predatory anthozoans that can form symbiotic relationships with fish and crustaceans
Hydrothermal vent communities	Specialized organisms living around deep-sea vents, including chemosynthetic bacteria and associated fauna
Cold seep communities	Organisms adapted to methane and sulfide-rich environments on the seafloor
Deep-sea glass sponges	Filter-feeding animals that create complex deep-water habitats and are important in silicon cycling
Sea cucumbers	Deposit-feeding echinoderms important in sediment processing and bioturbation

Continued on next page

Table S1 – Continued

Group Name	Description
Sea urchins	Herbivorous echinoderms that can control macroalgal abundance and affect reef structure
Crown-of-thorns starfish	Coral-eating sea stars that can significantly impact reef health during population outbreaks
Benthic filter feeders	Bottom-dwelling organisms that filter water for food, important in nutrient cycling and regulating water quality in various depths - bivalves, crinoids, sponges
Macrozoobenthos	Mobile large bottom-dwelling invertebrates in both shallow and deep waters, important in benthic food webs and bioturbation (predatory or omnivorous)
Benthic grazers	Bottom-dwelling organisms that graze on algae and detritus, influencing benthic community structure
Prawns	Small crustaceans that are important in benthic and pelagic food webs
Meiobenthos	Tiny bottom-dwelling organisms, important in sediment processes and as food for larger animals
Deposit feeders	Animals that feed on organic matter in sediments, important in nutrient cycling
Benthic infaunal carnivores	Predatory animals living within the seafloor sediments
Sedimentary Bacteria	Microscopic organisms crucial in nutrient cycling and the microbial loop in marine ecosystems
Large carnivorous zooplankton	Fish larvae, arrow worms and other large predatory zooplankton
Antarctic krill	Key species in Antarctic food webs, particularly important as prey for whales, seals, and seabirds
Ice-associated algae	Microalgae living within and on the underside of sea ice, important primary producers in polar regions
Ice-associated fauna	Specialized invertebrates living in association with sea ice, important in polar food webs
Mesozooplankton	Medium-sized zooplankton (200 µm to 2 cm) that feed on smaller plankton and serve as food for larger animals
Microzooplankton	Tiny zooplankton (20 µm to 200 µm) that graze on phytoplankton and bacteria, forming a crucial link in the microbial food web
Pelagic tunicates	Including larvaceans, salps, and pyrosomes, important in marine snow formation and carbon cycling
Jellyfish	Predatory gelatinous species
Diatoms	Larger phytoplankton (20 µm to 200 µm), silica dependent important primary producers in marine ecosystems
Dinoflagellates	Mixotrophic species (20 µm to 200 µm) that can switch between primary production and consumption as needed
Nanoplankton	Plankton ranging from 2 µm to 20 µm in size, including small algae and protozoans
Picoplankton	Plankton ranging from 0.2 µm to 2 µm in size, including both photosynthetic and heterotrophic organisms

Continued on next page

Table S1 – Continued

Group Name	Description
Microalgae (microphyto-benthos)	Microscopic algae that live on the seafloor or attached to other organisms
Pelagic bacteria	Watercolumn dwelling bacteria, consume marine snow amongst other things
Seagrass	Marine flowering plants that form important coastal habitats and nursery areas
Mangroves	Salt-tolerant trees forming critical coastal nursery habitats and protecting shorelines
Salt marsh plants	Coastal vegetation adapted to periodic flooding, important in nutrient cycling and shoreline protection
Macroalgae	Seaweeds of various sizes that provide habitat and food for many species, including both canopy and understory forms
Symbiotic zooxanthellae	Photosynthetic dinoflagellates living within coral and other marine invertebrates
Cleaner fish and shrimp	Species that remove parasites from other marine animals, important in reef health
Discards	Carriion and freshly discarded material from fisheries activities
Detritus	Labile components of natural death and waste

### 706 S3.2 Retrieval-Augmented Generation Implementation

707 We implement a retrieval-augmented generation system using ChromaDB for vector stor-  
 708 age and document management. Document processing begins with LlamaParse conversion  
 709 of source materials to markdown format, preserving structural elements while enabling  
 710 consistent text extraction across document types. We segment documents using a token-  
 711 aware chunking strategy with a 2000-token maximum size, determined through empirical  
 712 testing to balance context preservation with model limitations.

713 Document processing follows a two-phase approach. The initial phase generates em-  
 714 beddings for each document chunk using Azure OpenAI’s text-embedding-3-small model,  
 715 storing them in ChromaDB’s PersistentClient. The system maintains an indexed\_files.json  
 716 registry to track processed documents. The second phase handles incremental updates,  
 717 identifying and processing only new content when documents are added to the source  
 718 directory.

719 For diet composition analysis, we implement a two-stage query process. The first stage  
 720 employs a simple query to retrieve relevant document chunks:

What do [group] eat?

721 The system embeds this query using the same Azure OpenAI model and performs  
 722 vector similarity search to identify relevant document chunks. These results combine  
 723 with structured data sources including species occurrence frequencies, food category clas-  
 724 sifications, and GLOBI interaction data to form a comprehensive input for the second  
 725 stage.

726 We implement comprehensive error handling throughout the pipeline. The system em-  
 727 ploys exponential backoff retry logic for API interactions, with configurable parameters in-

728 cluding maximum retries (10), initial delay (1 second), and maximum delay (300 seconds).  
729 For model interactions, we utilize LlamaIndex’s query engine with zero-temperature sam-  
730 pling to ensure deterministic responses. The system supports multiple language model  
731 backends including Claude-3 Sonnet (200k token context), GPT-4, and AWS Claude,  
732 enabling flexible deployment based on availability and performance requirements.

733 The system maintains separate storage contexts for different document collections  
734 through ChromaDB’s collection management. This separation prevents cross-contamination  
735 between knowledge bases while enabling efficient parallel processing. We track document  
736 citations throughout the retrieval process, maintaining provenance information for all re-  
737 trieval content. The complete implementation, including embedding generation, chunking  
738 algorithms, and query processing functions, is available in the project repository.