

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

Optimizing functional groups in ecosystem models: Case study of the Great Barrier Reef

Vanessa Haller-Bull^{1,2} and Elena Rovenskaya^{2,3}

1. *School of Mathematical Sciences, Queensland University of Technology,*

2 George St, Brisbane, 4000, Qld, Australia;

2. *International Institute of Applied System Analysis (IIASA),*

Schlossplatz 1, Laxenburg, 2361, Austria

3. *Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State*

University, Leninskie Gory, 1(52), GSP-1, Moscow 119991, Russia

Corresponding author: Vanessa Haller

v.haller@hdr.qut.edu.au

+614 2099 3214

School of Mathematical Sciences, QUT, 2 George St, Brisbane, 4000, Qld,
Australia

34 **Abstract:**

35 Uncertainty is inherent in ecosystem modelling, however its effects on modelling results are often
36 poorly understood or ignored. This study addresses the issue of structural uncertainty or, more
37 specifically, model resolution and its impact on the analysis of ecosystem vulnerability to threats.
38 While guidelines for node assignments exist, they are not always underlined with quantitative
39 analysis. Different resolutions of a coral reef network are investigated by comparing the simulated
40 network dynamics over time in various threat scenarios. We demonstrate that the error between a
41 higher-resolution and a lower-resolution models increases, first slowly then rapidly with increased
42 degree of node aggregation. This informs the choice of an optimal model resolution whereby a finer
43 level of a food web representation yields only minimal additional accuracy, while increasing
44 computational cost substantially. Furthermore, our analysis shows that species biomass ratio and the
45 consumption ratio are important parameters to guide node aggregation to minimize the error.

46

47 **Keywords:** structural uncertainty, resolution, node aggregation, ecosystem models, foodweb model,
48 coral reefs

49

50

51 **1 Introduction**

52

53 Simplification of reality and related uncertainty is unavoidable in any applied research aiming to
54 support decision making; what influences the quality of analysis is how that uncertainty is
55 incorporated into management decision processes. This issue was highlighted by Ludwig et al. (1993),
56 who said: “effective policies are possible under conditions of uncertainty, but they must take
57 uncertainty into account”. Walker et al. (2003) defines uncertainty as “any deviation from the
58 unachievable idea of completely deterministic knowledge of the relevant system”. Another way of
59 thinking about uncertainty is that “uncertainty reflects the probability that a particular estimate, piece
60 of advice, or management action may be incorrect” (Lek, 2007). According to Menditto et al. (2007)

61 uncertainty is the quantitative expression of the performance characteristic accuracy with accuracy
62 being defined as “Closeness of agreement between a quantity value obtained by measurement and the
63 true value of the measureand”. While these definitions describe uncertainty well, they do not
64 differentiate between different types of uncertainty.

65

66 Uncertainty can arise at different instances in a modelling process. For the purpose of this paper, we
67 will concentrate on two types of uncertainty. *Parameter uncertainty* is defined as the difference
68 between the true value of a parameter and the mean value estimated using the data available and
69 statistical techniques (Skinner et al., 2014). *Structural uncertainty* refers to a mismatch between the
70 simplified mathematical equations of a model and the true complex ecological relationship observed
71 in situ (Refsgaard et al., 2006). One type of structural uncertainty, which is the one we focus on in this
72 study, is model resolution. In an ecological network, given a true number of species in the network,
73 the resolution refers to the level of aggregation within each node which is reflected in the total number
74 of nodes in the network. A node within the network represents an ecological unit that can be at
75 different aggregation levels incorporating one or more species. Generally, it is believed that low-
76 complexity models reduce parameter uncertainty by reducing the number of parameters, while more
77 complex models reduce structural uncertainty since they more closely describe the natural system
78 (Iwasa et al., 1987). This leads to a hump shape between the level of complexity and the accuracy,
79 with medium complexity models often performing best (Håkanson, 1995, Costanza and Sklar, 1985,
80 Jester, 1977).

81

82 Even though it has been acknowledged for a long time that different types of uncertainties are crucial,
83 their role is often not thoroughly understood, especially in complex models (Milner-Gulland and
84 Shea, 2017, Link et al., 2012). Previous attempts at addressing node aggregation in ecosystem models
85 (Fulton, 2001, Iwasa et al., 1987) have not been taken up by the average user of these kind of models.
86 As shown by Weijerman et al. (2015) who completed an extensive review of ecosystem models and of

87 the 27 ecosystem models reviewed only one addressed structural uncertainty as well as parameter
88 uncertainty. Instead, general guidelines are used to aggregate species (Heymans et al., 2016).

89

90 Ecosystems are complex due to the combination of multifaceted species interactions, which are often
91 nonlinear, which may result in multiple equilibria (Gordon, 2007, McClanahan et al., 2009).

92 Furthermore, ecosystem models are often faced with many threats which when included add to the
93 model complexity. Informing the management of systems under threat is often one of the main

94 modelling goals. Ecosystem models always involve high uncertainties, especially when attempting to
95 predict the effects of interventions and management actions (Costanza et al., 1993, Hill et al., 2007).

96 If uncertainty is larger than believed, the results are more likely to be misleading, and are therefore
97 more likely to generate an inefficient or incorrect management decision (Weijerman et al., 2015). Lek

98 (2007)'s definition of uncertainty emphasizes this possibility of decision-makers being misled by
99 models.

100

101 Ecosystem models are usually created at a functional group level to reduce the number of nodes and
102 therefore parameters (Fulton et al., 2003). A functional group refers to a group of species that are

103 assumed to be so similar in a defined set of characteristics that they can be investigated as one unit. In
104 the literature, species are commonly grouped according to their trophic status and diet (e.g., herbivore,

105 or detritivore) (Stoddart, 1969). In ecosystem modelling a functional group is often used as the basis
106 of one node: Instead of treating each species as different and assigning a different dynamic equation

107 and parameters to each species, the functional group is assumed to be homogenous enough to be
108 represented by a single equation. Reducing resolution by considering functional groups only of course

109 reduces the total number of nodes in the system, which consequently, reduces the complexity and the
110 number of parameters that needs to be estimated. Guidelines about which species to group together

111 have existed for a long time (Gardner and Ashby, 1970, Wiegert, 1975, O'Neill, 1975, Cale Jr and
112 Odell, 1980), however, they never used to have any quantitative foundations and were based purely

113 on experience. In the more recent time, there have been a few , but not always thorough studies

114 (Fulton, 2001, Fulton et al., 2003, Pinnegar et al., 2005) that explore the uncertainty that is caused by
115 the introduction of functional groups as suggested in the literature. Thus, we lack an understanding of
116 the magnitude and even distribution of structural uncertainty in the threat response of ecosystems
117 (Bellwood and Fulton, 2008).

118

119 One ecosystem under threat that benefits from a whole ecosystem analysis is the marine environment,
120 specifically coral reefs. Coral reefs (here we will consider the Great Barrier Reef) are often made up
121 of thousands of species, making this system a great example for high level aggregation. A common
122 tool for ecosystem modelling in the marine environment, especially for the evaluation of fisheries
123 management, is Ecopath with Ecosim (Coll  ter et al., 2015). Ecopath is based on the estimation of
124 biomasses and food consumption to create a mass-balanced food web. It has been reviewed and
125 extended on over the past 40 years to enable dynamic simulations (Ecosim) and spatial analysis
126 (Ecospace) (Christensen and Walters, 2004). Since its inception, over 800 studies have used it to
127 investigate questions related to fisheries management. However, even though used extensively it does
128 not include a formal strategy to include structural uncertainty in its outputs or management
129 recommendations. In terms of designing the structure, specifically the resolution, of an Ecopath the
130 only guidance available to a new user are a few published guidelines (Heymans et al., 2016). These
131 guidelines are mostly based on experience and only underlined by a few selected and non-
132 comprehensive analysis (Fulton, 2001).

133

134 In this study, we develop and apply a new approach of varying the resolution of an ecological network
135 while simulating the food web response to a species degradation. This will enable a comparison with
136 some of the guidelines and give them a quantitative foundation. The aim of this study is threefold:

- 137 1) We want to show that uncertainty introduced due to lowering the system resolution even
138 slightly can be substantial and should not be neglected.
- 139 2) We want to propose a basis for choosing an optimal resolution to balance parameter and
140 structural uncertainty.

141 3) We want to suggest ways to improve the guidelines and techniques used to group species in
142 ecosystem models while also generalising these techniques for other types of networks.
143 These objectives are achieved by extensive simulations of a reef within the Great Barrier reef,
144 considering different levels of aggregation and threat scenarios. Followed by the evaluation of these
145 using visual examination, basic statistics as well as a machine learning algorithm.

146

147 **2 Methods**

148

149 The goal of this study is to understand how the ecosystem resolution affects the model's predictive
150 power in what concerns the effects of species degradation onto the entire ecosystem. To do so, we will
151 examine a number of "threat" scenarios, in which a fraction of a biomass of a focal species is removed
152 giving rise to changes in biomasses of other species because of feeding relationships. We will
153 evaluate the variation between threat impacts modelled based on a network at species level versus a
154 network at a functional group level, which has a coarser resolution. For a single simulation we
155 calculate the error to reflect the accuracy of the simulation in relation to the reference model with full
156 resolution. Then we look at the distributions of the error over several simulations as a measurement of
157 uncertainty. To calculate errors, it is assumed that the high-resolution model reflects nature closely
158 and thus the error between the high- and low- resolution models can be used as an estimate of the
159 accuracy of a model structure. As the resolution of a model is a discrete characteristic, in this study,
160 nodes within a chosen functional group are being merged, one by one, to "imitate" a step-by-step
161 formation of the functional group. This enables a comparison of the response dynamics under the
162 same input (a particular threat scenario) and an estimation of the difference that each particular
163 merger entails. On this basis, a thorough understanding of the changes in the error and, consequently,
164 the uncertainty with a reduction in resolution can be obtained.

165

166 Naturally, one expects that a model should become more accurate with a finer resolution. However, it
167 comes at a cost. This cost is made up of two components. First, a finer scale could result in more

168 parameter uncertainty, because the data underlying the parameters is not known well enough to
169 support this scale. Second, there is a cost of computing power that is required. In this study, we refer
170 to an “optimum” in terms of good overall accuracy (including both, parameter and structural
171 uncertainty) without an explosion of computational power.

172

173 *2.1 Ecosystem simulation model and calibration*

174 The dynamic ecosystem model used in this study is based on a network of interacting nodes that
175 represent a single or a group of species. The interactions considered between the nodes resemble
176 predation, i.e., the consumption of biomass of one node by another. This means that the connection
177 between two nodes indicates organic matter being transported from one node to the other and
178 travelling generally up the food chain. Organic matter can be lost from the living nodes to the external
179 environment through energy spent in respiration. Egestion and other mortality that add to the detritus
180 close the loop from high to low trophic levels of the food chain.

181

182 The network model used for this study is based on the Ecopath with Ecosim toolbox. We use Ecopath
183 to calibrate our ecosystem network model and then we adopt Ecosim equations to simulate the
184 ecosystem dynamics in response to a shock. We re-code the model in MATLAB for convenience of
185 running simulations. In what follows, we present a general approach towards achieving the goal of
186 this study as discussed above, and then we apply it to Great Barrier Reef as a case study.

187

188 First, we describe the modelling framework. Ecopath is a static mass-balanced ecosystem model
189 describing the network through linear equations that connect biomasses, diet preferences, mortalities
190 and productions in an ecosystem. The static equations of Ecopath can then be extended to differential
191 (difference) equations to allow for dynamic manipulations of the system (Ecosim extension).

192

193 Ecopath is based on two master equations. The first master Equation (1) focuses on the mass balance
 194 within each nodes, while the second master Equation (2) describes the mass balance between nodes;
 195 the first master equation is as follows

$$196 \quad P_i = Q_i - R_i - U_i, \quad (1)$$

197 where P_i represents the total biomass production of node i ($t km^{-2}year^{-1}$) and Q_i represents the
 198 total biomass consumption of node i ($t km^{-2}year^{-1}$), R_i represents the respiration of node i
 199 ($t km^{-2}year^{-1}$), which is a loss of a part of biomass to the environment and U_i represents the
 200 amount of unassimilated food for each node i ($t km^{-2}year^{-1}$) for all $i = 1, \dots, n$ with n being the
 201 number of nodes in the ecosystem network. Unassimilated food refers to the amount of biomass lost
 202 through excretion, i.e., the amount of biomass input into the lowest trophic level node called detritus.

203

204 The second master equation unfolds the node production of all nodes (besides detritus) that consists of
 205 the amount of its biomass consumed by other nodes plus the amount of biomass that emigrates and the
 206 amount of biomass that is lost due to natural mortality, here called other mortality or excretion. The
 207 flow from one node to another is assumed to be proportional to the total consumption of the node-
 208 recipient, which is sometimes called top-down or recipient control:

$$209 \quad P_i = \sum_j Q_j DC_{ij} + E_i + P_i(1 - EE_i), \quad (2)$$

210 where DC_{ij} = proportion of the consumption by predator j that is made up of prey i (unitless), E_i =
 211 net emigration (emigration – immigration; $t km^{-2}year^{-1}$); and EE_i = ecotrophic efficiency
 212 (unitless). The ecotrophic efficiency is the proportion of production that is passed onto the next
 213 trophic level. This parameter is smaller than one due to natural mortality.

214

215 Compared to the standard form of the second master equation found in many publications
 216 (Christensen et al., 2005, Heymans et al., 2016), this formulation of the second master equation has
 217 been modified already to assume that there is no biomass accumulation (in Ecopath notations $BA_i =$
 218 0) to ensure a steady state and as a baseline it also assumes that there was no fishing ($Y_i = 0$).

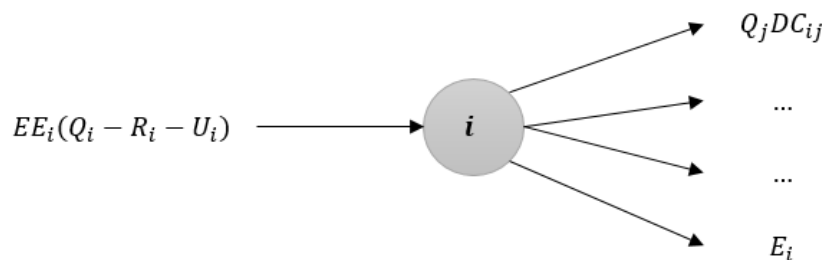
219

220 Combining (1) and (2) one obtains

$$221 \quad EE_i(Q_i - R_i - U_i) = \sum_j Q_j DC_{ij} + E_i \quad (3)$$

222 which means that the surviving production of one group (the left-hand side of this equation) is equal to
 223 the consumption of that group by all predators and the biomass that is leaving through emigration and
 224 other forms of losses not otherwise explicitly represented in the model (see Figure 1).

225



226

227 **Figure 1.** Visualisation of Equation (3), the mass-balance of a single node within the network

228

229 The data requirements for Ecopath encompass at a minimum the input of parameters DC_{ij} and E_i , as
 230 well as any three out of four variables P_i, Q_i, B_i and EE_i for each node in the model. The missing
 231 parameters are then estimated using the mass-balance Equations (1) and (2). This process is called
 232 balancing the Ecopath model.

233

234 For some nodes we might have information on all four parameters. However, since they have often
 235 been estimated during different studies and any such parameters contain uncertainty, we often cannot
 236 reach a mass-balance directly by just calculating the remaining unknown parameters (which are fewer
 237 than n). In this case, we conduct another step called the model calibration. For the model calibration,
 238 i.e. to find “free” parameters that ensure mass-balance, all parameters are varied according to the
 239 parameter uncertainty underlying their estimates until a mass-balance is reached. For example, the
 240 production of a species might have been investigated and confirmed by several studies, so the
 241 parameter uncertainty is low and the bound in which it is varied for calibration might be set at 10%.
 242 On the other hand, a production estimate of a different species might not be available for the study

243 region and needs to be inferred from a different region. This would result in a higher parameter
 244 uncertainty and hence a variation bound of 20% would be used. Rules to identify these uncertainty
 245 bounds have been established and are called pedigrees within the Ecopath software (Christensen et al.,
 246 2005). In practice this means that we can often find several possible parameterizations that ensure
 247 mass-balance. To identify robust conclusions for a study utilising these models several/ all of these
 248 parameterizations should be evaluated. We elaborate more on how we used them in this study in
 249 section 2.5.

250

251 Equation (2) motivates a differential equation in which species biomass evolves over time as a
 252 response to external shocks and pressures, such as harvesting. Let B_i be the biomass of a group i .
 253 From here on out we no longer consider the total production (P_i) or consumption (Q_i), but rather the
 254 production (p_i) and consumption (q_i) per unit of biomass. Since (2) has the assumption of a steady
 255 state, we can rearrange it to refer to the population change of zero as follows

$$256 \quad 0 = -\sum_j q_j B_j DC_{ij} + p_i B_i - E_i - p_i B_i (1 - EE_i); \quad (4)$$

257 or generally as change in population size over time

$$258 \quad \frac{dB_i}{dt} = -\sum_j q_j B_j DC_{ij} + p_i B_i - E_i - p_i B_i (1 - EE_i). \quad (5)$$

259 Furthermore, we consider that immigration I_i into the group is independent of the population sizes
 260 within the system, while emigration e_i out from the group depends on the population size of group
 261 i , due to the effect of crowding, which means

$$262 \quad E_i = e_i B_i - I_i \quad (6)$$

263 We denote

$$264 \quad M_i = p_i (1 - EE_i) . \quad (7)$$

265 which refers to the non-predation mortality. So, all together we obtain

$$266 \quad \frac{dB_i}{dt} = -\sum_j q_j B_j DC_{ij} + p_i B_i + I_i - (e_i + M_i) B_i . \quad (8)$$

267

268 In the literature, Equation (8) is often described in a more general form

$$269 \quad \frac{dB_i}{dt} = -\sum_j c_{ij}(B_i, B_j) + g_i \sum_j c_{ji}(B_i, B_j) + I_i - (e_i + M_i)B_i. \quad (9)$$

270 Equation (9) has the advantage that $c_{ij}(B_i, B_j)$ can be chosen to represent bottom-up, top-down or
 271 mixed control. In this study, we investigate bottom-up control due to its inherent stability (Hearon,
 272 1963) and consequently, assign

$$273 \quad c_{ij}(B_i, B_j) = \beta_{ij}B_i. \quad (10)$$

274 Since each flow from i to j needs to be consistent between the static and the dynamic approach we
 275 define

$$276 \quad \beta_{ij}B_i = Qb_jB_jDC_{ij}. \quad (11)$$

277 Consequently, β_{ij} is defined as

$$278 \quad \beta_{ij} = \frac{B_j}{B_i}Qb_jDC_{ij}, \quad (12)$$

279 where all entries in formula (12) are to be taken from the model calibration results.

280 Thus, we have

$$281 \quad \frac{dB_i}{dt} = -\sum_j \beta_{ij}B_i + g_i \sum_j \beta_{ji}B_j + I_i - (e_i + M_i)B_i - T_iB_i. \quad (13)$$

282 where g_i is the growth efficiency, which is a ratio between the production and consumption (unitless)
 283 and β_{ij} represent elements of the interaction matrix (Christensen et al., 2005). In (13), we also include
 284 threats (last term in the right-hand side) with T_i being the proportion of the total biomass removed by
 285 threat i in Ecopath usually only considered as fishing ($t km^{-2}year^{-1}$). Most threats remove part of a
 286 population so $0 < T_i < 1$, however it could also be $-1 < T_i < 0$ in case of a threat that increases an
 287 unwanted population, for example the increase of algae,

288

289 In the dynamic simulations, we initiate system (13) using initial values $B_{i,t=0}$ equal to the steady-state
 290 values B_i calculated using Ecopath. Emigration and immigration are also taken from Ecopath. As the
 291 dynamics is affected by a threat, the system is moving away from the initial steady state
 292 corresponding to an equilibrium in the absence of external disturbance. Due to the donor-control
 293 assumption, the dynamics converges to a new steady state.

294

295 2.2 *Case study system and data*

296 The Great Barrier Reef (GBR) is the biggest reef system in the world. It is currently under a severe
297 threat due to global warming and the resulting bleaching. One reef within this system is Rib reef (lat:-
298 18.48, long:146.88). It is located on the mid-shelf in the central Great Barrier Reef. Due to its
299 location, Rib reef is a good template for other reefs within the system. Rib reef was previously
300 investigated in a fisheries management assessment study using Ecopath (Tudman, 2001). This study
301 used an Ecosim network with 25 nodes representing functional groups. However, data in this paper is
302 simply a collection of studies and as such it is available from the original published sources at the
303 species level for most fish species, allowing for a more complex network of $n = 206$ nodes. This
304 makes Rib reef an ideal case study for evaluating the consequences of merging species into functional
305 groups for modelling and management. Note that in this case only fish species are identified at a
306 species level, while other functional groups (total of 15) are kept at a functional group level to reduce
307 the data requirements and the model complexity and consequently to keep computational time feasible
308 (in the absence of these limitations the model resolution could be increase to a few thousand nodes).
309 When interpreting our results here, this limitation needs to be remembered, i.e., the actual error due to
310 coarse model resolution is likely to be larger than estimated here.

311 Following Tudman (2001), we adopt input parameters P_i, Q_i, B_i for all $i = 1, \dots, n$ and the net
312 emigration E_i as zero (this is a common assumption that immigration is roughly equal to emigration, it
313 is also based on the original study the data is based on by Tudman (2001)). Parameters DC_{ij} for all
314 $i, j = 1, \dots, n$ are downloaded directly from Fishbase (Froese and Pauly, 2017), as these parameters
315 were only given at a functional group level in Tudman (2001). Assuming that the ecosystem is
316 initially in the steady state, using the available data for P_i, Q_i, B_i and DC_{ij} we derive the missing
317 parameters EE_i . This is possible if at least two parameters of each node are known. In this way, we
318 fully calibrate our dynamic model, the resulting parameters for the following simulations can be found
319 in the supplementary materials (S1).

320

321 2.3 *Threat implementation and threat scenarios*

322 We model threats as a fixed proportional reduction of biomass of a node at each time step. This is a
323 common way to represent fishing pressures in ecosystem models, and it can also represent other
324 threats such as bleaching or nitrification via reducing or increasing the biomass of a lower trophic
325 level.

326 We implement a total of six threat scenarios in this study to prototype a range of alternatives from
327 fishing on a variety of target species to coral bleaching and increased nutrition loading, choosing
328 those scenarios which represent most typical and critical threats for coral reefs. Importantly, these
329 scenarios represent threats impacting a variety of trophic levels from top predators to primary
330 producers and consequently, enable a detection of potential differences in terms of uncertainty spread
331 unevenly across trophic levels. However, it is important to keep in mind that the selected scenarios are
332 theoretical and not directly based on data. The full list of scenarios is as follows:

- 333 1. A reduction of the biomass of pelagic fish (fishing on a high trophic level)
- 334 2. A reduction of the biomass of coral trout (fishing on a high trophic level)
- 335 3. A reduction of the biomass of herbivores (fishing on a medium trophic level)
- 336 4. A reduction of the biomass of sharks (fishing on a high trophic level)
- 337 5. A reduction of the biomass of coral (indicative of bleaching, threat on low trophic level)
- 338 6. An increase of the biomass of algae (indicative of increased nutrients, threat on low trophic
339 level).

340 All of these scenarios are introduced at different intensities ranging from 10 to 90%
341 removal/introduction rate, i.e., $T_{\text{impacted node}} = 0.1, 0.2, \dots, 0.9$ (per year).

342

343 *2.4 Merging nodes and error calculation*

344 Merging nodes refers to pooling of biomasses of two nodes (here species or functional groups) into
345 one node and corresponding uniting/reassignment of incoming and outgoing flows. The following
346 describes the routine of consecutive mergers we use to evaluate the error for each of the 54 scenarios
347 (six different threats and nine intensities per threat). In the first instance, we focus on merging nodes
348 within each functional group from the original model (Tudman, 2001) individually, meaning that all

349 nodes outside the functional group in focus remain at species level. Then we merge the functional
 350 groups consecutively to fully recreate the original 25-node model.
 351 For each of the eight functional groups in the original model we run an “experiment”. Each
 352 experiment consists of a number of steps equal to the initial size of the functional group in terms of
 353 the number of nodes. At each experiment step, we reduce the number of nodes from the functional
 354 group by one from j to $j - 1$ (the calculations for all parameters after a merger can be found in S2).
 355 To do so, we implement all possible mergers of two nodes within this functional group into one, total
 356 $\frac{j!}{2!(j-2)!}$ mergers. For every merger at a given experiment step, we consecutively run all six threat
 357 scenarios with nine intensity levels, total 54 scenarios. For each scenario, we simulate the ecosystem
 358 dynamics using (13) and initial conditions from Ecopath for up to 500 time steps (which is equivalent
 359 to 5 years) and compare it with the dynamics corresponding to the case without any mergers. To
 360 quantify the comparison we use the total relative error as follows

361

$$362 \quad \text{Error} = \sqrt{\frac{\sum_i \frac{\sum_t (B_{i,t} - \hat{B}_{i,t})^2}{\sum_t B_{i,t}^2}}{n}}, \quad (14)$$

363

364 where $\hat{B}_{i,t}$ refers to the merged biomass of a node ($B_{n-1} + B_n$ assuming that you merge the two last
 365 nodes n and $n - 1$) and $B_{i,t}$ refers to the unmerged biomass of the two merged nodes added together
 366 after the simulations; n here is the size of the considered ecosystem model version. Threats are
 367 implemented in the merged simulations as

$$368 \quad \hat{T}_{n-1} = \frac{B_{n-1}T_{n-1} + B_n T_n}{B_{n-1} + B_n} \quad (15)$$

369 Where \hat{T}_{n-1} is the threat on the new node.

370

371 Since five years is a commonly used timeframe for management decision making, it is chosen as the
 372 length for the simulations.

373 Actually, due to the linearity of the model equations, for a given threat (including none) the initial (or
 374 any given) steady state of the full system can be reproduced exactly by the reduced system (see

375 supplementary materials S2) by a choice of coefficients DC_{ij} , as well as B_i , P_i and Q_i in the reduced
376 system. But modellers and decision makers are interested in models which are able to process a
377 variety of possible threats, hence we use a metric (14) to evaluate the “goodness” of a merger
378 averaging over all nodes in the system. Also (14) focuses on the transient path instead of just the
379 eventual steady state.

380 At each experiment step, we save the “best” merger and a corresponding network, resulting in the
381 lowest error (14). We use this “optimal” network configuration as a starting point for the next
382 experiment step. Experiments continue until the considered functional group becomes a single node.

383

384 *2.5 Parameter and structural uncertainty*

385 In a study that would like to estimate the response of this coral reef to a given threat, the stepwise
386 aggregation that we have done so far is of not much use. Managers would be interested in the total
387 amount of uncertainty around the population size estimates that are provided with a given model. This
388 situation is what we are exploring in this section. The total amount of uncertainty in the estimates is
389 partly due to parameter uncertainty and partly due to the structural uncertainty. Here, we are
390 comparing how much uncertainty we would assign depending on if we consider both the structural
391 and the parameter uncertainty or only one of them as it is often done in most previous studies
392 (Weijerman et al., 2013).

393 Parameter uncertainty can be easily estimated by simulating the system response with different
394 parameter sets. As described in the previous section 2.1 we calibrate the model with different
395 variations of the original parameters to create 10 separate parameter sets for the system. The
396 parameter sets are then each simulated over 500 time steps (5 years) for each scenario and intensity.
397 The resulting responses are compared to the response of original parameter set using the same error
398 estimates as for the structural uncertainty described in section 2.4. For these error estimates
399 (10 *parameter sets* * 6 *scenarios* * 9 *intensities* = 540 *estimates*) we then calculate the mean
400 and standard error.

401 Structural uncertainty here is based on errors that are calculated similar to the previous methods in
402 section 2.4. The only difference is that instead of calculating the error at each merger, we create three
403 resolutions and only compare the error between these. The three resolutions are as follows (S3): First,
404 we have the full resolution model (assumed to have no structural error) with 206 nodes. Second, we
405 have the medium resolution model identified as the optimum from the previous sections with 49
406 nodes. Third, we have the low resolution model based on the functional group in the original study by
407 Tudman. This results in 54 error estimates ($6 \text{ scenarios} * 9 \text{ intensities}$) for each resolution which
408 are then averaged to calculate the structural uncertainty.

409 Now we can combine both of these uncertainties by creating 10 parameter sets for each resolution and
410 comparing the errors across all of them ($10 \text{ parameter sets} * 6 \text{ scenarios} * 9 \text{ intensities} =$
411 $540 \text{ estimates for each resolution}$). The parameter sets were identified using the pedigree
412 approach build into Ecopath and Ecoranger. In this approach each variable is assigned a level of
413 confidence (based on the data collection) followed by randomly selecting parameters within these
414 confidence intervals and rebalancing of the model (Christensen et al., 2005).

415

416 *2.6 Regression tree analysis*

417 The previous sections show that there is a large spread between the errors introduced even when
418 reducing the resolution by a single step. This spread of errors is caused by the differed species/ nodes
419 that are merged, i.e. some species represent a “good” merger others a “bad’ merger. In order to make
420 better decisions on which species/ nodes to merge we need to know which characteristics of each
421 species/ node causes the error to increase. This is why we conduct this regression analysis.

422 Characteristics of interest are related to parameters of each node, i.e., biomass, production, or specific
423 to each combination of nodes in conjunction with the threat scenario, i.e., the difference in trophic
424 level between nodes merged and nodes experiencing a threat (Table 1). We use the bagged regression
425 tree analysis since input data, predictors and responses typically have different non-normal
426 distributions. A regression tree analysis is a supervised (ie. there is a response variable) machine
427 learning algorithm. At each step, the algorithm splits the samples according to one of the predictors to

428 form a more homogenous group of the response. This means that if we know the value for each
 429 predictor we can follow all of the splits and predict the likely response. An extension of this is used in
 430 this study by growing several trees (200) for more reliable results. In the bagged regression tree
 431 analysis we grow each regression tree while utilising all predictors but subsampling data from the
 432 initial sample (Prasad et al., 2006). Those trees are then combined in an ensemble to give more
 433 reliable predictions. We performed an analysis with replacement i.e. each subsample is placed back
 434 into the full sample after growing the tree and a new subsample is taken for the next tree. This
 435 provides an opportunity for internal validation at each tree level using the mean squared error (MSE)
 436 of predictions. Compared to other methods of growing ensembles of trees, bagged trees largely
 437 prevent overfitting. Another great feature about the bagged tree analysis is that it can calculate
 438 surrogate splits. At each split, when the tree is grown, the algorithm determines the next best split
 439 according to the MSE, this split is then known to be the surrogate split. Overall, this procedure
 440 enables comparing the optimal split (the split chosen by the algorithm to grow the tree due to the
 441 lowest MSE) and the surrogate splits for all variables. The variable association is then measured
 442 according to how different the MSE between the optimal and the surrogate split is. Variables that can
 443 be easily replaced by their surrogate splits are less important and could be excluded. Additionally, we
 444 also calculate the overall variable importance, which is the total reduction of the MSE that is due to
 445 splits based on that variable (Prasad et al., 2006). This means that if the variable importance is high,
 446 the splits based on that particular predictor have a large influence on the reduction of the MSE
 447 achieved by the analysis. Variables with a low importance only reduce the MSE a little, so could
 448 potentially be excluded from the analysis without much loss in overall performance. The variable
 449 importance calculated here is relative within the analysis, i.e. the magnitude of the importance can
 450 only be compared between variables of the same analysis not across different models.

451

Table 1: Predictors utilised in the bagged tress analysis

Variable name	Explanations
---------------	--------------

Threat distance	The difference in trophic level* between the threat and the average of the merged groups ($TL_{\text{threat}} - \frac{TL_1 - TL_2}{2}$)
Node distance	The absolute difference in trophic (TL*) level between the two groups to be merged ($ TL_1 - TL_2 $)
Predator overlap	The proportion of the outflows that are the same in nodes to be merged $\left(\frac{\text{Outflows that are the same}}{\text{Total number of outflows}}\right)$
Food overlap	The proportion of the inflows that are the same in nodes to be merged $\left(\frac{\text{Inflows that are the same}}{\text{Total number of inflows}}\right)$
Biomass ratio	Ratio of the node sizes between two merged nodes $\left(\frac{B_{1,0}}{B_{2,0}}\right)**$
Production ratio	Ratio of the total production between two merged nodes $\left(\frac{P_{1,0}}{P_{2,0}}\right)**$
Consumption ratio	Ratio of the total consumption between two merged nodes $\left(\frac{Q_{1,0}}{Q_{2,0}}\right)**$

452 *Trophic levels in Ecopath are continuous variables and based directly on the trophic levels of the
453 species consumed. For more information Christensen et al. (2005) section 6.4.4

454 **Considering that the designation of node 1 and 2 for these calculations is arbitrary, we allocate node
455 1 and 2 so that the ratios calculated are between zero and one. This is achieved by allocating the larger
456 value to node 1.

457

458 3 Results

459

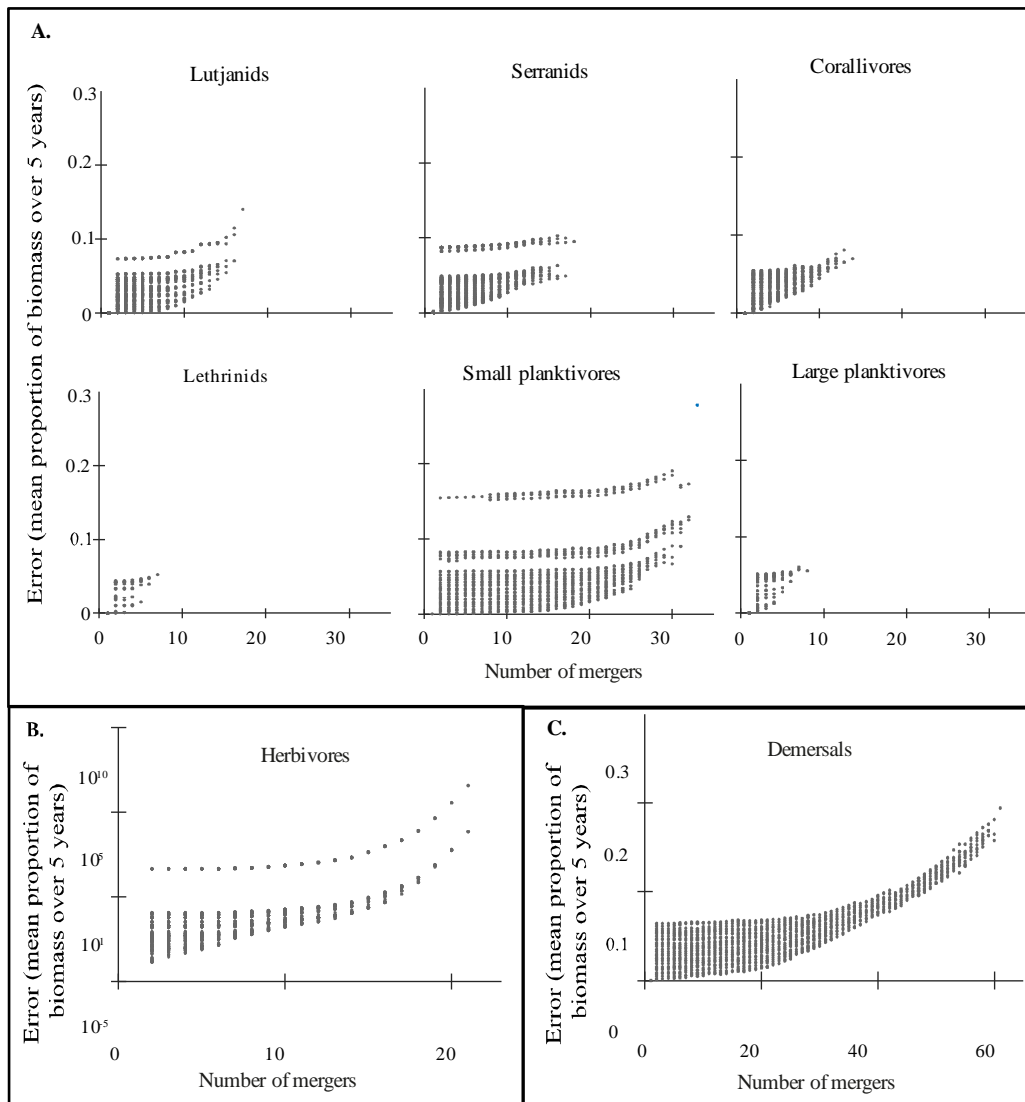
460 3.1 Error due to model coarse-graining

461 The error estimates after merging show a clear pattern, the error estimates start low then first slowly
462 then rapidly increase. However we are also interested in differentiating between the maximum and
463 minimum error at each merger. Since this represents the variation of errors that can occur depending
464 on if we merge two similar or two very different species or nodes. For most scenarios, the minimum
465 error is rather small when only few nodes are merged, however after a “tipping point” when a certain

466 “critical mass” of nodes is merged, it increases rapidly. The maximum error on the other hand starts
467 relatively high and remains mostly stable with just a small increase at the final mergers. This also
468 means that we can reduce the error when only merging a few species, however only if we are merging
469 the correct species. Finally, the spread of errors across the $\frac{i!}{2!(i-2)!}$ possible networks at each merger is
470 not even, but rather clustered. This clustering can aid the identification of the nodes (here species) that
471 should or should not be merged to avoid large losses of accuracy.

472 Figures 2 and 3 illustrate the typical patterns we observed in these experiments. Figure 2 depicts
473 errors in one (of six) threat scenario with fixed intensity (0.5) for all eight functional groups. Figure 3
474 illustrates the effects of all six threat scenarios with fixed intensity (0.5) on one selected functional
475 group (lutjanids). Both figures represents the tendencies concerning the minimum and maximum
476 errors described above from two different perspectives. The graphs of all functional groups across all
477 six scenarios and intensities can be found in the supplementary materials (S4). One feature that can be
478 observed in both figures and all panels is that there are two separate “clouds” of points. One is worse
479 in terms of errors than the other. This cloud of bad mergers is made up of a few species that whenever
480 they are added to the remainder spike the error. The regression analysis later in this section will
481 outline characteristics that can help determine these species in the hope to avoid them.

482



483

484 **Figure 2.** Example of results: Error estimates for all eight functional groups under threat scenario 1)

485 in which the biomass of pelagic fish is reduced with intensity 0.5. Panel A compares functional

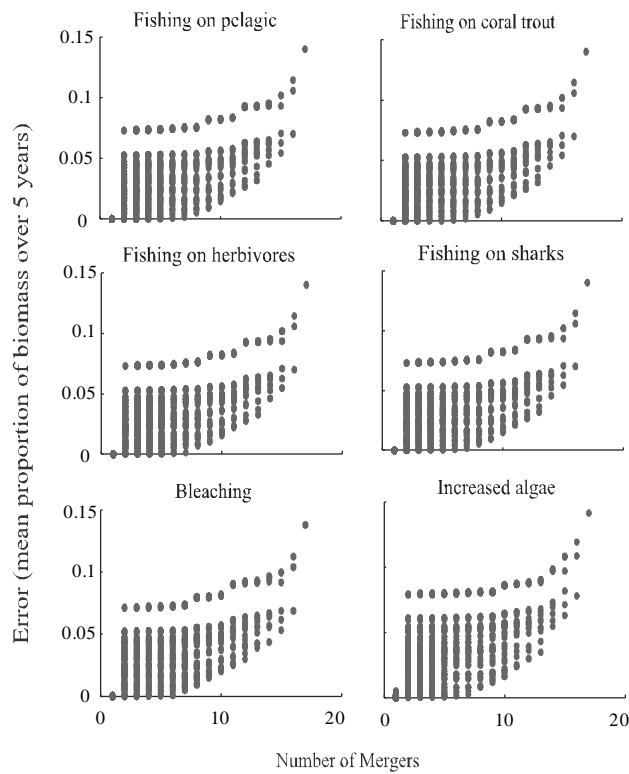
486 groups lutjanids, serranids, corallivores, lethrinids, small and large planktivores. Panel B shows

487 herbivores and panel C the functional group other demersals. Each point represents one possible

488 merger of two nodes originating from the model identified with the lowest error in the previous

489 mergers.

490



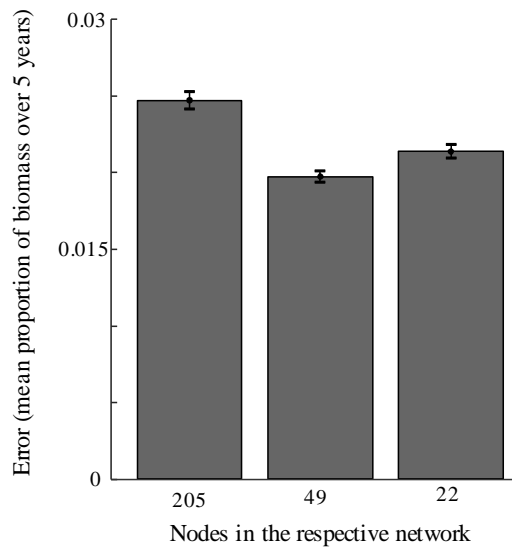
491

492 **Figure 3.** Error estimates for threat scenario 1) to 6) of intensity 0.5 for lutjanids. Each point
 493 represents one possible merger of two nodes originating from the model identified with the lowest
 494 error in the previous amount of mergers.

495

496 **3.2 Parameter versus structural uncertainty**

497 To compare the balance of parameter and structural uncertainty we are looking at the combined
 498 uncertainty (Fig.4). Since 205 nodes is the highest resolution considered here (reference point for the
 499 structural uncertainty), it only includes parameter uncertainty with the structural uncertainty being
 500 zero (according to our assumption). The 49 and the 22 node model both include structural and
 501 parameter uncertainty. We found that the 49 node model was the optimal model in the sense that the
 502 overall (structural and parameter) was the lowest. This supports the hypothesis that there is an optimal
 503 resolution even when taking into account the parameter uncertainty.



504

505 **Figure 4.** Average error estimates with standard errors for the different resolutions.

506

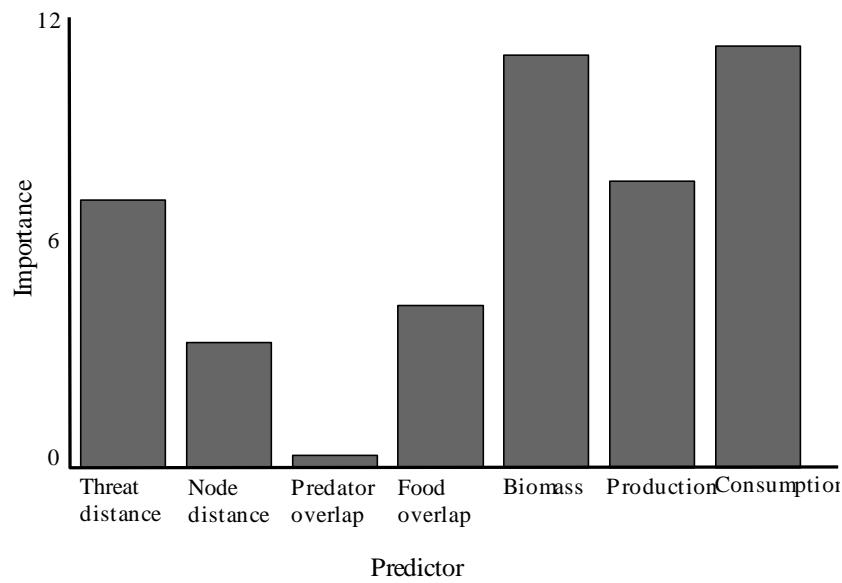
507 **3.3 Variable Importance**

508 The bagged trees analysis shows that we can explain most of the variation in the error of a particular
 509 merger when considering how similar in in the predictor variables the two nodes are ($R^2 = 0.89$).

510 The variable importance shows that the most effective predictors to determine a good merger are the
 511 biomass ratio (~ node size) and the consumption ratio (~total amount of inflow into the node) of the
 512 species (Fig. 5). The more similar (i.e. ratio of 1) biomass and consumption of the nodes are the lower
 513 the error, meaning the better the merger. Another variable of medium importance is the difference in
 514 trophic level between the threat scenario and the merged groups (Fig.5). In this case, the smaller the
 515 difference between the trophic level of both nodes the smaller the error, causing a better merger. It is
 516 interesting that the similarities in parental or child nodes between nodes to be merged (i.e. similarities
 517 in predators and food sources for the merged species) are of low value to determine the direction of
 518 merging. In fact, predator overlap is largely associated with all the other predictors (Fig.6), and hence
 519 it could be removed from the analysis without any loss of predictive power ($R^2 = 0.8$).

520

521

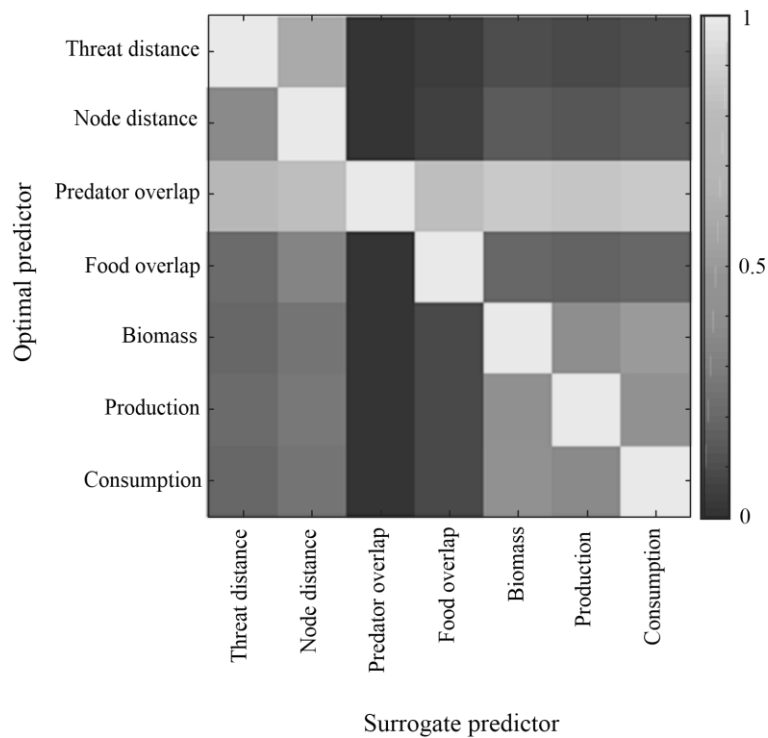


522

523 **Figure 5.** Predictor importance for the regression model

524 (relative measure of importance within the model, no units)

525



526

527 **Figure 6:** Association between predictors in the tree

528

529 **4 Conclusions and Discussion**

530

531 This study provides insights into structural uncertainty and more specifically into what role the
532 resolution plays in complex network-based models. Overall, the study gives a good indication of how
533 structural uncertainty in form of the trophic resolution could be better integrated into the process of
534 constructing the models based on complex networks. Here we considered an Ecopath model as a
535 prominent example in ecological modelling. We found that different levels of model resolution can
536 change the error in estimating model outcomes in response to exogenous shocks. While higher
537 resolutions always reduce the structural uncertainty this might not be the best resolution overall.
538 Besides the computational capacity needed for high resolution models, the overall (parameter and
539 structural) uncertainty is lowest at a medium resolution. This resolution can be considered as an
540 optimal resolution and can be found by merging species with the most similar parameters for biomass
541 and total consumption. Note that there is no direct cut-off that can be applied here. The acceptable
542 ratio between biomasses of both nodes depends on species and total network resolution. The study
543 confirms that the distance in trophic level between the merged nodes and the threat can be of
544 importance. Our results highlight that the common practice of putting high importance on the ratios of
545 the size of the merged nodes and their biomasses, rather than on the rest of the network the nodes are
546 connected to, and on the directions of in- and outflows may be warranted.

547

548 A result that holds true across all considered functional groups, independently from the number of
549 species or threats is that the minimum error grows with an increasing slope when the model resolution
550 becomes coarser (i.e., with increasing the number of mergers). This means that at a medium resolution
551 level the error is substantially lower than the error at the lowest resolution level (i.e., when a complete
552 merger of all species into one group is achieved). This is not surprising since some species have
553 common ecology, and hence they are even sometimes called “redundant species”, while others differ
554 dramatically in size, food source or predator protection. Redundancy refers to different species full-
555 filling similar ecological roles (Naeem, 1998), hence not much difference can be found if these
556 species are considered as one group in the analysis. While the existence of functional redundancy is
557 still debated (Hoey and Bellwood, 2009), this analysis seems to support the hypothesis. Similar

558 analysis on different systems other than coral reefs might be useful to explore this concept further.

559 The shape of the relationship between the degree of model coarse-graining and the model accuracy

560 introduces a notion of an optimal grouping of species in terms of structural uncertainty and

561 computational requirements.

562

563 This study supports the previous assumption of an opposite hump shape or seesaw between structural

564 and parameter uncertainty (Costanza and Sklar, 1985, Håkanson, 1995, Jester, 1977). When the

565 structural uncertainty is reduced (according to our assumptions entirely removed) and only parameter

566 uncertainty is considered (the 205 node resolution) then we have the highest overall uncertainty. The

567 medium resolution model produced here with 49 nodes shows the lowest overall uncertainty, i.e.

568 while the structural uncertainty is increased the lower amount of parameters resulted in a much lower

569 parameter uncertainty. On the other hand, once the resolution is reduced too much (22 nodes), the

570 structural uncertainty is so high that even combined with the now low parameter uncertainty, the

571 overall uncertainty is increased again. This in conjunction with the shape of the uncertainty estimates

572 across different numbers of mergers supports the hypothesis that an optimal resolution exists. This is

573 not just the optimal across structural uncertainty and computing power required, but also parameter

574 uncertainty. It should be considered here that we only considered errors introduced through the

575 resolution, not other sources such as the formulation of equation. However, the introduction of such

576 variations and their errors would generally only increase the structural uncertainty, not decrease it,

577 consequently the hump shape would be more and not less pronounced. Another interesting finding

578 here is that the species grouped in this optimal resolution model (here 49 nodes) are consistent for all

579 of the threat scenarios and intensities. This is especially important when the management questions

580 underlying the model are concerned with multiple threats.

581

582 When it comes to the practical implementation of these results, we would like to caution that the

583 errors calculated in this study should not be taken as an accurate measure of the true error. The error

584 calculated here is only based on resolution and due to the reference system itself having a reduced

585 resolution (lower trophic levels are grouped substantially), it represents a minimum error that we
586 would have to expect based on the merging of the fish species into functional groups. The error here is
587 supposed to guide decisions in reducing the overall error and highlight the importance of getting this
588 component “right”. The next few paragraphs will outline how we could more practically use the
589 results gathered in this study.

590

591 Common practice has often focused mainly on grouping together nodes with similar connections
592 (Cale Jr and Odell, 1980, Fulton, 2001, Gardner and Ashby, 1970, O'Neill, 1975, Wiegert, 1975). This
593 would be similar to the characteristic of predator and diet overlap in this study. However, this study
594 shows that the actual similarities of the predators is not very influential in terms of the aggregation of
595 species. Instead we found that similar abundances (here biomass) are the most important. One
596 example of considering connections over abundance, is Tudman (2001) who groups all herbivores
597 together irrespective of the large differences in their biomasses ranging from 0.01 to over 10
598 $t km^{-2} year^{-1}$. On the other hand, it has been recognized that for some very rare species it is better
599 to exclude them from the model than to merge with a species with another group (Fulton, 2001). The
600 results presented here have to be taken with some caution, however, since the data the analysis is
601 based on already assumes some similarity between then species merged, i.e. due to limitations related
602 to computing time the method does not allow a shark and a goby to be merged purely because they
603 have a similar biomass in the system. This restriction here was due to technical reasons mainly,
604 however, it also represents another commonly applied guideline: “do not aggregate serially linked
605 groups” (Fulton et al., 2003), i.e., expert knowledge should be used to provide the initial coarse
606 groupings.

607

608 In the literature, it is often suggested that Ecopath models should have as high as possible resolution
609 of the foodweb nodes that are of special interest to the question asked (Heymans et al., 2016,
610 Hollowed et al., 2000, Christensen et al., 2005). For example, if we are investigating fishing, we
611 should differentiate fish groups more explicitly than other parts of the foodweb such as, for example,

612 algae. While this is common practice and might suit the manager's needs, it has been pointed out that
613 this method could cause biased results (Fulton et al., 2003). Our study found some support to this
614 guideline, i.e., as we obtained that the difference between the trophic level of the threat and the
615 merged group had some importance, however, it needs to be recognised that it was lower than that of
616 the biomass and production. Furthermore, the optimal resolution, and specific species to group in the
617 medium sized model did not change in all of our threat scenarios. Since the threats that were used here
618 can represent anything from bleaching to high trophic level fishing, the results seem to indicate that it
619 is not important which question we are trying to answer when deciding on the species grouped within
620 each node. This contradicts some previous advice (Heymans et al., 2016) and should be further
621 investigated, especially, since it is often used as a justification to represent lower trophic levels in
622 massive groups that can represent hundreds or even thousands of species (Tudman, 2001).

623

624 Overall, when constructing a new network to underlay a foodweb model we should first and foremost
625 consider the biomass of species being merged. If a large difference exist we should rather exclude
626 than merge that respective species. Second, we should consider the amount that each species
627 consumes. Similar to biomass, we need to only merge species that are very similar in the amount they
628 consume, per biomass unit and overall. Only after these characteristics are considered we should
629 evaluate if the purpose of the individual study justifies a higher resolution at certain points within the
630 network. Please not we would recommend to only increase resolution close to points of interest rather
631 than decrease the resolution in other parts of the foodweb.

632

633 The future use of this study is twofold. The study can have a direct use for coral reef models
634 constructed in the future. The optimal groupings found here as well as the total amount of uncertainty
635 found can be utilised for any model of this system. However, the results from this study can reach
636 further since it provides information on how to aggregate nodes in any network model independent of
637 its use. In conclusion, this study is a good foundation for further investigation and the better
638 integration of structural uncertainty in ecosystem models, but also other network based models. As

639 long as merging nodes that are not serially linked, the most important determinant of uncertainty is the
640 size ratio of the merged nodes and their total outflow. This can give guidance to future models to
641 manage uncertainty caused by a coarser resolution which modellers have to accept in return for
642 feasible computing resources.

643

644 **5 Acknowledgements**

645

646 This research has been mostly supported by the YSSP summer fellowship of the International Institute
647 of Applied System Analysis (IIASA). It has also received some support from the HPC at James Cook
648 University and at Queensland University of Technology.

649

650

651 **6 References**

652

653 BELLWOOD, D. R. & FULTON, C. J. 2008. Sediment-mediated suppression of
654 herbivory on coral reefs: Decreasing resilience to rising sea-levels and climate
655 change? *Limnology and Oceanography*, 53, 2695-2701.

656 CALE JR, W. G. & ODELL, P. L. 1980. Behavior of aggregate state variables in
657 ecosystem models. *Mathematical Biosciences*, 49, 121-137.

658 CHRISTENSEN, V. & WALTERS, C. J. 2004. Ecopath with Ecosim: methods,
659 capabilities and limitations. *Ecological modelling*, 172, 109-139.

660 CHRISTENSEN, V., WALTERS, C. J. & PAULY, D. 2005. Ecopath with Ecosim: a user's
661 guide. *Fisheries Centre, University of British Columbia, Vancouver*, 154.

662 COLLÉTER, M., VALLS, A., GUITTON, J., GASCUEL, D., PAULY, D. &
663 CHRISTENSEN, V. 2015. Global overview of the applications of the Ecopath

- 664 with Ecosim modeling approach using the EcoBase models repository.
665 *Ecological Modelling*, 302, 42-53.
- 666 COSTANZA, R. & SKLAR, F. H. 1985. Articulation, accuracy and effectiveness of
667 mathematical models: a review of freshwater wetland applications. *Ecological*
668 *modelling*, 27, 45-68.
- 669 COSTANZA, R., WAINGER, L., FOLKE, C. & MÄLER, K.-G. 1993. Modeling complex
670 ecological economic systems: toward an evolutionary, dynamic understanding
671 of people and nature. *Ecosystem Management*. Springer.
- 672 FROESE, R. & PAULY, D. 2017. FishBase.
- 673 FULTON, E. A. 2001. *The effects of model structure and complexity on the behaviour and*
674 *performance of marine ecosystem models*. University of Tasmania.
- 675 FULTON, E. A., SMITH, A. D. & JOHNSON, C. R. 2003. Effect of complexity on marine
676 ecosystem models. *Marine Ecology Progress Series*, 253, 1-16.
- 677 GARDNER, M. R. & ASHBY, W. R. 1970. Connectance of Large Dynamic (Cybernetic)
678 Systems: Critical Values for Stability. *Nature*, 228, 784.
- 679 GORDON, I. J. 2007. Linking land to ocean: feedbacks in the management of socio-
680 ecological systems in the Great Barrier Reef catchments. *Hydrobiologia*, 591, 25-
681 33.
- 682 HÅKANSON, L. 1995. Optimal size of predictive models. *Ecological Modelling*, 78, 195-
683 204.
- 684 HEARON, J. Z. 1963. Theorems on linear systems. *Annals of the New York Academy of*
685 *Sciences*, 108, 36-68.

- 686 HEYMANS, J. J., COLL, M., LINK, J. S., MACKINSON, S., STEENBEEK, J., WALTERS,
687 C. & CHRISTENSEN, V. 2016. Best practice in Ecopath with Ecosim food-web
688 models for ecosystem-based management. *Ecological Modelling*, 331, 173-184.
- 689 HILL, S. L., WATTERS, G. M., PUNT, A. E., MCALLISTER, M. K., QUÉRÉ, C. L. &
690 TURNER, J. 2007. Model uncertainty in the ecosystem approach to fisheries.
691 *Fish and Fisheries*, 8, 315-336.
- 692 HOEY, A. S. & BELLWOOD, D. R. 2009. Limited functional redundancy in a high
693 diversity system: single species dominates key ecological process on coral reefs.
694 *Ecosystems*, 12, 1316-1328.
- 695 HOLLOWED, A. B., BAX, N., BEAMISH, R., COLLIE, J., FOGARTY, M.,
696 LIVINGSTON, P., POPE, J. & RICE, J. C. 2000. Are multispecies models an
697 improvement on single-species models for measuring fishing impacts on
698 marine ecosystems? *ICES Journal of Marine Science: Journal du Conseil*, 57, 707-
699 719.
- 700 IWASA, Y., ANDREASEN, V. & LEVIN, S. 1987. Aggregation in model ecosystems. I.
701 Perfect aggregation. *Ecological Modelling*, 37, 287-302.
- 702 JESTER, D. B. 1977. *A general population dynamics theory for largemouth bass fisheries*.
703 Virginia Polytechnic Institute and State University.
- 704 LEK, S. 2007. Uncertainty in ecological models. *Ecological Modelling*, 207, 1-2.
- 705 LINK, J. S., IHDE, T., HARVEY, C., GAICHAS, S. K., FIELD, J., BRODZIAK, J.,
706 TOWNSEND, H. & PETERMAN, R. 2012. Dealing with uncertainty in
707 ecosystem models: the paradox of use for living marine resource management.
708 *Progress in Oceanography*, 102, 102-114.

- 709 LUDWIG, D., HILBORN, R. & WALTERS, C. 1993. Uncertainty, resource exploitation,
710 and conservation: lessons from history. *Ecological Applications*, 3, 547-549.
- 711 MCCLANAHAN, T. R., CASTILLA, J. C., WHITE, A. T. & DEFEO, O. 2009. Healing
712 small-scale fisheries by facilitating complex socio-ecological systems. *Reviews*
713 *in Fish Biology and Fisheries*, 19, 33-47.
- 714 MENDITTO, A., PATRIARCA, M. & MAGNUSSON, B. 2007. Understanding the
715 meaning of accuracy, trueness and precision. *Accreditation and Quality*
716 *Assurance*, 12, 45-47.
- 717 MILNER-GULLAND, E. J. & SHEA, K. 2017. Embracing uncertainty in applied
718 ecology. *Journal of Applied Ecology*, n/a-n/a.
- 719 NAEEM, S. 1998. Species Redundancy and Ecosystem Reliability. *Conservation Biology*,
720 12, 39-45.
- 721 O'NEILL, R. V. 1975. Modeling in the eastern deciduous forest biome. *Systems analysis*
722 *and simulation in ecology*, 3, 49-72.
- 723 PINNEGAR, J. K., BLANCHARD, J. L., MACKINSON, S., SCOTT, R. D. & DUPLISEA,
724 D. E. 2005. Aggregation and removal of weak-links in food-web models: system
725 stability and recovery from disturbance. *Ecological Modelling*, 184, 229-248.
- 726 PRASAD, A. M., IVERSON, L. R. & LIAW, A. 2006. Newer Classification and
727 Regression Tree Techniques: Bagging and Random Forests for Ecological
728 Prediction. *Ecosystems*, 9, 181-199.
- 729 REFSGAARD, J. C., VAN DER SLUIJS, J. P., BROWN, J. & VAN DER KEUR, P. 2006.
730 A framework for dealing with uncertainty due to model structure error.
731 *Advances in Water Resources*, 29, 1586-1597.

- 732 SKINNER, D. J. C., ROCKS, S. A. & POLLARD, S. J. T. 2014. A review of uncertainty
733 in environmental risk: characterising potential natures, locations and levels.
734 *Journal of Risk Research*, 17, 195-219.
- 735 TUDMAN, P. 2001. Modelling the trophic effects of fishing on a mid-shelf coral reef
736 of the central Great Barrier Reef. *James Cook University*.
- 737 WALKER, W. E., HARREMOËS, P., ROTMANS, J., VAN DER SLUIJS, J. P., VAN
738 ASSELT, M. B., JANSSEN, P. & KRAYER VON KRAUSS, M. P. 2003. Defining
739 uncertainty: a conceptual basis for uncertainty management in model-based
740 decision support. *Integrated assessment*, 4, 5-17.
- 741 WEIJERMAN, M., FULTON, E. A., JANSSEN, A. B. G., KUIPER, J. J., LEEMANS, R.,
742 ROBSON, B. J., VAN DE LEEMPUT, I. A. & MOOIJ, W. M. 2015. How models
743 can support ecosystem-based management of coral reefs. *Progress in*
744 *Oceanography*, 138, 559-570.
- 745 WEIJERMAN, M., FULTON, E. A. & PARRISH, F. A. 2013. Comparison of coral reef
746 ecosystems along a fishing pressure gradient. *PloS one*, 8, e63797.
- 747 WIEGERT, R. G. 1975. Simulation Modeling of the Algal-Fly Components of a
748 Thermal Ecosystem: Effects of Spatial Heterogeneity, Time Delays, and Model
749 Condensation. *Systems Analysis and Simulation in Ecology, Volume III*. Elsevier.
750