

1

Supplementary Information:

2

Simultaneously estimating food web connectance and structure with uncertainty

3

Anubhav Gupta

Reinhard Furrer

Owen L. Petchey

4

21 November 2021

5

1 ABC methods

6 In the upcoming sections, we explain the Markov chain Monte Carlo and sequential Monte Carlo used to
7 parameterise the ADBM.

1.1 Markov chain Monte Carlo ABC

10 Rejection ABC does not learn anything from rejection or acceptance events. It can therefore be inefficient,
11 particularly when there is a high proportion of rejections. This may arise if prior distributions are wide and
12 the observed data are informative as they may constrain the posterior distributions relative to the priors.

13 MCMC ABC attempts to overcome this shortcoming by learning from rejections and acceptances, and thereby
14 focus tested sets of parameter values in regions of parameter space that have higher posterior probability, but
15 still explore other parts of parameter space.

16 Based on Marjoram et al. (2003), the MCMC method is as follows:

17 *Properties:*

18 As in rejection ABC except with the addition:

- 19 • A proposal distribution $K(\theta|\theta')$ which is a normal distribution

20 *Initialisation:*

- 21 • Run the rejection method to choose a set of parameter values from $\theta_0 \sim \pi(\theta)$ within the distance
 22 tolerance

23 *Sampling:*

24 for $i = 1 \dots n$

- 25 • Simulate $\theta' \sim K(\theta_i | \theta_{i-1})$
 26 • Compute model result $x_i = \text{model}(\theta')$
 27 • Compute summary statistics $s(x')$
 28 • Calculate $\alpha = \frac{K_h(d(s(x'), s(y)))\pi(\theta')K(\theta', \theta^{i-1})}{K_h(d(s(x^{i-1}), s(y)))\pi(\theta^{i-1})K(\theta^{i-1}, \theta')}$
 29 – Simulate $u \sim U(0, 1)$
 30 – If $u \leq \alpha$, $\theta^i = \theta'$ else, $\theta^i = \theta^{i-1}$

31 *Output:*

32 Construct posterior distribution using the correlated $\theta_1, \dots, \theta_n$.

33 **1.1.1 Burn in**

- 34 There is a possibility that the posterior distribution is influenced by the initial set of parameter values θ_0 .
 35 We therefore discard the first half of Markov chain and use only the second half.

36 **1.1.2 Thinning**

- 37 In the MCMC approach, the accepted parameter values are correlated as in some cases $\theta^i = \theta^{i-1}$; which
 38 reduces the independence of sampled parameter values in the posterior distribution. The Markov chains are
 39 thinned to reduce the dependence (autocorrelation < 0.05) which lowers number parameter values in the
 40 posterior distribution. The chain length was chosen such that the chain length was approximately 1000 after
 41 thinning.

42 **1.1.3 Convergence**

- 43 It is important to look into the convergence of the chains before drawing conclusions about the posterior
 44 distributions. The convergence was checked using Gelman-Rubin diagnostics. The value obtained from the
 45 Gelman-Rubin diagnostics should be less than 1.1 for convergence (Gelman and Rubin 1992).

46 **1.2 Sequential Monte Carlo ABC**

47 Sequential Monte Carlo ABC two main features: i. weighted resampling from the set of parameter values
 48 already drawn; ii. successive reduction in the distance threshold. This approach gives a set of posterior
 49 distribution where the successive posterior distribution is narrower than the previous one. SMC is more
 50 efficient than the rejection method when the posterior distribution is a lot narrower than the prior distribution.
 51 In SMC the parameters are uncorrelated and therefore do not require burn-in or assessment of convergence.

52 *Properties:*

53 As in rejection ABC except with the addition:

- 54 • A proposal distribution $K(\theta|\theta')$ which is a normal distribution
 55 • A decreasing sequence of tolerance thresholds $\epsilon_1, \dots, \epsilon_T$

56 *Initialisation:*

57 At iteration $t = 1$,

- 58 • for $i = 1, \dots, N$,
 59 – until distance $d(S(x), S(y)) < \epsilon_1$
 60 * simulate model parameter $\theta_i^{(1)} \sim \pi(\theta)$ and model output $x = \text{model}(\theta_i^{(1)})$.
 61 • Set equal probabilities $w_i^1 = 1/N$ for parameter values.
 62 • Take variance of proposal distribution τ_2^2 as twice the empirical variance of the $\theta_1^{(1)}$'s.

63 *Sampling:*

64 At iteration $2 \leq t \leq T$,

- 65 • for $i = 1, \dots, N$,
 66 – until distance $d(S(x), S(y)) < \epsilon_t$
 67 * pick θ_i^* from the $\theta_j^{(t-1)}$'s with probabilities $w_j^{(t-1)}$
 68 * generate model parameter $\theta_i^{(t)} \sim K(\theta|\theta_i^*; \tau_t^2)$ and model output $x = \text{model}(\theta_i^{(t)})$.
 69 – Set

$$w_i^{(t)} \propto \frac{\pi(\theta_i^{(t)})}{\sum_{j=1}^N w_j^{t-1} K(\theta_i^{(t)}|\theta_j^{(t-1)}; \tau_t^2)}$$

 70 – Take τ_{t+1}^2 as twice the weighted empirical variance of the $\theta_i^{(t)}$'s

71 Output:

- 72 Construct a sequence of converging posterior distributions using the parameter values: $\theta_1^t, \dots, \theta_n^t$ for all
73 $1 \leq t \leq T$

74 2 Choice of threshold value of distance

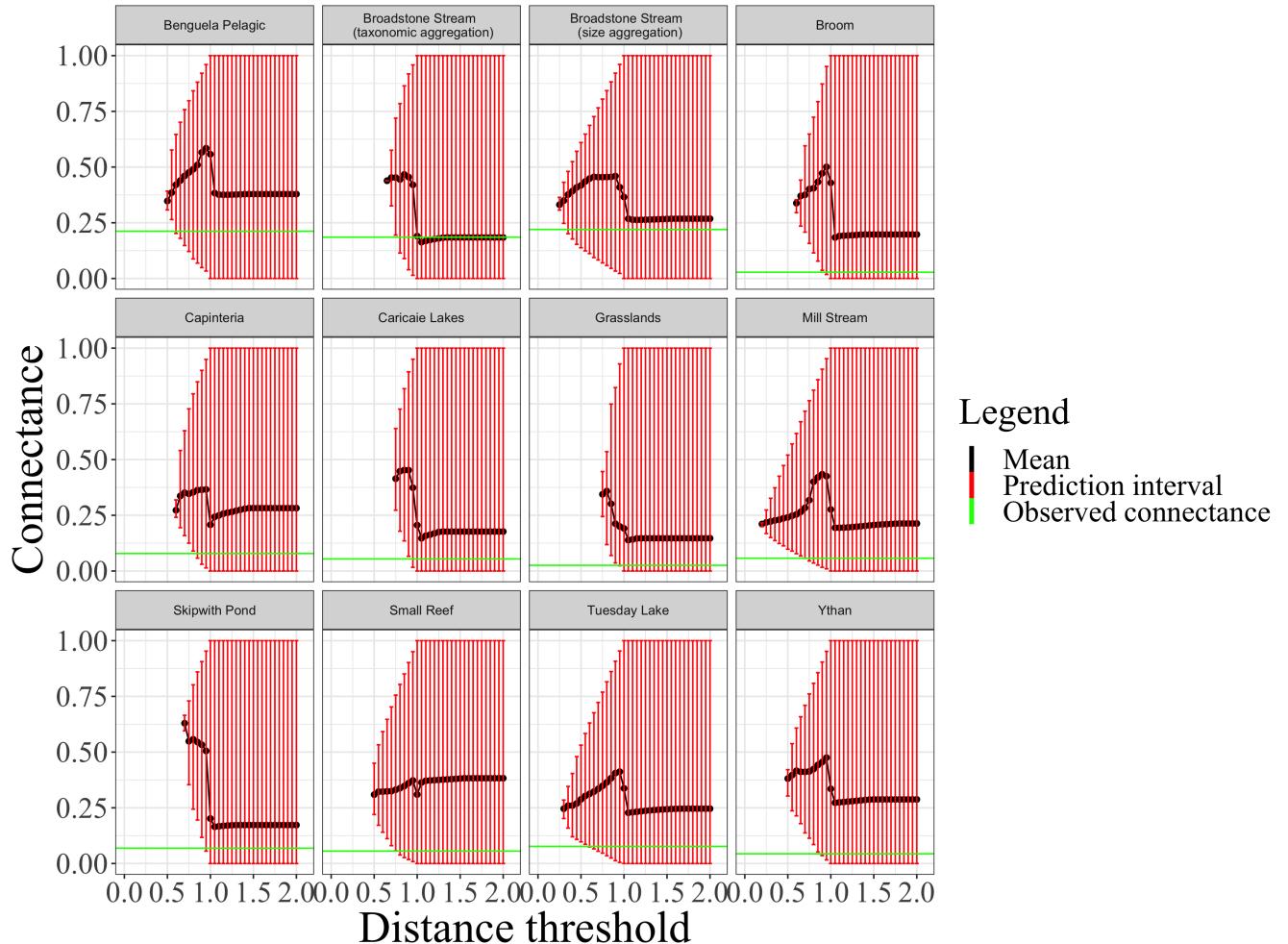


Figure S1: The prediction interval of the predicted connectance increases with increasing distance threshold. The green line and black line represent the observed connectance and mean of predicted connectance respectively.

⁷⁵ **3 Figure legends**

- ⁷⁶ Figure S1-S16: (a) Observed and predicted predation matrices of the given food web. Body size increases
⁷⁷ from left to right and top to bottom along the predation matrix. Black circles show where there were observed
⁷⁸ trophic links. The intensity of the pink circles shows the proportion of 1000 predicted food webs that had
⁷⁹ a trophic link between the corresponding species. This type of overlay is shown for two example predicted
⁸⁰ in panel (c). (b) The histogram of the number of times a link was predicted across the 1000 independently
⁸¹ predicted food webs. The red bar shows the number of pairs of species for which a trophic link was never
⁸² predicted. (c) Two predicted predation matrices correspond to the minimum and the maximum value of
⁸³ estimated b , and their sum.
- ⁸⁴ Figure S17-S32: Marginal prior and marginal posterior distribution of the ADBM parameters for 16 food
⁸⁵ webs estimated using rejection ABC.

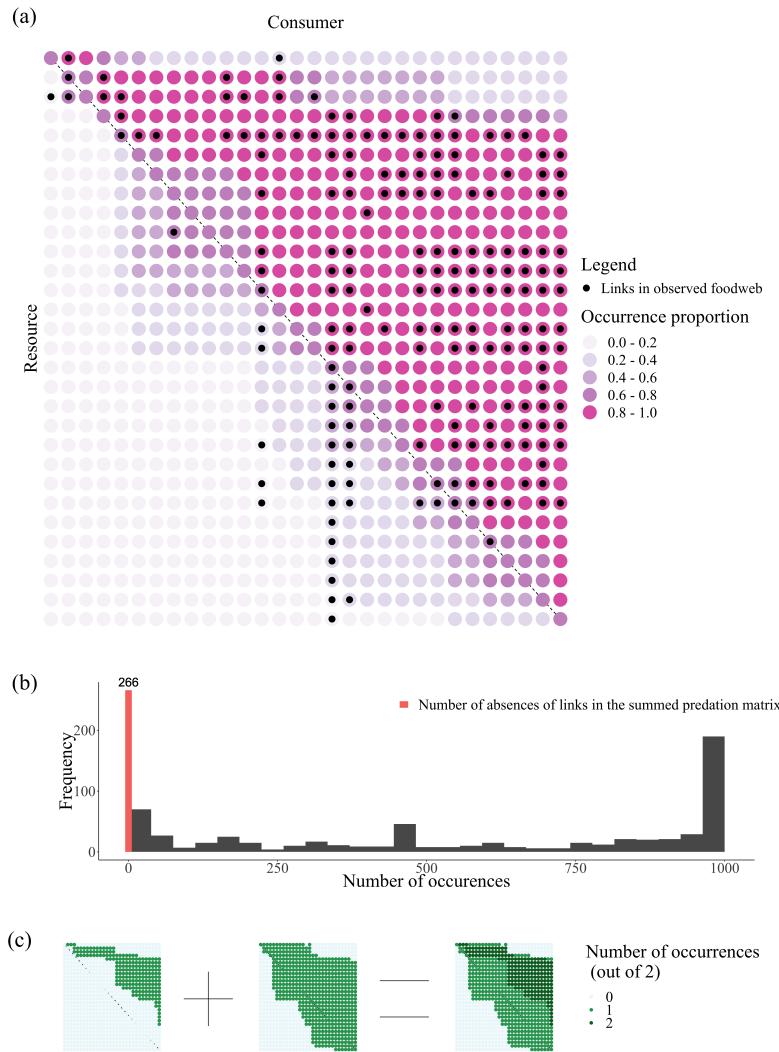


Figure S2: Benguela Pelagic

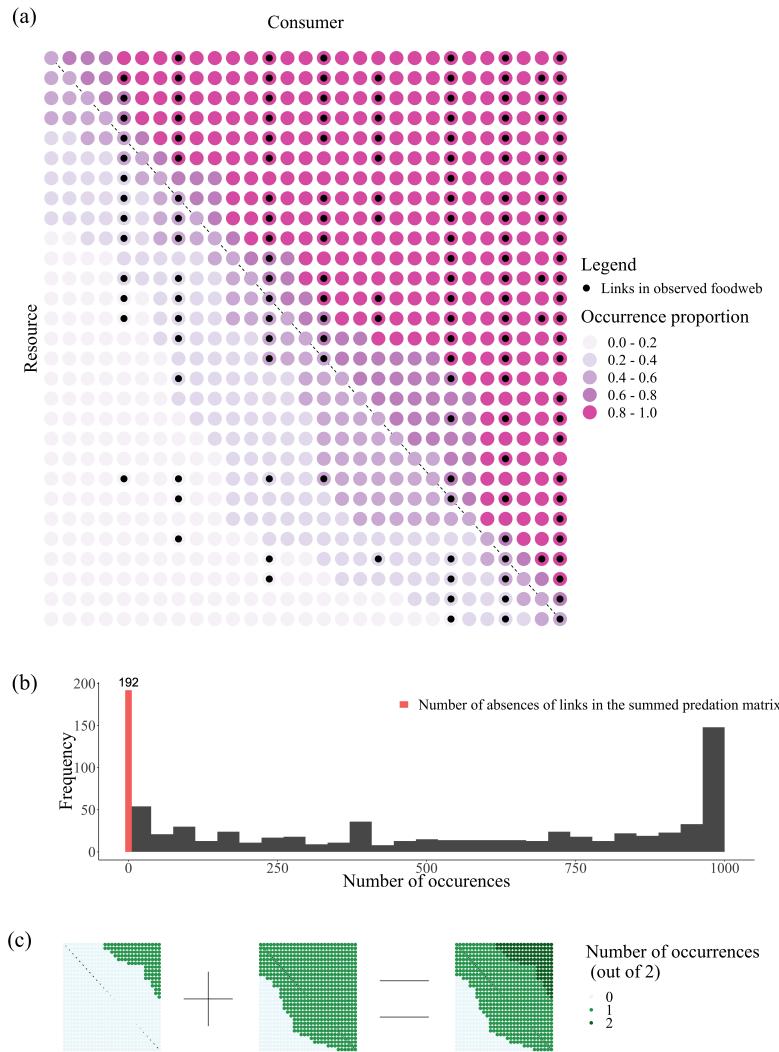


Figure S3: Broadstone Stream (taxonomic aggregation)

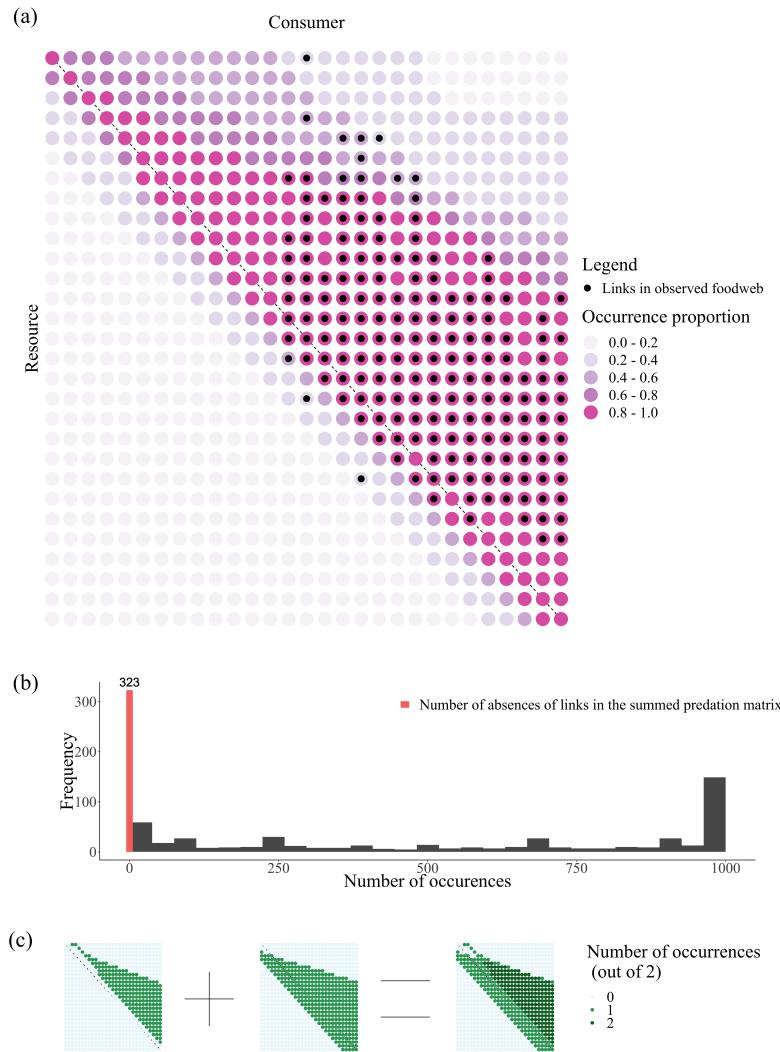


Figure S4: Broadstone Stream (size aggregation)

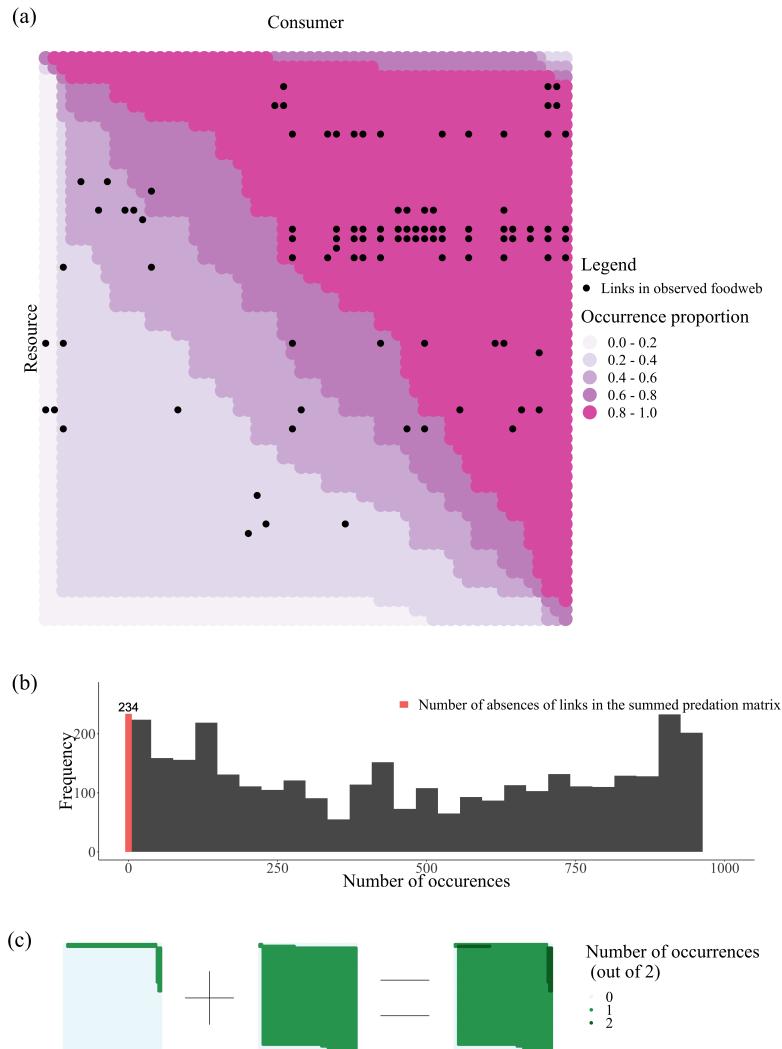


Figure S5: Broom

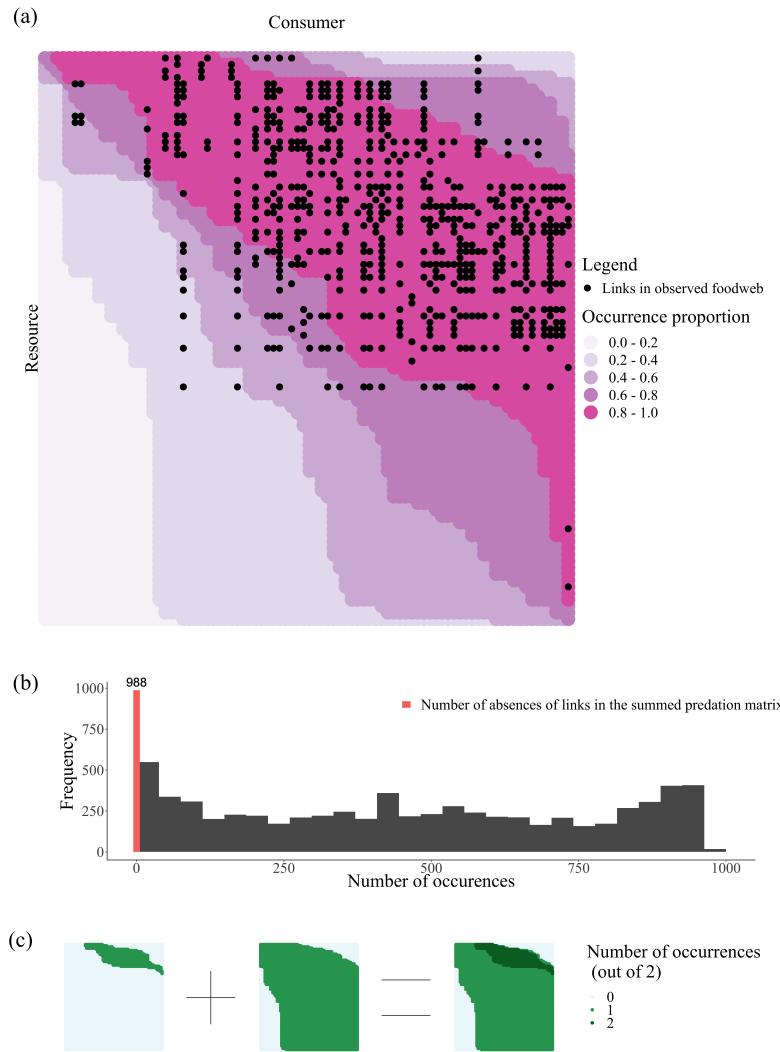


Figure S6: Capinteria

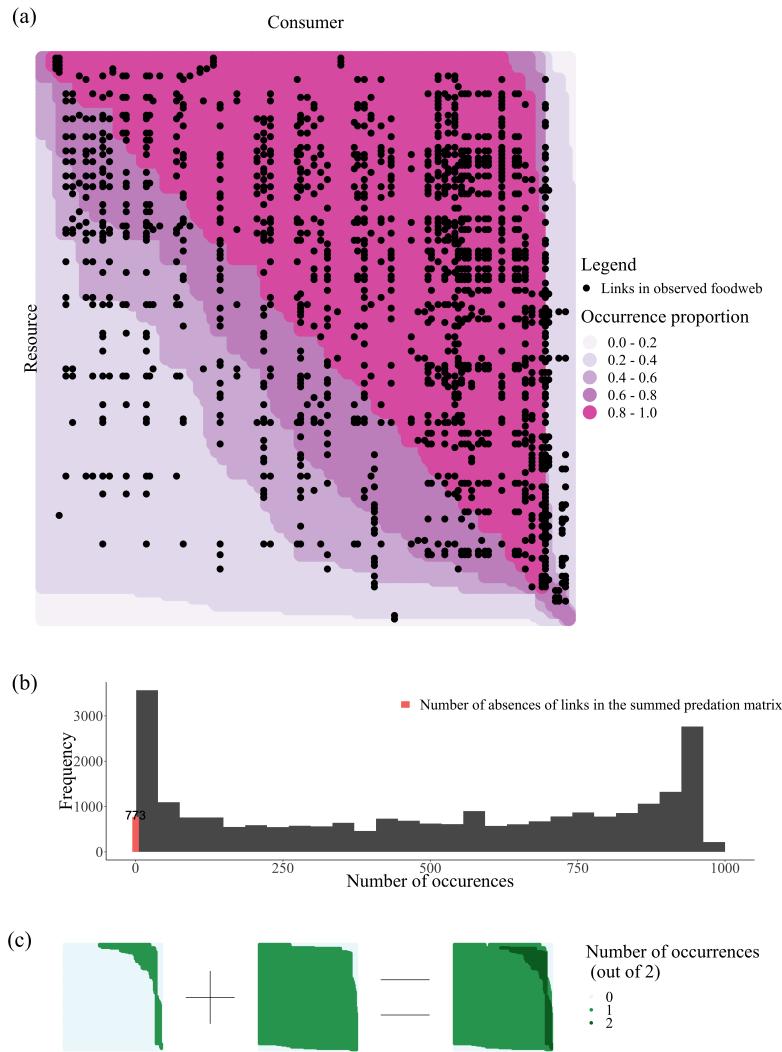


Figure S7: Caricaie Lakes

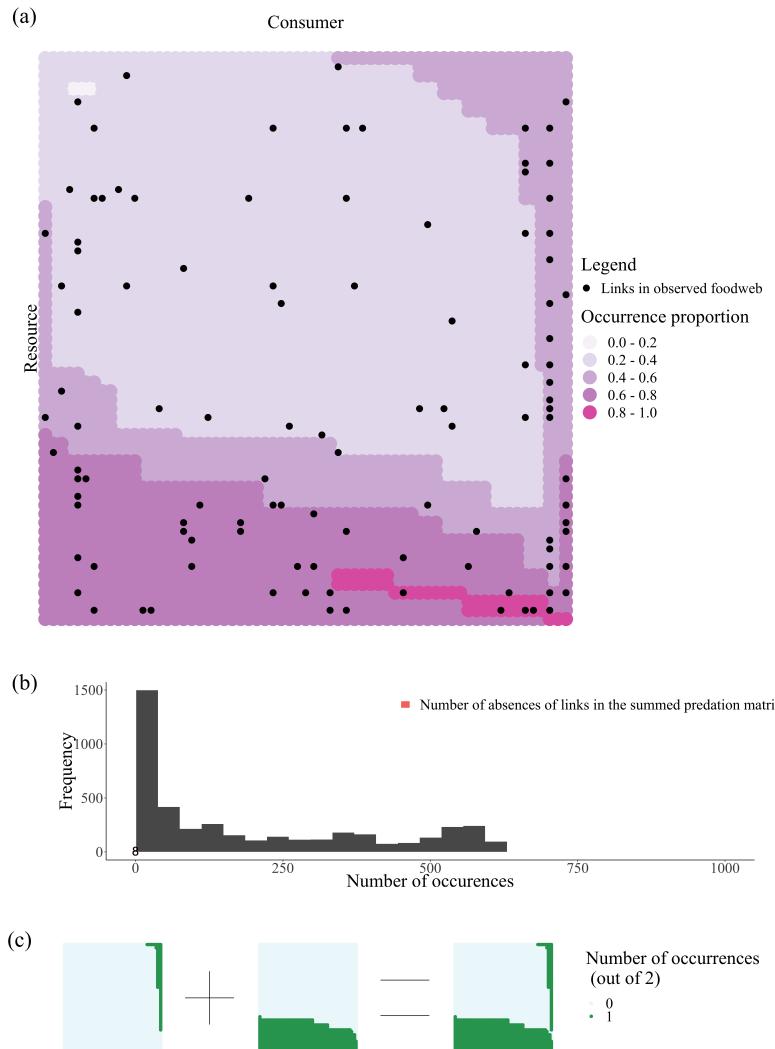


Figure S8: Grasslands

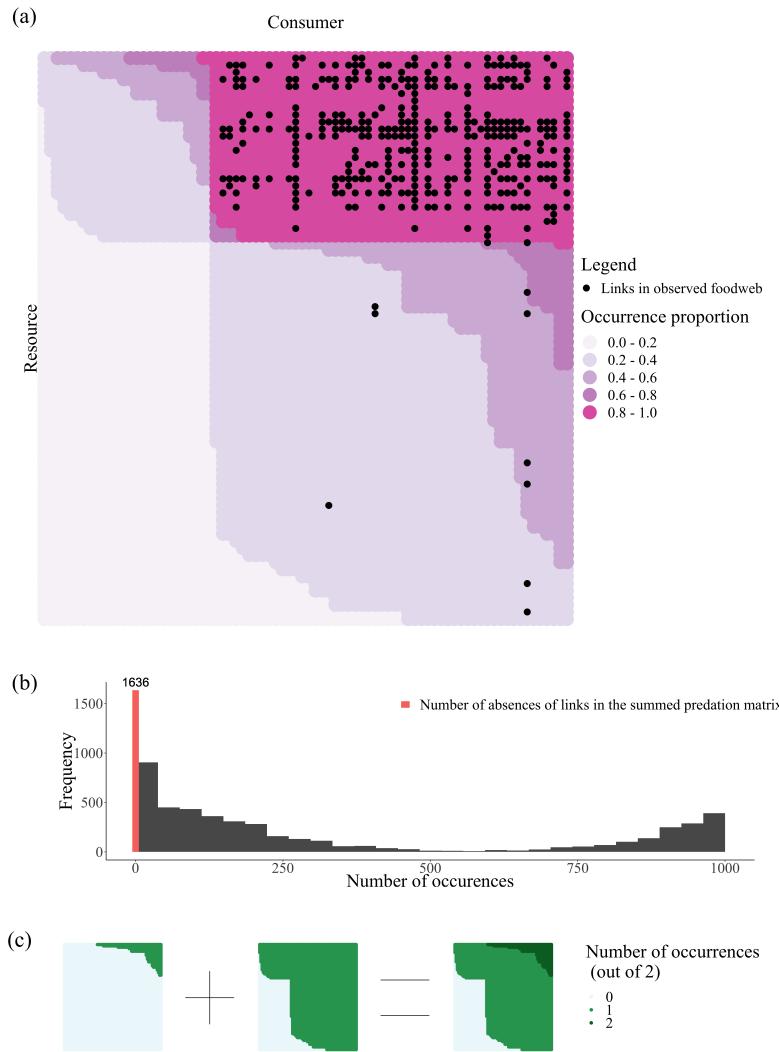


Figure S9: Mill Stream

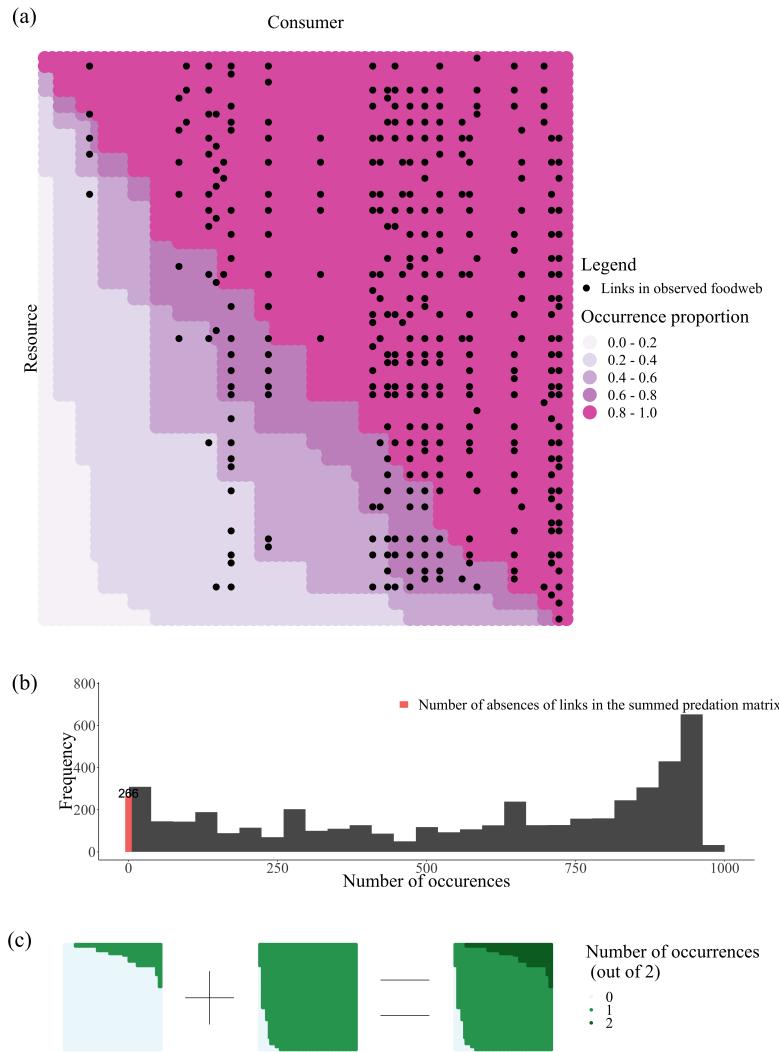


Figure S10: Skipwith Pond

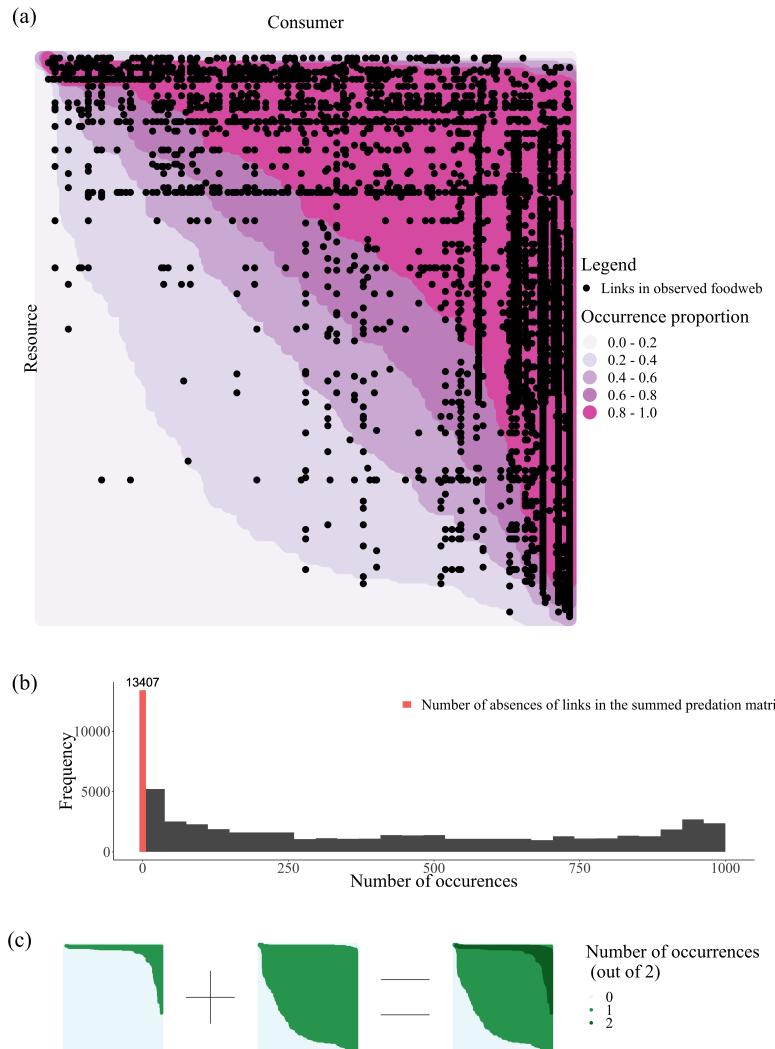


Figure S11: Small Reef

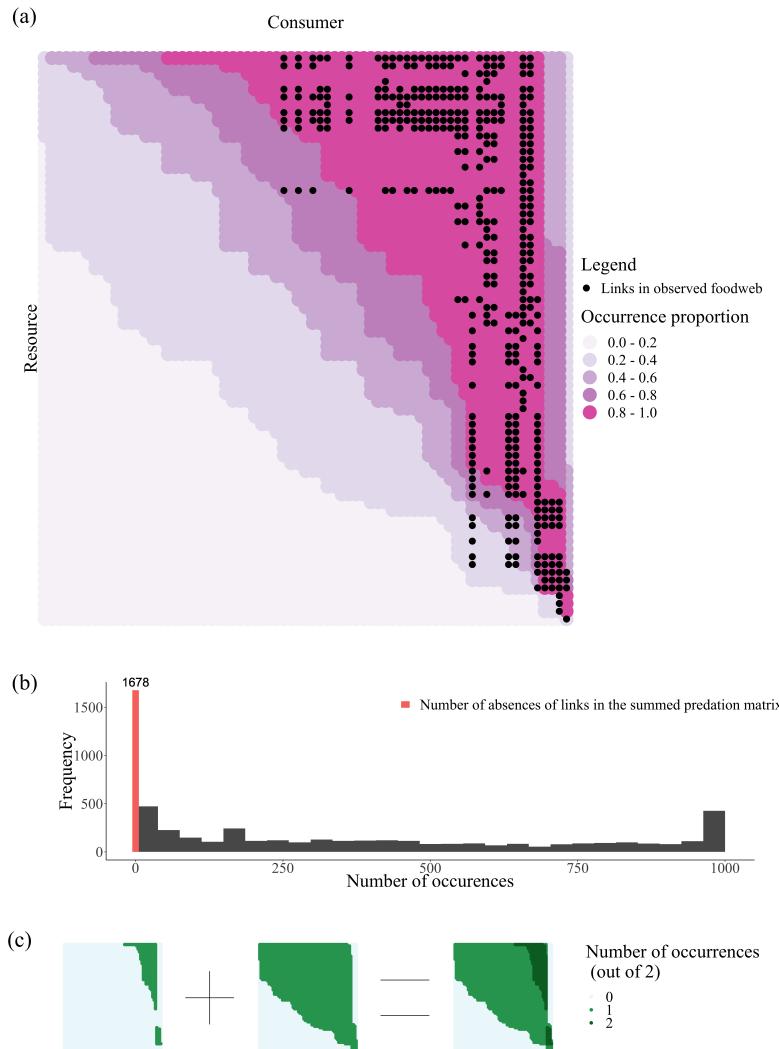


Figure S12: Tuesday Lake

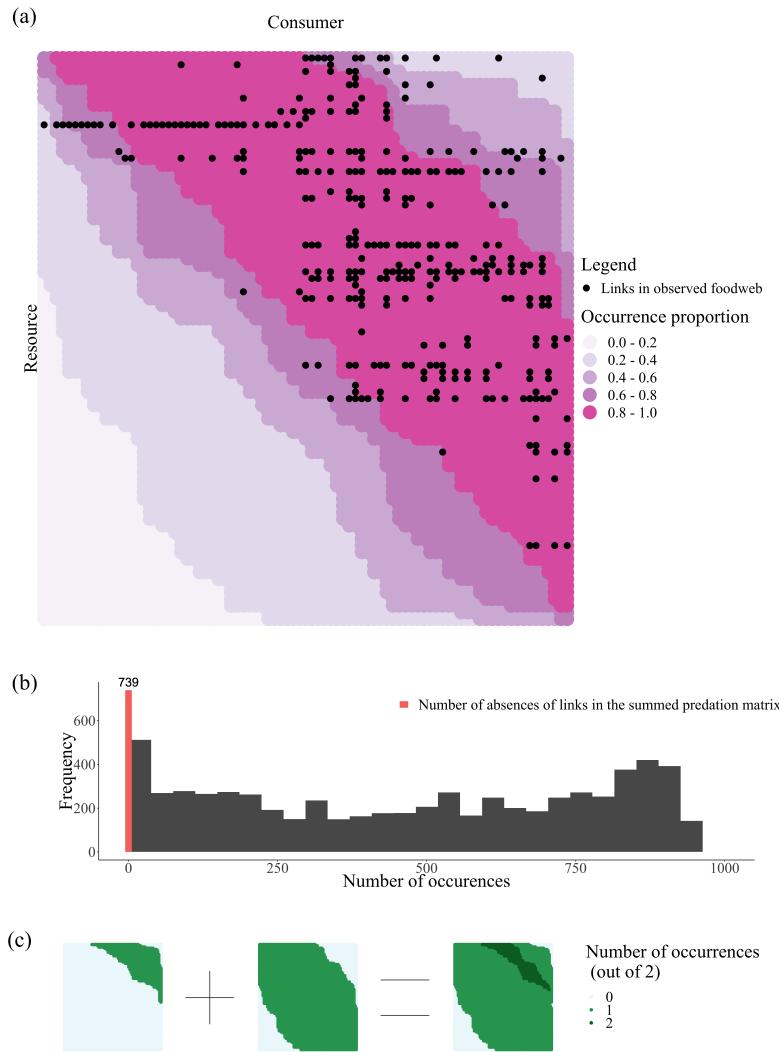


Figure S13: Ythan

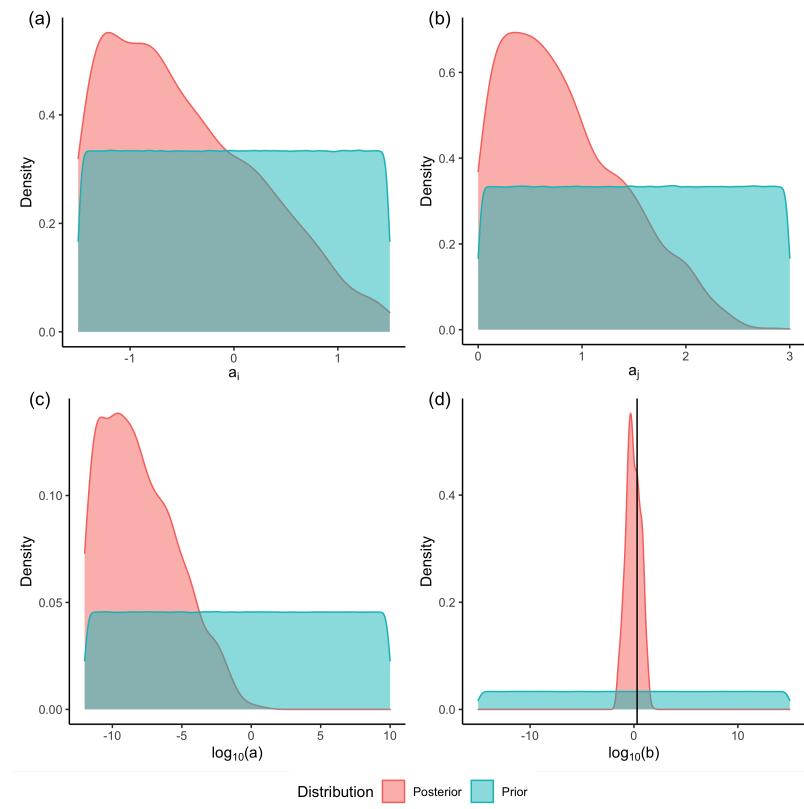


Figure S14: Benguela Pelagic

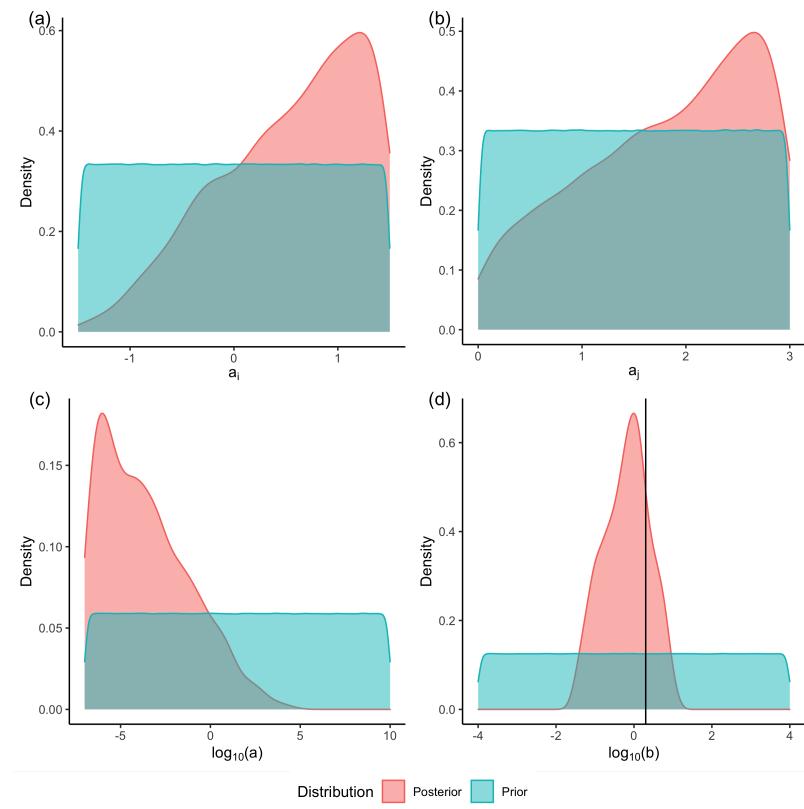


Figure S15: Broadstone Stream (taxonomic aggregation)

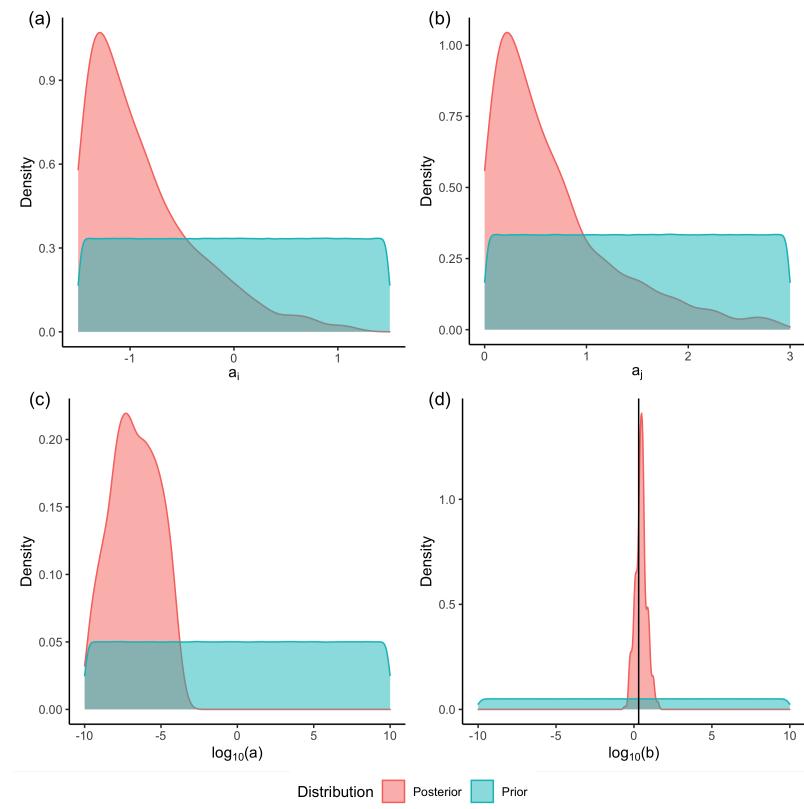


Figure S16: Broadstone Stream (size aggregation)

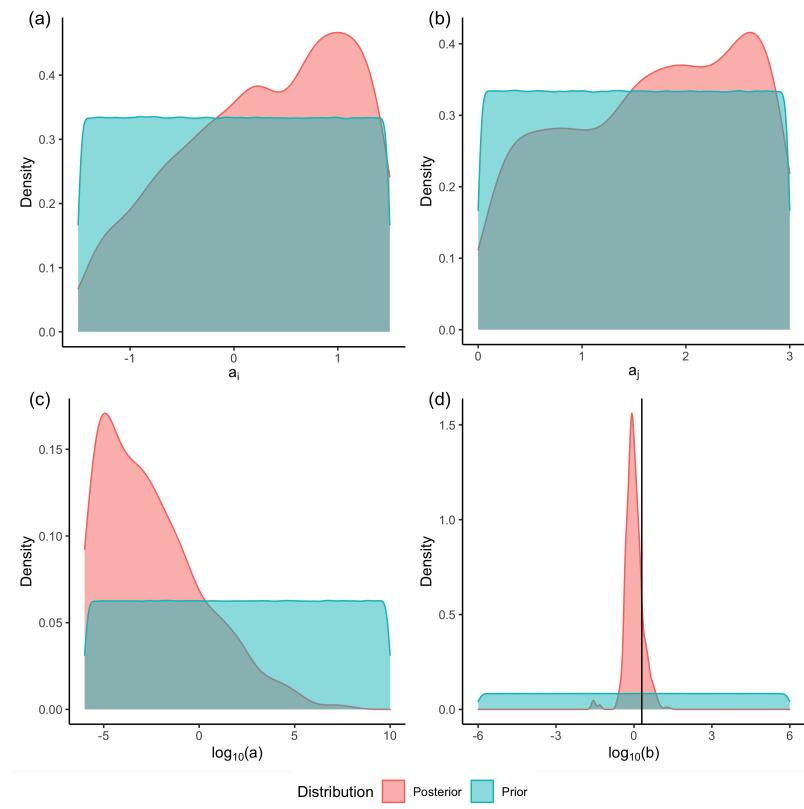


Figure S17: Broom

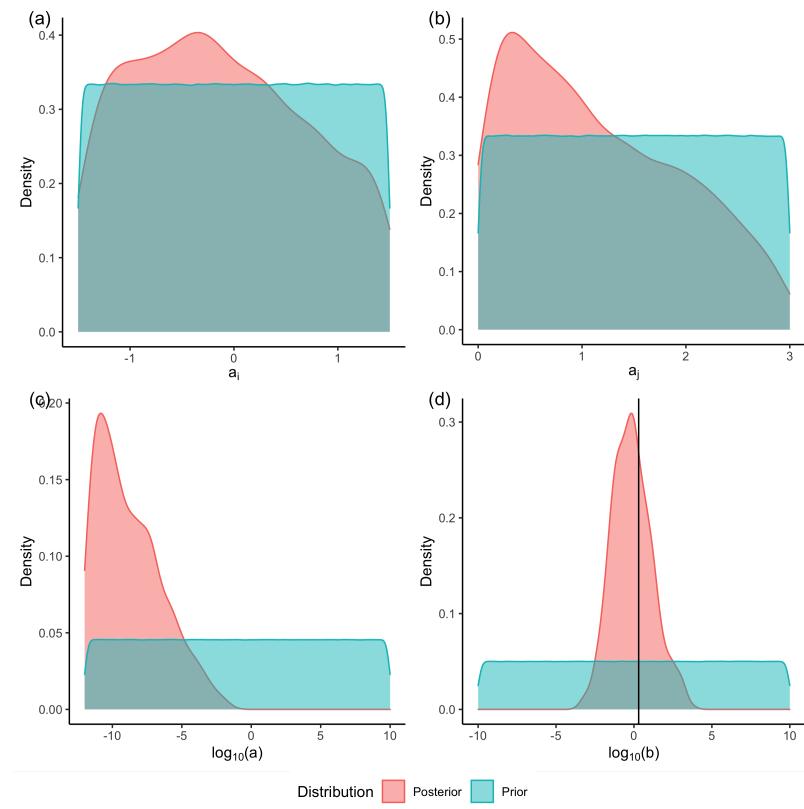


Figure S18: Capinteria

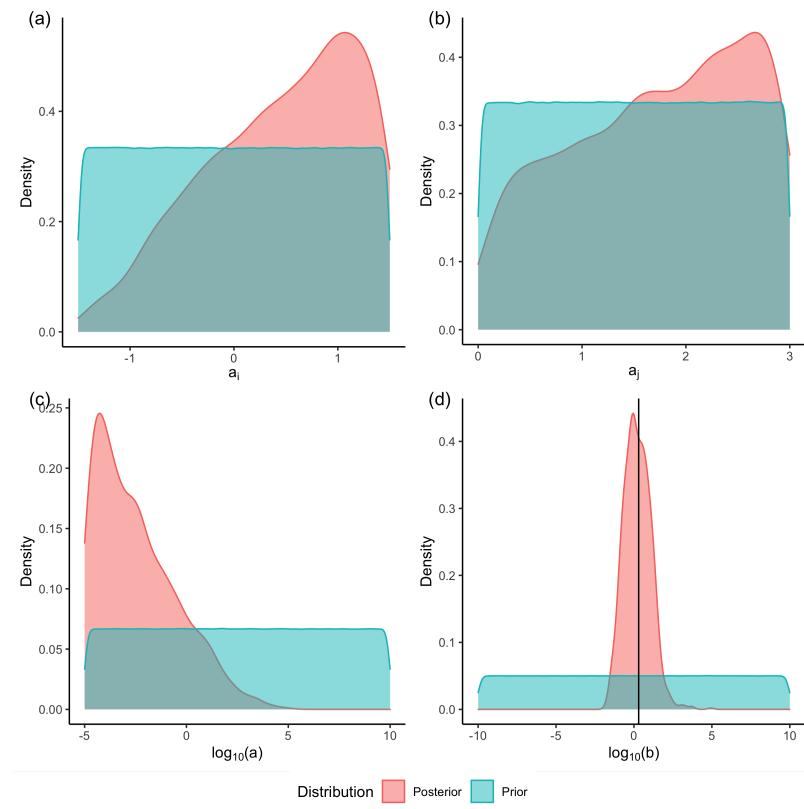


Figure S19: Caricaie Lakes

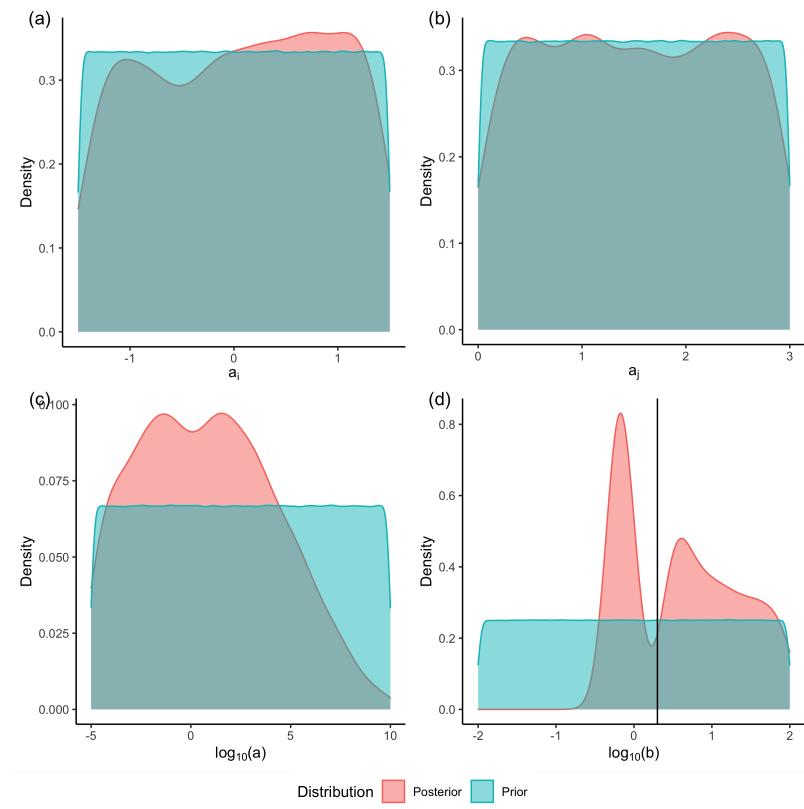


Figure S20: Grasslands

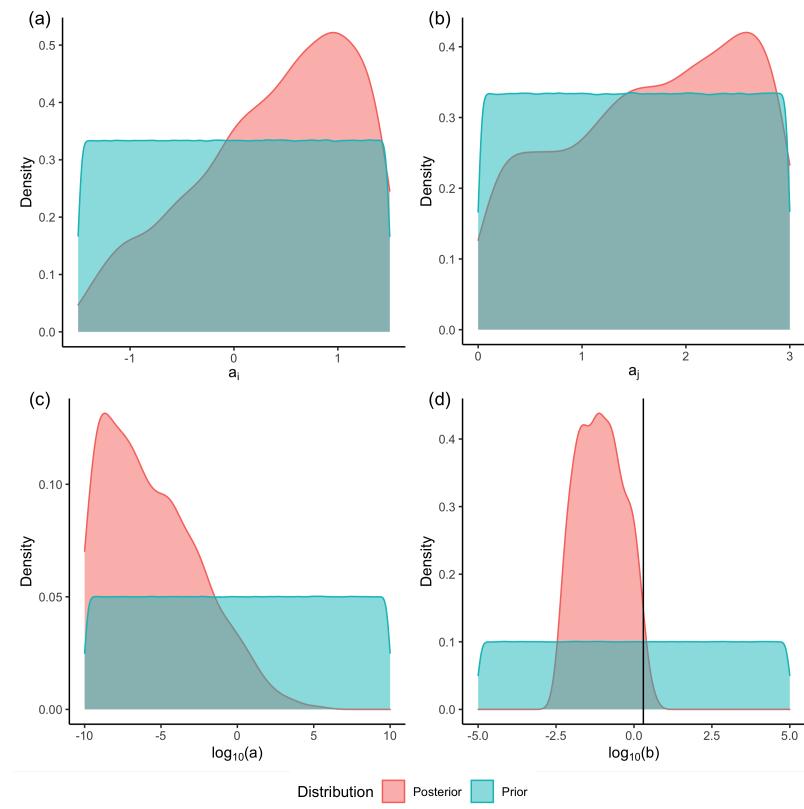


Figure S21: Mill Stream

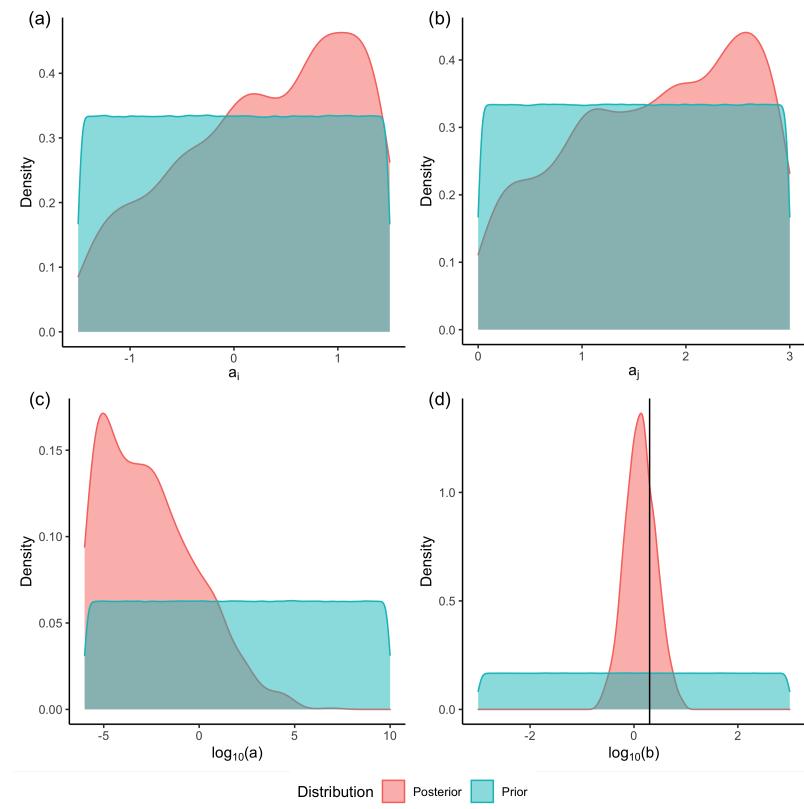


Figure S22: Skipwith Pond

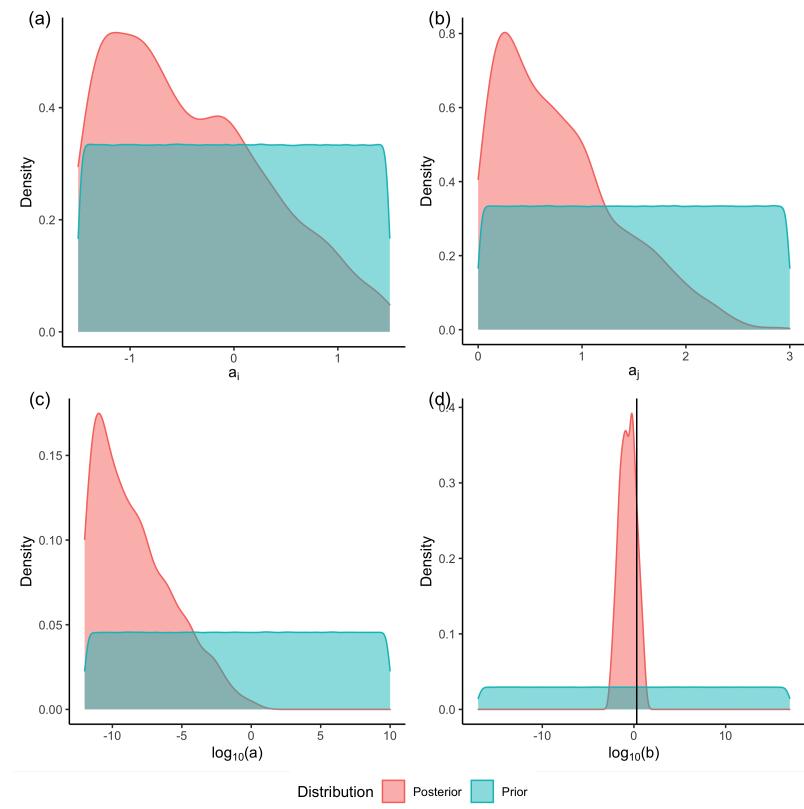


Figure S23: Small Reef

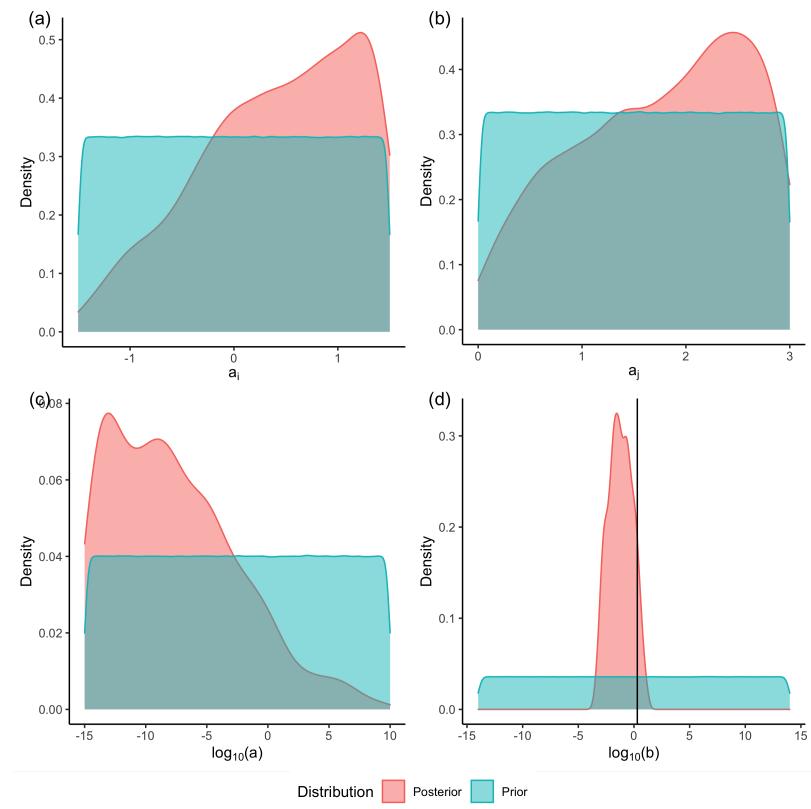


Figure S24: Tuesday Lake

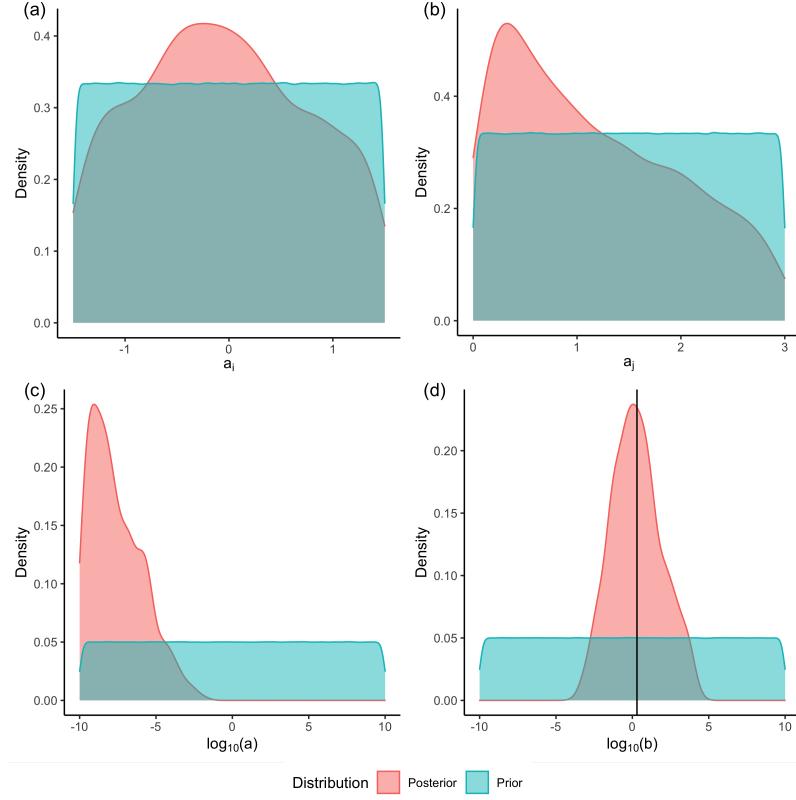


Figure S25: Ythan

86 4 Comparison between methods

87 We found consistent results among all three ABC methods (Fig. S27). The marginal posteriors of the ADBM
 88 parameters estimated using the three ABC methods for the Benguela Pelagic food web were quite similar in
 89 Fig. S26. In terms of simplicity, rejection ABC was the easiest to implement and resulted in similar results
 90 as the MCMC ABC and SMC ABC. To achieve similar number of accepted parameter values, MCMC ABC
 91 required a higher number of simulations, as MCMC ABC resulted in correlated parameter values and needed
 92 to be thinned. MCMC ABC was the most time consuming method. The SMC ABC method helped us to
 93 understand how the parameter distribution converged in response to the decrease in tolerance.

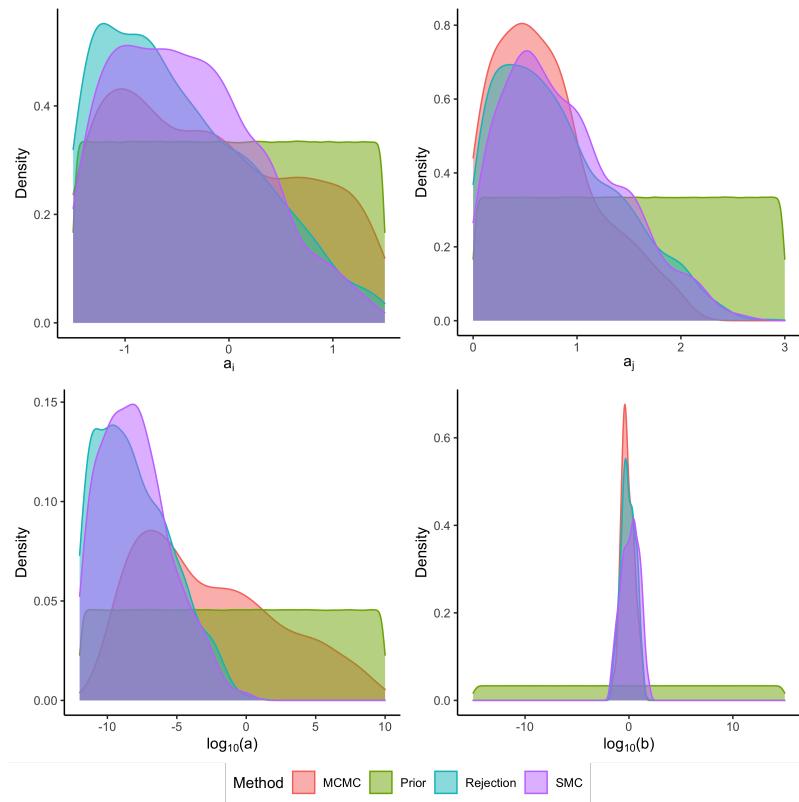


Figure S26: Marginal posterior distributions of the ADBM parameters estimated from rejection ABC, MCMC ABC and SMC ABC for Benguela Pelagic food web.

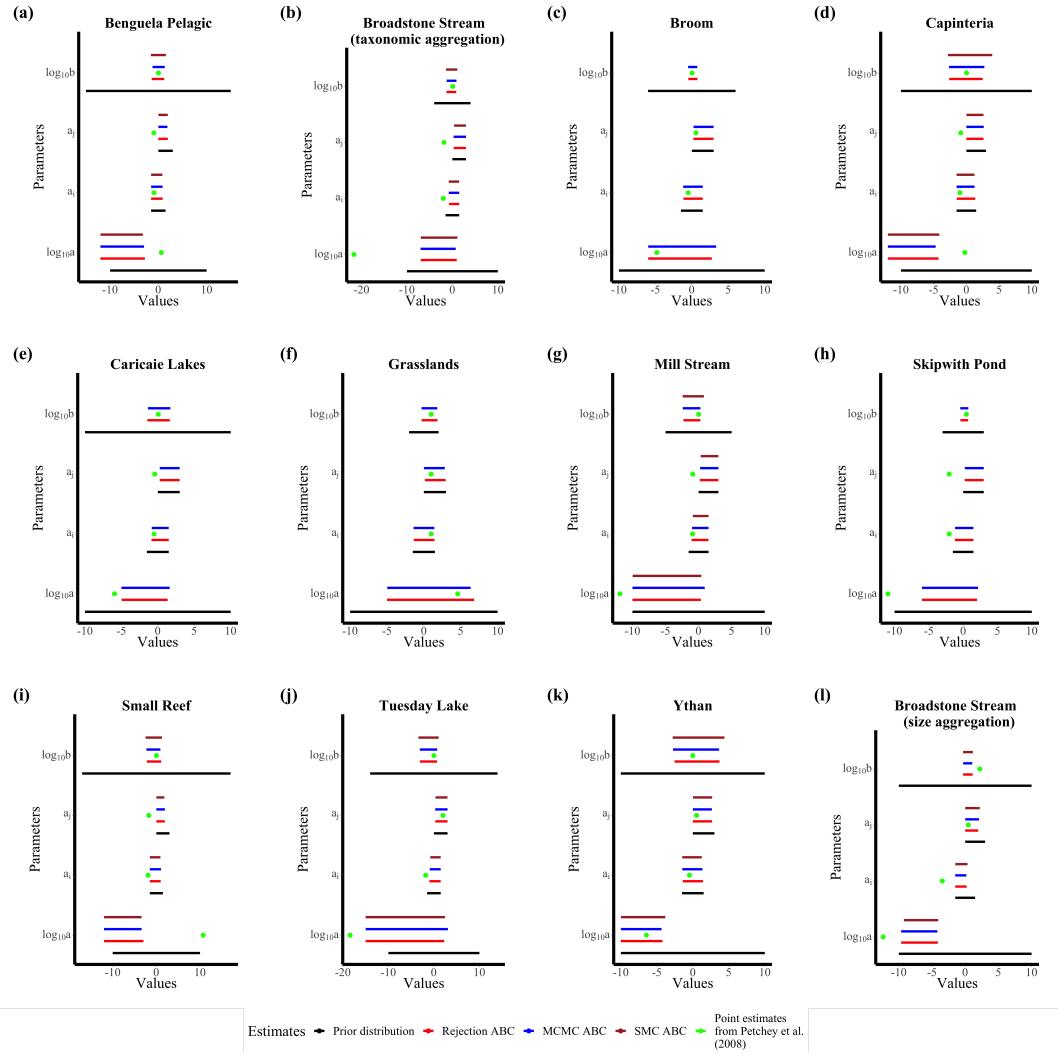


Figure S27: The 95% credible intervals of marginal posterior distributions of the ADBM parameters estimated from rejection ABC, MCMC ABC and SMC ABC for 16 food webs.

94 5 Model parameters

95 In the ADBM, the contingency model of optimal foraging predicts the diet k of each consumer j that
 96 maximises rate of energy intake given below:

$$97 \quad \frac{\sum_{i=1}^k \lambda_{ij} E_i}{1 + \sum_{i=1}^k \lambda_{ij} H_{ij}}$$

$$98 \quad \frac{\sum_{i=1}^k A_{ij} N_i E_i}{1 + \sum_{i=1}^k A_{ij} N_i H_{ij}}$$

99
$$\frac{\sum_{i=1}^k a M_i^{a_i} M_j^{a_j} n M_i^{n_i} e M_i}{1 + \sum_{i=1}^k a M_i^{a_i} M_j^{a_j} n M_i^{n_i} \frac{h}{b - \frac{M_i}{M_j}}}$$

100
$$\frac{\sum_{i=1}^k n a e M_i^{a_i + n_i + 1} M_j^{a_j}}{1 + \sum_{i=1}^k n a h M_i^{a_i + n_i} M_j^{a_j} \frac{1}{b - \frac{M_i}{M_j}}}$$

101 Since, the maximum of the above expression is affected by the product nah , it is sufficient to estimate only
 102 one of the parameters from nah , so we estimated only a . Also, parameter e does not affect the value of k
 103 where the above expression has a maximum so value of e can be arbitrary.

104 **6 Prior selection of a**

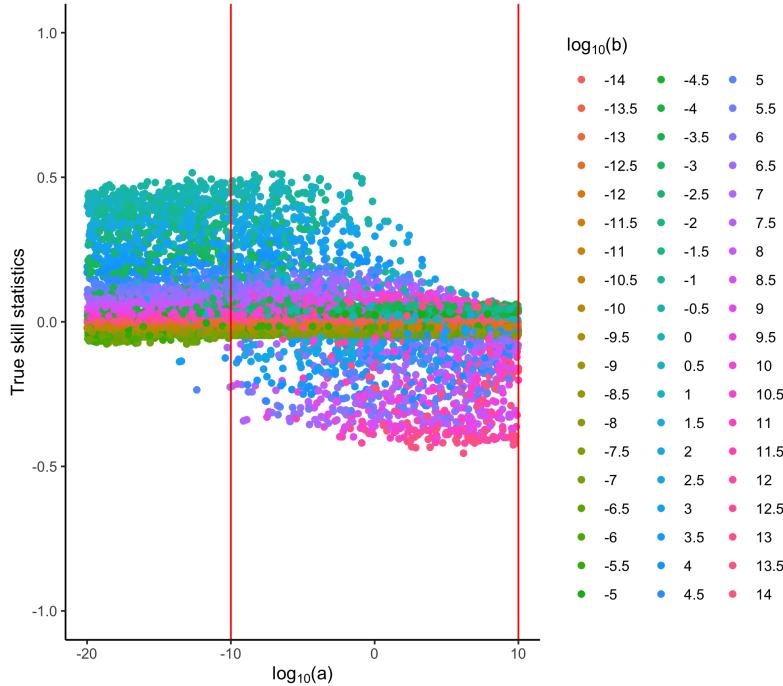


Figure S28: True skill statistics between the observed and predicted food web for Benguela Pelagic food web plotted against the attack rate scaling parameter a which is in log scale. The red lines represent the boundaries for the prior range of $\log_{10}(a)$.

105 **7 Parameter correlations**

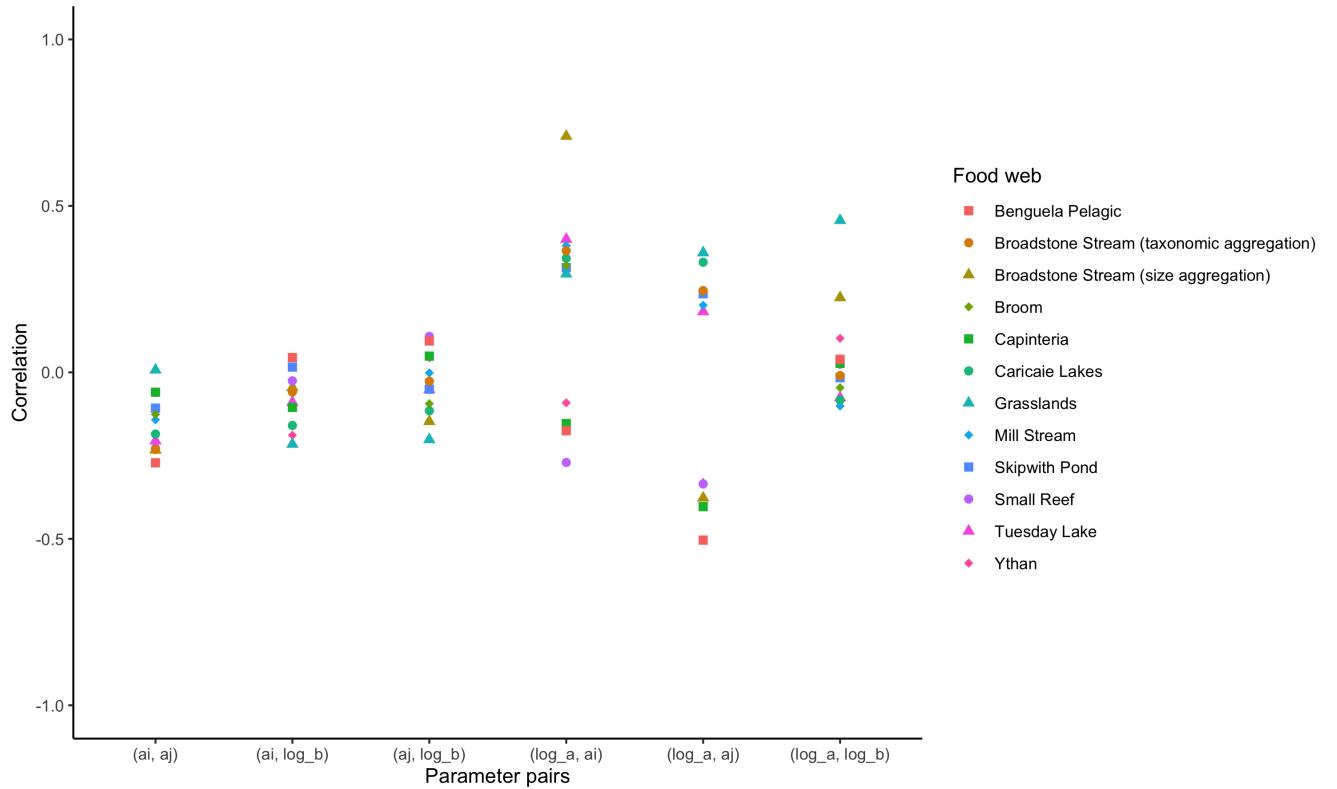


Figure S29: True skill statistics between the observed and predicted food web for Benguela Pelagic food web plotted against the attack rate scaling parameter a which is in log scale. The red lines represent the boundaries for the prior range of $\log_{10}(a)$.

106 **8 Principal Component Analysis on structural properties**

Table 1: Variance in structural properties explained by the first three principal components and correlation between connectance and first principal component.

Food web	Principal Component I	Principal Component II	Principal Component III	Correlation(PCI, Connectance)
Benguela Pelagic	0.66	0.20	0.07	-0.96
Broadstone Stream (taxonomic aggregation)	0.73	0.12	0.10	0.94
Broom	0.54	0.19	0.11	-0.89
Capinteria	0.65	0.13	0.10	0.92
Caricaie Lakes	0.59	0.18	0.10	0.85
Grasslands	0.57	0.20	0.10	0.74
Mill Stream	0.59	0.18	0.12	0.94
Skipwith Pond	0.61	0.19	0.11	0.89
Small Reef	0.67	0.11	0.09	0.95
Tuesday Lake	0.70	0.12	0.08	0.97
Ythan	0.64	0.18	0.08	0.86
Broadstone Stream (size aggregation)	0.47	0.25	0.17	0.60
Average	0.62	0.17	0.10	0.88

107 **9 Summary table of linear regression between average standard-**
108 **ized error and true skill statistic from Figs S31 and S32**

109 **9.1 Benguela Pelagic**

```
110 ##
111 ## Call:
112 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
113 ##
114 ## Residuals:
115 ##      Min       1Q   Median       3Q      Max
116 ## -0.07694 -0.02674 -0.01177  0.03389  0.07406
117 ##
118 ## Coefficients:
119 ##             Estimate Std. Error t value Pr(>|t|)
120 ## (Intercept)  0.32936   0.02082   15.82   <2e-16 ***
121 ## TSS        -0.50035   0.04517  -11.08   <2e-16 ***
122 ## ---
123 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
124 ##
125 ## Residual standard error: 0.03708 on 998 degrees of freedom
126 ## Multiple R-squared:  0.1095, Adjusted R-squared:  0.1086
127 ## F-statistic: 122.7 on 1 and 998 DF,  p-value: < 2.2e-16
```

128 **9.2 Broadstone Stream (taxonomic aggregation)**

```
129 ##
130 ## Call:
131 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
132 ##
133 ## Residuals:
134 ##      Min       1Q   Median       3Q      Max
135 ## -0.042217 -0.007716  0.000558  0.006924  0.027255
```

```

136  ##
137  ## Coefficients:
138  ##             Estimate Std. Error t value Pr(>|t|)
139  ## (Intercept)  0.34058    0.00334 101.97  <2e-16 ***
140  ## TSS        -0.36439    0.01151 -31.65  <2e-16 ***
141  ## ---
142  ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
143  ##
144  ## Residual standard error: 0.01135 on 998 degrees of freedom
145  ## Multiple R-squared:  0.5009, Adjusted R-squared:  0.5004
146  ## F-statistic: 1001 on 1 and 998 DF, p-value: < 2.2e-16

```

147 9.3 Broadstone Stream (size aggregation)

```

148  ##
149  ## Call:
150  ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
151  ##
152  ## Residuals:
153  ##      Min       1Q   Median       3Q      Max
154  ## -15.4077  -4.2887   0.1484   4.7925  13.4487
155  ##
156  ## Coefficients:
157  ##             Estimate Std. Error t value Pr(>|t|)
158  ## (Intercept)  73.714     5.097   14.46  <2e-16 ***
159  ## TSS        -83.629     7.283  -11.48  <2e-16 ***
160  ## ---
161  ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
162  ##
163  ## Residual standard error: 5.828 on 998 degrees of freedom
164  ## Multiple R-squared:  0.1167, Adjusted R-squared:  0.1158
165  ## F-statistic: 131.9 on 1 and 998 DF, p-value: < 2.2e-16

```

166 **9.4 Broom**

```
167 ##  
168 ## Call:  
169 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)  
170 ##  
171 ## Residuals:  
172 ##      Min       1Q   Median      3Q      Max  
173 ## -0.244483 -0.03674  0.01409  0.044485  0.08370  
174 ##  
175 ## Coefficients:  
176 ##              Estimate Std. Error t value Pr(>|t|)  
177 ## (Intercept)  0.437254  0.005208  83.97 < 2e-16 ***  
178 ## TSS        -0.152599  0.020292  -7.52 1.22e-13 ***  
179 ## ---  
180 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
181 ##  
182 ## Residual standard error: 0.05713 on 998 degrees of freedom  
183 ## Multiple R-squared:  0.05363,    Adjusted R-squared:  0.05268  
184 ## F-statistic: 56.55 on 1 and 998 DF,  p-value: 1.218e-13
```

185 **9.5 Capinteria**

```
186 ##  
187 ## Call:  
188 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)  
189 ##  
190 ## Residuals:  
191 ##      Min       1Q   Median      3Q      Max  
192 ## -0.21149 -0.07455  0.04157  0.07726  0.13695  
193 ##  
194 ## Coefficients:  
195 ##              Estimate Std. Error t value Pr(>|t|)
```

```

196 ## (Intercept) 0.36582    0.01153  31.729 < 2e-16 ***
197 ## TSS          -0.13714    0.02796  -4.906 1.09e-06 ***
198 ##
199 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
200 ##
201 ## Residual standard error: 0.09871 on 998 degrees of freedom
202 ## Multiple R-squared:  0.02354,   Adjusted R-squared:  0.02257
203 ## F-statistic: 24.06 on 1 and 998 DF,  p-value: 1.087e-06

```

204 9.6 Caricaie Lakes

```

205 ##
206 ## Call:
207 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
208 ##
209 ## Residuals:
210 ##      Min       1Q     Median      3Q      Max
211 ## -0.22367 -0.04988  0.02288  0.06424  0.10604
212 ##
213 ## Coefficients:
214 ##             Estimate Std. Error t value Pr(>|t|)
215 ## (Intercept) 0.34378    0.00892 38.541 < 2e-16 ***
216 ## TSS        0.28148    0.04888  5.758 1.13e-08 ***
217 ##
218 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
219 ##
220 ## Residual standard error: 0.07685 on 998 degrees of freedom
221 ## Multiple R-squared:  0.03216,   Adjusted R-squared:  0.03119
222 ## F-statistic: 33.16 on 1 and 998 DF,  p-value: 1.131e-08

```

223 9.7 Grasslands

```
224 ##
```

```

225 ## Call:
226 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
227 ##
228 ## Residuals:
229 ##      Min       1Q   Median      3Q      Max
230 ## -0.113414 -0.045792  0.001642  0.041063  0.181017
231 ##
232 ## Coefficients:
233 ##             Estimate Std. Error t value Pr(>|t|)
234 ## (Intercept) 0.190539  0.007333 25.982 < 2e-16 ***
235 ## TSS         0.373534  0.056238  6.642 5.08e-11 ***
236 ## ---
237 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
238 ##
239 ## Residual standard error: 0.05478 on 998 degrees of freedom
240 ## Multiple R-squared:  0.04233,    Adjusted R-squared:  0.04137
241 ## F-statistic: 44.12 on 1 and 998 DF,  p-value: 5.076e-11

```

242 9.8 Mill Stream

```

243 ##
244 ## Call:
245 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
246 ##
247 ## Residuals:
248 ##      Min       1Q   Median      3Q      Max
249 ## -0.31085 -0.06180 -0.03214  0.11952  0.15870
250 ##
251 ## Coefficients:
252 ##             Estimate Std. Error t value Pr(>|t|)
253 ## (Intercept) 0.49738   0.01835  27.10 <2e-16 ***
254 ## TSS        -0.42983   0.02642 -16.27 <2e-16 ***
255 ## ---

```

```

256 ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
257 ##
258 ## Residual standard error: 0.1155 on 998 degrees of freedom
259 ## Multiple R-squared: 0.2096, Adjusted R-squared: 0.2088
260 ## F-statistic: 264.6 on 1 and 998 DF, p-value: < 2.2e-16

```

261 9.9 Small Reef

```

262 ##
263 ## Call:
264 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
265 ##
266 ## Residuals:
267 ##      Min       1Q   Median       3Q      Max
268 ## -0.142108 -0.058311  0.006842  0.057750  0.090019
269 ##
270 ## Coefficients:
271 ##             Estimate Std. Error t value Pr(>|t|)
272 ## (Intercept) 0.24598    0.01060 23.202  <2e-16 ***
273 ## TSS        -0.02566    0.02490 -1.031    0.303
274 ## ---
275 ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
276 ##
277 ## Residual standard error: 0.06266 on 998 degrees of freedom
278 ## Multiple R-squared: 0.001063, Adjusted R-squared: 6.213e-05
279 ## F-statistic: 1.062 on 1 and 998 DF, p-value: 0.303

```

280 9.10 Tuesday Lake

```

281 ##
282 ## Call:
283 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
284 ##

```

```

285 ## Residuals:
286 ##      Min       1Q    Median      3Q      Max
287 ## -0.22724 -0.05970  0.02208  0.07776  0.10697
288 ##
289 ## Coefficients:
290 ##              Estimate Std. Error t value Pr(>|t|)
291 ## (Intercept) 0.44743   0.01961   22.82 <2e-16 ***
292 ## TSS        -0.38757   0.03284  -11.80 <2e-16 ***
293 ## ---
294 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
295 ##
296 ## Residual standard error: 0.08694 on 998 degrees of freedom
297 ## Multiple R-squared:  0.1225, Adjusted R-squared:  0.1216
298 ## F-statistic: 139.3 on 1 and 998 DF,  p-value: < 2.2e-16

```

299 9.11 Ythan

```

300 ##
301 ## Call:
302 ## lm(formula = mean_err ~ TSS, data = mean_err_TSS)
303 ##
304 ## Residuals:
305 ##      Min       1Q    Median      3Q      Max
306 ## -15.4077 -4.2887  0.1484  4.7925 13.4487
307 ##
308 ## Coefficients:
309 ##              Estimate Std. Error t value Pr(>|t|)
310 ## (Intercept) 73.714     5.097   14.46 <2e-16 ***
311 ## TSS        -83.629     7.283  -11.48 <2e-16 ***
312 ## ---
313 ## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
314 ##
315 ## Residual standard error: 5.828 on 998 degrees of freedom

```

```

316 ## Multiple R-squared: 0.1167, Adjusted R-squared: 0.1158
317 ## F-statistic: 131.9 on 1 and 998 DF, p-value: < 2.2e-16

```

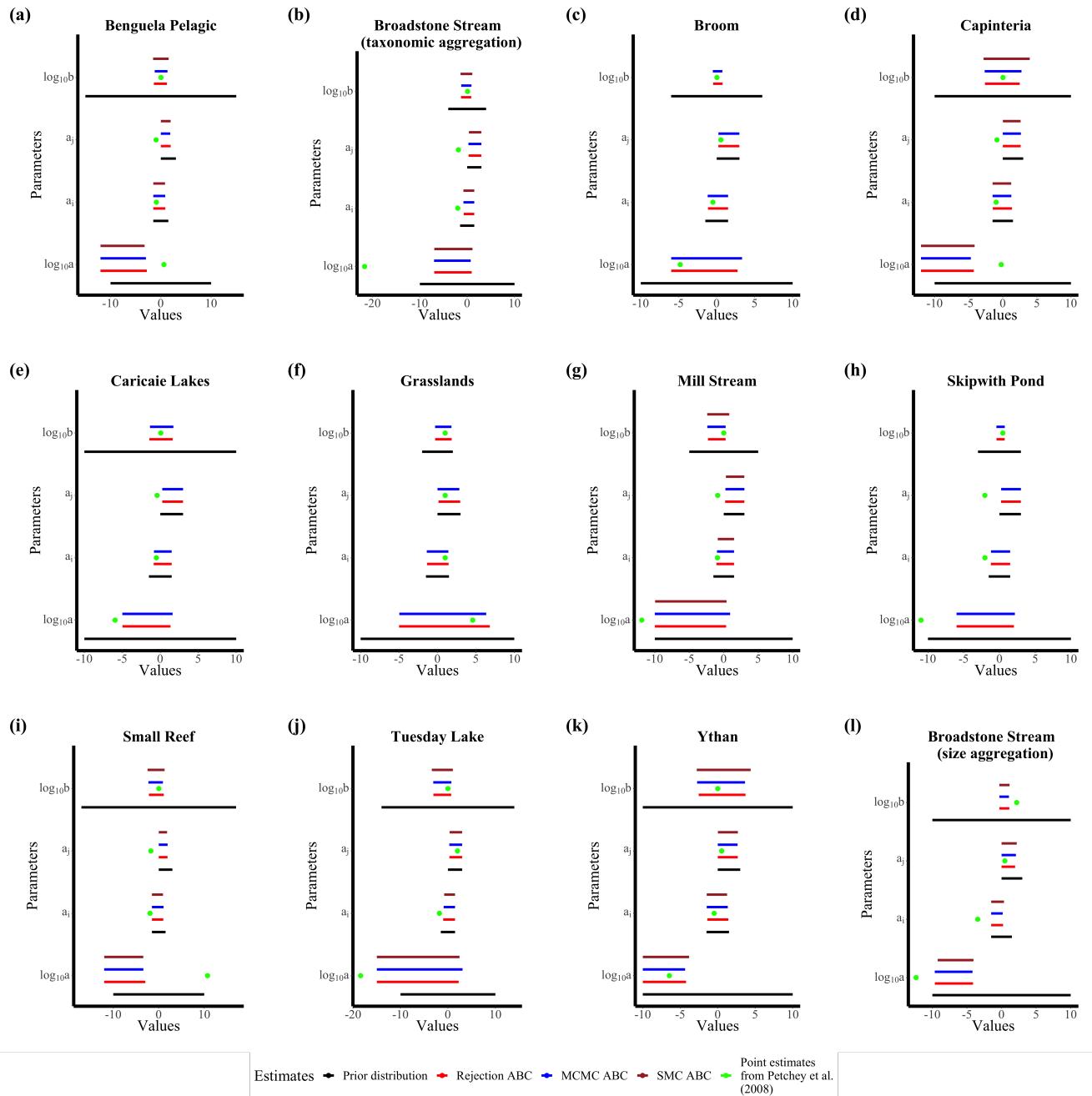


Figure S30: The 95% credible intervals of marginal posterior distributions of the ADBM parameters estimated from rejection ABC, MCMC ABC and SMC ABC for 16 food webs.

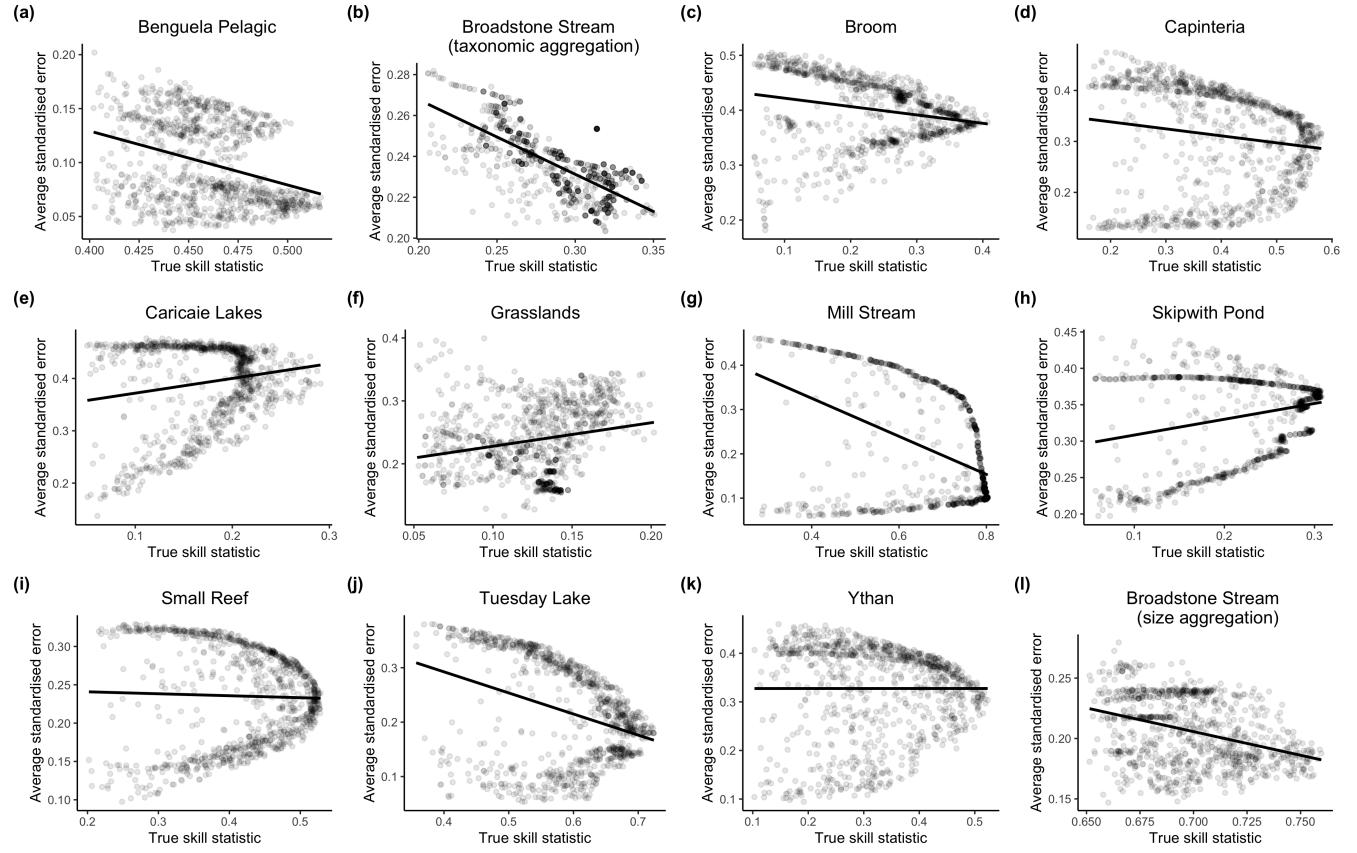


Figure S31: The mean standardised error of the food web properties predicted from the ADBM parameterised using rejection ABC plotted against the true skill statistic for each food webs. The solid lines are linear regressions (details in SI-S5).

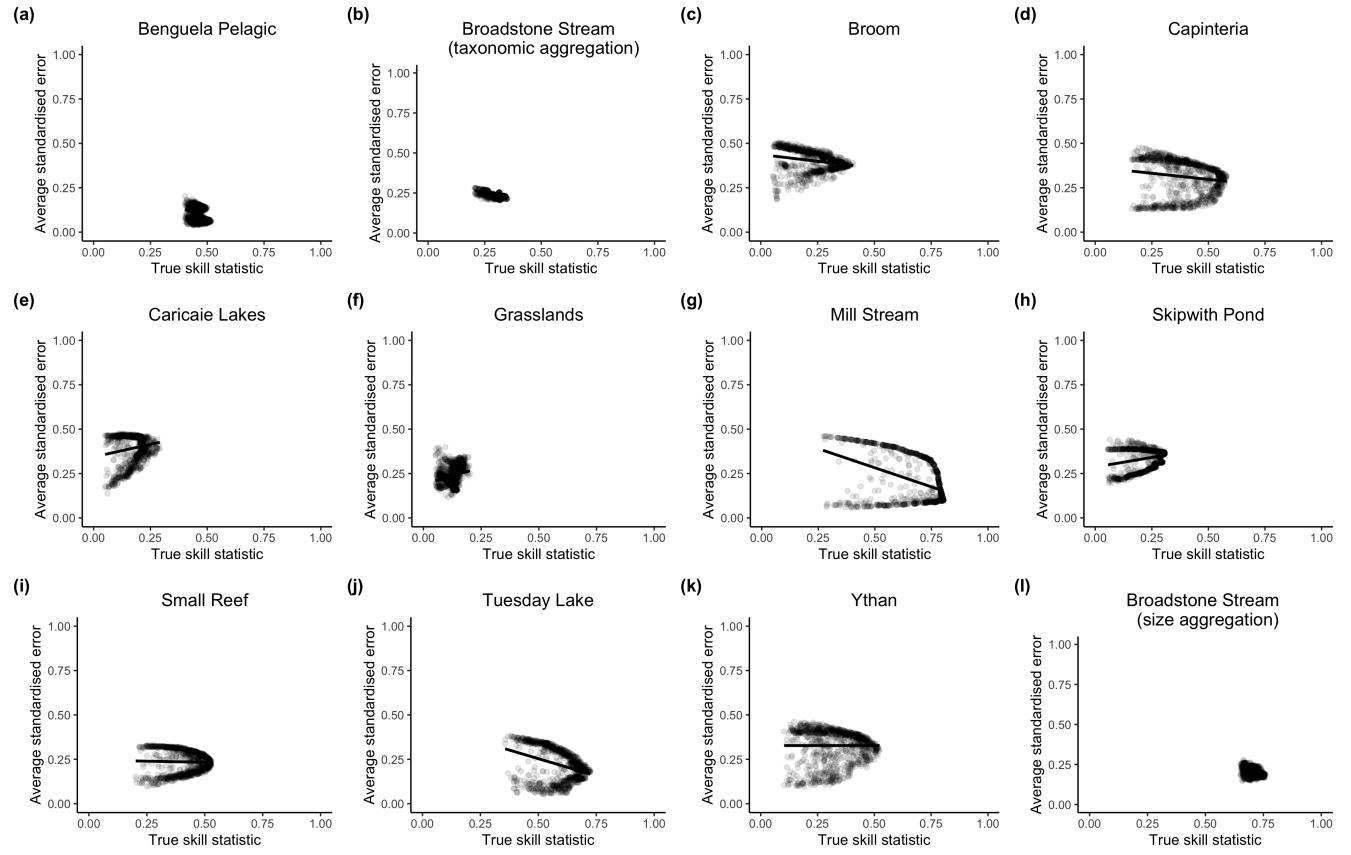


Figure S32: The mean standardised error of the food web properties predicted from the ADBM parameterised using rejection ABC plotted against the true skill statistic for each food webs. The solid lines are linear regressions (details in SI-S5). The limits of the horizontal and vertical axes have been set between 0 and 1.

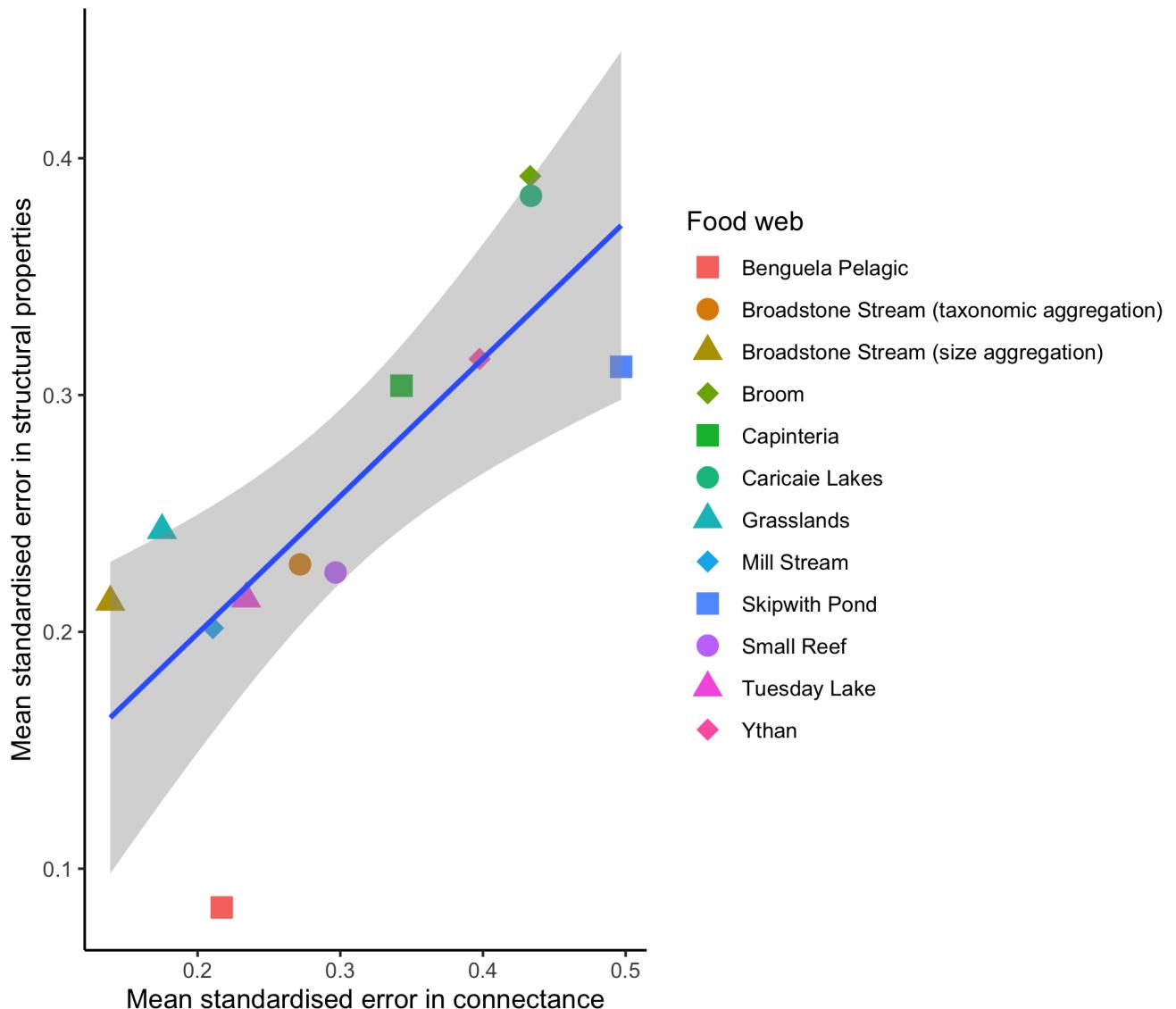


Figure S33: The mean standardised error in structural properties of food web versus mean standardised error in connectance. Solid blue line is linear regression ($t = 3.911$, $df = 10$, $P = 0.0029$)

318 References

- 319 Gelman, Andrew, and Donald B. Rubin. 1992. "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7 (4): 457–72. <https://doi.org/10.1214/ss/1177011136>.
- 320 Marjoram, P., J. Molitor, V. Plagnol, and S. Tavare. 2003. "Markov Chain Monte Carlo Without Likelihoods." *Proceedings of the National Academy of Sciences* 100 (26): 15324–28. <https://doi.org/10.1073/pnas.03068>

323 99100.