

UHI Research Database pdf download summary

A general framework for combining ecosystem models

Spence, Michael A.; Blanchard, Julia L.; Rossberg, A. G.; Heath, Michael R.; Heymans, Johanna; Mackinson, Steven; Serpetti, Natalia; Speirs, D. C.; Thorpe, Robert; Blackwell, Paul G.

Published in:
Fish and Fisheries

Publication date:
2018

Publisher rights:
© 2018 The Authors. Fish and Fisheries Published by John Wiley & Sons Ltd.

The re-use license for this item is:
CC BY

The Document Version you have downloaded here is:
Peer reviewed version

The final published version is available direct from the publisher website at:
[10.1111/faf.12310](https://doi.org/10.1111/faf.12310)

[Link to author version on UHI Research Database](#)

Citation for published version (APA):

Spence, M. A., Blanchard, J. L., Rossberg, A. G., Heath, M. R., Heymans, J., Mackinson, S., ... Blackwell, P. G. (2018). A general framework for combining ecosystem models. *Fish and Fisheries*, 1-12. DOI: 10.1111/faf.12310

General rights

Copyright and moral rights for the publications made accessible in the UHI Research Database are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights:

- 1) Users may download and print one copy of any publication from the UHI Research Database for the purpose of private study or research.
- 2) You may not further distribute the material or use it for any profit-making activity or commercial gain
- 3) You may freely distribute the URL identifying the publication in the UHI Research Database

Take down policy

If you believe that this document breaches copyright please contact us at RO@uhi.ac.uk providing details; we will remove access to the work immediately and investigate your claim.

A general framework for combining ecosystem models

Michael A. Spence^{1,2,3*}, Julia L. Blanchard⁴, Axel G. Rossberg^{3,5},
Michael R. Heath⁶, Johanna J Heymans⁷, Steven Mackinson^{3,8}, Natalia
Serpetti⁷, Douglas C. Speirs⁶, Robert B. Thorpe³ and Paul G. Blackwell¹

¹School of Mathematics and Statistics, University of Sheffield, Sheffield,

UK

²Department of Animal and Plant Sciences, University of Sheffield,

Sheffield, UK

³Centre for Environment, Fisheries and Aquaculture Science, Lowestoft,

Suffolk NR33 0HT, UK

⁴Institute for Marine and Antarctic Studies and Centre for Marine

Socioecology, University of Tasmania, 20 Castray Esplanade, Battery

Point. TAS. 7004

⁵Aquatic Ecology Group, Department of Organismal Biology, School of

Biological and Chemical Sciences, Queen Mary University of London,

17 Mile End Road, London E1 4NS

18 ⁶Department of Mathematics and Statistics, University of Strathclyde,

19 Glasgow G1 1XH, Scotland

20 ⁷Scottish Association for Marine Science, Scottish Marine Institute,

21 Oban, Argyll, PA371QA

22 ⁸Scottish Pelagic Fishermen's Association, Heritage House, 135 - 139

23 Shore Street, Fraserburgh, Aberdeenshire, AB43 9BP

24 *Corresponding author: michael.spence@cefas.co.uk

25 **Running title:** Combining ecosystem models

26 **Abstract**

27 When making predictions about ecosystems, we often have available a number of
28 different ecosystem models that attempt to represent their dynamics in a detailed
29 mechanistic way. Each of these can be used as a simulator of large-scale experiments
30 and make projections about the fate of ecosystems under different scenarios in or-
31 der to support the development of appropriate management strategies. However,
32 structural differences, systematic discrepancies and uncertainties lead to different
33 models giving different predictions. This is further complicated by the fact that the
34 models may not be run with the same functional groups, spatial structure or time
35 scale. Rather than simply trying to select a 'best' model, or taking some weighted
36 average, it is important to exploit the strengths of each of the models, while learn-
37 ing from the differences between them. To achieve this, we construct a flexible
38 statistical model of the relationships between a collection of mechanistic models

39 and their biases, allowing for structural and parameter uncertainty and for differ-
40 ent ways of representing reality. Using this statistical meta-model, we can combine
41 prior beliefs, model estimates and direct observations using Bayesian methods, and
42 make coherent predictions of future outcomes under different scenarios with robust
43 measures of uncertainty. In this paper we take a diverse ensemble of existing North
44 Sea ecosystem models and demonstrate the utility of our framework by applying it
45 to answer the question what would have happened to demersal fish if fishing was
46 to stop.

47 **Key-words:** Bayesian statistics, Complex models, Multi-model ensemble, Multi-
48 species models, Simulation models, Uncertainty analysis

49

50 **1 Introduction**

51 **2 General framework**

52 2.1 Uncertainty in simulator outputs

53 2.2 Individual discrepancy

54 2.3 Shared discrepancy

55 2.4 The truth

56 **3 Case Study**

57 3.1 Groups of species

58 3.2 Data and elements of the statistical model

59 3.3 Simulators

60 3.4 Ensemble model

61 3.5 Results

62 **4 Discussion**

63 4.1 General model features
64 4.2 Future work and extensions
65 4.3 Conclusion

66 1 Introduction

67 Ecosystem models are widely used to support policy decisions, including fisheries
68 and marine environmental policies (Hyder et al 2015). Any such model is imper-
69 fect, and in order to use it to inform policy making, it is important to quantify
70 the uncertainty of its predictions in a robust manner (Harwood and Stokes 2003;
71 Williams and Hooten 2016). Often several models are available, each embodying
72 some knowledge of a given ecosystem, but differing in their predictions. Choosing
73 to use one model’s prediction whilst excluding the others is limiting the amount
74 of information available and therefore increasing uncertainty. Our aim here is to
75 describe and demonstrate a framework for combining information from multiple
76 ecosystem models in a coherent way that, following Chandler (2013), exploits their
77 strengths and discounts their weaknesses.

78 Many methods of combining outputs from different models have been previously
79 proposed. One is to use a ‘democracy’ of simulators (Payne et al 2015; Knutti 2010),
80 where each model gets one vote, regardless of how well it represents the true system,
81 and a distribution of possible outputs comes from this. Similarly, one could take
82 an average of the model outputs, which often outperforms all the individual models
83 (Rougier 2016). However, some models are better at predicting some outputs than
84 others. An alternative approach is to try and find the ‘best’ model(s) (Payne et al
85 2015; Johnson and Omland 2004). These methods imply that at least one of the
86 models is ‘correct’, in the sense that it can predict the true output. Not only is

87 this a bold assumption, but the addition of another model may allow an area of the
88 output space to become probable when before it was not. Thus, by increasing the
89 number of models there is no guarantee that the uncertainty will reduce. One way
90 of deciding which model is the ‘best’ is to weight models using Bayes factors, also
91 known as Bayesian model averaging (Banner and Higgs 2017; Ianelli et al 2016).
92 As Chandler (2013) explains, there is generally no model better in all respects than
93 the others and so there is no natural way of assigning a single weight to each model.
94 Furthermore, if model outputs are not presented with uncertainty then, in the case
95 where the truth is a continuous quantity, a simulator will almost never be ‘correct’,
96 thus the probability of getting the true value from the ensemble is zero. Recently,
97 ‘ensemble models’ have been used to describe how model outputs related to reality
98 (Anderson et al 2017).

99 Applying the above methods to ecosystem models is not straightforward, as
100 different models have often been fitted to different data (Ianelli et al 2016), and
101 often their outputs are on different scales or represent different dynamical pro-
102 cesses, which are sometimes integrated out. A further difficulty in applying these
103 methods is that the ecosystem models can have different outputs that are not di-
104 rectly comparable. For example, whole ecosystem models often reduce complexity
105 through the use of functional groups (e.g. Heath 2012) whereas partial ecosystem
106 or multi-species models may focus on a reduced number of species (e.g. Blanchard
107 et al 2014). However, different ecosystem models are often developed with similar
108 underlying theory (e.g. food web interactions), could have similar dynamics and
109 may even be developed in the same research groups (e.g. Heath (2012) and Speirs
110 et al (2010)). They may also have similar forcing inputs, for example those com-
111 ing from global regional physical or biogeochemical models such as those used in
112 model inter-comparison studies (e.g. Tittensor et al 2017). When combining model

113 outputs, it is important to take these similarities into account rather than treating
114 the models as independent (Rougier et al 2013).

115 Another approach is to think of the ecosystem models as coming from a popu-
116 lation of such models (Tebaldi and Sansó 2009; Chandler 2013; Leith and Chandler
117 2010) and then describe how the population differs from reality. It makes sense that
118 several models in an ensemble model would inform one another. For example, one
119 model (m1) may contain several demersal fish species and the other (m2) a func-
120 tional group called “demersal fish”. Although m2 does not explicitly contain the
121 species Atlantic cod (*Gadus morhua*) its relationship with m1 may be able to tell
122 us something about Atlantic cod indirectly. In other words, modelling the models
123 allows us to sample the unobserved outputs, conditional on the models’ observed
124 outputs.

125 In this paper we describe an ensemble model which is based on the principles of
126 Chandler (2013) but which models the outputs themselves, varying in form between
127 the different ecosystem models, rather than statistical descriptors of the outputs.
128 Our approach involves statistical modelling of the relationship between an ‘ensem-
129 ble’ of ecosystem models. To avoid ambiguity, we will refer to the latter henceforth
130 as ‘simulators’ and we refer to the way in which a simulator output differs from real-
131 ity as its discrepancy. As we are interested in measuring uncertainty our statistical
132 modelling will apply Bayesian inference methods (Robert 2007), and our analysis
133 will consider any relevant prior knowledge as well as simulator outputs that pre-
134 dict what would happen in the future under different management scenarios. The
135 Bayesian approach is subjective; for an introduction to subjective uncertainty and
136 decision theory, see Berger (1985). Strictly speaking, any fully Bayesian analysis
137 involves obtaining the posterior beliefs of a particular individual, by combining
138 their prior beliefs with information from data and modelling. Depending on the

139 context, that individual may be, for example, either a scientist or a policy maker.
140 Our framework includes the elicitation of prior beliefs to combine with information
141 from the model ensemble, allowing different individuals' posterior distributions to
142 be obtained. For the purpose of our case study, the individual chosen is one of the
143 authors.

144 In Section 2 we set up the general framework and in Section 3 we demonstrate
145 the model by looking at a specific case study: what would have happened in the
146 North Sea if we had stopped fishing in 2014? We conclude by discussing wider
147 applications of the approach in Section 4.

148 2 General framework

149 We think of the available simulators as coming from some conceptual population.
150 Our *a priori* beliefs about each one are the same; we are treating the simulators as
151 unlabelled 'black boxes'. More formally, we regard the simulators as 'exchangeable';
152 see Gelman et al (2013). We consider relaxing this assumption in Section 4. This
153 idea is formalised by using a hierarchical model (for more information see Gelman
154 et al (2013)) to represent the ensemble of simulators. However, there is no reason to
155 believe that the population of simulators will either contain, or be centred on, the
156 truth (Chandler 2013) so we need to allow some difference between the population
157 of simulators and the truth.

158 To describe the relationship between the simulators and the truth we developed
159 an ensemble model that describes the population of simulators, its dynamics and its
160 relation with the true quantity of interest. We are interested in n true quantities,
161 $\mathbf{y}^{(t)} = (y_1^{(t)}, \dots, y_n^{(t)})'$, e.g. the biomass of n species at a time t , for times $t =$
162 $1, \dots, T$. We regard m simulators, each giving an output representing the quantities

163 of interest, $\mathbf{x}_i^{(t)} = (x_{i1}^{(t)}, \dots, x_{in}^{(t)})'$ for $i = 1, \dots, m$, as coming from a population with
 164 expected output $\boldsymbol{\mu}^{(t)} = (\mu_1^{(t)}, \dots, \mu_n^{(t)})'$, the simulator consensus. To define our
 165 ensemble model, we describe separately the difference between $\mathbf{y}^{(t)}$ and $\boldsymbol{\mu}^{(t)}$, the
 166 shared discrepancy, and the difference between $\mathbf{x}_i^{(t)}$ and $\boldsymbol{\mu}^{(t)}$, simulator i 's individual
 167 discrepancy. Figure 1 illustrates an example of the ensemble model at time t . It
 168 can be read as a geometrical representation of how the simulators and reality relate
 169 to one another (see also Chandler 2013). In the subsequent subsections we describe
 170 the specific details of the general ensemble model. A summary of the variables and
 171 the model can be found in Table 1.

172 2.1 Uncertainty in simulator outputs

173 The outputs from simulator i , an n_i dimensional vector $\mathbf{u}_i^{(t)}$, may not always rep-
 174 resent the elements of $\mathbf{x}_i^{(t)}$, its 'best guess', directly. For example, the elements of
 175 $\mathbf{x}_i^{(t)}$ may represent biomasses of individual fish species and the elements of $\mathbf{u}_i^{(t)}$ may
 176 represent the biomass of functional groups, e.g. biomass of demersal fish.

177 We say that

$$\mathbf{u}_i^{(t)} = f_i(\mathbf{x}_i^{(t)}),$$

178 for some simulator-specific function $f_i(\cdot)$. For example, if the elements of $\mathbf{u}_i^{(t)}$ are
 179 elements of $\mathbf{x}_i^{(t)}$ or are sums of those elements, perhaps with some rescaling, then
 180 the relationship is linear

$$\mathbf{u}_i^{(t)} = M_i \mathbf{x}_i^{(t)},$$

181 where M_i is an $n_i \times n$ matrix. For other examples see Table 2.

182 Generally the simulators are run with uncertain inputs and parameter values.
 183 This leads to uncertainty in the outputs and is commonly known as parameter

184 uncertainty. We say that

$$\mathbf{u}_i^{(t)} = \hat{\mathbf{u}}_i^{(t)} + \boldsymbol{\epsilon}_{u_i},$$

185 for $t \in S_i$, where $\boldsymbol{\epsilon}_{u_i}$ has expectation $\mathbf{0}$ and is sampled from a simulator-specific
 186 distribution and $\hat{\mathbf{u}}_i^{(t)}$ is the expectation of the i th simulator’s output at time t . The
 187 simulator-specific distribution is found from fitting the simulator to a finite dataset
 188 (e.g. Spence et al 2016; Thorpe et al 2015) or by performing sensitivity analysis of
 189 the simulator inputs (e.g. Morris et al 2014).

190 2.2 Individual discrepancy

191 At time t , the difference between simulator i ’s ‘best guess’, $\mathbf{x}_i^{(t)}$, and the simulator
 192 consensus, $\boldsymbol{\mu}^{(t)}$, is simulator i ’s individual discrepancy,

$$\mathbf{x}_i^{(t)} - \boldsymbol{\mu}^{(t)} = \boldsymbol{\gamma}_i + \mathbf{z}_i^{(t)}.$$

193 This divides the individual discrepancy between the long-term individual discrep-
 194 ancy, $\boldsymbol{\gamma}_i$, and the short-term individual discrepancy, $\mathbf{z}_i^{(t)}$. $\boldsymbol{\gamma}_i$ is an n dimensional
 195 random variable with expectation $\mathbf{0}$ and covariance C . It seems natural to allow
 196 $\mathbf{z}_i^{(t)}$ and $\mathbf{z}_i^{(t+1)}$ to be dependent on each other; for example, if at time t , $\mathbf{z}_i^{(t)}$ was
 197 less than $\mathbf{0}$, then we might also expect $\mathbf{z}_i^{(t+1)}$ to be less than $\mathbf{0}$. With this in mind,
 198 we say that $\mathbf{z}_i^{(t)}$ follows a stationary auto-regressive model of order 1,

$$\mathbf{z}_i^{(t)} = R_i \mathbf{z}_i^{(t-1)} + \boldsymbol{\epsilon}_{z,t,i}, \quad (1)$$

199 for $t > 1$, where each $\boldsymbol{\epsilon}_{z,t,i}$ is an independent n -dimensional random variable centred
 200 on $\mathbf{0}$ with covariance Λ_i and R_i is an $n \times n$ matrix with the constraint such that
 201 R_i is stable, i.e. $\lim_{k \rightarrow \infty} R_i^k = \mathbf{0}$. R_i and Λ_i describe the dynamics of simulator i
 202 with $R_i \sim g_R(\cdot)$ and $\Lambda_i \sim g_\Lambda(\cdot)$ for some distributions g_R and g_Λ . At $t = 1$, $\mathbf{z}_i^{(1)}$
 203 is sampled from the stationary distribution of the auto-regressive model described

204 in equation 1 (See Appendix A for more details). This formulation means that the
 205 expectation of the long-run behaviour of the individual discrepancy is the long-term
 206 individual discrepancy, i.e.

$$\begin{aligned}
 \lim_{k \rightarrow \infty} E(\gamma_i + z_i^{(t+k)} | \gamma_i + z_i^{(t)}) &= \gamma_i + \lim_{k \rightarrow \infty} E(z_i^{(t+k)} | z_i^{(t)}) \\
 &= \gamma_i + E(z_i^{(t)}) \\
 &= \gamma_i.
 \end{aligned}$$

207 2.3 Shared discrepancy

208 The shared discrepancy, the difference between the simulator consensus, $\boldsymbol{\mu}^{(t)}$, and
 209 truth, $\boldsymbol{y}^{(t)}$, is split up into the long-term shared discrepancy, $\boldsymbol{\delta}$, and the short-term
 210 shared discrepancy, $\boldsymbol{\eta}^{(t)}$, i.e.

$$\boldsymbol{y}^{(t)} - \boldsymbol{\mu}^{(t)} = \boldsymbol{\delta} + \boldsymbol{\eta}^{(t)}.$$

211 The short-term shared discrepancy is described by a stationary auto-regressive
 212 model of order 1

$$\boldsymbol{\eta}^{(t)} = R_\eta \boldsymbol{\eta}^{(t-1)} + \boldsymbol{\epsilon}_{\eta,t}, \quad (2)$$

213 for $t > 1$, where R_η is stable and $\boldsymbol{\epsilon}_{\eta,t}$ is an n dimensional random variable centred
 214 on $\mathbf{0}$ with covariance Δ . At $t = 1$, $\boldsymbol{\eta}^{(1)}$ is sampled from the stationary distribution
 215 of the auto-regressive model described in equation 2 (See Appendix A for more
 216 details). This formulation means that the expectation of the long-run behaviour of
 217 the shared discrepancy is the long-term shared discrepancy, i.e.

$$\begin{aligned}
 \lim_{k \rightarrow \infty} E(\boldsymbol{\delta} + \boldsymbol{\eta}^{(t+k)} | \boldsymbol{\delta} + \boldsymbol{\eta}^{(t)}) &= \boldsymbol{\delta} + \lim_{k \rightarrow \infty} E(\boldsymbol{\eta}^{(t+k)} | \boldsymbol{\eta}^{(t)}) \\
 &= \boldsymbol{\delta} + E(\boldsymbol{\eta}^{(t)}) \\
 &= \boldsymbol{\delta}.
 \end{aligned}$$

218

2.4 The truth

219

In the absence of any simulators, our prior beliefs for the truth at time t , $\mathbf{y}^{(t)}$, follow

220

a random walk,

$$\mathbf{y}^{(t)} = \mathbf{y}^{(t-1)} + \boldsymbol{\epsilon}_{\Lambda,t},$$

221

for $t > 1$, where each $\boldsymbol{\epsilon}_{\Lambda,t}$ is centred on $\mathbf{0}$ with covariance Λ_y . At $t = 1$, the truth,

222

$\mathbf{y}^{(1)}$, follows a generic prior distribution $p(\mathbf{y}^{(1)})$.

223

At times $t \in S_0$, there are n_y noisy and possibly indirect observations, $\hat{\mathbf{w}}^{(t)}$, of

224

the truth which come from some distribution, $p(\hat{\mathbf{w}}^{(t)}|\mathbf{y}^{(t)})$ that is problem specific

225

and is caused by data uncertainty (Li and Wu 2006). The elements of $\hat{\mathbf{w}}^{(t)}$ may

226

not be the same as that of $\mathbf{y}^{(t)}$, for example if observations are incomplete or

227

aggregated. We assume that the sampling distribution of observations depends on

228

the truth through some function $f_y(\cdot)$, such that

$$\mathbf{w}^{(t)} = f_y(\mathbf{y}^{(t)})$$

229

and $p(\hat{\mathbf{w}}^{(t)}|\mathbf{y}^{(t)}) = p(\hat{\mathbf{w}}^{(t)}|\mathbf{w}^{(t)})$.

230

For example if $\mathbf{w}^{(t)}$ is some linear transformation of $\mathbf{y}^{(t)}$, then

$$\mathbf{w}^{(t)} = M_y \mathbf{y}^{(t)}$$

231

where M_y is an $n_y \times n$ matrix.

232

3 Case Study

233

We illustrate our model by looking at a problem where a scientist needs to formally

234

summarise uncertain model results, for example to present to other scientists or to

235

decision makers about what would happen to the biomass of demersal species in

236

the North Sea if fishing were to stop completely in 2014. We use outputs from five

237

ecosystem simulators: Ecopath with Ecosim (EwE, Lynam and Mackinson 2015),

238 mizer (Blanchard et al 2014), FishSUMs (Speirs et al 2010), StrathE2E (Heath et al
239 2014) and LeMans (Thorpe et al 2015) (see Appendix B for more details about the
240 simulators), as well as data from the International Bottom Trawl Survey (IBTS)
241 (ICES Database of Trawl Surveys (DATRAS) 2015). In this example, one of the
242 authors, JLB, has taken this role. Her prior beliefs are elicited and expressed as a
243 prior distribution and the posterior distribution captures her uncertainty about the
244 future of the ecosystem in this scenario give the relationships among the simulators
245 and observations.

246 **3.1 Groups of species**

247 The five simulators represent demersal fish in different ways, with different species
248 resolution and coverage. While our main interest is in demersal fish collectively, we
249 need to represent the state of the ecosystem at a resolution that enables us to link
250 these simulator outputs together.

251 In representing the state of the ecosystem, it would be computationally ineffi-
252 cient to treat each species separately, given that we are interested in demersal fish
253 in aggregate. Instead, we can reduce the dimension of the problem by grouping
254 the species together. This grouping needs to have the property that any simulator
255 output that we can use can be expressed as the sum of one or more of our groups.
256 The groups do not necessarily need to have any direct biological interpretation;
257 provided the groups meet the criterion above, and allow us to represent the quan-
258 tities of interest—here, demersal fish, given by the sum of all groups—the precise
259 choice will not affect the answer obtained. For computational efficiency, we choose
260 the minimum number of groups that meets this criterion while covering all dem-
261 ersal species. For example we grouped together monkfish, long rough dab, lemon
262 sole and witch because they all occur in exactly the same simulators, as individual

263 species in EwE and LeMans and implicitly in StrathE2E, but are not contained
264 in any larger set of species for which this is true. This minimal set consists of 5
265 groups, which we will model explicitly. The groups are:

- 266 1. *Common demersal*: These are Atlantic cod (*Gadus morhua*), haddock (*Melanogram-*
267 *mus aeglefinus*), whiting (*Merlangius merlangus*), Norway pout (*Trisopterus*
268 *esmarkii*), European plaice (*Pleuronectes platessa*), common dab (*Limanda*
269 *limanda*) and grey gurnard (*Eutrigla gurnardus*).
- 270 2. *Sole*: This is common sole (*Solea solea*).
- 271 3. *Monkfish etc.*: These are monkfish (*Lophius piscatorius*), long rough dab (*Hip-*
272 *poglossoides platessoides*), lemon sole (*Microstomus kitt*) and witch (*Glypto-*
273 *cephalus cynoglossus*).
- 274 4. *Poor Cod and Rays*: These are poor cod (*Trisopterus minutus*), starry rays
275 (*Amblyraja radiata*) and cuckoo rays (*Leucoraja naevus*).
- 276 5. *Other demersal fish*: This consists of all other demersal fish.

277 We consider the total biomass densities for each of these groups, in tonnes per
278 square kilometre, modelled on the log scale (to base 10, for ease of interpretation).

279 **3.2 Data and elements of the statistical model**

280 The IBTS data were extracted as in Fung et al (2012), to reveal the total catch
281 on the survey for each of the 5 groups for the first (1986-2013) and third quarter
282 (1991-2013). How this value relates to the true biomass density in the North Sea is
283 not trivial, and these values are often multiplied by catchability coefficients (Walker
284 et al 2017) which are themselves uncertain and model-based. In this example we are
285 only interested in the biomass density relative to 2010, and therefore the total catch
286 from the IBTS survey is enough provided we assume that catchability coefficients

287 are constant over time. Thus each element of \mathbf{y}_t represents the log to base 10 of
 288 the total biomass (tonnes per kilometre squared) for one of our groups of species,
 289 averaged over year t , relative to 2010. Therefore,

$$\mathbf{w}^{(t)} = f_y(\mathbf{y}^{(t)}) = 10^{\mathbf{y}^{(t)}}.$$

290 The measurement error on the observations of the truth is assumed to be normally
 291 distributed on the \log_{10} scale such that

$$\log_{10} \left(\hat{\mathbf{w}}^{(t)} / \hat{\mathbf{w}}^{(2010)} \right) \sim \text{N}(\mathbf{y}^{(t)}, \Sigma_y),$$

292 for $t \neq 2010$. In this work we take Σ_y to be $2 \log_{10}(1.15)$ on the diagonal elements
 293 and 0 on the off diagonal elements. This was chosen so that it means that the
 294 standard deviation of the true biomass would be 15% of the actual amount caught.

295 3.3 Simulators

296 We have outputs from five different simulators all of which have been run with
 297 zero fishing pressure from 2014 onwards. A short summary of the simulators, their
 298 outputs with respect to this case study and their simulator-specific function, $f_i(\cdot)$,
 299 can be found in Table 2. The i th simulator's output is assumed to be normally
 300 distributed on the \log_{10} scale,

$$\log_{10} \mathbf{u}_i^{(t)} \sim \text{N}(\log_{10} \hat{\mathbf{u}}_i^{(t)}, \Sigma_i),$$

301 with Σ_i fitted based on running simulator i many times (Leith and Chandler 2010;
 302 Chandler 2013). However, if this was not the case Σ_i could be estimated within the
 303 hierarchical system.

3.4 Ensemble model

Each element of $\mathbf{x}_i^{(t)}$ is the “best guess” of simulator i of the elements of $\mathbf{y}^{(t)}$, for $t = 1968, \dots, 2100$, in log (base 10) tonnes per km squared of wet biomass. In this example we expect each of the simulators to converge to its own steady state, given that all external drivers are constant. This means that in equation 1 we expect R_i to tend towards 1 and Λ_i to tend towards 0. Furthermore, if a simulator reaches a stationary state before it has stopped running, then we know that it will be in that state forever. Simulator i 's individual discrepancy, $\gamma_i + \mathbf{z}_i^{(t)}$, is thus modelled as

$$\gamma_i \sim N(0, C)$$

and

$$\mathbf{z}_i^{(t)} \sim \begin{cases} N(R_i \mathbf{z}_i^{(t-1)}, \Lambda_i) & \text{if } t \leq 2013, \\ N(h_z(R_i, k_i, t) \mathbf{z}_i^{t-1}, h_\Lambda(t, k_i) \Lambda_i) & \text{if } 2014 \geq t. \end{cases}$$

where

$$h_z(R_i, k, t) = R_i + (1 - R_i)(1 - h_\Lambda(t, k_i))$$

and

$$h_\Lambda(t, k_i) = \exp\{-k_i(t - 2013)\}.$$

This is saying that, after the end of fishing, the variance of the truth of model i reduces and the amount that the last value of $\mathbf{z}_i^{(t)}$ relates to the next moves towards 1 by a factor of $\exp(k_i)$ each year. We take $k_i \in [0, 6]$, as there is not much difference numerically if k_i goes above 6, with

$$k_i/6 \sim \text{Beta}(a_k, b_k).$$

The diagonal elements of R_i fall between -1 and 1 with

$$\frac{R_i + 1}{2} \sim \text{Beta}(\mathbf{a}_R, \mathbf{b}_R)$$

320 and the off-diagonal elements are set to 0. The simulator-specific variance param-
 321 eter, Λ_i , is decomposed into a diagonal matrix of variances, Π_i , and a correlation
 322 matrix, P_i , such that

$$\Lambda_i = \Pi_i P_i \Pi_i. \quad (3)$$

323 The form of the prior distribution for the j th diagonal element of Π_i was

$$\pi_{ij} \sim \text{Gamma}(\alpha_{\pi,j}, \beta_{\pi,j}).$$

324 Distributions over correlation matrices are complicated by the mathematical re-
 325 quirement of positive definiteness. In practice, we specify separate priors on the
 326 elements, and then condition on positive definiteness; the unconditional prior for
 327 the j, k th element of P_i is given by

$$\frac{\rho_{ijk} + 1}{2} \sim \begin{cases} \text{Beta}(a_{\rho jk}, b_{\rho jk}) & \text{if } j \neq k, \\ 1 & \text{otherwise.} \end{cases}$$

328 The difference between the truth at time t and the corresponding simulator con-
 329 sensus, $\boldsymbol{\mu}^{(t)}$, is then

$$\left(\mathbf{y}^{(t)}\right) - \left(\boldsymbol{\mu}^{(t)} - \boldsymbol{\mu}^{(2010)}\right) = \boldsymbol{\eta}^{(t)} + \boldsymbol{\delta}$$

330 with

$$\boldsymbol{\eta}^{(t)} \sim \text{N}(R_\eta \boldsymbol{\eta}^{(t-1)}, \Delta_\eta). \quad (4)$$

331 When the fishing is turned off, we are particularly uncertain about what will hap-
 332 pen; thus we will remove any direct relation between \mathbf{y}_t and \mathbf{y}_{t+1} beyond that time.

333 We will say that

$$\boldsymbol{\mu}^{(t)} \sim \text{N}(\boldsymbol{\mu}^{(t-1)}, h_\Lambda(t, k_\mu) \Delta_\mu) \quad (5)$$

334 where $k_\mu \in [0, 6]$, so that the simulator consensus reaches a stationary point, as the
 335 individual simulators do.

336 We focus on the subjective probabilities of a particular individual, in this case
337 JLB. Her prior beliefs were elicited using the method described in O’Hagan et al
338 (2006) and Alhussain and Oakley (2017). Details of the prior elicitation can be
339 found in Appendix C. Due to the dimensionality and correlation of the uncertain
340 parameter space, we fitted the model using No U-turn Hamiltonian Monte Carlo
341 (Hoffman and Gelman 2014) in the package Stan (Gelman et al 2015).

342 3.5 Results

343 The ensemble model predictions show changes in the uncertainty of relative biomass
344 over time for each group of species, including projections following a fishing closure
345 in 2014 (Figure 2). Each plot shows the marginal posterior distributions of each
346 element of $\mathbf{y}^{(t)}$, for all t . Unsurprisingly, the ensemble model predicts *common*
347 *demersal* fish increase following the fishery closure, as this group contains a lot of
348 species targeted by fisheries.

349 According to the ensemble model the probability that there will be a greater
350 total biomass of *common demersal* in 2050 than in 2010 is 0.90. There is a similar
351 number for *sole* (0.93) and for *monkfish etc.* (0.88) but it is lower for *poor cod and*
352 *rays* (0.55) and for the *other demersal* species (0.17).

353 The ensemble model also ‘predicts’ what happened before the data; that is, it
354 gives posterior distributions for the actual values given the imperfect data and the
355 simulator runs. Only *sole* and *common demersal* are output by simulators prior to
356 1986 and this is reflected in the increased uncertainty as we move further back in
357 time from 1986.

358 The uncertainty in the prediction increases the further away from the obser-
359 vations of the truth, both when projecting and hindcasting. The uncertainty also
360 increases when there are fewer simulators that give outputs. All of the simulators

361 give outputs for the *common demersal* group, four explicitly and one implicitly,
362 and therefore we are more certain about what will happen to it in the future than
363 for *poor cod and rays*, where only three simulators predict values for the future and
364 only one explicitly. The uncertainty is highest for *other demersal* species. This
365 is understandable as only two simulators predict values for this group of species,
366 neither of which does so explicitly.

367 The absolute total biomass of demersal species is difficult to calculate here with-
368 out information on the discrepancy between the simulator consensus and the truth.
369 Although survey data are available, their relationship with the truth depends on the
370 varying, and unknown, catchability coefficients for each of the groups. Although
371 catchabilities can be estimated, for simplicity here we examine the total demersal
372 biomass under the assumption that the groups had the same catchability coeffi-
373 cients (Figure 3). Again there is high uncertainty about whether the biomass will
374 grow relative to the biomass in 2010. However, what it was before 1986 is also quite
375 uncertain. This is because of the uncertainty in the populations of *Other demersal*
376 *species*.

377 The median “best guess” of each of the simulators can also be compared across
378 the different simulators (Figure 4). StrathE2E predicts quite a large increase in
379 *common demersal* despite not explicitly outputting it. Mizer does not do a very
380 good job of predicting the dynamics of *sole*, therefore the dynamics of the simulator
381 consensus do not match the dynamics of mizer.

382 The posterior predictive distribution for the relative truth in 2025 for *common*
383 *demersal* and *monkfish etc.* are positively correlated with each other (0.28), albeit
384 weakly. This suggests that learning something about the *common demersal* group
385 would tell you something about *monkfish etc.* Hence the mizer simulator gives
386 some information regarding the *monkfish etc.* despite not actually predicting it.

387 See Appendix D for the other correlations between the groups.

388 4 Discussion

389 By treating the simulator outputs as coming from a population of simulators and
390 modelling this population, we have presented in this paper a general way of com-
391 bining ecosystem simulators to inform scientists and decision makers about the
392 consequences of management strategies. Our model combines many different simu-
393 lators, exploiting their strengths and discounting their weaknesses (Chandler 2013)
394 to provide synthetic and comprehensive information to support decision making.

395 4.1 General model features

396 One of the difficulties in building an ensemble model with ecosystem simulators
397 is that the simulator outputs are often done on different scales and are not di-
398 rectly comparable, for example StrathE2E models groups of species (e.g. pelagic,
399 demersal) whereas mizer models major species individually. Our approach, unlike
400 existing methods of combining simulators (e.g. Bayesian model averaging (Banner
401 and Higgs 2017; Ianelli et al 2016)), allows us to combine outputs from these widely
402 differing simulators. We achieve this by modelling what each simulator would pre-
403 dict for each of the groups of species we are interested in, whether it is explicitly
404 modelled or not by the simulator. For example, in the case study, StrathE2E only
405 models the total demersal species. Using information from the other simulators re-
406 garding the breakdown of demersal species and how the dynamics between species
407 work, the ensemble model can say what StrathE2E would predict on a species level.
408 In the case study, EwE and StrathE2E both implicitly predict groups of species.
409 For EwE it is the sum of *poor cod and rays* and *other demersal* and for StrathE2E

410 it is the sums of all of the groups. As with the simulators that do not predict
411 specific groups, we are able to infer what these simulators predict about implicit
412 groups through correlations learned from other simulators. In this sense, the mizer
413 model, which only predicts *common demersal* and *sole*, gives information about
414 how StrathE2E divides its demersal species and therefore gives some information
415 about other groups. Therefore, if we were interested in what would happen to the
416 other demersals if we were to stop fishing, we should include all the simulators
417 despite only two of them predicting it.

418 Simulators that are predictably wrong are more informative than those that are
419 unpredictably wrong, even if the latter are less wrong in the absolute sense. In our
420 framework, we distinguish between short-term and long-term individual discrep-
421 ancies, which allows us to distinguish between predictably wrong simulators with
422 small short-term individual discrepancies, z_i , and unpredictably wrong simulators.
423 Furthermore, we allow the short-term individual discrepancies to be different for
424 each group, thus allowing a simulator to contribute to the ensemble model for
425 groups that it is informative about and be ignored for groups that it is not. In
426 the case study, mizer does not predict the dynamics of *sole* very well and so the
427 simulator consensus, μ , only weakly follows the mizer predictions. On the other
428 hand, mizer does a reasonable job of predicting the dynamics of *common demersal*
429 and therefore it contributes more to the simulator consensus for this group. Thus
430 the ensemble model exploits mizer's strengths, *common demersal*, and discounts its
431 weaknesses, *sole*.

432 The ensemble model enables formal quantification of uncertainty. This uncer-
433 tainty reflects a specific individual's updated beliefs having observed the simulators
434 and the observation data (Robert 2007). The individual could be a scientist or a de-
435 cision maker and could be informed by multiple experts (Albert et al 2012). Such

436 a framework could be used to help communicate uncertainty or enable decision-
437 makers to directly quantify risks and therefore evaluate management trade-offs more
438 rigorously (Harwood and Stokes 2003; Finkle 1990). The ensemble model takes ac-
439 count of uncertainty from each of the simulators, through parameter uncertainty
440 and structural uncertainty, data uncertainty, through noisy and possibly indirect
441 observations of the truth, and uncertainty in the ensemble model parameters.

442 As the simulators are describing the same system, we might expect the dynamics
443 in the individual discrepancies to be similar. To reflect this, we allow the short-term
444 individual discrepancies to come from some underlying distribution. Furthermore,
445 in ecosystems simulators, the dynamics may be similar in direction but likely not in
446 magnitude. To include this information in the case study, we split the short-term
447 individual discrepancies, Λ_i , into correlations and magnitude (equation 3), allowing
448 different levels of confidence for each. We used beta distributions for each of the
449 off-diagonal elements of the correlation matrix and then conditioned on positive
450 definiteness. This enabled us to learn about each element of the correlation matrix
451 separately which is not possible in other formulations of the covariance matrix
452 (Alvarez et al 2014). By acknowledging these features of simulators, we were able
453 to better quantify the uncertainty.

454 It was also important to use informative priors as none of the simulators explic-
455 itly model *other demersal*. As there is no lower bound (on the log scale) for the
456 values of the “best guess” of *other demersal*, we required some prior information
457 about the distribution of the standard deviations, Π . This does suggest that the
458 ensemble prediction is somewhat based on that of the priors for Λ_i . In practice,
459 we suggest checking that your ensemble model predicts in a way that the decision
460 maker believes before observing the truth, similar to the hypothetical data method
461 of Kadane et al (1980). In the case study described here, we checked that the

462 dynamics of the biomasses prior to 1986 followed JLB’s beliefs.

463 When building the ensemble model, how the species groups are decided depends
464 on the question being asked. In the case study, we were interested in what would
465 happen to demersal fish if we were to stop fishing, so we grouped the species into
466 as few groups as possible. However, if we were interested in another question, for
467 example if we had been interested in what would happen to commercial fish, we
468 would divide the species into groups with commercial and non-commercial fish con-
469 ditioned on species in each group being presented in exactly the same simulators.
470 As the number of groups increases, the dimensions of the covariance matrices in-
471 creases, so we advise that the number of groups be kept to a minimum as this would
472 aid computation time and require less simulators and prior elicitation.

473 Using the ensemble model developed here, there is no need to identify the “best
474 model” driven by the question being asked (Dickey-Collas et al 2014), but one
475 should include all available simulators. Rather than developing many simulation
476 models to answer different specific questions, the ensemble model can be designed
477 to answer the question at hand thus reducing computational costs. Furthermore,
478 as the ensemble model implicitly weights the simulators by their strengths and
479 weaknesses, it is better for a simulator to be good at modelling one aspect of the
480 ecosystem rather than being average at modelling a lot of things (Anderson et al
481 2017). Due to tractability it is not possible to explicitly show these weightings in
482 the case study presented here, for an example of weightings in a more tractable
483 example see Chandler (2013).

484 The nature of the different ecosystem simulators capturing different processes
485 can limit the number of models available to run certain scenarios (e.g. in climate
486 scenarios where some but not all the simulators contain links to temperature). If we
487 were interested in one of the scenarios that a specific simulator was unable to run,

488 we should still include that simulators in the ensemble model as it gives information
489 about how species interact with one another as well as the state of the ecosystem
490 up until the current time. To include this simulator in the ensemble, we could learn
491 about how it differs from the simulators that were able to run the specific scenario
492 and increase a simulator's parameter uncertainty, Σ_i , as a function of time with in
493 the future (Szuwalski and Thorson 2017).

494 4.2 Future work and extensions

495 Some ecosystem simulators are more similar than others, for example there are a
496 number of size-based simulators in the marine literature (e.g. Blanchard et al 2009;
497 Scott et al 2014) that are very similar, which may violate the exchangeability as-
498 sumption made in Section 2. Additional hierarchy could be added to the ensemble
499 model that would allow such simulators to have more similar discrepancies. In
500 climate science, where the simulators are very similar to one another and phylo-
501 genetic trees show the development history of each simulator (Knutti et al 2013),
502 Demetriou (2016) added additional hierarchy allowing closely related simulators to
503 have similar discrepancies. They found that the major source of uncertainty was
504 due to the shared discrepancy, and the results of the ensemble model were close to
505 when all the simulators were assumed to be exchangeable.

506 In this paper, we have demonstrated the ideas and methods in cases where
507 the quantities of interest are of fairly low dimension and have joint Gaussian dis-
508 tributions. However, with the increased efficiency of new statistical software and
509 algorithms (see e.g. Girolami and Calderhead 2011), it is possible to address larger
510 problems involving more general distributions.

511 The framework presented here is not exclusive to ecosystem simulators in fish-
512 eries, but can be used to combine any mechanistic simulators in many areas of

513 ecology (e.g. Individual-based models, Railsback and Grimm 2012) or even other
514 areas of research such as systems biology (Kuepfer et al 2007) and epidemiology
515 (Lessler et al 2016).

516 **4.3 Conclusion**

517 This work allows for a synthesis of many modelling studies that have been and are
518 being conducted in such a way that we can obtain more holistic knowledge over
519 a wide scope of complex ecological systems. It also allows for including a formal
520 quantitative understanding of uncertainties and knowledge gaps. This enables us
521 to make comprehensive model projections that take into account all that we have
522 learnt from the simulators collectively.

523 **Acknowledgments**

524 The work was supported by the Natural Environment Research Council and De-
525 partment for Environment, Food and Rural Affairs [grant number NE/L003279/1,
526 Marine Ecosystems Research Programme]. The authors would like to thank Tom
527 Webb, Remi Vergnon, Yuri Artioli, Sévrine Saillery, Paul Somerfield, Melanie
528 Austen, Nicola Beaumont and Stefanie Broszeit for participating in early elicitation
529 exercises. We thank Tony Pitcher and two anonymous reviewers for comments on
530 an earlier version of the paper.

531 **Author contribution**

532 MAS, PGB and JLB conceived the ideas and designed the methodology; JLB ex-
533 tracted the data for the main case study; MAS, MRH, SM, DS, AGR, RBT, JJH

534 and NS ran the simulators for the case study; MAS implemented the methodol-
535 ogy; MAS and PGB analysed the data; MAS and PGB led the writing of the
536 manuscript. All authors contributed critically to the drafts and gave final approval
537 for publication.

538 **References**

539 Albert I, Donnet S, Guihenneuc-Jouyaux C, Low-Choy S, Mengersen K, Rousseau
540 J (2012) Combining Expert Opinions in Prior Elicitation. *Bayesian Analysis*
541 7(3):503–532

542 Alhussain ZA, Oakley JE (2017) Eliciting judgements about uncertain population
543 means and variances. arXiv:170200978 URL <https://arxiv.org/abs/1702.00978>

544 Alvarez I, Niemi J, Simpson M (2014) Bayesian inference for a covariance matrix.
545 arXiv:14084050 URL <https://arxiv.org/abs/1408.4050>

546 Anderson SC, Cooper AB, Jensen OP, Minto C, Thorson JT, Walsh JC, Afflerbach
547 J, Dickey-Collas M, Kleisner KM, Longo C, Osio GC, Ovando D, Mosqueira I,
548 Rosenberg AA, Selig ER (2017) Improving estimates of population status and
549 trend with superensemble models. *Fish and Fisheries* DOI 10.1111/faf.12200

550 Banner KM, Higgs MD (2017) Considerations for Assessing Model Averaging of Re-
551 gression Coefficients. *Ecological Applications* 27(1):78–93, DOI 10.1002/eap.1419

552 Berger JO (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd edn.
553 Springer Series in Statistics, Springer-Verlag

554 Blanchard JL, Jennings S, Law R, Castle MD, McCloghrie P, Rochet MJ, Benoît

555 E (2009) How does abundance scale with body size coupled size-structured food
556 webs? *Journal of Animal Ecology* 78(270-280)

557 Blanchard JL, Andersen KH, Scott F, Hintzen NT, Piet G, Jennings S (2014)
558 Evaluating targets and trade-offs among fisheries and conservation objectives
559 using multispecies size spectrum model. *Journal of Applied Ecology* 51(3):612–
560 662, DOI 10.1111/1365-2664.12238

561 Chandler RE (2013) Exploiting strength, discounting weakness: combining informa-
562 tion from multiple climate simulators. *Philosophical Transactions of the Royal*
563 *Society A: Mathematical, Physical and Engineering Sciences* 371(1991), DOI
564 DOI: 10.1098/rsta.2012.0388

565 Demetriou D (2016) A Bayesian approach to the interpretation of climate model
566 ensembles. PhD thesis, University College London

567 Dickey-Collas M, Payne MR, Trenkel VM, Nash RDM (2014) Hazard warning:
568 model misuse ahead. *ICES Journal of Marine Science: Journal du Conseil*
569 72(8):2300–2306

570 Finkle AM (1990) Confronting uncertainty in risk management: A guide for
571 decision-makers: a report. Tech. rep., Centre for Risk Management, Resources
572 for the Future

573 Fung T, Farnsworth KD, Reid DG, Rossberg AG (2012) Recent data suggests no
574 further recovery in North Sea Large Fish Indicator. *ICES Journal of Marine*
575 *Science* 69:235–239, DOI 10.1093/icesjms/fsr206

576 Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian
577 Data Analysis, 3rd edn. Chapman and Hall

578 Gelman A, Lee D, Guo J (2015) Stan: A probabilistic programming language.
579 Journal of Educational and Behavioural Statistics 40:530–543

580 Girolami M, Calderhead B (2011) Riemann manifold Langevin and Hamiltonian
581 Monte Carlo methods. Journal of Royal Statistical Society B 73:1–37, DOI
582 10.1111/j.1467-9868.2010.00765.x

583 Harwood J, Stokes K (2003) Coping with uncertainty in ecological advice: lessons
584 from fisheries. Trends in Ecology and Evolution 18(12):617–622

585 Heath MR (2012) Ecosystem limits to food web fluxes and fisheries yields in the
586 north sea simulated with an end-to-end food web model. Progress in Oceanogra-
587 phy 102:42 – 66, DOI 10.1016/j.pocean.2012.03.004

588 Heath MR, Speirs DC, Steele JH (2014) Understanding patterns and processes in
589 models of trophic cascades. Ecology Letters 17:101–114, DOI 10.1111/ele.12200

590 Hoffman MD, Gelman A (2014) The No-U-Turn Sampler: Adaptively Setting Path
591 Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research
592 15:1593–1623

593 Hyder K, Rossberg AG, Allen JI, Austen MC, Barciela RM, Bannister HJ, Blackwell
594 PG, Blanchard JL, Burrows MT, Defriez E, Dorrington T, Edwards KP, Garcia-
595 Carreras B, Heath MR, Hembury DJ, Heymans JJ, Holt J, Houle JE, Jennings S,
596 Mackinson S, Malcolm SJ, McPike R, Mee L, Mills DK, Montgomery C, Pearson
597 D, Pinnegar JK, Pollicino M, Popova EE, Rae L, Rogers SI, Speirs D, Spence
598 MA, Thorpe R, Turner RK, van der Molen J, Yool A, Paterson DM (2015)
599 Making modelling count - increasing the contribution of shelf-seas community
600 and ecosystem models to policy development and management. Marine Policy
601 61:291 – 302, DOI 10.1016/j.marpol.2015.07.015

602 Ianelli J, Holsman KK, Punt AE, Aydin K (2016) Multi-model inference for in-
603 corporating trophic and climate uncertainty into stock assessments. *Deep Sea*
604 *Research Part II: Topical Studies in Oceanography* 134:379–389

605 ICES Database of Trawl Surveys (DATRAS) (2015) International Bottom Trawl
606 Survey (IBTS) data 1985-2014. URL <http://datras.ices.dk>

607 Johnson JB, Omland KS (2004) Model selection in ecology and evolution. *Trends*
608 *in Ecology & Evolution* 19(2):101–108, DOI 10.1016/j.tree.2003.10.013

609 Kadane J, Dickey J, Winkler J, Smith W, Peters S (1980) Interactive elicitation of
610 opinion for a normal linear-model. *Journal of American Statistical Association*
611 75(372):845–854

612 Knutti R (2010) The end of model democracy? *Climate Change* 102:395–404

613 Knutti R, Masson D, Gettelman A (2013) Climate model genealogy: Generation
614 CMIP5 and how we got there. *Geophysical Research Letters* 40(6):1194–1199,
615 DOI 10.1002/grl.50256

616 Kuepfer L, Peter M, Sauer U, Stelling J (2007) Ensemble modeling for anal-
617 ysis of cell signaling dynamics. *Nature Biotechnology* 25(9):1001–1006, DOI
618 10.1038/nbt1330

619 Leith NA, Chandler RE (2010) A framework for interpreting climate model outputs.
620 *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59(2):279–
621 296, DOI 10.1111/j.1467-9876.2009.00694.x

622 Lessler J, Azman AS, Grabowski MK, Salje H, Rodriguez-Barraquer I (2016) Trends
623 in the mechanistic and dynamic modeling of infectious diseases. *Current Epidemi-*
624 *ology Reports* 3(3):212–222, DOI 10.1007/s40471-016-0078-4

625 Li H, Wu J (2006) Uncertainty analysis in ecological studies. In: Wu J, Jones KB,
626 Li H, Loucks OL (eds) *Scaling and Uncertainty Analysis in Ecology: Methods*
627 *and Applications*, 43-64, Springer, pp 43–64

628 Lynam CP, Mackinson S (2015) How will fisheries management measures con-
629 tribute towards the attainment of good environmental status for the North
630 Sea ecosystem? *Global Ecology and Conservation* 4(0):160–175, DOI
631 10.1016/j.gecco.2015.06.005

632 Mackinson S, Platts M, Garcia C, Lynam CP (2018) Evaluating the fishery and
633 ecological consequences of the proposed North Sea multi-annual plan. *PLOS ONE*
634 13(1):e0190015. DOI 10.1371/journal.pone.0190015

635 Morris DJ, Speirs DC, Cameron AI, Heath MR (2014) Global sensitivity analysis
636 of an end-to-end marine ecosystem model of the North Sea: Factors affecting the
637 biomass of fish and benthos. *Ecological Modelling* 273:251–263

638 O’Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ,
639 Oakley JE, Rakow T (2006) *Uncertain judgements: eliciting experts’ probabili-*
640 *ties*. John Wiley and Sons

641 Payne MR, Barange M, Cheung WWL, MacKenzie BR, Batchelder HP, Cormon X,
642 Eddy TD, Fernandes JA, Hollowed AB, Jones MC, Link JS, Neubauer P, Ortiz I,
643 Queirós AM, Paula J (2015) Uncertainties in projecting climate-change impacts
644 in marine ecosystems. *ICES Journal of Marine Science: Journal du Conseil* DOI
645 10.1093/icesjms/fsv231

646 Railsback SF, Grimm V (2012) *Agent-based and individual-based modeling a prac-*
647 *tical introduction*. Princeton University Press

648 Robert CP (2007) *The Bayesian Choice*, 2nd edn. Springer, New York

649 Rougier J (2016) Ensemble Averaging and Mean Squared Error. *Journal of Climate*
650 29(24):8865–8870, DOI 10.1175/JCLI-D-16-0012.1

651 Rougier J, Goldstein M, House L (2013) Second-Order Exchangeability Anal-
652 ysis for Multimodel Ensembles. *Journal of American Statistical Association*
653 108(503):852–863

654 Scott F, Blanchard JL, Andersen KH (2014) mizer: an R package for multispecies,
655 trait-based and community size spectrum ecological modelling. *Methods in Ecol-*
656 *ogy and Evolution* 5(10):1121–1125, DOI 10.1111/2041-210X.12256

657 Speirs D, Guirey E, Gurney W, Heath M (2010) A length-structured partial ecosys-
658 tem model for cod in the north sea. *Fisheries Research* 106(3):474 – 494, DOI
659 10.1016/j.fishres.2010.09.023

660 Spence MA, Blackwell PG, Blanchard JL (2016) Parameter uncertainty of a dy-
661 namic multi-species size spectrum model. *Canadian Journal of Fisheries and*
662 *Aquatic Sciences* 73(4):589–597

663 Szuwalski CS, Thorson JT (2017) Global fishery dynamics are poorly predicted by
664 classical models. *Fish and Fisheries* DOI 10.1111/faf.12226

665 Tebaldi C, Sansó B (2009) Joint projections of temperature and precipita-
666 tion change from multiple climate models: a hierarchical Bayesian approach.
667 *Journal of Royal Statistics Society A* 172(1):83–106, DOI 10.1111/j.1467-
668 985X.2008.00545.x

669 Thorpe RB, Le Quesne WJF, Luxford F, Collie JS, Jennings S (2015) Evaluation
670 and management implications of uncertainty in a multi-species size-structured
671 model of population and community responses to fishing. *Methods in Ecology*
672 *and Evolution* 6(1):49–58

673 Tittensor DP, Eddy TD, Lotze HK, Galbraith ED, Cheung W, Barange M, Blan-
674 chard JL, Bopp L, Bryndum-Buchholz A, Büchner M, Bulman C, Carozza DA,
675 Christensen V, Coll M, Dunne JP, Fernandes JA, Fulton EA, Hobday AJ, Huber
676 V, Jennings S, Jones M, Lehodey P, Link JS, Mackinson S, Maury O, Niiranen S,
677 Oliveros-Ramos R, Roy T, Schewe J, Shin YJ, Stock CA, Underwood P, Volkholz
678 J, Watson JR, Walker ND (2017) A protocol for the intercomparison of marine
679 fishery and ecosystem models: Fish-MIP v1.0. *Geoscientific Model Development*
680 *Discussions* 2017:1–39, DOI 10.5194/gmd-2017-209

681 Walker ND, Maxwell DL, Le Quesne WJF, Jennings S (2017) Estimating efficiency
682 of survey and commercial trawl gears from comparisons of catch-ratios. *ICES*
683 *Journal of Marine Science* 74(5):1448–1457, DOI 10.1093/icesjms/fsw250

684 Williams PJ, Hooten MB (2016) Combining statistical inference and decisions in
685 ecology. *Ecological Applications* 26(6):1930–1942, DOI 10.1890/15-1593.1

686

Tables

687

1 A summary of the variables in the ensemble model. The ensemble
 model is run for $t = 1 \dots T$ 33

688

689

2 A summary of the simulators, their outputs used in the case study,
 the simulator-specific function, $\mathbf{u}_i^{(t)} = f_i \mathbf{x}_i^{(t)} = M_i 10^{\mathbf{x}_i^{(t)}}$ and a refer-
 ence to where the parameter uncertainty, Σ_i , was calculated. 34

690

691

Table 1: A summary of the variables in the ensemble model. The ensemble model is run for $t = 1 \dots T$.

Variable	Dimension	Times	Description	Relationship
$\mathbf{y}^{(t)}$	n	$t = 1 \dots T$	The truth	$\mathbf{y}^{(t)} = \mathbf{y}^{(t-1)} + \boldsymbol{\epsilon}_{\Lambda,t}$
$\mathbf{w}^{(t)}$	n_y	$t = 1 \dots T$	Possibly incomplete version of the truth	$\mathbf{w}^{(t)} = f_y(\mathbf{y}^{(t)})$
$\hat{\mathbf{w}}^{(t)}$	n_y	$t \in S_0$	Noisy observation of $\mathbf{w}^{(t)}$	$\hat{\mathbf{w}}^{(t)} \sim p(\hat{\mathbf{w}}^{(t)} \mathbf{w}^{(t)})$
$\boldsymbol{\delta}$	n	NA	Long-term shared discrepancy	
$\boldsymbol{\eta}^{(t)}$	n	$t = 1 \dots T$	Short-term shared discrepancy	$\boldsymbol{\eta}^{(t)} = R_\eta \boldsymbol{\eta}^{(t-1)} + \boldsymbol{\epsilon}_{\eta,t}$
$\boldsymbol{\mu}^{(t)}$	n	$t = 1 \dots T$	Simulator consensus	$\boldsymbol{\mu}^{(t)} = \mathbf{y}^{(t)} + \boldsymbol{\delta} + \boldsymbol{\eta}^{(t)}$
γ_i	n	NA	Simulator i 's long-term individual discrepancy	
$\mathbf{z}_i^{(t)}$	n	$t = 1 \dots T$	Simulator i 's short-term individual discrepancy	$\mathbf{z}_i^{(t)} = R_i \mathbf{z}_i^{(t-1)} + \boldsymbol{\epsilon}_{z,t,i}$
$\mathbf{x}_i^{(t)}$	n	$t = 1 \dots T$	Simulator i 's best guess	$\mathbf{x}_i^{(t)} = \boldsymbol{\mu}^{(t)} + \gamma_i + \mathbf{z}_i^{(t)}$
$\mathbf{u}_i^{(t)}$	n_i	$t = 1 \dots T$	Simulator i 's incomplete version of $\mathbf{x}_i^{(t)}$	$\mathbf{u}_i^{(t)} = f_i(\mathbf{x}_i^{(t)})$
$\hat{\mathbf{u}}_i^{(t)}$	n_i	$t \in S_i$	The expectation of simulator i 's output $\mathbf{u}_i^{(t)}$	$\hat{\mathbf{u}}_i^{(t)} = \hat{\mathbf{u}}_i^{(t)} + \boldsymbol{\epsilon}_{u_i}$

Table 2: A summary of the simulators, their outputs used in the case study, the simulator-specific function, $\mathbf{u}_i^{(t)} = f_i \mathbf{x}_i^{(t)} = M_i 10^{\mathbf{x}_i^{(t)}}$ and a reference to where the parameter uncertainty, Σ_i , was calculated.

Simulator	Description	Outputs	M_i	Reference for Σ_i
EcoPath	Total biomass is modelled at	1) <i>Common demersal</i>		Mackinson et al (2018)
EcoSim (EwE)	the species level	2) <i>Sole</i> 3) <i>Monkfish etc.</i> 4) Sum of <i>Poor cod and Rays</i> and <i>Other demersal fish</i>	$M_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$	
mizer	Total weight is modelled in weight classes by species	1) <i>Common demersal</i> 2) <i>Sole</i> for $t = 1968 - 2100$.	$M_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$	Spence et al (2016)

This study, see Appendix B

Abundance in length classes is
modelled by species

- 1) *Common demersal*
for $t = 1990 - 2098$.

$$M_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

This study, see Appendix B

Biomass is modelled for different functional groups

- 1) Sum of *Common demersal*, *Sole*, *Monkfish etc.*, *Poor cod and Rays* and *Other demersal fish*
for $t = 1983 - 2050$.

$$M_4 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Thorpe et al (2015)

Abundance in length classes is
modelled by species

- 1) *Common demersal*
- 2) *Sole*
- 3) *Monkfish etc.*
- 4) *Poor cod and Rays*
for $t = 2000 - 2099$

$$M_5 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Figures

692

693

1 A schematic that shows an example of the ensemble model at time t .

694

In this example we have four simulators that are all able to predict

695

the elements of $\mathbf{y}^{(t)}$. Each simulator's 'best guess', $\mathbf{x}_i^{(t)}$, is observed

696

with parameter uncertainty where $\hat{\mathbf{u}}_i^{(t)}$ is the expected output of the

697

i th simulator (see Section 2.1). The difference between the i th simu-

698

lator's 'best guess', $\mathbf{x}_i^{(t)}$, and the simulator consensus, $\boldsymbol{\mu}^{(t)}$, is known

699

as simulator i 's individual discrepancy and is split between its long-

700

term, $\boldsymbol{\gamma}_i$, and short-term, $\mathbf{z}_i^{(t)}$, individual discrepancy (see Section

701

2.2). The difference between the truth, $\mathbf{y}^{(t)}$ and the simulator con-

702

sensus, $\boldsymbol{\mu}^{(t)}$, is known as the shared discrepancy and is divided into

703

long-term, $\boldsymbol{\delta}$, and short-term, $\boldsymbol{\eta}^{(t)}$, shared discrepancy (see Section

704

2.3). In addition, we do not directly observe the truth but we do

705

observe a noisy version of it, $\hat{\mathbf{w}}^{(t)}$ (see Section 2.4). 37

706

2 Estimates of the log biomass of each group of species relative to 2010.

707

The solid line is the median and the dotted lines are the upper and

708

lower quartiles. The first vertical line is at 1986, the year that we

709

first have data, and the second line is in 2013, the simulated cessation

710

of fishing. 39

711

3 The total biomass of demersal species as predicted by the models

712

relative to 2010. 40

713

4 The median best guess for the simulators (\mathbf{x}_i) for mizer (black), Fish-

714

SUMs (purple), LeMans (green), EwE (red) and StrathE2E (pink)

715

and the median simulator consensus ($\boldsymbol{\mu}$) and its quartiles in solid

716

grey and dotted grey respectively. 41

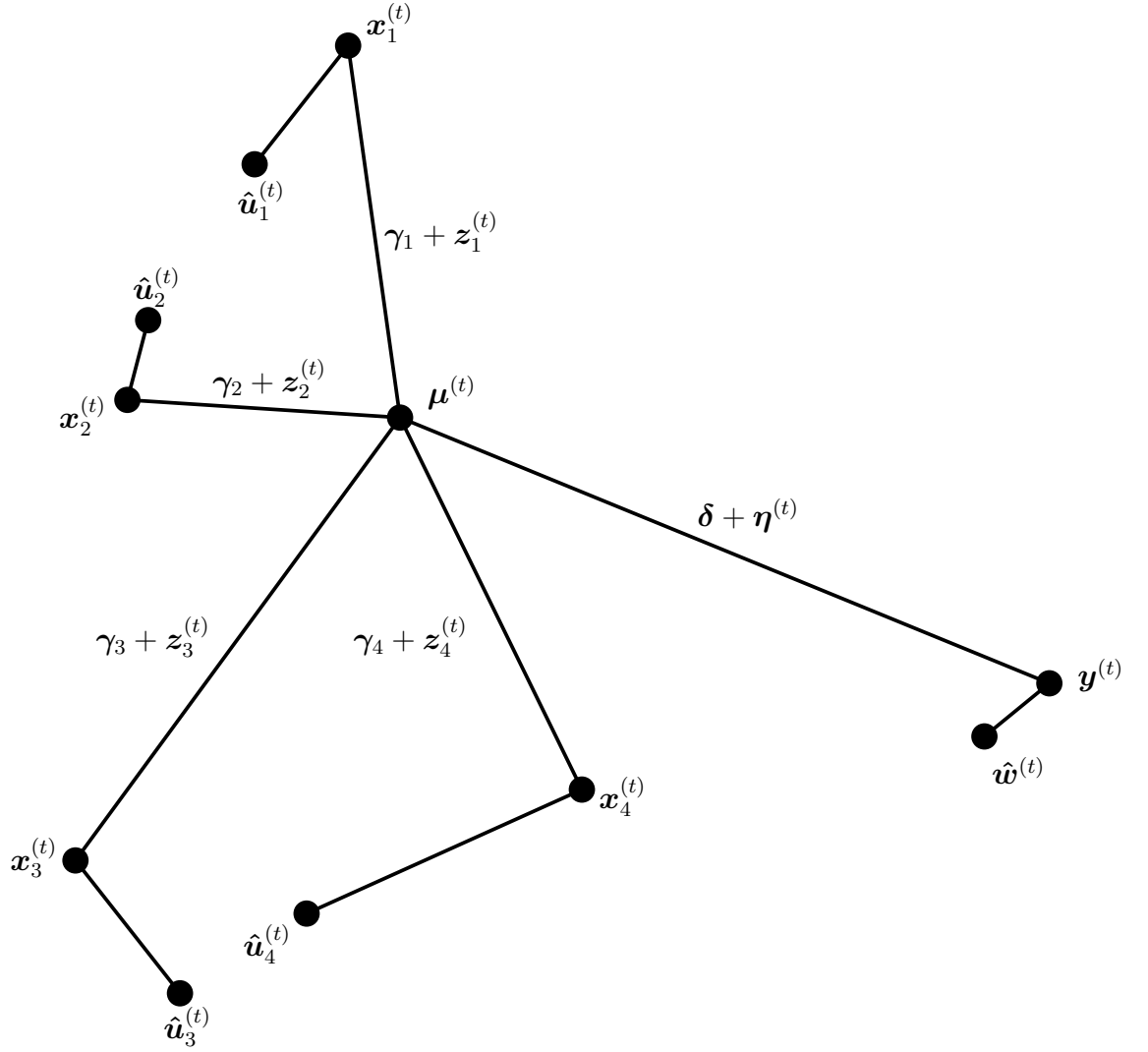


Figure 1: A schematic that shows an example of the ensemble model at time

t . In this example we have four simulators that are all able to predict the

elements of $\mathbf{y}^{(t)}$. Each simulator's 'best guess', $\mathbf{x}_i^{(t)}$, is observed with param-

eter uncertainty where $\hat{\mathbf{u}}_i^{(t)}$ is the expected output of the i th simulator (see

Section 2.1). The difference between the i th simulator's 'best guess', $\mathbf{x}_i^{(t)}$, and

the simulator consensus, $\boldsymbol{\mu}^{(t)}$, is known as simulator i 's individual discrep-

ancy and is split between its long-term, γ_i , and short-term, $\mathbf{z}_i^{(t)}$, individual

discrepancy (see Section 2.2). The difference between the truth, $\mathbf{y}^{(t)}$ and the

727 simulator consensus, $\boldsymbol{\mu}^{(t)}$, is known as the shared discrepancy and is divided
728 into long-term, $\boldsymbol{\delta}$, and short-term, $\boldsymbol{\eta}^{(t)}$, shared discrepancy (see Section 2.3).
729 In addition, we do not directly observe the truth but we do observe a noisy
730 version of it, $\hat{\boldsymbol{w}}^{(t)}$ (see Section 2.4).

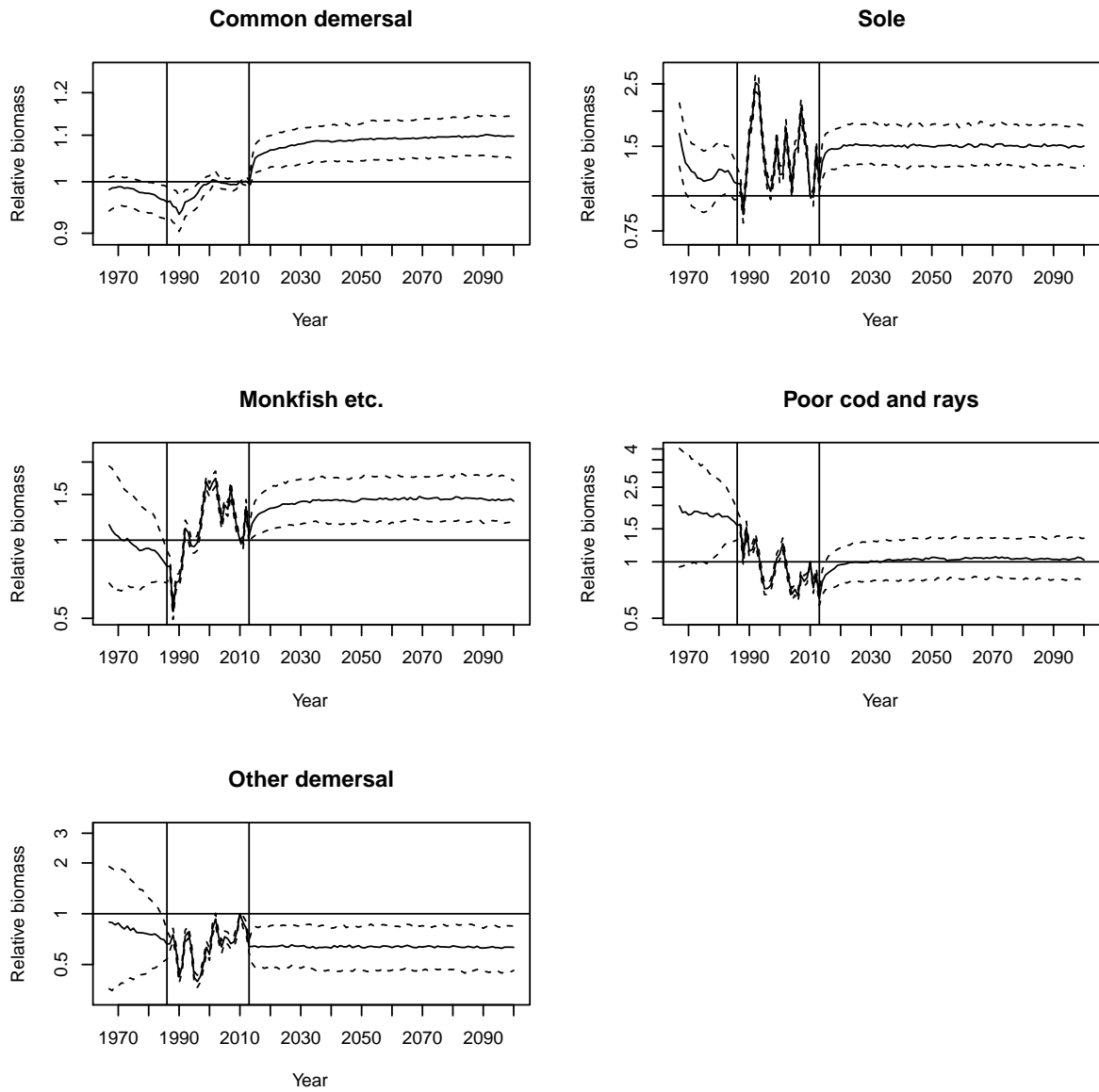


Figure 2: Estimates of the log biomass of each group of species relative to 2010. The solid line is the median and the dotted lines are the upper and lower quartiles. The first vertical line is at 1986, the year that we first have data, and the second line is in 2013, the simulated cessation of fishing.

Total demersal

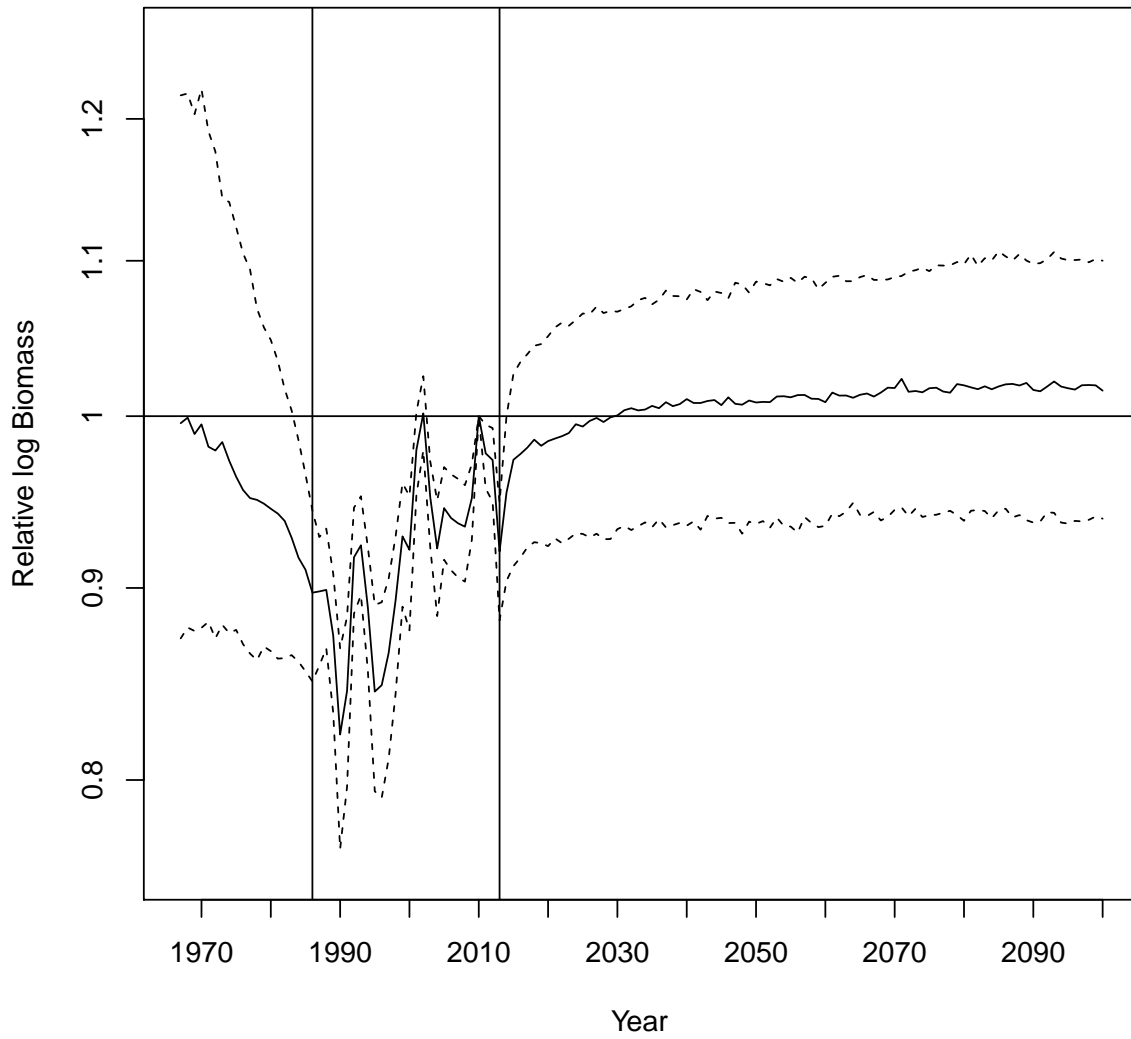


Figure 3: The total biomass of demersal species as predicted by the models relative to 2010.

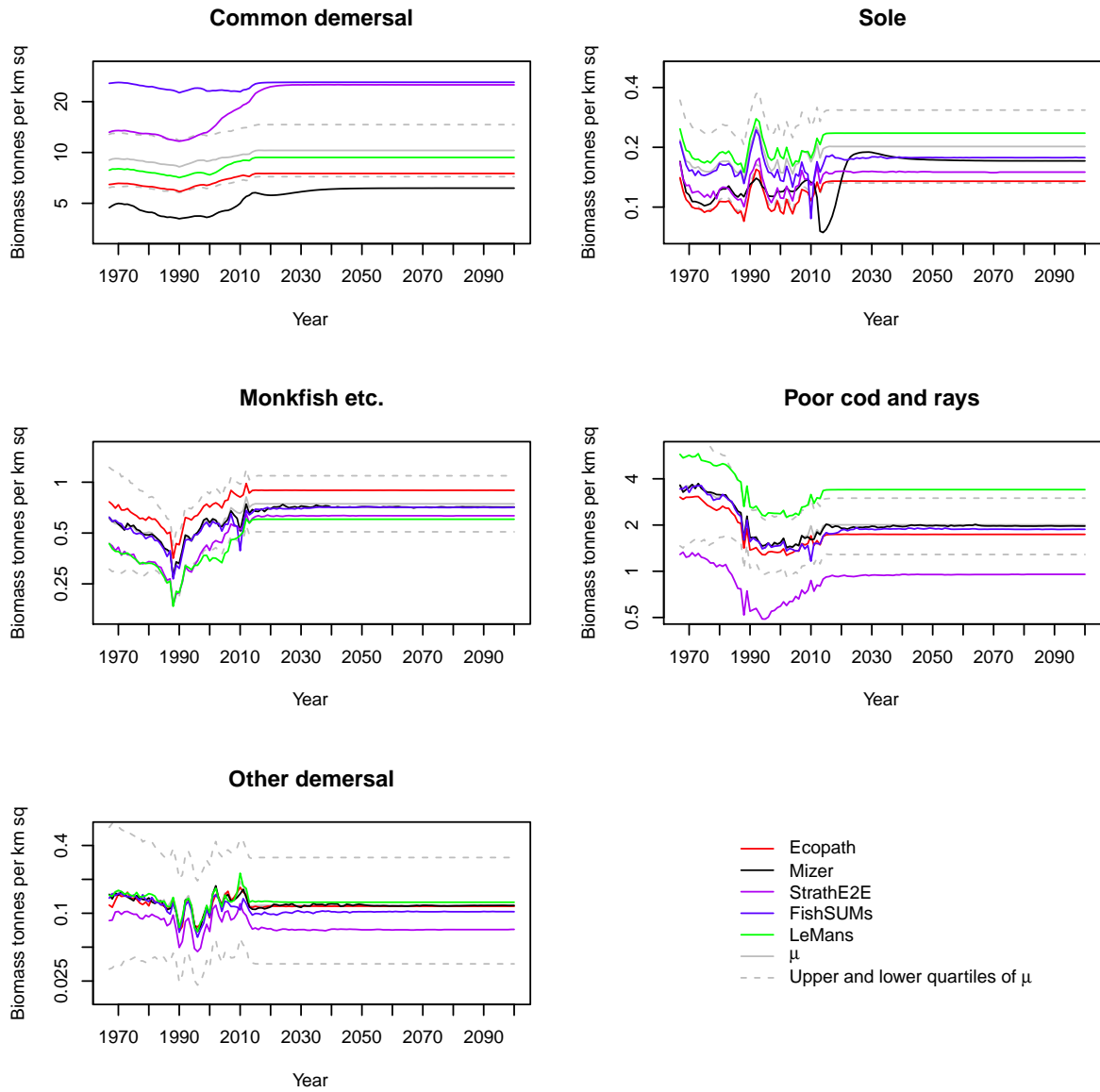


Figure 4: The median best guess for the simulators (\mathbf{x}_i) for mizer (black), FishSUMs (purple), LeMans (green), EwE (red) and StrathE2E (pink) and the median simulator consensus (μ) and its quartiles in solid grey and dotted grey respectively.