

1 Predicting biodiversity dynamics in response to environmental  
2 change

3 Can we do it? A report from assess.sim.basic.R

4 Ryan Batt

5 2015-08-23

6 **Abstract**

7 I use simulated data to evaluate the ability of a multispecies occupancy model (MSOM) to estimate  
8 species richness. An emphasis is placed on the potential to use MSOMs to estimate richness over time  
9 and space, although the current version of the simulation and of the MSOM is somewhat simplistic in  
10 this regard. The principal finding is that the MSOM can be sensitive to changes in detectability. Not  
11 emphasized explicitly in this document, but discovered through experimenting in different versions, is  
12 that these sensitivities can depend quite heavily on sample size. Therefore, it may be prudent to analyze  
13 all years of trawl data at once, rather than separately. Finally, some of the methods for evaluating the  
14 performance of the MSOM on simulated data may provide a useful guide for how to gauge the reliability  
15 of the results when the MSOM is applied to empirical data. This document represents a first step towards  
16 what I think we want to do in a paper, and also serves to get everyone on the same page. I look forward  
17 to further progress and conversation.

# <sup>18</sup> Contents

<sup>19</sup> <b>Introduction</b>	<b>3</b>
Overview . . . . .	3
The Simulation . . . . .	3
Multispecies Occupancy Models (MSOMs) . . . . .	4
<sup>23</sup> <b>Conventions and Settings</b>	<b>5</b>
Dimension Conventions . . . . .	5
Settings . . . . .	6
<sup>26</sup> <b>Species Richness</b>	<b>8</b>
Definition of species richness . . . . .	8
Regional Richness . . . . .	8
Site Specific Richness ( $N_{site}$ ) . . . . .	11
<sup>30</sup> <b>Occupancy Probability, <math>\psi</math></b>	<b>14</b>
Definition of $\psi$ . . . . .	14
Scatter Plot of $\psi$ . . . . .	14
Scatter Plots for Each $\psi_{t,r}$ . . . . .	15
Occupancy Response Curves . . . . .	17
<sup>35</sup> <b>Probability of Detection, <math>p</math></b>	<b>20</b>
Definition of $p$ . . . . .	20
Scatter Plot of $p$ . . . . .	20
Scatter Plots for Each $p_{t,r}$ . . . . .	21
<sup>39</sup> <b>Assessment with Mixed Effects Models</b>	<b>24</b>
Describe Motivation for Mixed Effects Models . . . . .	24
Example LMER Analysis for $\psi$ . . . . .	24
<sup>42</sup> <b>Conclusion</b>	<b>28</b>
Discussion of Results . . . . .	28
Next Steps . . . . .	28
Concluding Remarks . . . . .	28
<sup>46</sup> <b>Report Generation Notes</b>	<b>29</b>
R Session Information . . . . .	29
Date Document Last Compiled . . . . .	29

50 **Introduction**

51 **Overview**

52 As water temperatures change, species may shift the size and location of their geographical ranges, bearing  
53 consequences for the food webs and economies linked to those species. However, species don't always respond  
54 similarly to shifting temperatures (different thermal tolerances, e.g.), which means that changing temperature  
55 may remix the composition and diversity of ecological communities.

56 The biological, spatial, and temporal scale of community diversity shifting in response to climate is massive.  
57 A functional definition of a community may consist of 100's or 1000's of species, each of which may be  
58 shifting its range at a scale of decades and 100's kilometers. As a result, we need statistical methods for  
59 estimating biodiversity that don't rely on heavy replication and that make efficient use of available data.  
60 Enter the superstars: on the data side, the trawl data set has amazing spatiotemporal and taxonomic extent  
61 and resolution; on the statistical side, multispecies occupancy models (MSOMs) are hierarchical state space  
62 models that are designed to estimate species richness and don't require consistent or extensive "replication".  
63 Although they're superstars, even these data and models have their limitations and pitfalls.

64 Can we estimate the dynamics of species richness from trawl data using an MSOM? It's a hard question to  
65 answer because we can never know the "truth" for sure, but we can get an idea of how reliable our analysis  
66 is by simulating fake data, for which we know true values because we created them. The trawl data set is  
67 generated by two distinct processes: Nature's data generating process (NDGP), and the process by which  
68 humans observe the result of NDGP. So we ask: to what extent is the accuracy of estimates from an MSOM  
69 dependent on characteristics of NDGP, and in particular, the way in which we observe the result of NDGP?  
70 The strategy for answering this question is to simulate fake data where we approximate Nature but gain  
71 knowledge of "truth", "observe" the results of the true process, then try to recover the true species richness  
72 from these simulated data.

73 **The Simulation**

74 The goal of this simulation was to use a very basic process to generate presences and absences of species  
75 in space and time. In this version of the simulation, there is no explicit connection between years (they  
76 are independent). There is a modest spatial connection, because in the simulation a spatially-explicit  
77 environmental variable determines habitat suitability. I think of this environmental variable as temperature,  
78 and I filled a grid with temperatures that ranged from the coldest at the top of the grid (north) and  
79 the warmest at the bottom (south) and added random variation among columns in the same row (among  
80 longitudes at the same latitude).

81 One level of the simulation mimics NDGP. In this level, NDGP is best characterized by  $\psi$ , which is the product  
82 of a temperature and species' response curves. I.e., temperatures were used to determine the suitability of  
83 each grid cell to each simulated species. This suitability is known as  $\psi$  throughout this document.

84 A second level of the simulation mimics human observation of NDGP — what we do when we collect data.  
85 This process was simulated by assigning each species has a unique probability of being observed or "detected"  
86 (this variable is  $p$ ). The observation process gets several attempts at observing a given species in a given grid  
87 cell; think of this as subdividing each site into subsites, and when you visit each subsite you have probability  
88  $p$  of observing a particular species (each species has its own  $p$ ). Depending on the settings used in the analysis  
89 that this document summarizes, the maximum number of subsites can vary, as can the number of subsites per  
90 site (OK, fine; the maximum number of subsites in this version is 4, the number of subsites per site varied  
91 between 1 and 4, and overall 50% of total possible subsites were sampled).

92 As previously mentioned, the simulation included "time". In this basic version, not much changes between  
93 the "years" for the true process (temperature doesn't change, nor do the response curves), but the mean of  $p$   
94 does change. In a given year, the entire community has an overall mean probability of being detected, and  
95 each species randomly deviates from that mean.

96 The simulation also has replicates. To understand the replicates, it needs to be clear that even when a  
97 parameter in the simulation does not change, the outcome can change. The replicates hold the realization  
98 of the simulated NDGP constant, and draw new realizations of the observation process. I.e., both  $\psi$  and  
99  $p$  are constant among replicates, and the binary *outcome* of  $\psi$  is also held constant, but the outcome for  $p$   
100 can change. Furthermore, although each replicate has same values of  $p$  (both the mean  $p$  and each species'  
101 individualized random draw from that distribution), each replicate switches which year is associated with  
102 which  $p$ 's. In this way we can observe each outcome of Nature's data generating process under a series of  
103 settings for the human observation process.

104 **Multispecies Occupancy Models (MSOMs)**

105 Multispecies occupancy models are Bayesian statespace hierarchical models. They distinguish between truth  
106 and observation of the truth, and many parameters share a common “parent” distribution. They are very  
107 flexible models, and can be adapted to include new types of processes. The MSOM being used here is a  
108 relatively simple version of these models. It predicts the probability of each species existing in a grid cell from  
109 a logistic regression equation that uses a second-order polynomial of the environmental variable as a covariate.  
110 The parameters in this level of the model are hierarchical, with species having their own parameter values,  
111 but these individual parameters are not wholly independent in the sense that they share a common parent  
112 distribution, which sort of acts to both limit how different they can be and to inform one another. The model  
113 also has an observation level, which only has a hierarchical intercept (just a mean) as a predictor variable.  
114 The MSOM makes guesses of the true state of the system (whether a species is actually present or not). It  
115 then makes guesses at how the observation of that true state might turn out, which is effectively a prediction  
116 of what our data will be. The Bayesian model fitting process then uses this comparison of the observed data  
117 to the estimate of the observation to tweak the parameters in the MSOM. This process is repeated until the  
118 choice of parameters boils down to what is essentially the posterior distribution of the estimated parameters.  
119 Right now the MSOM model is fit separately to each year and each replicate. So the model never gets to see  
120 multiple years or multiple replicates at the same time. Furthermore, when referring to a parameter value  
121 fitted in the MSOM, it is implied that it can be subscripted with time or replicate (because all years and  
122 replicates are fit independently).  
123 The parameters in the logistic regression that predicts the value of  $\psi$  vary among species, although  $\psi$  itself  
124 varies among species and space, because the regression parameters (subscripted by species) are multiplied by  
125 the environmental variable (subscripted by space). More or less, it can be said that, for a given species,  $\psi$   
126 varies among space because of the environmental variable, and in a given location it varies among species  
127 because of the regression parameters.

128

129

---

130 **Conventions and Settings**

131 In this section I outline the subscripting and notation used in the MSOM analysis and for the simulation. I  
132 also outline various settings (number of species simulated, replicates, etc.). Most of the numbers you see  
133 (and some of the text) is dynamically generated based on the code that produced the statistics and figures.  
134 Therefore, you can refer back to these sections to see what settings may have changed since the last version  
135 of this document.

136 **Note:** *I've often found myself having to get creative with subscripts and superscripts. I've tried to be clear an*  
137 *consistent, but small inconsistencies likely exist, so don't be confused by them. For example, if you see  $\max(Z)$*   
138 *and  $Z_{\max}$  in two different sections, they are probably referring to the same thing. If you see something*  
139 *confusing, let me know (preferably by [creating an issue on GitHub](#)), and I'll fix it.*

140 **Dimension Conventions**

141 **Summary**

142 1. Site ( $j = 1, 2, \dots, j_{\max} = 9 \times 9 = 81$ )

- 143 • Sites are unique combinations of latitude and longitude  
144 • The spatial arrangement of sites is *not* arbitrary, although importance depends on settings (see  
145 `dynamic` below)  
146 • The environmental variable  $X$  varies among sites (and years, below)

147 2. Sub-sites ( $k = 1, 2, \dots$ )

- 148 • Sub-sites are only relevant to the “observation” process  
149 • Each site has the same number of possible sub-sites, but the number of sub-sites observed can vary  
150 • In this simulation,  $k_{\max} = 4$ ,  $k_{\min}^{\text{observed}} = 1$ , and  $k_{\max}^{\text{observed}} = 4$   
151 • Substrata are primarily useful for determining  $p$ , the **detection probability**

152 3. Species ( $i = 1, 2, \dots, i_{\max} = R = 30$ )

- 153 • Does not include “augmented” species  
154 • For this MSOM analysis, the species array was padded with 10 0’s

155 4. Time ( $t = 1, 2, \dots, 4$ )

- 156 • Time is primarily used to vary the parameters controlling the “true” process  
157 • When those parameters don’t change, time provides independent\*realizations of the same “true”  
158 process  
159 — \*Note: only when `dynamic=FALSE` in `sim.spp.proc` (`dynamic` is FALSE in the basic simulation)

160 5. Replicates ( $r = 4$ )

- 161 • Replicates are *simulated* repeated human observations of the same *realization* of the “true” process  
162 at Time $_t$   
163 • Replicates are used to vary the parameters that control the “observation” process  
164 • When those parameters don’t change, each replicate provides an independent\*realization of the  
165 same “observation” process

169 **In Code**

170 The MSOM analyzes each  $year_t$ -replicate $_r$  combination independently. Parameters subscripted by these  
171 dimensions are derived from separate analyses.

172 In my code, I've tried to be consistent in my use of these indices to describe arrays, matrices, and rasters.  
173 Rows are dimension 1, columns dimension 2, etc. The precedent of subscripted dimensions follows the  
174 numeric ordering of the list above. E.g., in the matrix  $X_{i,t}$  each row will refer to a different species, and  
175 each column a different year (note that site $_j$  is skipped, so species $_i$  is "promoted" to dimension 1, the row.).  
176 By default, R fills matrices and arrays by column, whereas the **raster** package fills them by row. In most  
177 cases where an R object needs to split sites into the lat/ lot components, I make use of the **raster** package.  
178 Therefore, the numbering of the sites proceeds row-wise, where each site is numbered according to the order

179 in which it is filled, as in this  $2 \times 3$  matrix:  $J = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$

180 Note that even though this matrix is numbered row-wise, it is still indexed as  $J_{row,column}$ , such that  $J_{1,2} = 2$ .  
181 As mentioned previously, this information is primarily important for understanding the code involved with  
182 this project, and in most cases it is not crucial to be explicitly aware of the spatial arrangement of sites.

183

184

---

185 **Settings**

186 **Simulation Settings**

187 I created a class called "spp", which has methods for **print()**. The **Dimensions** are the number of sites, the  
188 number of species, then the number of years.

189 Also printed are some richness summary statistics. **All cells** refers to the collective richness over all  $j$   
190 taken together. The meaning of **One cell** differs slightly between the true and observed printouts: in the  
191 true printout the richness is of a particular site ( $j$ ), and in the observed printout it is of a particular sub-site  
192 ( $k$ ).

193 **## Dimensions:** 81, 30, 4  
194 **## grid.h** = 9  
195 **## grid.w** = 9  
196 **## grid.t** = 4  
197 **##**  
198 **## Number Species Possible (ns):**  
199 **## 30**  
200 **## Total Species Richness:**  
201 **## 30**  
202 **## Total Observed Species Richness:**  
203 **## 30**  
204 **##**  
205 **## Annual Species Richness:**  
206 **##** Min. 1st Qu. Median Mean 3rd Qu. Max.  
207 **## All cells** 29 29 29 29.25 29.25 30  
208 **## One cell** 5 10 14 13.63 17.00 24  
209 **##**  
210 **##**  
211 **## Observed Annual Species Richness:**  
212 **##** Min. 1st Qu. Median Mean 3rd Qu. Max.  
213 **## All cells** 27 27.75 28.5 28.250 29 29  
214 **## One cell** 0 0.00 0.0 4.008 7 23

215 In the MSOM, detectability ( $p_i$ ) is determined in the form of a logistic regression, which currently only  
 216 has an intercept ( $v_0$ ) as a predictor (so just a mean). That intercept varies among species (i.e.,  $v_{0,i}$ ), and  
 217 that variation is generated by drawing each individual species's intercept ( $v_{0,i}$ ) from a parent distribution:  
 218  $v_{0,i} \sim \mathcal{N}(\mu_{v_0}, \sigma_{v_0}^2)$ . See [section about  \$p\$](#)  for more info.

year	mu.v0	sigma.v0
1	-2	2
2	0	2
3	2	2
4	4	2

219

---

220 **Settings for JAGS & MSOM**

nChains	nIter	n0s	nSamples
3	50000	10	500

221 In the table above, **nChains**, **nIter**, and **nSamples** are all variables that are strictly pertinent to the Bayesian  
 222 analysis carried out in JAGS. The **n0s** value refers to the the degree of “data augmentation”. In this process,  
 223 you add extra species to the data set, and say that they were never observed. For our purposes, this is  
 224 employed for purely technical reasons, although it can be used to extra further inferences about species  
 225 richness.  
 226 The posterior samples from JAGS consist of 500 samples (above). However, to save memory and hard drive  
 227 space, I have often only saved measures of central tendency for each of these. In this assessment, I have  
 228 performed all calculations on the **centralT=median** of the posterior samples.

229

230

---

231 **Species Richness**

232 **Definition of species richness**

233 Species richness is the number of different species, or more generically, unique taxa. The point is moot in the  
234 simulation study, and in the empirical trawl data it refers to species.

235 Estimates of richness ( $R$ ) can be made spatially or temporally explicit (or neither, or both). In the following  
236 figures, different levels of aggregation are performed – for most figures  $R$  is split by year (this is true for  
237 all figures but the Boxplot Figure). The Time Series of Richness Figure emphasizes temporal dynamics  
238 and keeps replicates separated, but aggregates over space (the  $j$  sites). The Nsite Scatter Figure doesn't  
239 aggregate over space or time, but it does aggregate over “replicate” observations; importantly, while the  
240 figure does present any spatial aggregation, it does not retain the spatial relationship (you can't tell which  
241 sites are next to others). The final figure of the section (Heatmap of Richness Figure) is similar to the  
242 previous figure, except that spatial relationship among points is retained via a heatmap representation.

243 None of these estimates of richness include the 10 species that were part of the “data augmented”/ “adding  
244 0's” process. Richness values can either be true (true simulated NDGP;  $R^{true}$ ), observed (true simulated  
245 human observation of NDGP;  $R^{obs}$ ), or MSOM estimates of one of those two ( $\hat{R}^{true}$  or  $\hat{R}^{obs}$ ).

246

---

247 **Regional Richness**

248 These estimates of species richness only distinguish between replicates and years. They do not contain any  
249 site-specific information.

250 **Richness Boxplots**

251 With the boxplots we're mostly looking to see if the estimates of richness vary with the mean probability of  
252 detection,  $p$ . In the empirical data, we know that taxonomic identification changed over time (it improved;  
253 generally, more species were ID'd in later years). We also suspect that gear might change, which affects  
254 the probability of observing a species. The “Average Detection Probability” category in the boxplots is the  
255 cross-species average of  $p$  (which with large sample size approach the hyperparameter  $p_\mu$ ).

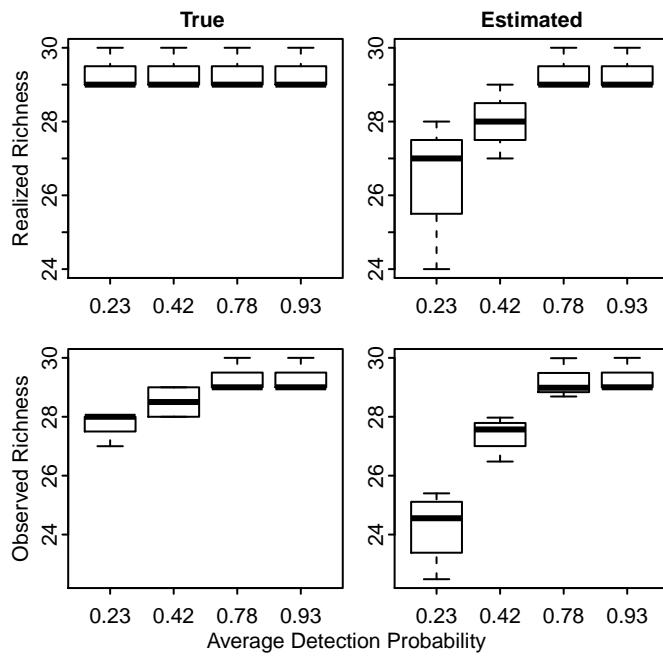


Figure 1: Boxplots of species richness. Numeric groupings indicate the average value of  $p$  across species during a given year–replicate combination. The panels in the left column are the true simulated values, and the panels on the right are the corresponding MSOM estimates. The top row indicates the latent realized species richness or MSOM estimates of the richness. The bottom row’s panels are the simulated observed values of richness (the response variable in the MSOM) and the MSOM estimates of the observed values.

## 257 Richness Time Series

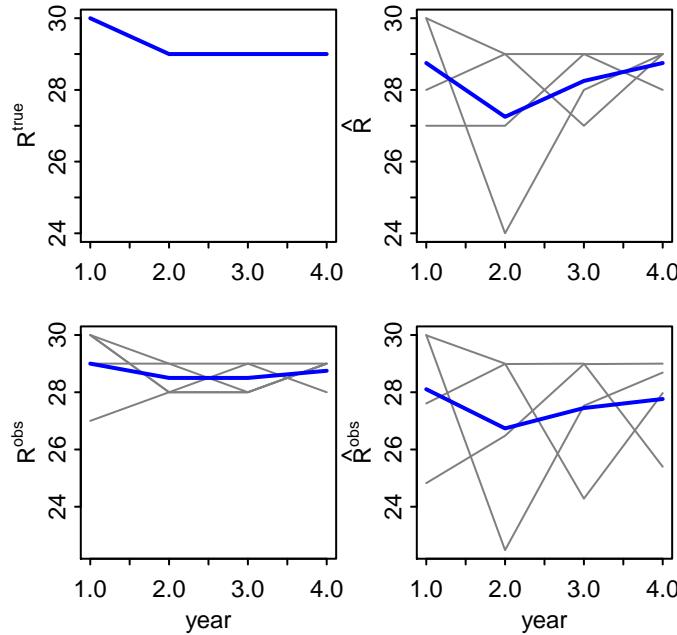


Figure 2: Time series of richness. Gray lines are individual replicates. Blue lines are averages. Note that detection probabilities ( $p_{i,t,r}$ , see [simulation settings above](#), as well as [definition of  \$p\$  below](#)) change over time, and their temporal ordering differs among replicates.

258 1.  $R^{true}$  is just true richness (among all sites) in each year

259  
260 2.  $\hat{R}$  is from  $\sum_{i=1}^{R^{true}} \max(\hat{Z}_{i,\nabla j \in J})$ ;

- 261     • i.e., basically same as for  $R^{true}$ , except from  $\hat{Z}$  instead of  $Z$   
 262     • only difference is that  $\hat{Z}$  contains extra (fake, always absent) species compared to  $Z$  (see next)  
 263     •  $\hat{R}$  doesn't include the augmented species introduced to the MSOM occurrence matrix ( $Y$ )

264  
265 3.  $R^{obs}$  are the values of  $R^{true}$  after they pass through the observation process

- 266  
267     • Note that “true” and “observed” and “estimated” can be confusing here; the last term can prefix either  
 268       of the first 2 terms. This notation needs some work still.

269  
270 4.  $\hat{R}^{obs}$  is the MSOM estimate of what was observed

- 271  
272     • this is the estimate that is compared to the data in the fitting process

- 276 • true values are latent (unobserved)
- 277
- 278 • thus, we need this extra step to connect our estimates of what's actually going on to our data

279

  - 280 • although, it should be noted that our data don't arrive in terms of "richness", but in presences/absences

281 In these "time series" plots (so short!), the replicates (grey lines) look so jagged and inconsistent because each  
282 replicate shifts which year gets which value of  $p_\mu$  (average detectability).

283

284

---

## 285 Site Specific Richness (`Nsite`)

### 286 Scatter Plots of `Nsite` Split by Year

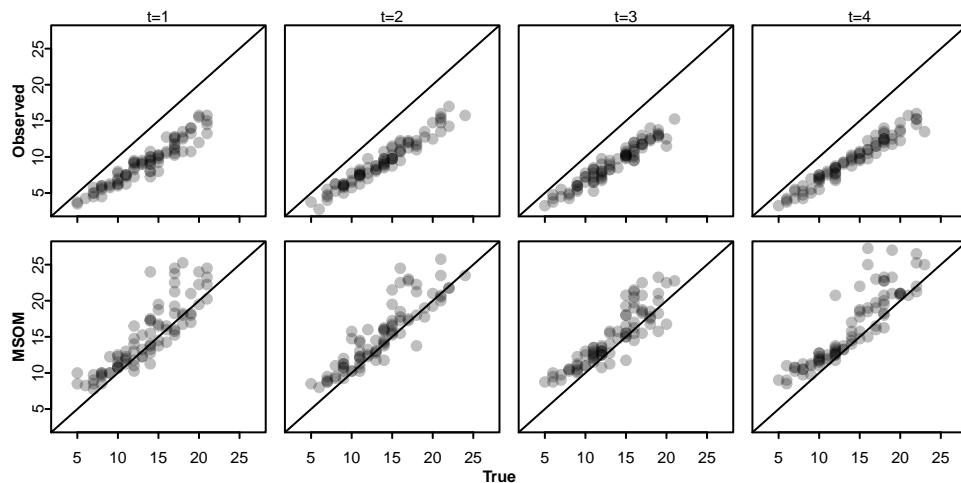


Figure 3: Site-specific richness (`Nsite`,  $N_j$ ) from simulated observations (vertical axis, top row;  $N_j^{obs}$ ) and from MSOM estimates (vertical axis, bottom row,  $\hat{N}_j$ ) vs true site-specific richness (horizontal axis;  $N_j^*$ ). The panel columns delineate the years of the simulation. Each point is site-specific species richness that has been averaged over the simulated replicate observations.

287 The first thing I notice in the `Nsite` scatter plot figure is that the observations tend to underestimate true  
288 richness, and the absolute magnitude of this underestimate increases as true richness increases (i.e., with the  
289 observation as the response and the true value as the predictor, the intercept is fine, but the slope is too  
290 shallow). The MSOM tends to do a much better job of staying near the 1:1 line, but there's more variance  
291 in the residuals at high richness. I'm not sure I understand why (in later figures,  $\psi$  estimates don't get worse  
292 or more uncertain at high values of true  $\psi$ ).

293

294

---

## 295 Maps of Richness (space and time)

296 In these maps, the environmental variable  $X$  changes linearly across the y-axis, and randomly (and much  
297 less) across the x-axis. The different columns represent separate years. The environmental variable changes

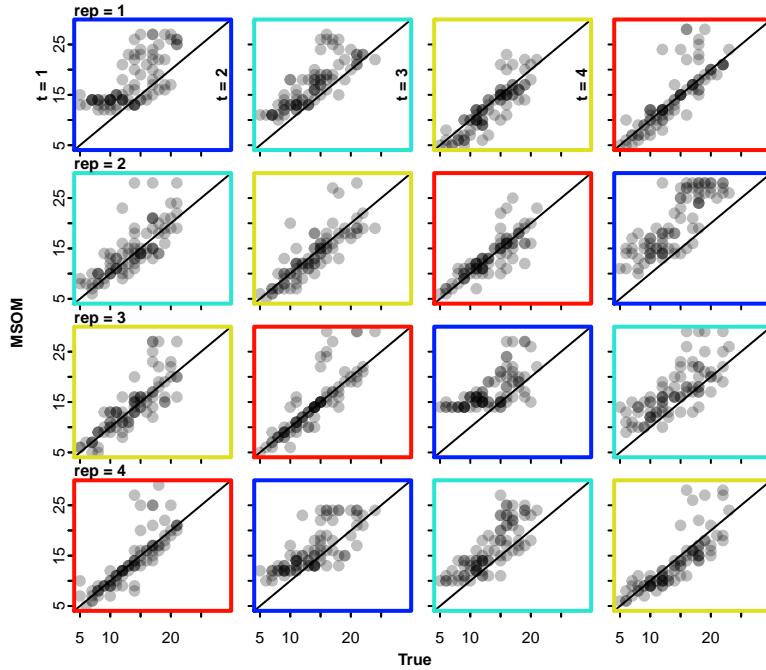


Figure 4: Scatter plots of MSOM-estimated species richness (y-axis) and the true species richness (x-axis), with each panel representing a particular year-replicate combination. The color of the panel border indicates the value of  $p_\mu$ , the community-wide mean detectability.

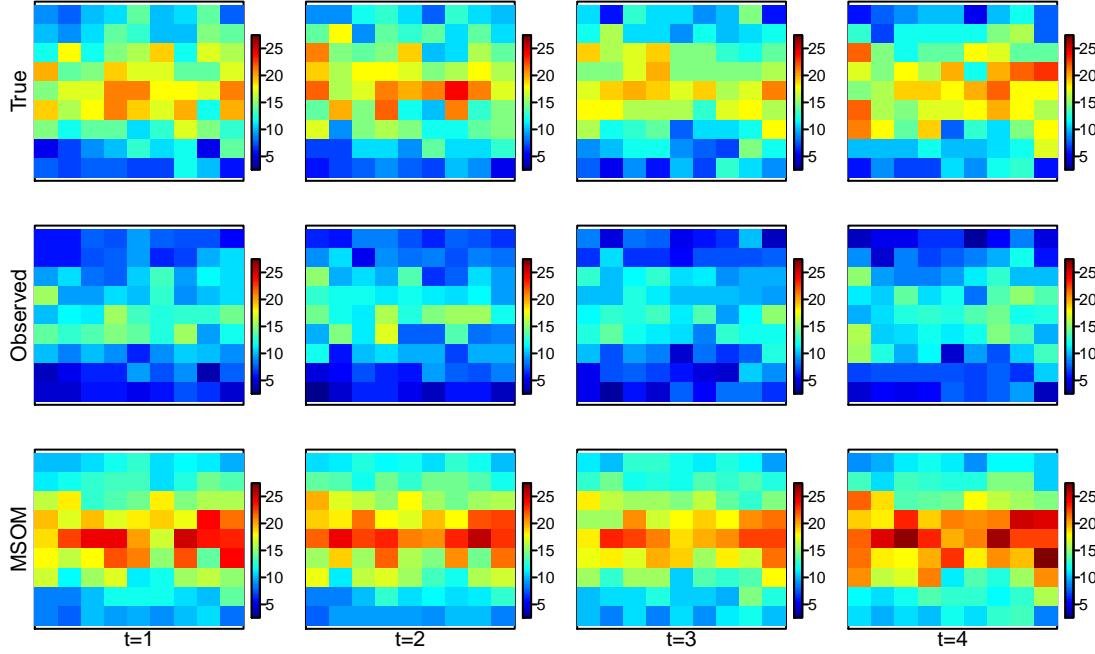


Figure 5: Maps of site- and year-specific species richness ( $N_{site}$ ) from the simulation of the True process (top row), simulation of the Observed process (middle row), and the MSOM estimates (bottom row). X-axis and Y-axis indicate position in 2 dimensional space. Colors indicate species richness (warm colors are higher richness than cool colors), averaged over the simulated replicate observations. The color scale is the same among all panels.

298 linearly among years (the rate of change is the same for all x-y locations), and in this basic simulation that  
299 rate of change is actually just 0 units/year.

300 Because so many things are static in this simulation the maps of `Nsite` are not much more informative than  
301 the scatter plots of `Nsite`. However, making the spatial aspect of richness visually explicit does emphasize  
302 that richness is highly dependent on the environmental variable — looking at the occupancy response curves  
303 below indicates that for the community as a whole the probability of occupancy is highest in the middle of  
304 the environmental gradient, which also happens to be the middle of the “latitudinal” gradient in this figure.  
305 Within a latitudinal band, the MSOM doesn’t do as good a job of parsing the variability in richness.

307 **Occupancy Probability,  $\psi$**

308 **Definition of  $\psi$**

The probability that a particular location  $j$  will be occupied by species  $i$  is  $\psi$  (omitting subscripts). This probability is a function of the environment and 3 species-specific parameters. To calculate  $\psi_i$  under a set of known conditions in an environmental variable  $X$  at time  $t$  and site  $j$ , we can express it as the log of the odds ratio (logit link) resulting from the linear combination of three terms:

$$\text{logit}(\psi_{j,i,t}) = \begin{pmatrix} 1 \\ x_{j,t} \\ x_{j,t}^2 \end{pmatrix}^\top \begin{pmatrix} a_{0,i} \\ a_{3,i} \\ a_{4,i} \end{pmatrix}$$

and

$$\begin{aligned} a_{0,i} &\sim \mathcal{N}(\mu_{a0}, \sigma_{a0}^2) \\ a_{3,i} &\sim \mathcal{N}(\mu_{a3}, \sigma_{a3}^2) \\ a_{4,i} &\sim \mathcal{N}(\mu_{a4}, \sigma_{a4}^2) \end{aligned}$$

309 Thus, the parameters are hierarchical, and for this reason the response curves vary but are somewhat clustered  
310 around a central value.

311 In my mind,  $\psi$  is the Holy Grail of parameters to recover. It tells us the odds that a species will be in a  
312 certain place at a certain time. If I could know this perfectly, I would be very pleased (and we'd all be very  
313 famous).

314 When this analysis (using an MSOM on the trawl data) was originally crafted,  $\psi$  was more of a means to an  
315 end than it was the objective –  $\psi$  lets us get at richness, and we have hypotheses about how richness should  
316 change with climate that we'd like to test. But if you know what controls  $\psi$ , you know what controls richness.  
317

318

---

319 **Scatter Plot of  $\psi$**

320 In a general sense, the MSOM can distinguish between instances (sites/ years) when a species is likely to be  
321 present, and when it's not. However, in every simulation I've done (varying many parameters that aren't  
322 compared in this document), the scatter plot of  $\psi$  always makes it apparent that

- 323 1. There is a lot of variability around the 1:1 line  
324  
325 2. The residuals are not normal, and they are not independent  
326     i. In general, I've found that  $\hat{\psi}$  exhibits an upward bias, overestimating  $\psi^{true}$   
327  
328     ii. Smoothly-curving excursions from the 1:1 line are often prominent

329 These patterns are somewhat concerning. The curve-like sequence of residuals is probably a byproduct of  
330 slightly incorrect estimates of the parameters in the logistic regression ( $[a_0, a_1, a_2]$ ), resulting in estimated  
331 **response curves** that deviate non-randomly from the true response curve. For a heuristic of how these  
332 smooth excursions can occur, in R try something as simple as `d <- rnorm(100); plot(dnorm(d), dt(d,  
333 1))` to see the relationship between the density estimate from the correct distribution and that from  
334 the wrong distribution (the density is analogous to  $\psi$ ); or for really crazy patterns, try `d <- rnorm(100);  
335 plot(dnorm(d), do.call(approxfun, density(d)[c("x", "y")])(d))`. So the curves are explainable, but  
336 I cannot explain the consistent overestimation; I could understand how underestimating detectability ( $p$ )

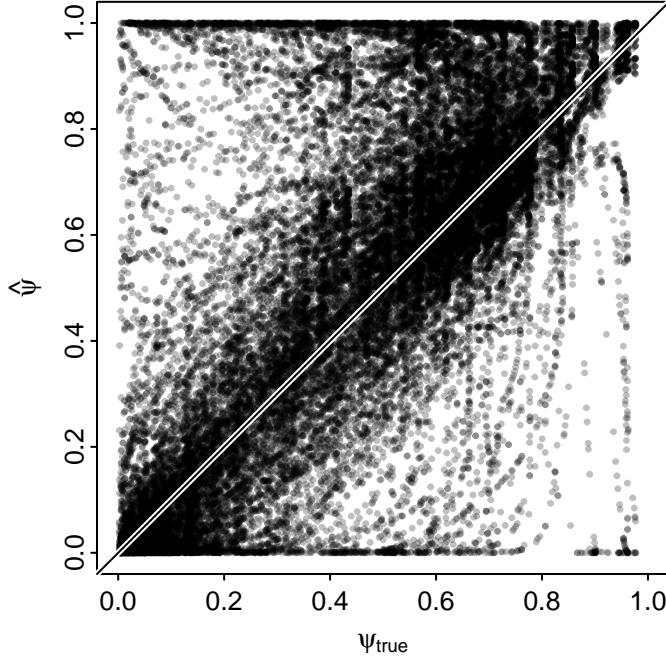


Figure 6: MSOM estimates of  $\psi$  ( $\hat{\psi}$ ) vs. true values of  $\psi$  ( $\psi_{true}$ ). Each point is a  $\psi$  value for a particular site-species-year-replicate. The white and black line is the 1:1 line.

337 would result in overestimating  $\psi$ , but the MSOM appears to recover true  $p$  values rather well (e.g., see [p Scatter Figure](#)), so that's not a satisfying explanation.

339 In the next section I drill into  $\psi$  a bit more to try and understand what causes the largest deviations from  
340 true values.

341

342

---

### 343 Scatter Plots for Each $\psi_{t,r}$

344 The estimates and true values of  $\psi$  are best correlated when  $p$  is high. When the average species has a low  
345 chance of being detected (when  $p_\mu$  is, say, 20%), the estimates of  $\psi$  are a mess.

346 Does this help explain an apparent positive bias in the aggregated scatter plot of  $\psi$ ? Maybe. When  
347 detectability ( $p_\mu$ , the parameter indicated by panel colors) is low, that's when we run into trouble. With  
348 low detectability, you have fewer observations. You have a poorer sense of what's going on. So that adds  
349 uncertainty. Perhaps when detectability is super low, it's entirely too easy to conflate an absence with an  
350 undetected presence — you start assuming that 0's are just because you didn't look hard enough, not because  
351 it's really absent. I'm not convinced by this logic, though; I'd want to see a better explanation. Alternatively,  
352 maybe the chains aren't converging yet; I didn't run diagnostics. So for now this is a mystery to me.

353 Note: what I refer to as  $p$  here is the probability that a species will be detected if an occupied site is sampled,  
354 so the number of substrata sampled per site isn't reflected in  $p$ . In this simulation, 50% of substrata were  
355 sampled, and while being below 100% doesn't this doesn't influence  $p$ , visiting few sites could add noise to  
356 various parameter estimates.

357

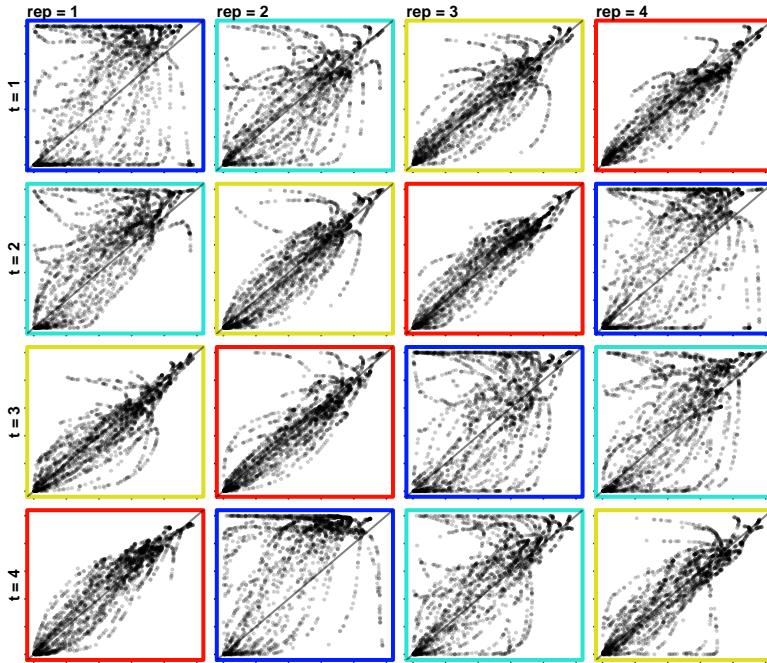


Figure 7: True (horizontal axes) and MSOM estimates (vertical axes) of occupancy probabilities ( $\psi_{j,i,t,r}$ ) of species  $i$  occupying a location  $j$ . Years ( $t$ ) are separated by rows, replicates ( $r$ ) are separated by columns. The border color of each panel indicates the community-level mean probability of detection ( $p_\mu$ ; where  $p_i \sim \mathcal{N}(p_\mu, \sigma^2)$ ), with warm colors indicating high detectability, and cool colors low. The species-specific detectabilities are **not** re-randomized among replicates, but even when the probabilities associated with the observation process do not change, the outcome of the process can change. The year  $t$  of the simulated true process changes across the rows of panels, and the simulated replicate observation  $r$  changes across columns.

---

359    **Occupancy Response Curves**

Occupancy response curves are calculated as  $\text{logit}(\psi_i) = \mathbf{X} \times \mathbf{a}_i$ , where

$$\mathbf{X} = \begin{pmatrix} 1 & X_{min} & X_{min}^2 \\ \vdots & \vdots & \vdots \\ 1 & X_{max} & X_{max}^2 \end{pmatrix}; \mathbf{a}_i = \begin{pmatrix} a_{0,i} \\ a_{3,i} \\ a_{4,i} \end{pmatrix}$$

360 Therefore, these curves are tantamount to values of  $\psi$ , except that  $\psi$  generally pertains to a simulated,  
361 observed, or true occupancy probability, whereas the occupancy probability in the response curves is calculated  
362 over hypothetical conditions (i.e., over hypothetical values of the environmental gradient  $X$ ). This formulation  
363 yields values of  $\psi$  for each species at all temperatures over the interval  $(X_{min}, X_{max})$  (this implies that the  
364 number of rows in  $\mathbf{X}$  is infinite, which it isn't, but it is quite large). Also, note that I'm making a distinction  
365 between  $X$  (environmental variable) and  $\mathbf{X}$  (the matrix, one column of which is  $X$ ).

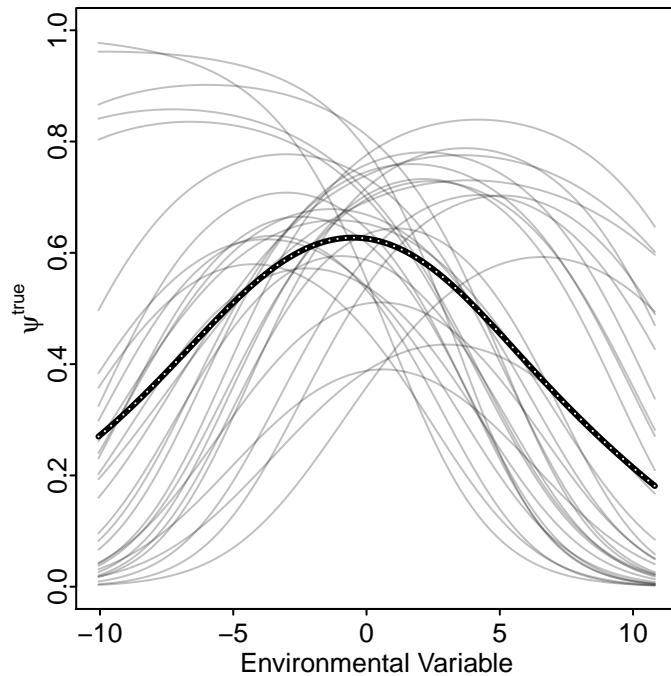
366    **True Occupancy Response Curves**


Figure 8: True simulated response curves. Vertical axis is the value of  $\psi^{true}$ , horizontal axis is the value of the environmental variable that, along with species-specific regression parameters, determines  $\psi^{true}$ . The thick line is the among-species mean value of  $\psi^{true}$  at a given value of the environmental variable.

367 In the response curve, the values of the environmental variable are an arbitrary gradient, and do not  
368 necessarily correspond to what was observed in the simulated environment (although they are intended to  
369 cover the same range). The formulation  $\text{logit}(\psi_i) = \mathbf{X} \times \mathbf{a}_i$  is useful for producing the response curve, but it  
370 is inherently discrete. This is actually how I simulated the true  $\psi$  in the model, but it's not how the MSOM  
371 analyzes it (although the difference is negligible because I use so many rows in  $\mathbf{X}$ ).  
372

374    **Estimated Occupancy Response Curves**

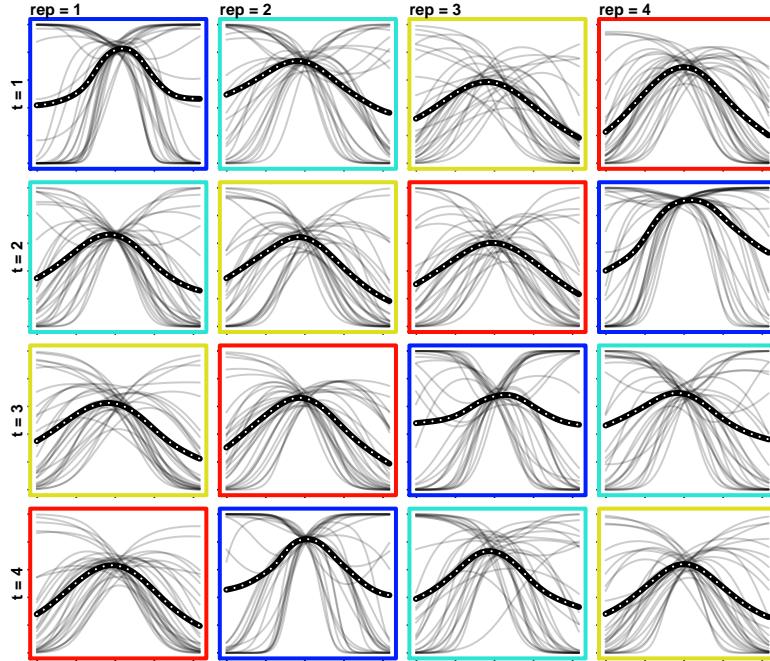


Figure 9: Response curves of species' probability of occupancy ( $\psi_i$ , vertical axis) across the full range of temperatures in the simulation. The color of the boxes around each panel refer to the among-species average of the probability of detection; warm colors indicate that the mean detection probability is high, whereas cool colors indicate that  $p$  was low. The year  $t$  of the simulated true process changes across the rows of panels, and the simulated replicate observation  $r$  changes across columns.

375 It is rather evident in these figures that the estimated response curves look less like the true response curves  
 376 for low detection probabilities (low  $p_\mu$ ). It is important to remember this effect when evaluating the esimates  
 377 of `Nsite`, and especially when looking at the heat maps of `Nsite` in [Figures 4 and 5](#) because in those figures  
 378 each grid cell represents the richness averaged over replicates, and the value of  $p_\mu$  varies among replicates.  
 379 Furthermore, the inaccurate response curves at low  $p_\mu$  are consistent with the relationships between  $\hat{\psi}$  and  
 380  $\psi^{\text{true}}$  in [Figure 7](#).

381 It is my suspicion that, at a large enough sample size of both preseneces and abseneces across a full gradient  
 382 of temperatures, these deficiencies will diminish. However, this analysis, in its basic form, was not designed  
 383 to test the influence of sample size (I've run separate versions of this model, not presented here, and can  
 384 informally confirm my suspicion – sample size matters a lot). Furthermore, most regions in the trawl data set  
 385 have a few dozen sites (site being defined on a 1° grid), so the  $9 \times 9$  grid simulated here is approximately the  
 386 spatial sample size we'd have to work with in the empirical analysis.

387 Two ways to change the sample size in the empirical analysis would be to

- 388    1. reduce grid size to increase the number of sites
- 389        • could do a half-degree grid
- 390        • make substrata 1/4 degree
- 391        • downside is that each site would be represented in a fewer proportion of years
- 392
- 393    2. Use multiple years of data

- 396     • drastically increases sample size  
397  
398     • would require a new model  
399  
400     • the model would need to consider factors that change among years  
401

402

---

403 **Probability of Detection,  $p$**

404 **Definition of  $p$**

405 The probability of detection ( $p$ ), is a species specific parameter in the MSOM model. The MSOM analyzes  
406 all years ( $t$ ) and replicates ( $r$ ) separately, so I am going to leave those subscripts out of this description. In  
407 the simulation, the probability of observing a species is a function of two independent factors:

- 408 1. The probability that site  $j$  is occupied by species  $i$ ; this is  $\psi_{j,i}$
- 409   •  $\psi_{j,i}$  is a function of species-specific niche and an environmental variable that changes over space  
410   and time
- 411   •  $Z_{j,i}$  is the species- and site-specific richness, which is a function of  $\psi$  (given that we're only talking  
412   about species that are in the pool of possible species, determined by  $w_i$ )
- 413
- 414 2. A species-specific ( $i$ ) chance of being identified (`taxChance`), given that it is present in a location that  
415 was sampled (i.e., detectability does not reflect the probability of sampling a place); this detectability  
416 parameter is  $p_i$
- 417   • Detectability changed between years.
- 418   • In a given year,  $\text{logit}(p_i) \sim \mathcal{N}(p_\mu, \sigma^2)$ .  $p_\mu$  changed between years (taking on values of -2, 0, 2, and  
419 4),  $\sigma^2 = 2$  in all years.
- 420   • The value of  $p$  only changes between species (and years), but the observation process occurs at the  
421 substratum ( $k$ ) level. Thus, the parameter is really  $p_{j,k,i}$ , but for a given  $i$ , all  $p_{j,k}$  are constant. I  
422 represent this probability as  $p_i$  with the understanding that this value is repeated over space.
- 423   •  $Y_{j,i}$  is the observed version of  $Z_{j,i}$ .
- 424   •  $Y_{j,i} \sim \text{Bern}(p_i \times Z_{j,i})$ .
- 425     – Note: Because  $p$  is actually subscripted to  $k$ , the  $Y$  are also actually subscripted to  $k$ . Maybe  
426 leaving these subscripts out is making things more confusing. I've only excluded them to  
427 emphasize how parameters are estimated.
- 428   • Our data about species presence/ absence correspond to  $Y_{j,i}$ . So it might be useful to think of the  
429 MSOM as estimating  $\hat{Y}_{j,i}$ , which is compared to the observed data  $Y_{j,i}^{obs}$ .
- 430

431

---

432 **Scatter Plot of  $p$**

433 The MSOM does a pretty good job of recovering  $p$ . This is true in other conditions I've simulated —  
434 conditions where it failed miserably to recover  $\psi$  (mainly because of small sample sizes). It's worth noting  
435 here that most of the points are either very close to 1 or very close to 0. In the next figures, the reason for  
436 this will become more apparent.

437

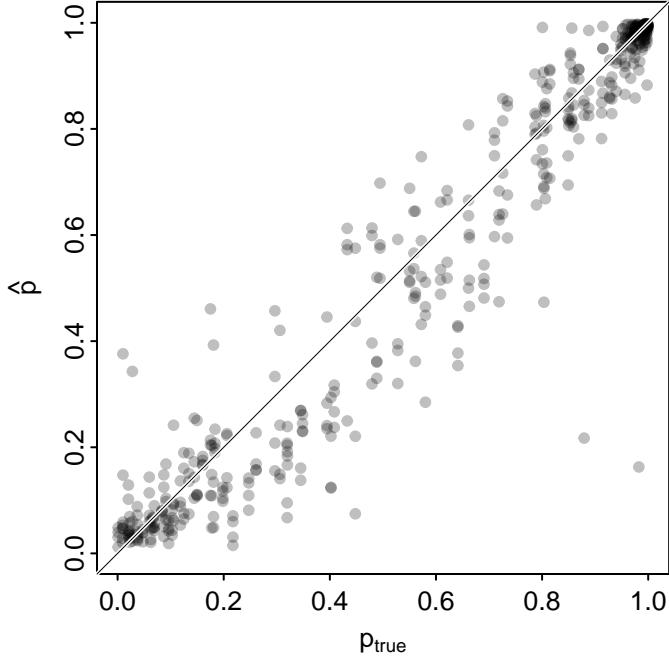


Figure 10: MSOM estimates (vertical axis) and true values of  $p_i$ , the species-specific ( $i$ ) detection probability. Each point is subscripted by species  $i$ , year  $t$ , and observation replicate  $r$ .

438

---

439 **Scatter Plots for Each  $p_{t,r}$**

- 440 The grouping in this figure is a bit odd because the panels with the highest detectability also don't show very  
 441 much range (because both axes are sitting at the high detectability!). However, in general, it looks like the  
 442 estimates of  $p$  are pretty good overall. There are some weird points, but overall, both the multi-panel and  
 443 the combined scatter plot indicate that we are recovering  $p$  fairly reliably.
- 444 In the previous section I pointed out that points group close to 0 or 1. That's just due to the different  
 445 values of  $p_\mu$ . I also wonder what a real value for  $p$  would look like. I have no idea which of these values are  
 446 reasonable.

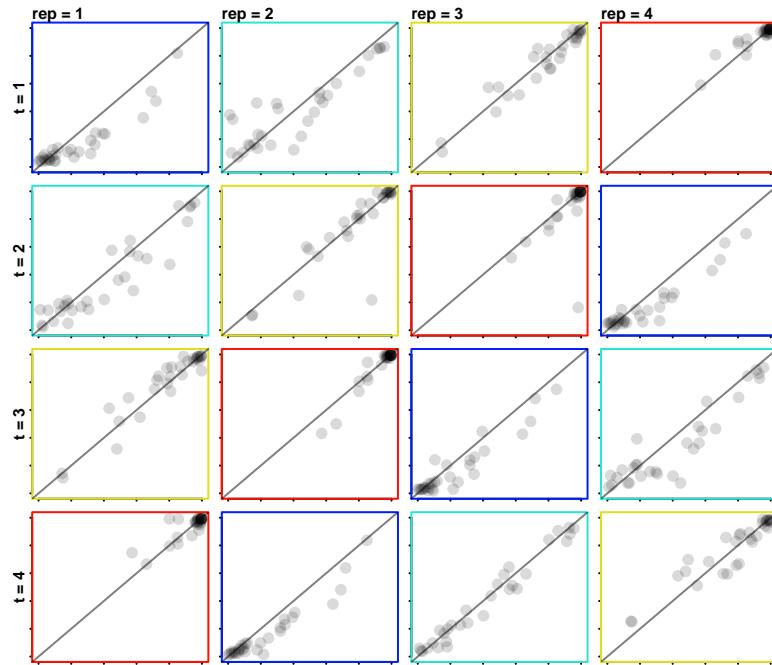


Figure 11: Probability of detection ( $p$ ), separated by year and replicate. True values ( $p^{true}$ ) are on the horizontal axis, MSOM estimates ( $\hat{p}$ ) are on the vertical axis. The color of the panel borders corresponds to the value of  $p_\mu$ , with warm colors being high and cool colors being low values of  $p_\mu$ . In a given panel, each point is a different species. The rows are different years, the columns are the simulated replicate observations of those years. The same set of true values for  $p$  is used for each replicate, just in a different order. If this simulation shows the same color panel box more than once per column, those panels share the same  $p_\mu$ , but they have independent realizations of  $p^{true}$ . See the section entitled [The Simulation](#) for further clarification.



## 448 Assessment with Mixed Effects Models

### 449 Describe Motivation for Mixed Effects Models

- 450 **Motivation:** MSOM skill might differ across dimensions, trying to figure out what patterns I should expect  
451 to pick out (spatial patterns in richness, temporal?) E.g., Is the correlation between MSOM and True the  
452 same comparing across sites as comparing across years? Species, reps, also.
- 453 **Motivation:** What factors influence MSOM skill in a given dimension? E.g., Skill in finding differences in  $\psi$   
454 across species may depend on  $p$ , the chance of being identified. If  $p$  changes among years, might also explain  
455 Read more about [specifying mixed effects models using lmer in R here](#)
- 456 This example is looking at  $\psi$ , probability of an individual species being present

### 457 Example LMER Analysis for $\psi$

```
# Just exploration/ starting point
library(car)
library(lme4)

# psi
# psi true
dat.psi.true <- reshape2:::melt.array(
  psi.true,
  varnames=c("site","spp","time","rep"),
  value.name="psi.true",
  as.is=T
)
# psi hat
dat.psi.hat <- reshape2:::melt.array(
  psi.hat,
  varnames=c("site","spp","time","rep"),
  value.name="psi.hat",
  as.is=T
)

# p
# p true
dat.p.true <- reshape2:::melt.array(
  aperm(array(p.true, dim=c(dim(p.true), dim(psi.true)[1])),c(4,1,2,3)),
  varnames=c("site","spp","time","rep"),
  value.name="p.true",
  as.is=T
)
# p hat
dat.p.hat <- reshape2:::melt.array(
  aperm(array(p.hat, dim=c(dim(p.hat), dim(psi.hat)[1])),c(4,1,2,3)),
  varnames=c("site","spp","time","rep"),
  value.name="p.hat",
  as.is=T
)
```

```

# n.hauls
n.hauls <- sapply(big.out.obs, function(x)attributes(x)$n.haul)
n.hauls.dim <- c(grid.w*grid.h, n.obs.reps, ns, grid.t)
dat.n.hauls <- reshape2:::melt.array(
  aperm(array(n.hauls, dim=n.hauls.dim), c(1,3,4,2)),
  varnames=c("site","spp","time","rep"),
  value.name="n.hauls",
  as.is=T
)

# grid.X
# same structure (dims) as n.hauls
temp <- values(grid.X)
temp.dim <- c(grid.w*grid.h, n.obs.reps, ns, grid.t)
dat.temp <- reshape2:::melt.array(
  aperm(array(temp, dim=temp.dim), c(1,3,4,2)),
  varnames=c("site","spp","time","rep"),
  value.name="temp",
  as.is=T
)

# tax chance
tax.chance <- simplify2array(
  lapply(big.out.obs, function(x)(attributes(x)$obs.params)$tax.chance)
)
tax.chance.dim <- c(grid.t, ns, n.obs.reps, grid.w*grid.h)
dat.tax.chance <- reshape2:::melt.array(
  aperm(array(tax.chance, dim=tax.chance.dim), c(4,2,1,3)),
  varnames=c("site","spp","time","rep"),
  value.name="tax.chance",
  as.is=T
)

mod.dat <- cbind(
  dat.psi.true,
  psi.hat=dat.psi.hat[, "psi.hat"],
  p.true=dat.p.true[, "p.true"],
  p.hat=dat.p.hat[, "p.hat"],
  n.hauls=dat.n.hauls[, "n.hauls"],
  tax.chance=dat.tax.chance[, "tax.chance"],
  temp=dat.temp[, "temp"]
)
mod.dat[, "psi.error"] <- mod.dat[, "psi.hat"]-mod.dat[, "psi.true"]
mod.dat[, "p.error"] <- mod.dat[, "p.hat"] - mod.dat[, "p.true"]

mod.dat$site <- as.factor(mod.dat$site)
mod.dat$spp <- as.factor(mod.dat$spp)
mod.dat$time <- as.factor(mod.dat$time)
mod.dat$rep <- as.factor(mod.dat$rep)

# =====

```

Table 3: Mixed effect models assessing sensitivity of  $\psi_\epsilon$  to simulation conditions

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	0.062*** (0.010)	0.060*** (0.011)	0.046*** (0.008)	0.062*** (0.010)	0.040*** (0.009)	0.040*** (0.009)
temp	0.001*** (0.000)			0.001 (0.002)		
n.hauls		0.001 (0.001)				
p.error			-0.618*** (0.012)		-0.867*** (0.136)	-0.868*** (0.136)
AIC	-6331.371	-6314.550	-8840.187	-8685.580	-11637.684	-11636.548
BIC	-6288.530	-6271.709	-8797.346	-8634.170	-11586.274	-11576.570
Log Likelihood	3170.686	3162.275	4425.094	4348.790	5824.842	5825.274
Num. obs.	38880	38880	38880	38880	38880	38880
Num. groups: site	81	81	81	81	81	81
Num. groups: spp	30	30	30	30	30	30
Variance: site.(Intercept)	0.000	0.000	0.000	0.000	0.000	0.000
Variance: spp.(Intercept)	0.003	0.003	0.002			0.002
Variance: Residual	0.049	0.049	0.046	0.046	0.043	0.043
Variance: spp.temp				0.000		
Variance: spp.1.(Intercept)				0.003	0.002	
Variance: spp.p.error					0.543	0.543

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

```
# = Do LMER Analysis =
# =====
mod1 <- lmer(psi.error~temp+(1|spp)+(1|site), data=mod.dat)
mod2 <- lmer(psi.error~n.hauls+(1|spp)+(1|site), data=mod.dat)
mod3 <- lmer(psi.error~p.error+(1|spp)+(1|site), data=mod.dat)
# mod4 <- lmer(psi.error~n.hauls-1+(1|spp)+(n.hauls-1|spp)+(1|site), data=mod.dat)
mod4 <- lmer(psi.error~temp+(1|spp)+(temp-1|spp)+(1|site), data=mod.dat)
mod5 <- lmer(psi.error~p.error+(1|spp)+(p.error-1|spp)+(1|site), data=mod.dat)
mod6 <- lmer(psi.error~p.error+(p.error|spp)+(1|site), data=mod.dat)

# Calculate covariance matrix (for spp)
mod6.varcor.spp <- attr(summary(mod6)$varcor$spp, "correlation")
mod6.varcor.spp <- format(mod6.varcor.spp, digits=2)
mod6.varcor.spp[!lower.tri(mod6.varcor.spp)] <- ""

mod.cap <- c(
  "Mixed effect models assessing sensitivity of $\psi_\epsilon$ to simulation conditions"
)
```

458 The goal with the mixed effects models was to understand what causes errors in  $\psi$ . I focused on  $\psi$  because it  
 459 has all the information needed to understand variability in richness, but it has more information than the  
 460 actual richness (richness is a community level statistic,  $\psi$  is species-specific). In these models, the response  
 461 variable is  $\hat{\psi} - \psi^{true}$ , which we'll call  $\psi_\epsilon$ . If we understand the source of variability in  $\psi_\epsilon$ , then we can  
 462 understand what leads to inaccuracies in our model.

463 When analyzing the trawl data, we will not know  $\psi_\epsilon$  – we can obtain model residuals, but these are distinct

464 from  $\psi_\epsilon$ , because calculating  $\psi_\epsilon$  requires knowing  $\psi^{true}$  which is a latent, unobserved variable. An empirical  
465 analysis would also lack some of the explanatory variables made available to us in the simulation. However, if  
466 we can explain variability in  $\psi_\epsilon$  using simulated information that will also be available to the trawl analysis,  
467 then we can build intuition about the sources of error in our estimate of species richness even when we don't  
468 know the true value. And that is the fundamental goal of this document.

469 I'll highlight some of the things I learned from this analysis:

470 1. Neither the environmental variable (**temp**) nor the number of subsites sampled per site (**n.hauls**) were  
471 strongly related to  $\psi_\epsilon$

- 472     i. But **temp** might be a better predictor if transformed into an absolute value  
473     ii. When the model does not contain a site-specific random intercept (not shown here), **n.hauls**  
474         accounts for more variability

475  
476 2. The **spp** random intercept explains much more variability than **site** equivalent

477  
478 3. **p.error** explains a ton of variability

- 479     i. inversely related to  $\psi_\epsilon$ ; intuition: if you have perfect detectability but you think it's terrible, you'll  
480         overestimate the true value  
481     ii. model 6 is a worse fit than model 5, meaning that adding covariance structure between **spp**-specific  
482         intercept and **spp**-specific **p.error** does not explain much variability

483 As stated above, a lot of variance is explained by the model term (**p.error|spp**), which allows the parameter  
484 associated with **p.error** and the intercept (both fixed effects) to vary randomly among species. The  
485 interpretation of these model terms is that

- 486     • Each species gets to draw its own intercept ("Variance: spp.1.(Intercept)" in the table) from a parent  
487         distribution of intercepts  
488     • A unit of error in the estimate of  $p$  has an influence on  $\psi_\epsilon$  that, similar to the intercept, varies among  
489         species ("Variance: spp.p.error" in the table).

490 Therefore, the effect that a bad estimate of  $p$  has on  $\psi_\epsilon$  is not the same among species. It is not clear what  
491 causes some species to be more sensitive to a poorly estimated  $p$  than others; one possibility is that  $p$  is  
492 poorly estimated for species that have not been observed much, and it is this lack of observation that is also  
493 responsible for generating uncertainty in  $\hat{\psi}$ . Regardless, a bad (good) estimate of  $p$  is a good predictor of a  
494 bad (good) estimate of  $\hat{\psi}$ .

495

496

---

497 **Conclusion**

498 **Discussion of Results**

499 Overall, the MSOM performed well. It is definitely data-hungry in the sense that it needs to observe 1's  
500 and 0's for each species in many places under different conditions. This may seem counter to the goal of the  
501 MSOM – to make efficient use of hard-won data. Would other richness methods be better? But remember,  
502 what we're getting is site-specific richness, and even species-specific presences and absences. Also, we don't  
503 have the “replicates” in the trawl data set needed for the other richness methods.

504 This seems obvious in retrospect, but Figure 1 (Boxplots) shows us that the sensitivity of our richness  
505 estimates to detectability is highly nonlinear. It'll be important to gauge where we think we are along that  
506 spectrum. Regardless, we were never too far off the real richness — e.g., when 29 species exist, we estimate  
507 24. On the upside, we're not *too* much worse off if we want to know  $\psi$  instead of just R. Of greater concern is  
508 the tendency to overestimate R or  $\psi$  when  $p$  is low. However, these low values for  $p$  might be *really* low, so  
509 perhaps that “problem” is not in a relevant region of parameter space.

510 **Next Steps**

511 I think the next important step is to decide if we want to analyze all of the years together. It'll require  
512 a more complicated model, but it'll give a lot more statistical power (and better estimates). I'm leaning  
513 towards doing this.

514 We also need to decide what regions from the trawl data set we'll want to use for this analysis. Right now the  
515 Alaskan regions are the front runners (just Eastern Bering Sea, or Aleutians and Gulf of Alaska as well?)  
516 simply because they have pdf file ranking each species on a scale 1-3 for each year according how likely it was  
517 to be identified. This isn't  $p$  per se (because other factors affect detectability as well), but this could give  
518 insight into a big contributor to temporal shifts in  $p$ .

519 We may also want to upgrade the realism of the simulation. I'm already set up to link years together if we  
520 decide to do a multi-year MSOM (MY MSOM?).

521 We should also think about what exactly our story/ selling points will be. I tried to word the introduction to  
522 this document as something that might sound like a paper Introduction. But this is tough when you don't  
523 actually know what you're introducing. I think the key to this paper may to emphasize an ecological finding  
524 the Alaskan data sets, but back up that finding with a careful evaluation of the method we're using. My goal  
525 would be to highlight how it's important but difficult to understand biodiversity dynamics, and then discover  
526 something about biodiversity dynamics while showing that the MSOM is a valid tool for doing so with trawl  
527 (and similar) data.

528 **Concluding Remarks**

529 I think we have a good group of people to discover some cool stuff in the trawl data set. What we have in  
530 this document is just the prelude – I think the real exciting stuf will come once we have estimates of richness  
531 for the trawl data that we trust. I'm looking foward to hearing your thoughts and working with all of you on  
532 this project!

534 **Report Generation Notes**

535 **R Session Information**

```
536 ## R version 3.2.0 (2015-04-16)
537 ## Platform: x86_64-apple-darwin13.4.0 (64-bit)
538 ## Running under: OS X 10.10.5 (Yosemite)
539 ##
540 ## locale:
541 ## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
542 ##
543 ## attached base packages:
544 ## [1] parallel   grid      stats     graphics   grDevices utils     datasets
545 ## [8] methods    base
546 ##
547 ## other attached packages:
548 ## [1] lme4_1.1-9      Matrix_1.2-0      car_2.0-26      rbLib_0.0.2
549 ## [5] doParallel_1.0.8 iterators_1.0.7 foreach_1.4.2   R2jags_0.5-6
550 ## [9] rjags_3-15      coda_0.17-1     igraph_0.7.1    fields_8.2-1
551 ## [13] maps_2.3-9      spam_1.0-1      data.table_1.9.4 raster_2.3-40
552 ## [17] sp_1.1-0       texreg_1.35    stargazer_5.2   kfigr_1.2
553 ## [21] xtable_1.7-4    rmarkdown_0.7   knitr_1.11     taxize_0.5.2
554 ##
555 ## loaded via a namespace (and not attached):
556 ## [1] reshape2_1.4.1   splines_3.2.0    lattice_0.20-31 htmltools_0.2.6
557 ## [5] yaml_2.1.13     mgcv_1.8-6      chron_2.3-45    XML_3.98-1.3
558 ## [9] nloptr_1.0.4    plyr_1.8.2     stringr_1.0.0   codetools_0.2-11
559 ## [13] evaluate_0.7.2 Taxonstand_1.7 SparseM_1.7    permute_0.8-4
560 ## [17] quantreg_5.11  pbkrtest_0.4-2 numbers_0.6-1   highr_0.5
561 ## [21] Rcpp_0.11.6    formatR_1.2    vegan_2.3-0    jsonlite_0.9.16
562 ## [25] abind_1.4-3    digest_0.6.8   stringi_0.5-5  ssh.utils_1.0
563 ## [29] tools_3.2.0    bitops_1.0-6   magrittr_1.5   RCurl_1.95-4.6
564 ## [33] bold_0.2.6    cluster_2.0.1  ape_3.3        MASS_7.3-40
565 ## [37] minqa_1.2.4   assertthat_0.1 reshape_0.8.5   httr_0.6.1
566 ## [41] boot_1.3-16    nnet_7.3-9     R2WinBUGS_2.1-21 nlme_3.1-120
```

567 **Date Document Last Compiled**

```
568 ## Last compiled on: 2015-08-30
```

569

---