

On extrapolating past the range of observed data when making statistical predictions in
ecology

PAUL B. CONN^{1,2} AND DEVIN S. JOHNSON¹, PETER L. BOVENG¹

¹*National Marine Mammal Laboratory, NOAA, National Marine Fisheries Service, Alaska
Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115 USA*

¹ *Abstract.* We'll do this later

² *Key words:* *Abundance, Extrapolation, General additive models, Generalized linear
models, Independent Variable Hull, Leverage, Prediction variance, Spatial regression,
Species distribution model*

INTRODUCTION

⁵ In ecology and conservation, a common goal is to make predictions about an
⁶ unsampled random variable given a limited sample from the target population. For
⁷ instance, given a model (\mathcal{M}), estimated parameters ($\hat{\theta}$), and a covariate vector \mathbf{x}_s , we often
⁸ desire to predict a new observation y_s at s . For instance, we might use a generalized linear
⁹ model (McCullagh and Nelder, 1989) or one of its extensions to predict species density or
¹⁰ occurrence in a new location, or to predict the future trend of a population.

¹¹ Early in their training, ecologists and statisticians are warned against extrapolating
¹² statistical relationships past the range of observed data. This caution is easily interpreted
¹³ in the context of single-variable linear regression analysis; one should be cautious in using
¹⁴ the fitted relationship to make predictions at some new point y_s whenever $x_s < \min(\mathbf{x})$ or
¹⁵ $x_s > \max(\mathbf{x})$. But what about more complicated situations where there are multiple
¹⁶ explanatory variables, or when one uses a spatial regression model to account for the

²Email: paul.conn@noaa.gov

17 residual spatial autocorrelation that is inevitably present in patchy ecological data
18 (Lichstein et al., 2002)? How reliable are spatially- or temporally-explicit predictions in
19 sophisticated models for animal abundance and occurrence?

20 Statisticians have long struggled with the conditions under which fitted regression
21 models are capable of making robust predictions at new combinations of explanatory
22 variables. The issue is sometimes considered more of a philosophical problem than a
23 statistical one, and has even been likened to soothsaying (Ehrenberg and Bound, 1993). To
24 our mind, the reliability of predictions from statistical models is likely a function of several
25 factors, including (i) the intensity of sampling, (ii) spatial or temporal proximity of the
26 prediction location to locations where there are data, (iii) variability of the ecological
27 process, and (iv) the similarity of explanatory covariates in prediction locations when
28 compared to the ensemble of covariates for observed data locations.

29 Our aim in this paper is to investigate one possibility for defining extrapolation in the
30 generalized linear model and its extensions, including generalized additive models (GAMs;
31 Hastie and Tibshirani, 1999; Wood, 2006) and spatial, temporal, or spatio-temporal
32 regression models (STRMs). In particular, we exploit some of the same ideas used in
33 multiple linear regression regarding leverage and outliers (Cook, 1979) to operationally
34 define “extrapolation” as making predictions that occur outside of a generalized
35 independent variable hull (gIVH) of observed data points. Application of the gIVH and
36 related criterion (e.g. prediction variance) can provide intuition regarding the reliability of
37 predictions in unobserved locations, and can aid in model construction and survey design.

38 We illustrate use of the gIVH on simulated count data, on a species distribution model
39 (SDM) for ribbon seals (*Histriophoca fasciata*) in the eastern Bering Sea, and on a
40 population trend model for Steller Sea Lions (*Phoca fasciata*).

41 THE GENERALIZED INDEPENDENT VARIABLE HULL (GIVH)

42 Extrapolation is often distinguished from interpolation. In a prediction context, we might
43 define (admittedly quite imprecisely) that extrapolation consists of making predictions that
44 are “outside the range of observed data” while interpolation consists of making predictions
45 “inside the range of observed data.” But what exactly do we mean by “outside the range of
46 observed data”? Predictions outside the range of observed covariates? Predictions for
47 locations that are so far from places where data are gathered that we are skeptical that the
48 estimated statistical relationship still holds? To help guide our choice of an operational
49 definition, we turn to early work on outlier detection in simple linear regression analysis.

50 In the context of outlier detection, Cook (1979) defined an independent variable hull
51 (IVH) as the smallest convex set containing all design points of a full-rank linear regression
52 model. Linear regression models are often written in matrix form, i.e.

53
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

54 where \mathbf{Y} give observed data, \mathbf{X} is a so-called design matrix that includes explanatory
55 variables (see e.g. Draper and Smith, 1966), and $\boldsymbol{\epsilon}$ represent normally distributed residuals
56 (here and throughout the paper, bold symbols will be used to denote vectors and
57 matrices). Under this formulation, the IVH is defined relative to the hat matrix,
58 $\mathbf{V}_{LR} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (where the subscript “LR” denotes linear regression). Letting v
59 denote the maximum diagonal element of \mathbf{V}_{LR} , one can examine whether a new design
60 point, \mathbf{x}_0 is within the IVH. In particular, \mathbf{x}_0 is within the IVH whenever

61
$$\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \leq v. \quad (1)$$

62 Cook (1979) used this concept to identify influential observations and possible outliers,
63 arguing that design points near the edge of the IVH are deserving of special attention.
64 Similarly, points outside the IVH should be interpreted with caution.

65 We simulated two sets of design data to help illustrate application of the IVH (Fig. 2).

66 In simple linear regression with one predictor variable, predictions on a hypothetical
67 response variable obtained at covariate values below the lowest observed value or above the
68 highest observed value are primarily outside of the IVH. We suspect this result conforms to
69 most ecologists intuition about what constitutes “extrapolating past the observed data.”

70 However, fitting a quadratic model exhibits more nuance; if there is a large gap between
71 design points, it is entirely possible that intermediate covariate values will also be outside
72 of the IVH and thus more likely to result in problematic predictions. Fitting a model with
73 two covariates and both linear and quadratic effects, the shape of the IVH is somewhat
74 more irregular, and even includes a hole in the middle of the surface when interactions are
75 modeled (Fig. 2). These simple examples highlight the sometimes counterintuitive nature
76 of predictive inference, a problem that can only become worse as models with more
77 dimensions are contemplated (including those with temporal or spatial structure).

78 Fortunately, the ideas behind the IVH provide a potential way forward.

79 Cook’s (1979) formulation for the IVH is particular to linear regression analysis, which
80 assumes iid Gaussian error. Thus, it is not directly applicable to generalized models, such
81 as those including alternative error distributions (e.g., Poisson, binomial) or spatial random
82 effects. Further, the hat matrix is not necessarily well defined for models with more general
83 spatial structure. However, since the hat matrix is proportional to prediction variance,
84 Cook (1979) notes that design points with maximum prediction variance will be located on
85 the boundary of the IVH. We therefore define a generalized independent variable hull

86 (gIVH) as the set of all potential design points \mathcal{S}_0 (note that \mathcal{S}_0 can include both observed
 87 and unobserved design points) such that

$$88 \quad \text{var}(\lambda_s) \leq \max[\text{var}(\boldsymbol{\lambda}_{\mathcal{S}})], \quad (2)$$

89 where $s \in \mathcal{S}_0$, λ_s corresponds to the mean prediction at s , \mathcal{S} denotes the set of observed
 90 design points, and $\boldsymbol{\lambda}_{\mathcal{S}}$ denotes predictions at \mathcal{S} .

91 Generalizations of the linear model are often written in the form

$$92 \quad Y_s \sim f_Y(g(\mu_s)), \quad (3)$$

93 where f_Y denotes a probability density or mass function (e.g. Bernoulli, Poisson), g gives
 94 an inverse link function, and μ_s is a linear predictor. For many such generalizations, it is
 95 possible to specify the μ_s as

$$96 \quad \boldsymbol{\mu} = \mathbf{X}_{aug}\boldsymbol{\beta}_{aug} + \boldsymbol{\epsilon}, \quad (4)$$

97 where the $\boldsymbol{\epsilon}$ represent Gaussian error, \mathbf{X}_{aug} denotes an augmented design matrix, and $\boldsymbol{\beta}_{aug}$
 98 denote an augmented vector of parameters. For instance, in a spatial model, $\boldsymbol{\beta}_{aug}$ might
 99 include both fixed effect parameters and spatial random effects in a reduced dimension
 100 subspace (see Appendix A for examples of how numerous types of models can be written in
 101 this form).

102 When models are specified as in Eq. 4, we can write prediction variance generically as

$$103 \quad \text{var}(\boldsymbol{\mu}_x) = \mathbf{x}\text{var}(\hat{\boldsymbol{\beta}}_{aug})\mathbf{x}', \quad (5)$$

104 where it is understood that the exact form of \mathbf{x} and $\text{var}(\hat{\boldsymbol{\beta}}_{\text{aug}})$ depends on the model chosen
 105 (i.e., GLM, GAM, or STRM; Appendix A). This expression for prediction variance is on
 106 the linear predictor scale; if a non-identity link function is used, we can use the delta
 107 method (Dorfman, 1938; Ver Hoef, 2012) to approximate prediction variance on the real
 108 scale (i.e. the scale in which data are measured) as

$$109 \quad \text{var}(\boldsymbol{\lambda}_x) = \text{var}(g(\boldsymbol{\mu}_x)) \quad (6)$$

$$110 \quad \approx \Delta \text{var}(\boldsymbol{\mu}_x) \Delta', \quad (7)$$

111 where Δ is a matrix of partial derivatives $\partial g(\mu_r)/\partial \mu_c$ (r and c denoting rows and columns
 112 of Δ , respectively). Under common univariate link functions (e.g. log, logit, probit), Δ has
 113 a diagonal form, while for multivariate links (e.g. multinomial logit) Δ will be dense.

114 Unlike Cook's original formulation, the gIVH will often involve unknown parameters.

115 Further, the exact form of $\text{var}(\hat{\boldsymbol{\beta}}_{\text{aug}})$ depends on the underlying model structure and
 116 estimation procedure. For most of the following treatment, we shall assume that data have
 117 already been collected (see Discussion for comments on the potential use of the gIVH in
 118 survey planning), and that informative probability distributions can be specified for these
 119 parameters. Specifically, we impose a probability distribution $[\boldsymbol{\theta}]$ for unknown parameters,
 120 and compute prediction variance as

$$121 \quad \text{var}(\boldsymbol{\lambda}_x | \mathbf{Y}) = \int_{\boldsymbol{\theta}} \text{var}(\boldsymbol{\lambda}_x | \mathbf{Y}, \boldsymbol{\theta}) [\boldsymbol{\theta}] d\boldsymbol{\theta}. \quad (8)$$

122 Our approach in this paper is to conduct a Bayesian analysis and substitute the posterior
 123 distribution $[\boldsymbol{\theta} | \mathbf{Y}]$ for $[\boldsymbol{\theta}]$ in Eq. 8. However, one could also consider substituting

124 parametric approximations based on output from standard likelihood-based analyses.

125 We propose to use the gIVH in much the same manner as Cook (1979). In particular,
126 we use the gIVH to differentiate whether spatial predictions are interpolations (predictive
127 design points lying inside the gIVH) or extrapolations (predictive design points lying
128 outside the gIVH). The gIVH, together with posterior prediction variance, seem ideally
129 situated to diagnosing potential extrapolation issues. Further, one can compare prediction
130 variance in places one has data to places where predictions are desired to gauge the relative
131 amount of information that predictions will be based on.

132 *Computing*

133 We developed a package **SpatPred** in the R statistical programming environment (R
134 Development Core Team, 2012) to simulate data and conduct all analyses. The seal and
135 sea lion datasets are included as part of this package, and are available at
136 <https://github.com/pconn/SpatPred>. The R package has also been archived via figshare
137 [WE PLAN TO DO THIS WHEN THE MANUSCRIPT IS ACCEPTED FOR
138 PUBLICATION].

139 EXAMPLES

140 We used simulation to investigate the usefulness of the gIVH for limiting the scope of
141 inference when fitting models to count data that have already been collected. We also
142 applied the gIVH to examining the robustness of a species distribution model (SDM) of
143 ribbon seals in the eastern Bering Sea, and to time series projections of Steller sea lion
144 population trends.

145

Simulation study

146 We conducted a simulation study to investigate whether the gIVH (and accompanying
147 prediction variance) was useful in diagnosing prediction biases when analyzing animal
148 count data. For each of 100 simulations, we generated animal abundance over a 30×30
149 grid assuming that animal density was homogeneous in each grid cell. Animal abundance
150 was generated as a function of three hypothetical spatially autocorrelated habitat
151 covariates (Appendix B). For each simulated landscape, we conducted virtual surveys of
152 $n = 45$ survey units using two different designs: (1) a spatially balanced sample using a
153 random tessellation design (Stevens Jr. and Olsen, 2004), and (2) a convenience sample
154 where the probability of sampling was greater for cells closer to a “base of operations”
155 located in the middle of the survey grid. The former approach preserves randomness while
156 seeking a degree of regularity when distributing sampling locations across the landscape,
157 while the latter may be easier to implement.

158 We configured virtual sampling quadrats such that they encompassed 10% of the area
159 of each selected grid cell. For ease of presentation and analysis, we assumed detection
160 probability was 100% in each quadrat. Once animal counts were simulated, three different
161 estimation models were fitted to the data: a GLM, a GAM, and an STRM (Appendix A).
162 The GLM and STRM expressed log-density as a function of linear and quadratic covariate
163 effects (as well as first order interactions), while the GAM used a kernel smoother with 4
164 knots for each covariate (Appendix A). Each model was provided with two of the three
165 covariates used to generate the data.

166 For each simulated data set and model structure, we calculated the posterior predictive
167 variance and resulting gIVH as in Eq. 2. We then calculated posterior predictions of

168 animal abundance within and outside of each gIVH in order to gauge bias as a function of
169 this restriction. Specifically, the performance of gIVH may help decide its utility in limiting
170 the scope of inference once data have been collected and analyzed, and perhaps point out
171 areas worthy of additional sampling. A fuller, technical description of the simulation study
172 design is provided in Appendix B; for illustration of a single simulation replicate see Fig. 3.

173 Posterior predictions from simulations indicated that the distribution for proportional
174 error in total abundance was right skewed when statistical inference was made with regard
175 to the entire survey area (Fig. 4). This was particularly true for GLM and STRM models,
176 and was exacerbated when convenience sampling was employed. The magnitude of mean
177 absolute bias was reduced when inference was constrained to the gIVH for all 6
178 configurations of survey design and estimation model. Interestingly, the convolution kernel
179 GAMs we employed often underestimated total abundance, perhaps because of dampening
180 “edge effects” at extreme ranges of observed covariates, or possibly because interactions
181 among covariates were not modeled. For GLMs and STRMs, positive proportional bias was
182 the rule, and was of concerning magnitude (e.g. ≈ 0.3 ; Fig. 4) for GLMs and STRMs when
183 convenience sampling was employed and inference was not restricted to the gIVH.

184 *Ribbon seal SDM*

185 As part of an international effort, researchers with the U.S. National Marine Fisheries
186 Service conducted aerial surveys over the eastern Bering Sea in 2012 and 2013. Agency
187 scientists used infrared video to detect seals that were on ice, and collected simultaneous
188 digital photographs to provide information on species identity. Here, we use spatially
189 referenced count data from photographed ribbon seals, *Phoca fasciata* (Fig. 1) on a subset
190 of 10 flights flown over the Bering Sea from April 20-27, 2012. We limited flights to a one

¹⁹¹ week period because sea ice melts rapidly in the Bering Sea in the spring, and modeling
¹⁹² counts over a longer duration would likely require addressing how sea ice and seal
¹⁹³ abundance changes over both time and space (see Conn et al., In Press). However, limiting
¹⁹⁴ analysis to a one week period makes the assumption of static sea ice and seal densities
¹⁹⁵ tenable (Conn et al., 2014).

¹⁹⁶ Our objective with this dataset will be to model seal counts on transects through 25km
¹⁹⁷ by 25km grid cells as a function of habitat covariates and possible spatial autocorrelation.
¹⁹⁸ Estimates of apparent abundance can then be obtained by summing predictions across grid
¹⁹⁹ cells. Figure 5 shows the transects flown and the number of ribbon seals encountered in
²⁰⁰ each cell, and Figure 6 show explanatory covariates gathered to help predict ribbon seal
²⁰¹ abundance. These data are described in fuller detail by Conn et al. (2014), who extend the
²⁰² modeling framework of STRMs to account for incomplete detection and species
²⁰³ misidentification errors. Since our focus in this paper is on illustrating spatial modeling
²⁰⁴ concepts, we devote our efforts to the comparably easier problem of estimating apparent
²⁰⁵ abundance (i.e., uncorrected for vagaries of the detection process).

²⁰⁶ Inspection of ribbon seal data (Fig. 5) immediately reveals a potential issue with
²⁰⁷ spatial prediction: abundance of ribbon seals appears to be maximized in the southern
²⁰⁸ and/or southeast quadrant of the surveyed area. Predicting abundance in areas further
²⁰⁹ south and east may thus prove problematic, as the values of several explanatory covariates
²¹⁰ (Fig. 6) are also maximized in these regions. To illustrate, let Y_i denote the ribbon seal
²¹¹ count (Y_i) obtained in sampled grid cell i . Suppose that counts arise according to a

212 log-Gaussian Cox process, such that

$$Y_i \sim \text{Poisson}(\lambda_i) \text{ and} \quad (9)$$

$$\log(\lambda_i) = \log(P_i) + \log(A_i) + \theta_i + \epsilon_i,$$

213 where P_i gives the proportion of area photographed in grid cell i , A_i gives the proportion of
214 suitable habitat in cell i (ribbon seals do not often utilize land), θ_i is a linear predictor, and
215 ϵ_i is normally distributed *iid* error. By formulating θ_i differently, we can arrive at
216 representations characteristic of GLMs, GAMs, and STRMs (see Appendix S1).

217 We start by fitting a hierarchical GLM, GAM, and STRM to the full suite of predictor
218 covariates (Fig. 6). For the GLM and STRM, we fit a model with linear effects of all
219 predictor variables, and with an additional quadratic term for ice concentration (seal
220 density is often maximized at an intermediate value of ice concentration; see Ver Hoef
221 et al., 2013; Conn et al., 2014). For the STRM, we imposed a restricted spatial regression
222 (RSR) formulation for spatially autocorrelated random effects, where dimension reduction
223 was accomplished by only selecting eigenvectors of the spectral decomposition associated
224 with eigenvalues that were greater than 0.5 (see Appendix S1 for additional information on
225 model structure). For the GAM, we employed smooth terms with 8 knots for each
226 covariate. This included 4 knots at the minimum, 1/3 and 2/3 quantiles, and the
227 maximum observed covariate values, in addition to 4 knots above and below the range of
228 observed data to reduce the magnitude of biased endpoint behavior.

229 As with simulated data examples, we adopted a Bayesian perspective and conducted
230 estimation using MCMC (see Appendix S1 for algorithm details) with 60,000 iterations
231 where the first 10,000 iterations were discarded as a burn-in. We generated posterior

232 predictions of ribbon seal abundance across the landscape as

$$N_i \sim \text{Poisson}(A_i \lambda_i), \quad (10)$$

233 and calculated the gIVH as in Eqn. 2, with delta method modifications as specified in
234 Eqns. 5-8. Initial spatial predictions using the three models (GLM, GAM, STRM) all
235 produced extremely high, unbelievable predictions along the southern boundary of the
236 study area 7. Predictions in this region were also largely out of the gIVH, indicating the
237 potential utility for the gIVH in revealing problematic extrapolations.

238 We considered several possible alternatives for trying to obtain more robust abundance
239 estimates before settling on a preferred alternative. First, one could refine the study area to
240 eliminate predictions outside of the gIVH (as in the simulation study). However, this is not
241 ideal in that one does not get an abundance estimate for the whole study area, and it may
242 be difficult to compare abundance from one year to the next using this approach. Second,
243 one could consider a simpler estimation model, such as one post-stratifying by ecoregion as
244 was used in a previous analysis of ribbon seal data to try to avoid this problem (Conn
245 et al., 2014). Finally, one could build in a priori knowledge of habitat preferences into the
246 model structure. We adopted the latter solution, incorporating “pseudo-absences” (i.e.,
247 zero counts where sampling was not conducted) in locations where it would have been
248 (nearly) impossible to detect seals. Specifically, we inserted pseudo-absences in cells where
249 ice concentrations were < 0.1%. This solution seemed the most logical, as many of the
250 large, anomalous predictions were over open water along the southern edge of the study
251 area, where we would have obtained zero counts had they been surveyed.

252 Fitting our three models to the augmented dataset with pseudo-absences, most

253 predictions occurred within the gIVH (Fig. 7). The largest difference among the three
254 models was in the southwest corner of the study area, coinciding with the few cells that
255 were outside the gIVH in the three models. Posterior summaries of abundance across the
256 study area were very comparable (albeit with a longer right hand tail for the STRM), with
257 5%, 50%, and 95% posterior prediction quantiles as follows: GLM (48686,64836,93927);
258 GAM(46120,61214,90873); STRM(41039,63717,194095). We note that these estimates are
259 for example illustration only, as they are uncorrected for imperfect detection (e.g.,
260 incomplete detection of thermal cameras, animals that were unavailable for sampling
261 because they were in the water, species misidentification; see Conn et al., 2014).

262 *Steller sea lion trends*

263 DISCUSSION

264 We have demonstrated the capacity for certain classes of statistical models to produce
265 biased predictions of animal abundance when extrapolating past the range of observed
266 data. In simulations, commonly used models (GLMs, STRMs) exhibited positive bias when
267 predictions were required for the entire study area, particularly when convenience sampling
268 was employed; in the ribbon seal example, naive extrapolation of fitted statistical
269 relationships produced high positive bias along the southern boundary of the study area;
270 and in the Steller sea lion example, projected abundance trends were increasingly
271 implausible the farther the predictions were from the realm of observed data. However, the
272 gIVH appeared useful in diagnosing places where extrapolations from the fitted statistical
273 model were problematic. For instance, in the simulation study, limiting the scope of
274 inference to locations falling within the gIVH reduced or eliminated bias. For ribbon seal

275 relative abundance, it was useful for confirming that the naive models needed to be
276 reformulated, and that reformulated models (with pseudo-absence data) produced
277 predictions in unsurveyed areas that were largely within the realm of observed data.

278 Use of the gIVH in seal and sea lion applications to identify extrapolation issues was
279 not particularly profound, as experts could easily use ecological knowledge of these
280 populations to identify biologically implausible predictions and come to the conclusion that
281 models either needed to be reformulated or inference limited to a smaller set of design
282 points. Likewise, one could argue that the high positive bias in the simulation study could
283 likely have been reduced had an expert meticulously examined diagnostics from each
284 simulation replicate and reconfigured estimation models to better match the range of
285 covariates observed. However, such determinations are likely to be quite subjective, and
286 may prove insufficient when there are large number of regression coefficients and
287 interaction terms. Even relatively simple regression models may exhibit non-intuitive
288 patterns (e.g. Fig. 2). Further, relying on expert opinion alone in successive rounds of
289 model formulation and fitting may lead to investigators choosing models based on how
290 much they like the results, which is clearly not ideal scientific practice.

291 Our intent in writing this paper is to raise awareness of potential problems with
292 extrapolation bias in statistical models, and to provide an additional tool (the gIVH) to
293 help diagnose its presence. Other approaches for selecting models to reduce overfitting,
294 such as cross validation (e.g. Picard and Cook, 1984), are also useful for this purpose, but
295 may not entirely eliminate the problem (particularly for sparse datasets). In fields such as
296 species distribution modeling and model-based abundance estimation, the goal for analysts
297 is often to build predictive maps of species abundance and/or occurrence using a limited
298 number of sample locations. In such applications, the ultimate aim of analysts should be to

299 build models that have low bias and maximal precision. However, traditional approaches to
300 quantifying bias (e.g., goodness-of-fit statistics) only work with observed data points.

301 When inference is extended to unsampled locations, the gIVH appears to be a useful
302 diagnostic for whether bias for predictions in unsampled locations can be expected.

303 The analyses in this paper focused on abundance estimation, which is necessarily
304 non-negative. As such, counts are usually analyzed with a log link function, and there is a
305 much greater potential for positive bias than negative. However, one could also apply the
306 gIVH when predicting species occurrence from presence/absence data. In this case,
307 common link functions (e.g. probit or logit) are symmetric, and potential for positive and
308 negative bias in predictive maps seem equally likely. Future research should be directed to
309 examine conditions under which the gIVH is a useful diagnostic in such applications.

310 One area that gIVH ideas may also prove useful is in formulating survey designs. The
311 topic of optimal (or near-optimal) spatial design has received considerable attention in the
312 statistical literature, often in the context of designing environmental monitoring systems
313 (Diggle and Lophaven, 2006; Xia et al., 2006; Müller, 2007). Optimal designs can differ
314 depend on whether one wishes to minimize a criterion such as integrated prediction
315 variance, or a summary of the information matrix (i.e. if one wishes to optimize regression
316 parameter performance). Optimal designs can also be sensitive to the structure of the
317 estimation model that is used, so that tailoring a survey design to a particular model can
318 be somewhat dangerous if there is uncertainty about the ultimate “best” structure for the
319 model used to relate animal abundance and occurrence to landscape or habitat covariates.

320 Nevertheless, one could still think about augmenting a spatially balanced sampling design
321 with a number of locations which are known or thought to have high leverage as a function
322 of available covariates. We are excited about this prospect, and it is a subject of current

323 research.

324 With regards to ribbon seal abundance, we once again caution that the estimates
325 presented here are for illustration only, as numerous factors that influence the detectability
326 of seals was not included in count models used in this paper (but see Conn et al., 2014).

327 Our approach here was to examine extrapolation and prediction error using relatively
328 simple models, with the understanding that such effects are also likely to occur in complex
329 models with more realistic observation components. Indeed, our experience is that the
330 magnitude of extrapolation bias can increase when additional data on imperfect detection
331 or species misclassification are included.

332 LITERATURE CITED

333 ???? .

334 Conn, P. B., D. S. Johnson, J. M. Ver Hoef, M. B. Hooten, J. M. London, and P. L.
335 Boveng. In Press. Using spatio-temporal statistical models to estimate animal
336 abundance and infer ecological dynamics from survey counts. Ecological Monographs .

337 Conn, P. B., J. M. Ver Hoef, B. T. McClintock, E. E. Moreland, J. M. London, M. F.
338 Cameron, S. P. Dahle, and P. L. Boveng. 2014. Estimating multi-species abundance
339 using automated detection systems: ice-associated seals in the eastern Bering Sea.
340 Methods in Ecology and Evolution 5:1280–1293.

341 Cook, R. D. 1979. Influential observations in linear regression. Journal of the American
342 Statistical Association 74:169–174.

- 343 Diggle, P., and S. Lophaven. 2006. Geostatistical Design. Scandinavian Journal of
344 Statistics **33**:53–64.
- 345 Dorfman, R. 1938. A note on the delta-method for finding variance formulae. The
346 Biometric Bulletin **1**:129–137.
- 347 Draper, N. R., and H. Smith. 1966. Applied Regression Analysis. John Wiley & Sons, New
348 York.
- 349 Ehrenberg, A., and J. Bound. 1993. Predicability and prediction. Journal of the Royal
350 Statistical Society A **156**:167–206.
- 351 Hastie, T. J., and R. J. Tibshirani. 1999. Generalized Additive Models. Chapman &
352 Hall/CRC, Boca Raton, Florida.
- 353 Lichstein, J., T. Simons, S. Shriner, and K. E. Franzreb. 2002. Spatial autocorrelation and
354 autoregressive models in ecology. Ecological Monographs **72**:445–463.
- 355 McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models. Chapman and Hall,
356 New York.
- 357 Müller, W. G. 2007. Collecting Spatial Data; Optimum Design of Experiments for Random
358 Fields, Third Edition. Springer, Berlin.
- 359 Picard, R. R., and R. D. Cook. 1984. Cross-validation of regression models. Journal of the
360 American Statistical Association **79**:575–583.
- 361 R Development Core Team, 2012. R: A Language and Environment for Statistical
362 Computing. R Foundation for Statistical Computing, Vienna, Austria. URL
363 <http://www.R-project.org>.

- ³⁶⁴ Stevens Jr., D., and A. Olsen. 2004. Spatially balanced sampling of natural resources.
- ³⁶⁵ Journal of the American Statistical Association **99**:262–278.
- ³⁶⁶ Ver Hoef, J. M., 2012. Aerial Survey Data. Page (online) doi:
³⁶⁷ 10.1002/9780470057339.vnn144 *in* A. H. El-Shaarawi and W. W. Piegorsch, editors.
- ³⁶⁸ Encyclopedia of Environmetrics, 2nd Edition. Wiley, Chichester, U.K.
- ³⁶⁹ Ver Hoef, J. M., M. F. Cameron, P. L. Boveng, J. M. London, and E. E. Moreland. 2013. A
³⁷⁰ hierarchical model for abundance of three ice-associated seal species in the eastern Bering
³⁷¹ Sea. Statistical Methodology page <http://dx.doi.org/10.1016/j.stamet.2013.03.001>.
- ³⁷² Wood, S. N. 2006. Generalized additive models. Chapman & Hall/CRC, Boca Raton,
³⁷³ Florida.
- ³⁷⁴ Xia, G., M. L. Miranda, and A. E. Gelfand. 2006. Approximately optimal spatial design
³⁷⁵ approaches for environmental health data. Environmetrics **17**:363–385. URL
³⁷⁶ <http://dx.doi.org/10.1002/env.775>.

377 FIGURE 1. A ribbon seal, *Phoca fasciata*; the focus of spatial modeling efforts in this
378 paper.

379 FIGURE 2. Example IVHs constructed from simulated data. In panels A and B, the
380 investigator plans to model a hypothetical (unmeasured) response variable using a linear
381 regression model as a function of a single covariate, x , obtained at a number of design
382 points (denoted with an “x”). Using x as a simple linear effect (A), only predictions less
383 than the minimum observed value of x or greater than the maximum value of x are outside
384 the IVH (shaded area), as scaled prediction variance in these areas (solid line) are greater
385 than the maximum scaled prediction variance for observed data (dashed line). Using both
386 linear and quadratic effects of x , some intermediate points are also outside the IVH;
387 predictions at these points should also be viewed with some caution. Panels C & D show a
388 more complicated IVH when the investigator wishes to relate an unmeasured response
389 variable to linear and quadratic effects of two covariates, x and y , either without
390 interactions (C) or with interactions (D). Any potential predictions in the shaded area are
391 outside of the IVH.

392 FIGURE 3. Depiction of a single simulation scenario. Panels (A-C) give simulated
393 covariate values, panel D gives true animal abundance, (E) gives estimated abundance from
394 a GLM run on count data from a spatially balanced survey design, and (F) gives
395 abundance from a GLM applied to count data from a convenience survey. In (E-F),
396 predictions outside the gIVH are represented by black boxes, and sampling locations are
397 represented with an x. For the convenience sample, the median posterior abundance
398 prediction for the entire survey area is 57% greater than true abundance when inference is
399 made to the whole study area. When inference is limited to the gIVH, median posterior
400 abundance was just 16% greater than true abundance.

401 FIGURE 4. Boxplots summarizing distribution of proportional error in the posterior
402 predictive median of abundance for the simulation study as a function of estimation model
403 (x-axis), survey design (columns) and whether or not inference was restricted to the gIVH
404 (rows). The lower and upper limits of each box correspond to first and third quartiles,
405 while whiskers extend to the lowest and highest observed bias within 1.5 interquartile
406 range units from the box. Outliers outside of this range are denoted with points.
407 Horizontal lines within boxes denote median bias. The two numbers located below each
408 boxplot indicate mean bias (upper number) and the number of additional outliers for which
409 proportional bias was greater than 2.0 (lower number).

410 FIGURE 5. Aerial survey tracks over the Bering Sea, April 22-29, 2012 (blue lines)
411 overlayed on a tessellated study area consisting of 25km by 25km grid cells (gray lines).
412 Dark gray indicates land, while the brown line indicates a 1000m depth contour, and the
413 orange line shows the U.S Exclusive Economic Zone (EEZ) boundary (which was used to
414 construct the easterly edge of the survey area). Colored pixels indicate ribbon seal counts
415 along aerial transects. The average effective area surveyed in each grid cell was
416 approximately 2.6km^2 (0.4%). Note that surveys were designed to target multiple seal
417 species, several of which had high densities further north (results not shown).

418 FIGURE 6. Potential covariates gathered to help explain and predict ribbon seal
419 abundance in the eastern Bering Sea. Covariates include distance from mainland
420 (`dist_mainland`), distance from 1000m depth contour (`dist_shelf`), average remotely
421 sensed sea ice concentration while surveys were being conducted (`ice_conc`), and distance
422 from the southern sea ice edge (`dist_edge`). All covariates except ice concentration were
423 standardized to have a mean of 1.0 prior to plotting and analysis.

424 FIGURE 7 Posterior mean predictions of ribbon seal apparent abundance across the

425 eastern Bering sea from models fit to survey data. Each row gives result for different model
426 types (GLM, GAM, or STRM, respectively); left column plots give results for naive runs
427 without pseudo-absences, while plots in the right column give predictions for runs where
428 pseudo-absence data (i.e., 0 counts in cells with less than 0.1% ice) were included. Cells
429 highlighted in black indicate those where predictions were outside the generalized
430 independent variable hull (gIVH).

FIGURES



FIG 1

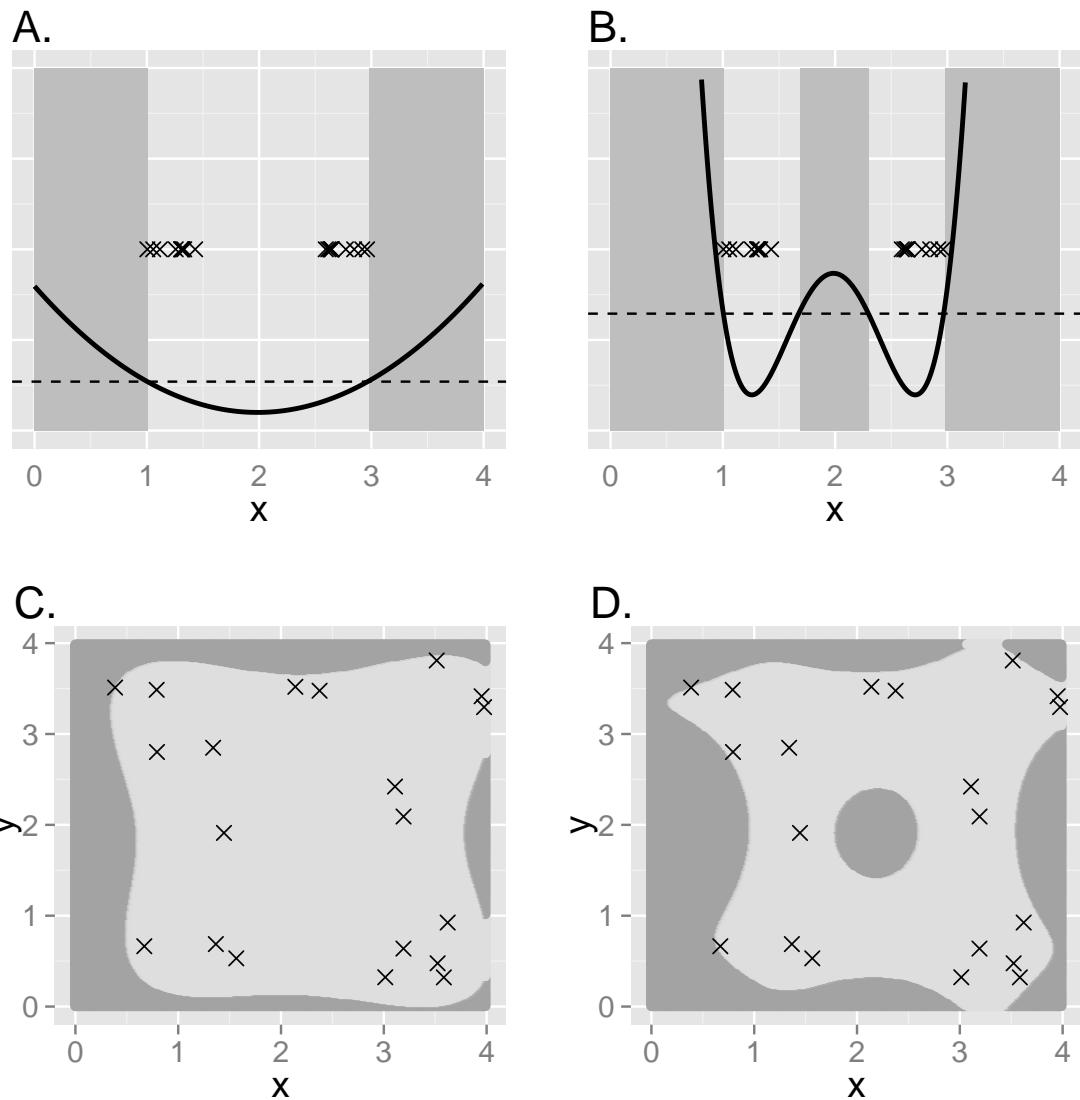


FIG 2

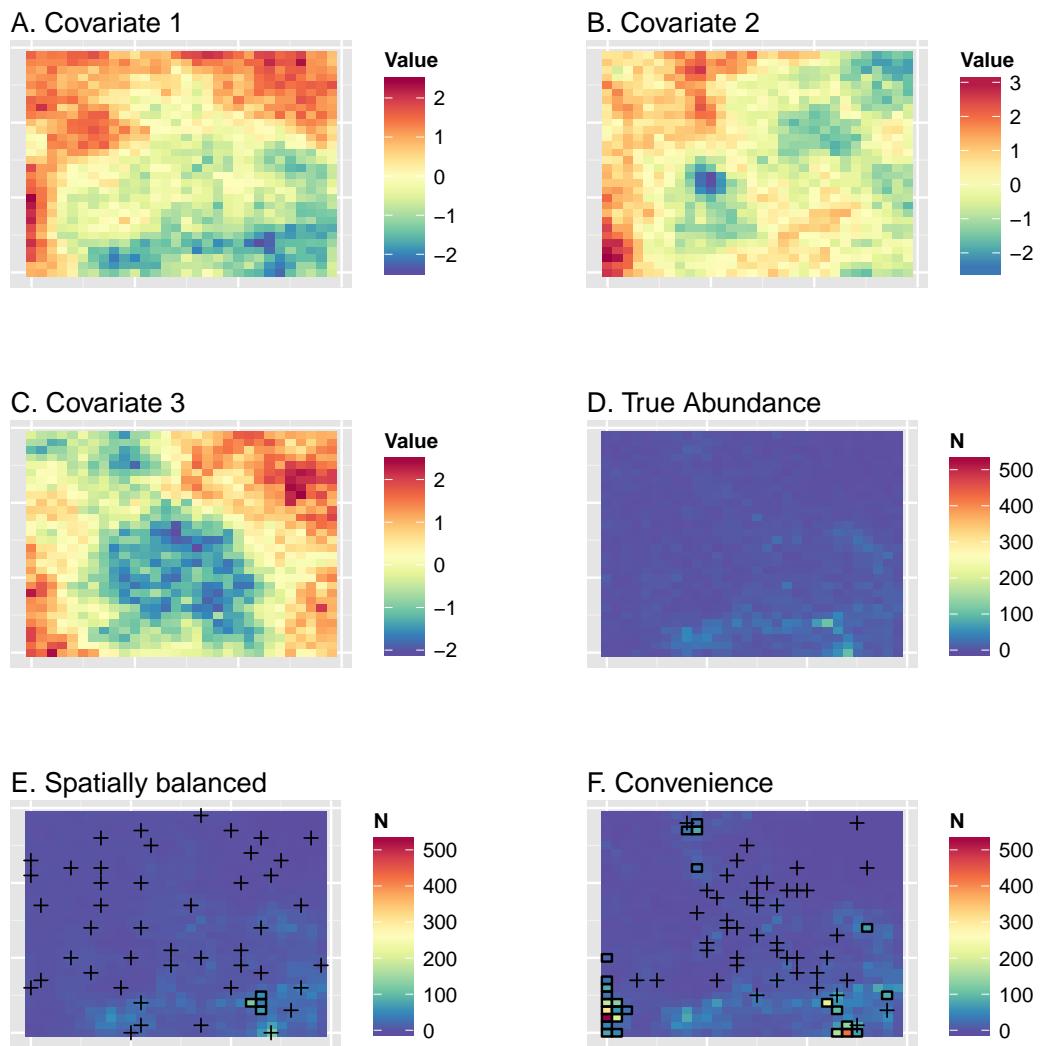


FIG 3

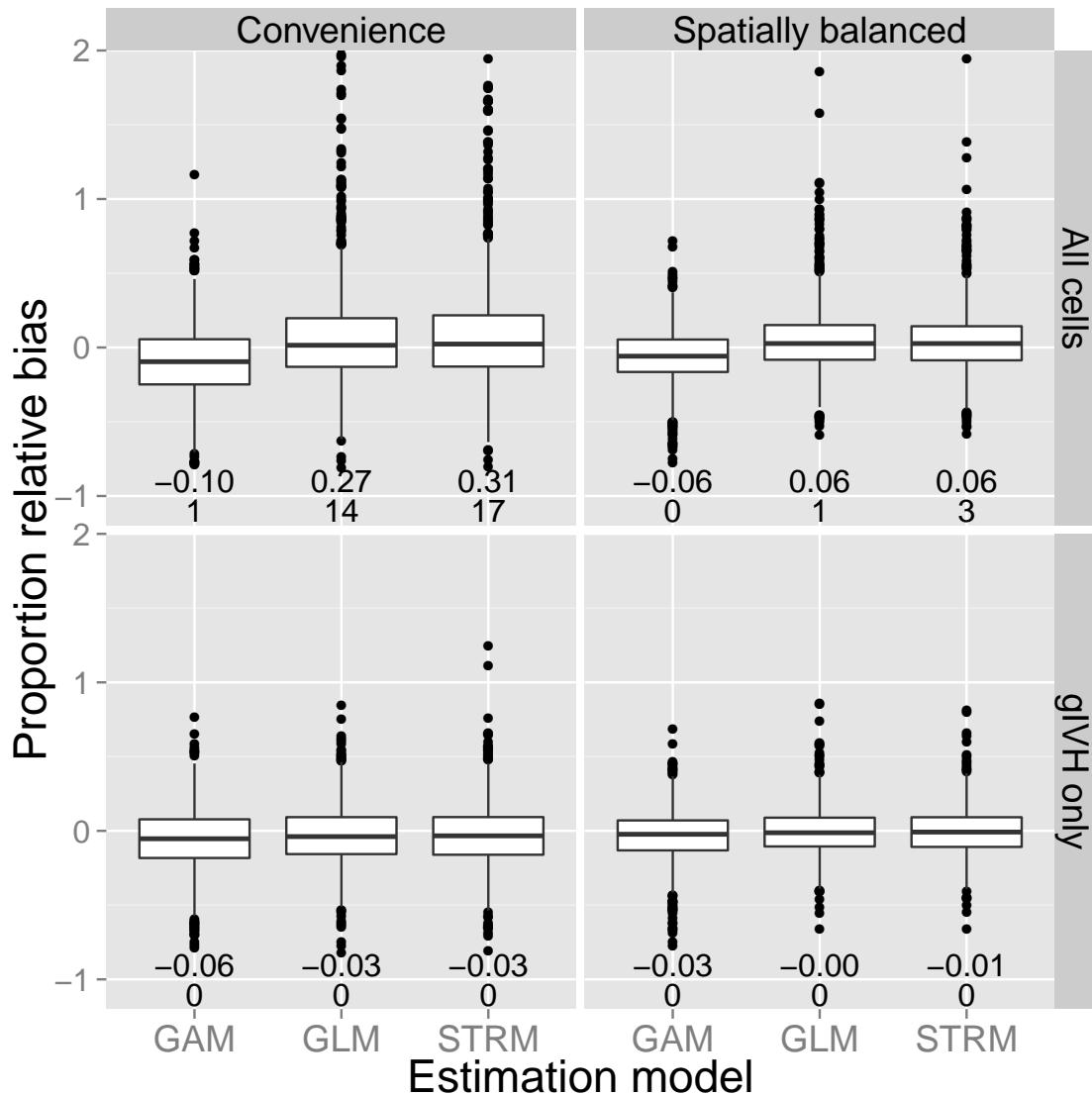


FIG 4

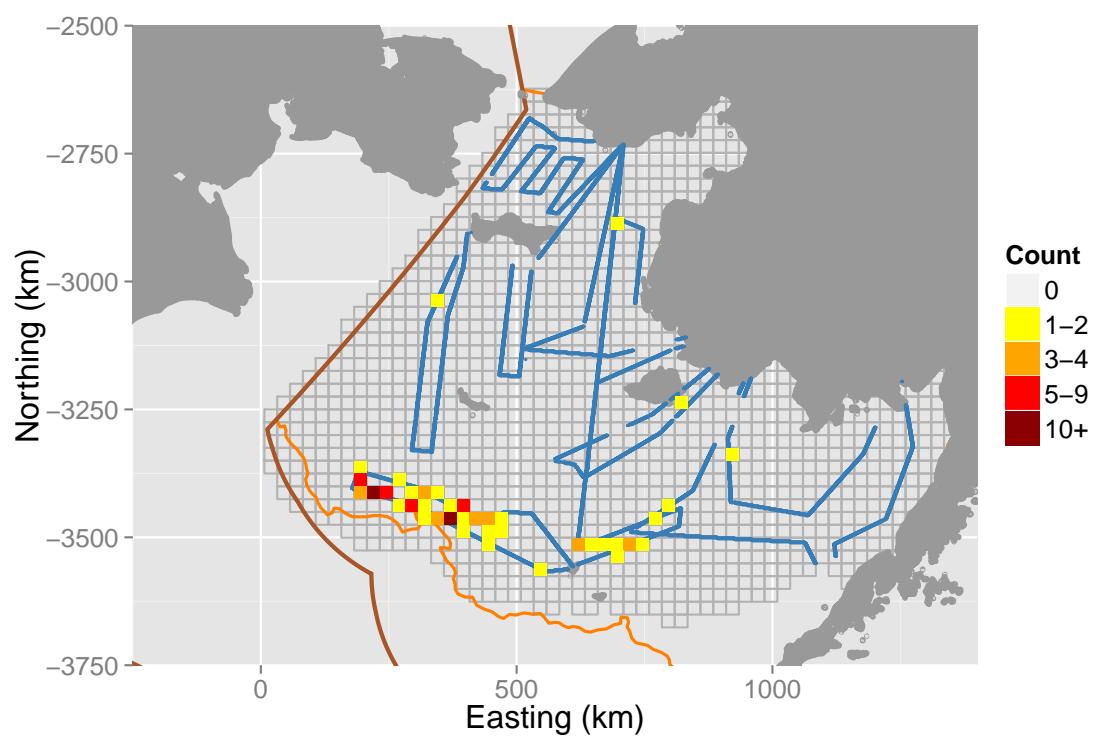


FIG 5

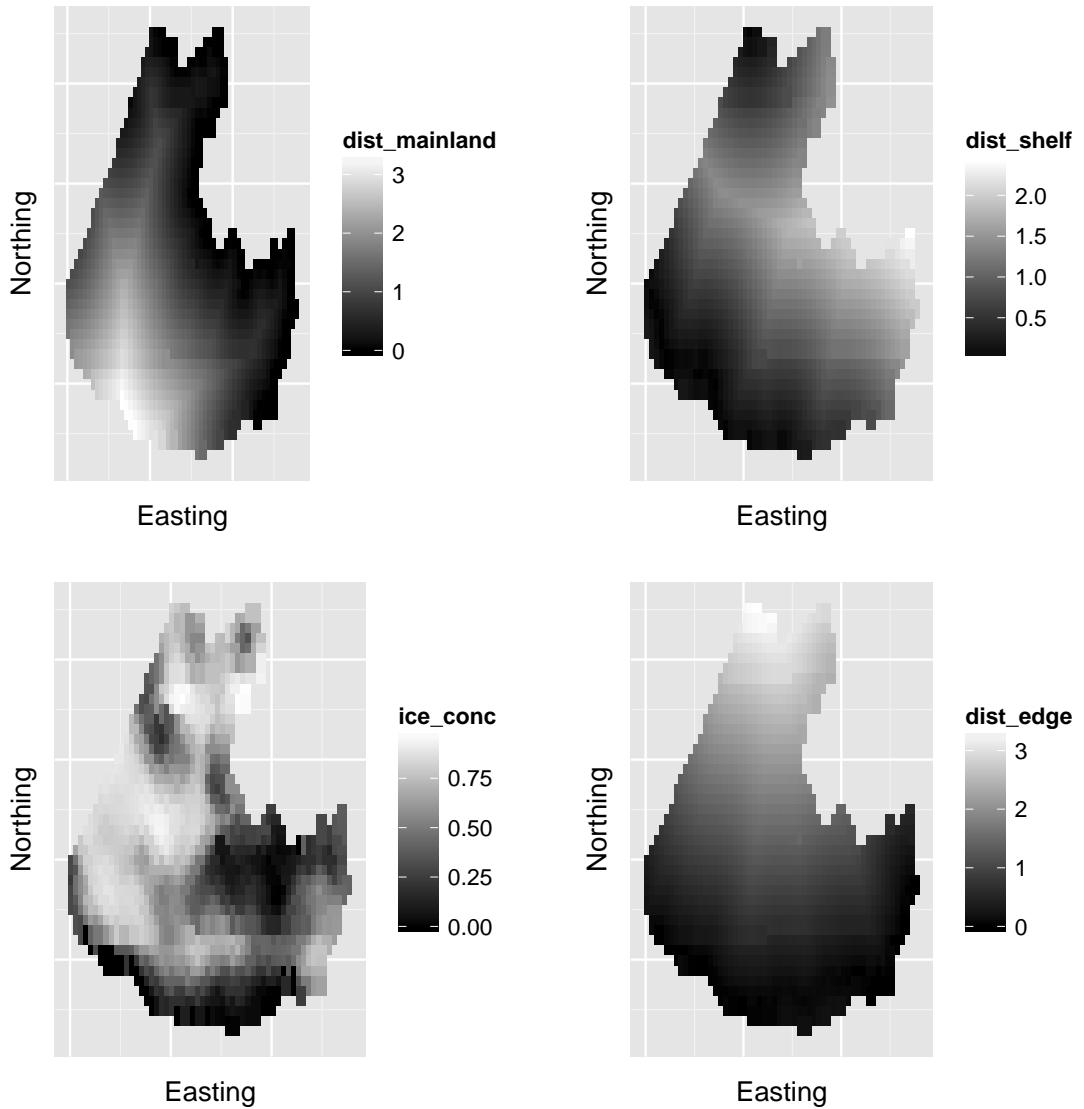


FIG 6

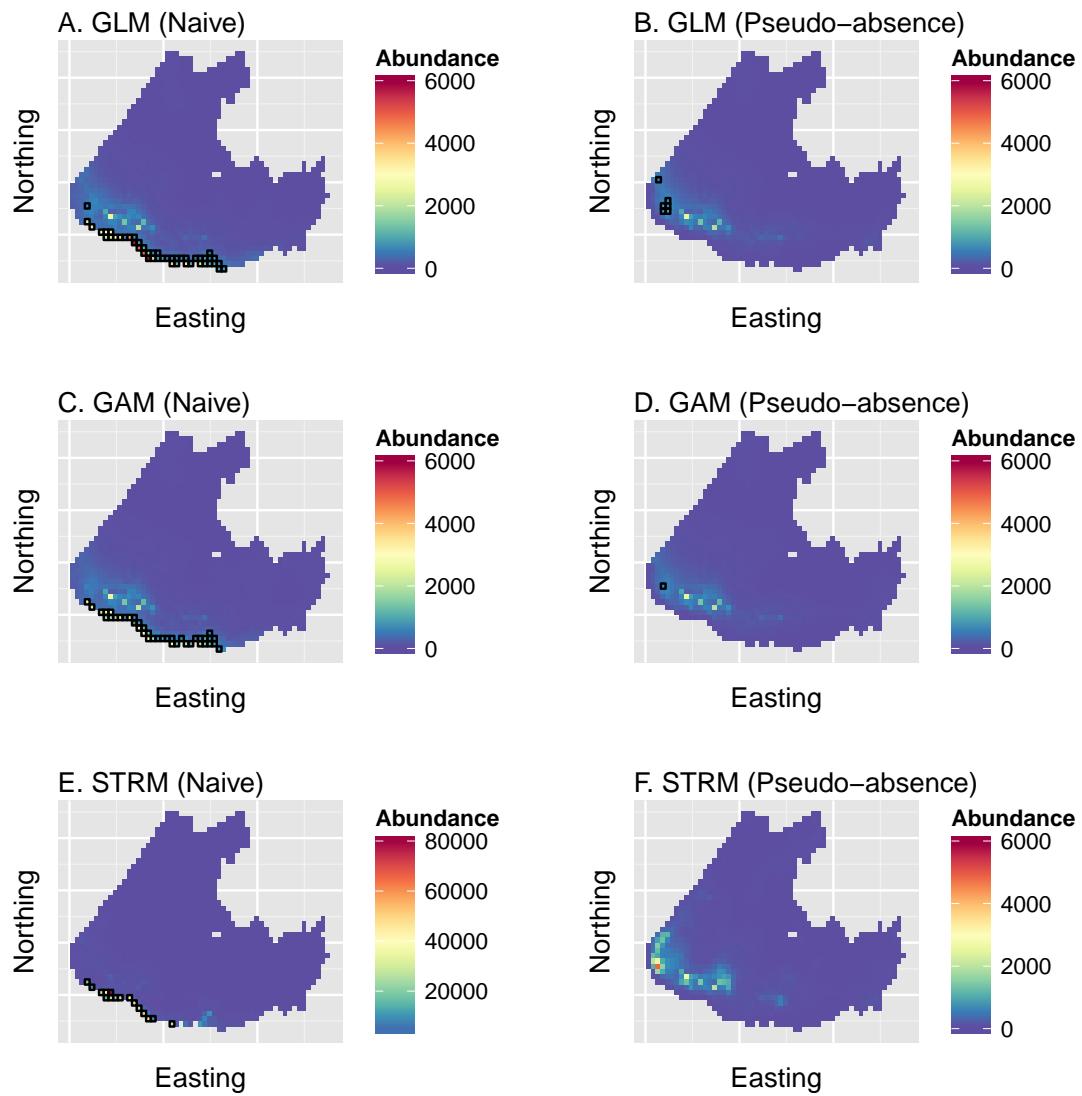


FIG 7