1   Title: Hotspot prioritizations show sensitivity to data type

2   Authors: Kari EA Norman[1], Ethan P White[2]

3

4   1. Department of Environmental Science, Policy, and Management, University of California

5   Berkeley, Mulford Hall, 130 Hilgard Way, Berkeley, CA 94720

6   2. Department of Wildlife Ecology and Conservation, University of Florida, 110 Newins-Ziegler

7   Hall, Gainesville, FL 32611.

8   3. Informatics Institute, University of Florida, 432 Newell Drive, Gainesville, FL 32611.

9   4. Biodiversity Institute, University of Florida, 432 Newell Drive, Gainesville, FL 32611.

10   ABSTRACT

11   Prioritizing regions for conservation is essential for effectively allocating limited

12   conservation resources. One of the most common approaches to prioritization is identifying

13   regions with the highest biodiversity, or hotspots, typically using global range map data. Range

14   maps are readily available at large scales for an array of taxa, but are also known to differ from

15   local-scale survey data in the same regions. We examined how prioritizations may differ between

16   range map and survey data using the North American Breeding Bird survey (BBS) and BirdLife

17   International range maps as a case study. Hotspot prioritizations were generated for species

18   richness and the richness of rare species at two scales.

19   Total species richness patterns differed substantially between data types with at most a

20   41% overlap in identified hotspots. Some regions had few or no hotspots for one data type and a

21   significant number for the other. Hotspots for rare species were more similar across the data

22   types with 44% overlap at the larger scale. Future efforts to prioritize areas for conservation

23   should consider differences between local-scale survey data and range maps, match data to the

24   scale of interest, and develop methods to better downscale range map-based prioritizations to the

25   scale of conservation decisions.

26

27   Keywords: conservation prioritization, range maps, survey data, hotspots, biodiversity

28   INTRODUCTION

29   A key component of allocating limited conservation resources is identifying high priority

30   regions for protection or management. Doing this effectively is essential to address threats to

31   natural systems and the services they provide, given growing human populations, rising global

32   temperatures, and widespread land use change. Approaches for systematically identifying the

33   most important areas to conserve is broadly referred to as conservation prioritization[1].

34   Conserving biodiversity is both a goal of conservation and a metric associated with other

35   positive conservation outcomes[2]. As a result, many conservation prioritization analyses focus on

36   maximizing the number of species (i.e., species richness) in a given area to be conserved. Other

37   desirable conservation criteria include endemism (the number of species occurring only in a

38   particular area), vulnerability (species designation as threatened or endangered), and level of

39   threat (likelihood of future habitat loss).

40   Myers' (2000) now-classic paper established biodiversity-based conservation

41   prioritization with a global assessment that identified 25 global biodiversity hotspots based on

42   vascular plant endemism and threat[3]. Building on this work, hotspot prioritizations have been

43   created for a number of different taxa[4–7], compared to current reserve networks to evaluate their

44   effectiveness in protecting biodiversity[8–11], and used to assess the scope of human impact on

45   biodiversity centers[12].

46   In addition, recent developments in software and methods (e.g., Zonation[13] and

47   Marxan[14]) provide improved prioritizations that incorporate greater levels of ecological and

48   human complexity. Such advances have transformed conservation prioritization from a technique

49   exclusively used in global-scale categorizations to a viable tool for local managers. Managers

1

50   now use conservation prioritization to inform decisions such as the allocation of conservation

51   money and expansion of local reserve networks[1].

52   As use of conservation prioritization proliferates, a largely unacknowledged

53   methodological divide has emerged between studies using two distinct kinds of data. Large-scale

54   hotspot analyses rely almost exclusively on geographic range map data[3,4,7,8,15] . These data are

55   heavily informed by expert opinion and potential habitat[16]. The relatively low cost of that

56   information means range maps are accessible for a wide array of taxa at continental to global

57   scales. However, range map data have two potential weaknesses: 1) typically, they are

58   temporally static; and 2) they reflect biodiversity at spatial scales of nearly 2x2 degrees (~40,000

59   km^2)[16]. In contrast, managers working at smaller scales typically use survey data. Though costly

60   to collect, these data provide direct observations of species richness or abundance in a particular

61   region. These regions are often much smaller than the 2x2 degree grid cells approximated using

62   range map data.

63   Despite this dichotomy between range map and survey-based approaches, there have

64   been no analyses to examine how differences in the underlying data influence the regions

65   prioritized for conservation. While range map suitability for use in IUCN classification has been

66   questioned[17], and comparisons of range map and survey data for biodiversity patterns more

67   generally show significant disparities[18], it is unclear how those discrepancies impact

68   prioritization results.

69   Here we explore the sensitivity of biodiversity hotspots, the simplest conservation

70   prioritization analysis, to the type of data (surveys or range maps) used in the analysis. We do

71   this by combining data from a continental scale survey of breeding landbirds with data from the

72   most widely used digital range maps. We compare the hotspots identified by these two data

73   sources at two scales using both total species richness and the richness of rare species and

74   discuss the implications of the results for future conservation prioritization efforts.

75

76   METHODS

77   *Data Sources*

78   We compared biodiversity patterns of North American breeding landbird species based

79   on survey and range map data. Digital breeding range maps were obtained from Birdlife

80   International[4,19]. Survey data were from 2,769 routes of the North American Breeding Bird

81   Survey (BBS; acquired using the Data Retriever[20]), collected from 2005-2015[21]. Each route is

82   24.5 miles long and surveyed annually in June. Three-minute point counts are made along the

83   route every 0.5 miles, in which every bird seen or heard within 400 m is recorded. Difficult to

84   survey groups including waterbirds, shorebirds, nightjars, and owls were excluded from survey

85   and range analyses.

86   Birds with less than 50% of their range in the survey area (based on range map data) were

87   also excluded to prevent boundary effects for maps of rare species. Rarity was determined by the

88   percentage of sites occupied for each species, with fewer occupied sites considered more rare.

89   Number of occupied sites only acts as an appropriate surrogate for range size when the survey

90   area covers the entire range. For example, species occurring primarily in Mexico with a portion

91   of their range along the US-Mexico border would be consistently classified as rare despite

92   having potentially large unsampled ranges. These species were excluded from both richness and

93   rarity analyses to allow for comparison across analyses, but richness maps retaining excluded

94   species have only marginal differences (Supplement). A complete list of species excluded from

95   analysis can be found in the supplement.

3

96

Analysis

98   The first step in comparing prioritizations based on survey and range map data was to

99 ensure a fair comparison between inherently different data types and methodological approaches.

100 Survey data are discrete point estimates of richness at each site. In contrast, range maps data are

101 typically aggregated into cells and richness patterns analyzed across those cells. To

102 accommodate both standard approaches we compared survey and range map based

103 prioritizations at both site and cell levels for total species richness and the richness of rare

104 species. Since all range maps (except those previously discussed) were included in the analyses

105 regardless of whether the species was observed in survey data, total range richness is higher for

106 range map data than survey data.

107   Point richness estimates for both data types were made at the starting position of each

108 BBS survey route. For survey data, estimates were the number of species recorded at any time on

109 each route over the ten-year survey period (2005-2015). All sites were included regardless of the

110 number of times they were sampled to most closely reflect range maps, which are a

111 representation of occurrence. The majority of sites (69%) were surveyed more than half of the

112 years in the survey period. Range map estimates were calculated by counting the number of

113 individual species ranges intersecting with each point. Cell level richness was calculated for 100

114 km^2 cells across North America[22]. This is a typical cell area that accounts for range map

115 resolution[16] while still producing a large number of cells for analysis.

116   Maps for rare species were created using the same methods for site and cell level richness

117 estimates. We classified a species as rare when it occurred at a proportion of sites less than the

118 median proportion of site occurrence[4], which was less than 4% or 5% of sites for survey and

119   range data respectively. As previously discussed, the rarity proportion is sensitive to variation in

120   the intensity of spatial sampling, as a species could be considered rare due to a lack of sampling

121   locations within its range. Richness estimates for rare species were therefore based on a subset of

122   sites adjusted to have consistent sampling intensity across the study area. This subset was made

123   up of three randomly selected sites from within 100 km^2 cells across North America. Cells

124   containing less than three sites were excluded. This combination of cell size and number of

125   samples was sufficient to address the bias while retaining a sufficient percentage of the data for a

126   meaningful analysis.

127   The sites or cells with species richness in the top 5% were identified as hotspots, a

128   commonly used cutoff for such analyses[4,16]. Hotspot locations based on survey data were

129   compared to those based on range map data. The percent of hotspots that were shared between

130   range map and survey based approaches was calculated by direct site-to-site and cell-to-cell

131   comparisons.

132   All code for this project is available on GitHub

133   (https://github.com/weecology/diversity-conservation) and archived on Zenodo

134   (https://doi.org/10.5281/zenodo.3258263). Data for Breeding Bird Survey is downloaded by this

135   code. Range map data can be obtained from http://datazone.birdlife.org/species/requestdis.

136

137   RESULTS

138   *Richness*

139   Maps of hotspot locations for species richness showed both similarities and notable

140   differences at the site and cell level (Figure 1). At the site level, survey data show hotspots across

141   the northeastern United States to the Great Lakes region up into Lake Winnipeg, and stretching

142   down to southern states. Smaller hotspot contingents are also scattered across the Rockies, the

143   Colorado Plateau, and California. Range map data show generally less dispersed priority regions

144   likely due at least in part to the much higher spatial autocorrelation inherent in range map data[18].

145   Heavily concentrated priority areas occur primarily in the northern Rockies, the Colorado

146   Plateau, and the Great Lakes Region into Lake Winnipeg. Smaller hotspots were also identified

147   in the mountains of northern California and New York. In general range map data show a much

148   higher concentration of hotspots in the Rocky Mountains and a general absence of the hotspots

149   identified by survey data in the Appalachian Mountains. As a result, range map analyses gave

150   higher priority to regions in western North America, while analyses of survey data gave higher

151   priority to eastern North America. Hotspot similarity was low at only 22% (Figure 3).

152   At the cell level, hotspots for survey data remain largely concentrated in the Great Lakes

153   region and New England. Scattered hotspots in the Rockies, California and the Colorado Plateau

154   aggregated to maintain hotspot representation in those areas while hotspots in the Midwestern

155   and southern states disappeared. Range map data show similar hotspot distributions to those at

156   the site level, though the small number of hotspots in the central Great Lakes region and northern

157   Californian disappeared. Concentrations of hotspots remained around Lake Winnipeg, Lake

158   Superior, Lake Ontario, the northern Rockies and the Colorado Plateau (Figure 1). This resulted

159   in more similar maps and greater hotspot similarity, 41%, between the survey and range map

160   relative to the site level (Figure 3). However, range maps gave higher priority to the Colorado

161   Plateau than the survey maps which prioritized the Great Lakes region and northern states more

162   heavily.

163   *Rarity*

6

164   Richness patterns of rare species show more congruence across level and data type

165   (Figure 2, Figure 3). All four prioritizations identify hotspots along the western coast and the

166   Colorado Plateau, with a larger concentration of hotspots on the Colorado Plateau for range maps

167   and British Columbia for survey maps. Hotspots for rare species had 42% overlap at the site

168   level and 44% at the cell level (Figure 3).

169

170

171   DISCUSSION

172   Comparisons of conservation priority areas based on survey and range map data show

173   both similarities and notable differences between data types. Some of the discrepancies are

174   substantial, particularly when evaluating total species richness at the site level. While some

175   hotspots were in qualitatively similar regions for both data types, direct site comparison showed

176   only a 22% overlap (Figure 3) and there were large regions for which there was high biodiversity

177   for one data type and low biodiversity for the other. Aggregation to the cell level increased

178   regional consistency qualitatively but still resulted in only 42% overlap between identified

179   hotspots for the two data types (Figure 3). Even for regions prioritized by both data types, maps

180   differ notably in which regions are most heavily emphasized. The Colorado Plateau has the

181   largest concentration of hotspots for range maps and only a few for survey maps, while the

182   reverse is true of the northeast with many hotspots for survey data and few hotspots for range

183   map data. These discrepancies highlight the necessity for understanding the drivers of difference

184   between data types and identifying the appropriate contexts for each to ensure the most informed

185   conservation decisions. This need is especially pressing given uncertainty in how discrepancies

186 may propagate through the increasingly complex prioritization analyses that are becoming more

187 common.

188   Priority areas based on rare species richness were more congruent (Figure 2) with 42%

189 and 44% overlap for site and cell level comparisons respectively (Figure 3). Hotspot regions

190 were qualitatively consistent across data type and level, with hotspots along the western coast

191 and on the Colorado Plateau. Greater overlap in hotspots for rare species compared to overall

192 richness could be attributed to the different drivers that total richness and richness of rare species

193 respond to. Species richness patterns are driven primarily by species with wide ranges and

194 therefore drivers such as area, habitat heterogeneity and productivity[23]. In contrast, the most

195 important driver for rare species is topographic heterogeneity[23]. Topographic heterogeneity is

196 also generally positively associated with range map richness but negatively associated with

197 survey richness[18]. Since range map richness and the richness of rare species are both positively

198 influenced by topographic heterogeneity, similarity in drivers offer one possible explanation for

199 higher congruence between data types for rare species than for common ones. To use both data

200 types effectively further work is needed to understand the drivers of consistencies and

201 inconsistencies between data types.

202   In choosing a data type for future prioritizations, one approach is to match the scale of the

203 data type to the scale at which decisions are being made. Our cell level analyses were performed

204 at a resolution of approximately 1 degree (100 km^2 cells), a resolution typical for range map-

205 based prioritizations. However, this is a coarser grain than would typically be used for local

206 management decisions. As described in Jenkins et al.[4], an area of that size in some parts of the

207 world contains multiple mountain ranges and valleys. Attempts to simply analyze range map data

208 at a scale more appropriate for conservation (e.g., Jenkins et al.[4]) may be misleading, as the

209   inherent resolution of most range maps by some assessments is only 2 by 2 degrees[16]. Analysis

210   of range map data below this scale results in overestimates of biodiversity and distorted spatial

211   patterns[16].

212   The mismatch between the resolution of range map data and the scale of conservation

213   questions leaves survey data as a natural replacement. The local scale of survey data captures

214   biodiversity patterns more representative of reality at the scales important for conservation.

215   However, the use of survey data does present challenges. Important considerations for survey

216   data use include the difficulty in accounting for incomplete surveys due to imperfect

217   detectability[24] and the availability of high quality data for a variety of taxa and locations[25].

218   Methods developed to adjust for imperfect detectability show estimated richness patterns that are

219   highly correlated with observed richness in birds, indicating that detectability likely does not

220   have a large influence on spatial patterns of richness[26]. The increased availability of high-quality

221   survey data through community science and large-scale government efforts means that using

222   survey data in place of range map data is increasingly possible for a greater variety of taxa.

223   Efforts like the National Ecological Observatory Network and burgeoning community science

224   programs will assure that more data for survey based assessments is available in the future. The

225   potential for survey data to improve the accuracy of biodiversity-based conservation indicates

226   that its further availability is a worthwhile investment for the conservation community.

227   Despite the many benefits of survey data, conservation decisions cannot wait for more

228   comprehensive data availability. It is therefore important to explore approaches for using

229   currently available survey and range data as effectively as possible. Developments in single

230   species and joint species distribution modeling offer promising approaches for filling geographic

231   and temporal gaps in survey data using environmental information[27,28]. Spatial interpolation has

232   been shown to perform well for smoothing spatial gaps in survey data[29]. Potential improvements

233   in range maps include methods for downscaling range map data to more useful scales[30,31],

234   updating old maps to reflect changes such as range shifts and land use changes, and new

235   approaches for addressing range map porosity[18]. Establishing relationships between richness

236   drivers and congruence between data types, as previously described, will also play an important

237   role in making methods for accurate range map data use possible. Development of new methods

238   that integrate survey and range map data to inform occurrence extent would greatly improve use

239   of all available species data.

240   Our findings underline the importance of understanding the implications of data type

241   when prioritizing areas for conservation. Discrepancies in hotspot location and overall

242   biodiversity patterns between data types give evidence of the current tradeoff between accuracy

243   and availability in data type selection. Further exploration of biodiversity analyses' sensitivity to

244   data type is essential for partitioning the appropriate roles of each data type and ensuring the

245   most effective conservation planning into the future.

246

254   1. Lehtomäki, J., Tomppo, E., Kuokkanen, P., Hanski, I. & Moilanen, A. Applying spatial

255   conservation prioritization software and high-resolution GIS data to a national-scale study in

256   forest conservation. *Forest Ecology and Management* **258**, 2439–2449 (2009).

257   2. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67

258   (2012).

259   3. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, G. A. B. da & Kent, J.

260   Biodiversity hotspots for conservation priorities. *Nature* **403**, 853 (2000).

261   4. Jenkins, C. N., Pimm, S. L. & Joppa, L. N. Global patterns of terrestrial vertebrate diversity

262   and conservation. *Proceedings of the National Academy of Sciences* **110**, E2602–E2610

263   (2013).

264   5. Ceballos, G. Global Mammal Conservation: What Must We Manage? *Science* **309**, 603–607

265   (2005).

266   6. Brum, F. T. *et al.* Global priorities for conservation across multiple dimensions of

267   mammalian diversity. *Proceedings of the National Academy of Sciences* **114**, 7641–7646

268   (2017).

269   7. Brooks, T. M. Global Biodiversity Conservation Priorities. *Science* **313**, 58–61 (2006).

270   8. Jenkins, C. N., Van Houtan, K. S., Pimm, S. L. & Sexton, J. O. US protected lands mismatch

271   biodiversity priorities. *Proceedings of the National Academy of Sciences* **112**, 5081–5086

272   (2015).

273   9. Brooks, T. M. *et al.* Coverage Provided by the Global Protected-Area System: Is It Enough?

274   *BioScience* **54**, 1081 (2004).

275   10. Rodrigues, A. S. L. *et al.* Effectiveness of the global protected area network in representing

276   species diversity. *Nature* **428**, 640–643 (2004).

277 11. Venter, O. *et al.* Bias in protected-area location and its effects on long-term aspirations of

278   biodiversity conventions. *Conservation Biology* (2017).

279 12. Gaston, K. J., Blackburn, T. M. & Goldewijk, K. K. Habitat conversion and global avian

280   biodiversity loss. *Proceedings of the Royal Society B: Biological Sciences* **270**, 1293–1300

281   (2003).

282 13. Lehtomäki, J. & Moilanen, A. Methods and workflow for spatial conservation prioritization

283   using Zonation. *Environmental Modelling & Software* **47**, 128–137 (2013).

284 14. Moilanen, A. & Wilson, K. A. *Spatial conservation prioritization : quantitative methods*

285   *and computational tools*. (Oxford University Press, 2009).

286 15. Brooks, T. M. *et al.* Habitat Loss and Extinction in the Hotspots of Biodiversity.

287   *Conservation Biology* **16**, 909–923 (2002).

288 16. Hurlbert, A. H. & Jetz, W. Species richness, hotspots, and the scale dependence of range

289   maps in ecology and conservation. *Proceedings of the National Academy of Sciences* **104**,

290   13384–13389 (2007).

291 17. Ramesh, V., Gopalakrishna, T., Barve, S. & Melnick, D. J. IUCN greatly underestimates

292   threat levels of endemic birds in the Western Ghats. *Biological Conservation* **210**, 205–221

293   (2017).

294 18. Hurlbert, A. H. & White, E. P. Disparity between range map- and survey-based analyses of

295   species richness: patterns, processes and implications: Range map- vs. survey-based species

296   richness. *Ecology Letters* **8**, 319–327 (2005).

297 19. BirdLife International and Handbook of the Birds of the World (2016) Bird species

298   distribution maps of the world. Version 6.0. Available at

299   http://datazone.birdlife.org/species/requestdis.

300   20. Morris, B. D. & White, E. P. The EcoData Retriever: Improving Access to Existing

301   Ecological Data. *PLoS ONE* **8**, e65848 (2013).

302   21. Pardieck, K.L., D.J. Ziolkowski Jr., M. Lutmerding and M.-A.R. Hudson. 2018. North

303   American Breeding Bird Survey Dataset 1966 - 2017, version 2017.0. U.S. Geological

304   Survey, Patuxent Wildlife Research Center. https://doi.org/10.5066/F76972V8.

305   22. Orme, C. D. L. *et al.* Global hotspots of species richness are not congruent with endemism or

306   threat. *Nature* **436**, 1016–1019 (2005).

307   23. Jetz, W. Geographic Range Size and Determinants of Avian Species Richness. *Science* **297**,

308   1548–1551 (2002).

309   24. Boulinier, T., Nichols, J. D., Sauer, J. R., Hines, J. E. & Pollock, K. H. Estimating species

310   richness: the importance of heterogeneity in species detectability. *Ecology* **79**, 1018–1028

311   (1998).

312   25. Cantú-Salazar, L. & Gaston, K. J. Species richness and representation in protected areas of

313   the Western hemisphere: discrepancies between checklists and range maps. *Diversity and*

314   *Distributions* **19**, 782–793 (2013).

315   26. Evans, K. L., Newson, S. E., Storch, D., Greenwood, J. J. D. & Gaston, K. J. Spatial scale,

316   abundance and the species–energy relationship in British birds. *Journal of Animal Ecology*

317   **77**, 395–405 (2008).

318   27. Elith, J. & Leathwick, J. R. Species Distribution Models: Ecological Explanation and

319   Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*

320   **40**, 677–697 (2009).

321   28. Peterson, A. T., Navarro-Sigüenza, A. G. & Gordillo, A. Assumption-versus data-based

322   approaches to summarizing species' ranges: Summarizing Species' Ranges. *Conservation*

323   *Biology* (2016).

324   29. Bahn, V. & McGill, B. J. Can niche-based distribution models outperform spatial

325   interpolation? *Global Ecology and Biogeography* **16**, 733–742 (2007).

326   30. Keil, P., Belmaker, J., Wilson, A. M., Unitt, P. & Jetz, W. Downscaling of species

327   distribution models: a hierarchical approach. *Methods in Ecology and Evolution* **4**, 82–94

328   (2013).

329   31. Keil, P., Wilson, A. M. & Jetz, W. Uncertainty, priors, autocorrelation and disparate data in

330   downscaling of species distributions. *Diversity and Distributions* **20**, 797–812 (2014).

331

332   **Figure 1.** Maps of species richness based on survey and range map data at the site and cell level.

333   Cell level richness is a count of species recorded at any site within the cell. Darker regions

334   correspond to higher biodiversity, with hotspots (richest 5%) marked in red.

335

336   **Figure 2.** Maps of the richness of rare species based on survey and range map data at the site and

337   cell level. Cell level richness is a count of rare species recorded at any site within the cell. Darker

338   regions correspond with more rare species, with hotspots (richest 5%) marked in red.

339

340   **Figure 3.** Percent hotspot similarity between range map and survey data types for richness and

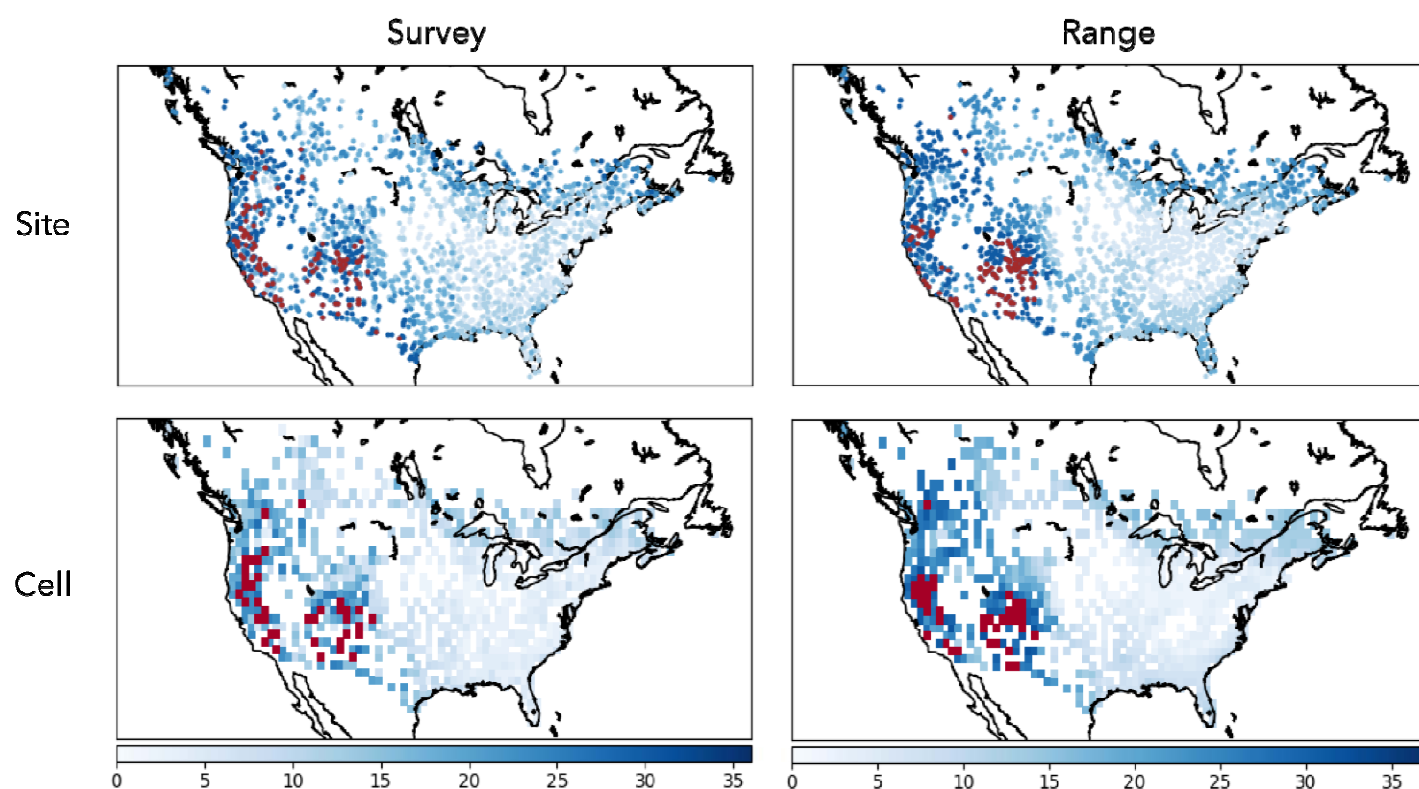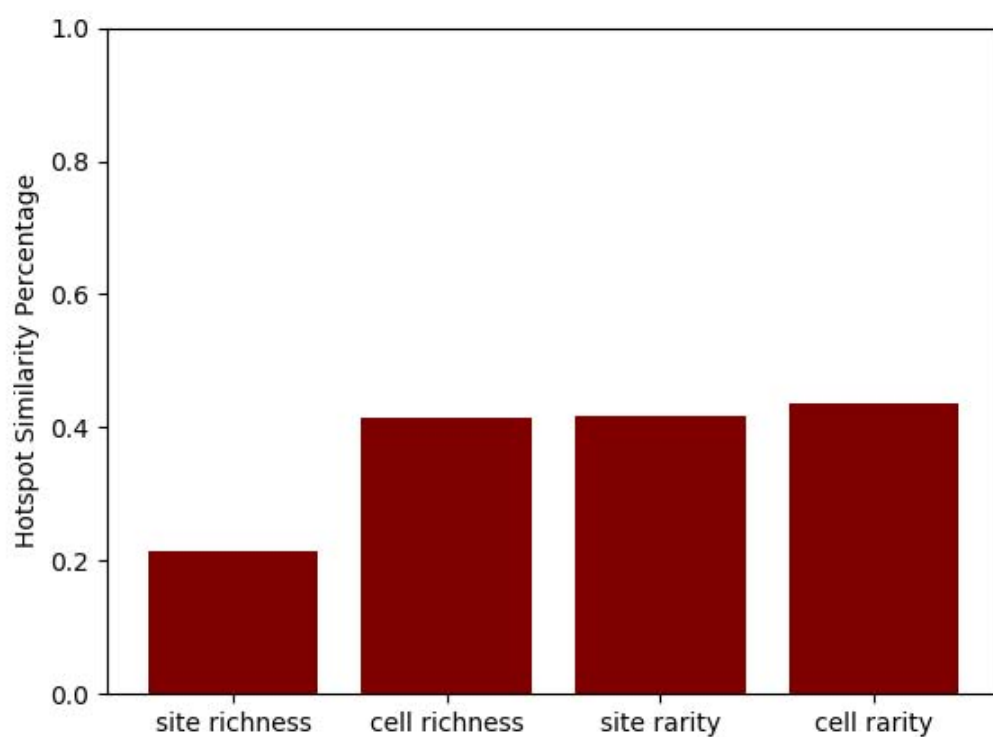341   rare species at the site and cell level.

342 **Figure 1**
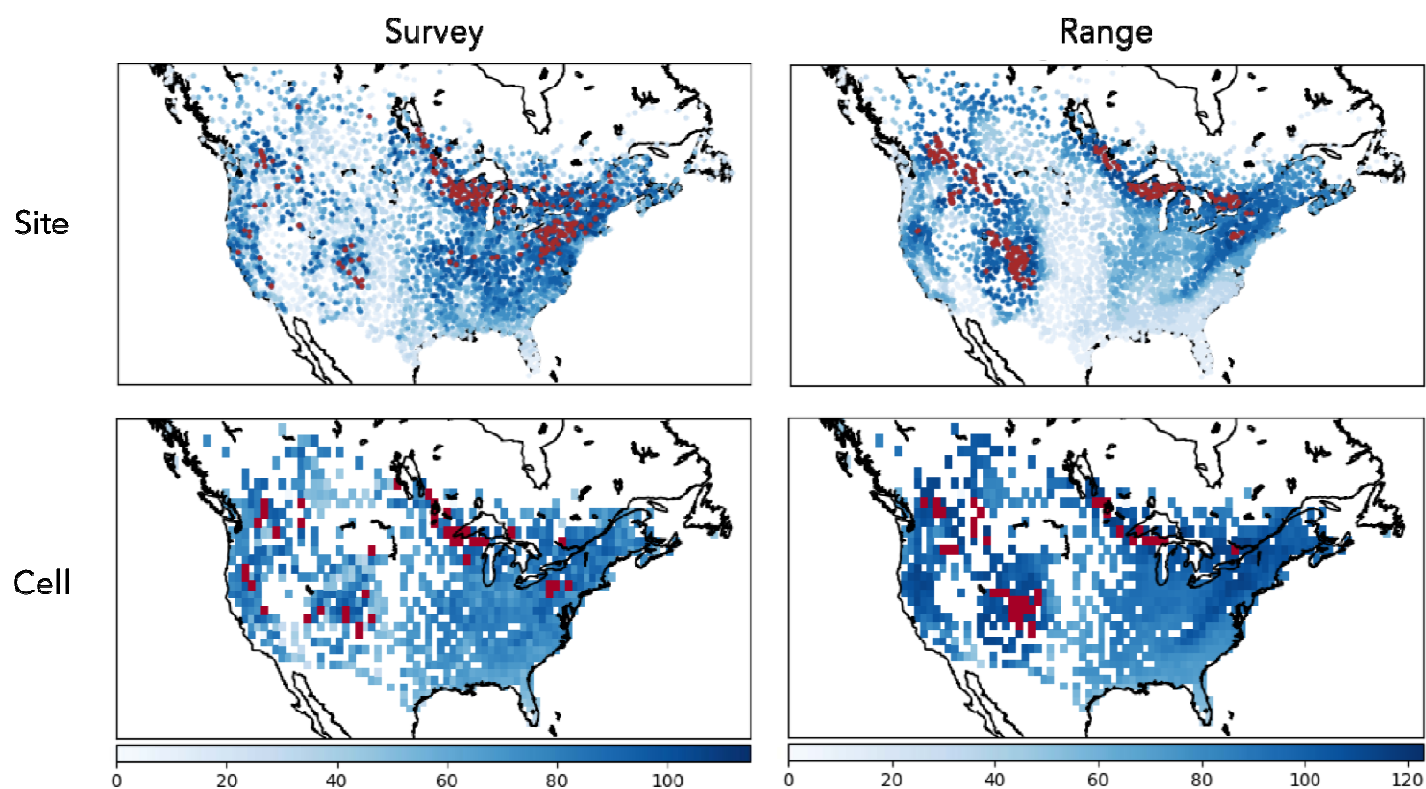
343

344   **Figure 2**

345

346 **Figure 3**

347

18