DOCKET
9-Reneweo-1
DATE June 27 2011
RECD. June 27 2011

prbo

Comments on "Proposed Species Habitat Modeling Approach Draft Memorandum of June 14, 2011" DRECP Docket Number 09-RENEW EO-01 provided by Christine A. Howell, PhD from PRBO Conservation Science on June 27, 2011

Below are my comments on the memorandum. In summary, I strongly suggest that the contractor use statistically-based modeling methods as the first step in creating distribution models; this was also the recommendation of the Independent Science Advisory Panel to the DRECP. Statistically-based models are widely used and accepted by academic and applied conservation scientists; they are highly transparent and allow for model evaluation and validation (as opposed to expert-based models). If statistical models perform poorly based on model evaluation diagnostics, then expert opinion models could be used as a follow-up effort. Overall statistical modeling will be more robust, accurate, transparent, and defensible than an expert opinion model. In my experience statistical models using algorithms such as Maxent can be developed, run, and evaluated for model performance within one week for a species. If the contractor is encountering difficulties with statically-based model methods because of lack of technical expertise, then they should obtain assistance from those more knowledgeable with those methods. Lack of technical expertise should not be used as rationale to ignore important conservation tools such as statistically-based models. Finally, spatial resolution is not addressed in the memo, but care should be given to selecting the appropriate spatial scale for the proposed modeling exercises.

Sincerely,
Christine A. Howell, PhD

Christine A Howell

Sr. Conservation Scientist
PRBO Conservation Science
3820 Cypress Dr. #11
Petaluma, CA 94954
707-781-2555 ext 315 / chowell@prbo.org

- **Comment 1**: Regarding the sentence on page 2 of the memo which states: "We considered the development of statistically-based models for the DRECP but rejected this approach for several reasons including technical difficulties, existing data limitations, and lack of transparency."

  Issue of technical difficulties: The memo does not elaborate on the technical difficulties encountered, but numerous ecological studies have successfully applied statistically-based model algorithms (Elith et al. 2006, Elith et al. 2011) including algorithms appropriate for situations where there are only records of a species' presence ("presence-only" algorithms), as well as algorithms that are appropriate for situations where there are both species' presence and absence records ("presence-absence" algorithms). Maxent is an algorithm that works with presence-only data and has been highly competitive when

compared to other presence-only or presence-absence algorithms (see Elith et al. 2006 for an extensive comparison of algorithms). As described in Elith et al. 2011 (p. 44):

> "MaxEnt's predictive performance is consistently competitive with the highest performing methods (Elith et al., 2006). Since becoming available in 2004, it has been utilized extensively for modelling species distributions. Published examples cover diverse aims (finding correlates of species occurrences, mapping current distributions, and predicting to new times and places) across many ecological, evolutionary, conservation and biosecurity applications (Table 1). Government and nongovernment organizations have also adopted MaxEnt for large-scale, real-world biodiversity mapping applications, including the Point Reyes Bird Observatory online application (http://www.prbo.org/) and the Atlas of Living Australia (http://www.ala.org.au/)."

If the contractor is encountering technical difficulties because of lack of expertise with statistically-based modeling methods then they should obtain assistance from those more knowledgeable in these methods. Lack of expertise should not be used as rationale to not use important conservation tools such as statistically-based models. In fact the DRECP ISA (2010) states (pg 54, bold emphasis added here):

> "Species distribution (or occupancy) modeling is a very active and constantly evolving research field with numerous recent advances (Elith et al. 2006, Elith and Leathwick 2009; http://biodiversityinformatics.amnh.org/index.php?section=sdm_guide). SDMs use environmental variables characterizing places where a species does (or does not) occur based on survey data to develop sophisticated correlative models. SDMs may also be extrapolated to project future occurrences in places where the correlated environmental features are projected to be present in the future (Wiens et al. 2009). Care should be taken to select a modeling approach and SDM algorithm that performs well based on recent peer-reviewed literature and which is appropriate for the organism being modeled. It may be prudent to model the data with more than one SDM algorithm and examine overlap among model outputs ("consensus modeling"), as well as the amount of uncertainty among model outputs (see Wiens 2009 for an example of uncertainty analysis).
> **We emphasize the importance of expertise and rigor in applying these highly technical models. In our collective experience, this expertise is generally lacking at environmental consulting firms that prepare HCPs, NCCPs, and NEPA and CEQA documents. However, there is a growing pool of appropriate expertise at academic research institutions, science-based NGOs, and science-based government agencies, such as USGS. *We urge DRECP to tap appropriate expertise for the application of any scientific models,*** *because learning-while-doing is inefficient and error-ridden.*"

Conclusion: There is not adequate justification to list "technical difficulties" as rationale to dismiss the use of statistically-based methods.

Issue of existing data limitations: The memo cites a study by Wisz et al. 2008; this study evaluates predictions from 12 algorithms for 46 species (from six different regions of the world) at three sample sizes (100, 30, and 10 records). In the abstract of the paper Wisz et al. state (p. 763, bold emphasis added here):

> "Other algorithms were much less sensitive to sample size, including an algorithm based on maximum entropy **(MAXENT) that had among the best predictive power across all sample sizes**. Relative to other algorithms, a distance metric algorithm (DOMAIN) and a genetic algorithm (OM-GARP) had intermediate performance at the largest sample size and among the best performance at the lowest sample size. **No algorithm predicted consistently well with small sample size (*n*< 30) and this should encourage highly conservative use of predictions based on small sample size and restrict their use to exploratory modelling."**

Based on the Wisz et al. study, Maxent is an appropriate algorithm when there are 30 or more records. Wisz et al. (p 771) also note the potential for exploring data using Maxent with sample sizes less than 30:

> **"We have shown that no modeling approach tested was fully robust to small sample sizes, but that for exploratory modelling with sample sizes between 10 and 30 records, MAXENT,** OM-GARP, and possibly DOMAIN **may be the best available."**

It should also be noted that data limitations also apply to expert opinion models. For example, if there is limited or spatially biased sampling for a species, the fact that the majority of known localities may be in a particular vegetation or soil type could potentially bias the expert's opinion just as it would the statistical model. However, statistical modeling also allows one to conduct sensitivity analyses in order to determine how potential bias might be affecting the model output. Overall statistical modeling will be more robust, accurate, transparent, and defensible than an expert opinion model.

Conclusion: While various studies have shown that statistical models are more robust with larger sample sizes, a sample size of 30 is robust for modeling data with at least one algorithm (Maxent). Exploratory analyses are also possible with smaller data sets. If 30 records are available then statistically-based models should be constructed and model performance should be evaluated. Non-statistical modeling methods should be considered for sample sizes less than 30 and when statistical models created with 30 or more records have demonstrated poor model performance after undergoing model validation diagnostics. The consultants should determine what the sample sizes are for each of the taxa of interest and then consult with species experts to determine if there are additional data sources. There are numerous potential sources of species occurrence data beyond the CNDDB (many of these are listed in the DRECP ISA 2010) including eBird.

<u>Issue of lack of transparency</u>: It is not clear how the use of statistical algorithms is not transparent. In fact the bottom of page one of the memo states that " Species distribution modeling can provide an objective, transparent, and repeatable means of assessing species habitat distribution where the species distribution or distribution of suitable habitat for a species is not well known." Moreover, species distribution models have a range of validation and evaluation diagnostics that allow the user to better understand the results and the validity of the model outputs. These tools allow the user to determine how much of the variance is explained by the model, as well as which variables are most important in the model, thus providing greater transparency. These diagnostics also allow the user to assess the uncertainty associated with the model because the outputs are created in terms of a probability of occurrence (ranging from zero to one). These probability of occurrence surfaces may reflect the dynamic nature of a species' range and can be used to prioritize future conservation actions or survey efforts. Expert-opinion models do not provide a means to evaluate model performance or model uncertainty.

Conclusion: It is not justifiable to list "lack of transparency" as a reason not to use statistically-based methods.

- **Comment 2**: Regarding sentence on page 3 of the memo that states: "As described by Elith and Leathwick (2009), "*modelers are still coming to terms with how best to model presence-only data*." " [italics added]

  Issues with presence only modeling: The phrase italicized above gives the impression that presence-only model methods are inadequate. However in the context of the entire paragraph of their paper, Elith and Leathwick (2009) are highlighting the range of views by researchers of presence-only modeling approaches and they conclude by indicating the need for further progress in the field. In a paper published two years later (Elith et al. 2011) provide additional details (p. 43) on the progress of presence only modeling as well as details on implementing Maxent. The paper is:

  > "…written for ecologists and describes the MaxEnt model from a statistical perspective, making explicit links between the structure of the model, decisions required in producing a modelled distribution, and knowledge about the species and the data that might affect those decisions."

  Conclusion: Presence only algorithms can provide robust model results if judiciously applied.

- **Comment 3**: Regarding the last sentence on page 3 of the memo which states: "Additionally, the spatial bias inherent in the occurrence data can result in skewed modeling results with statistically-based models."

  Issue of spatial bias: Elith et al. 2011 (p. 43) describes methods for dealing with biased species data using the Maxent algorithm:

  > "Appropriate treatments for survey bias, unprojected data, locally restricted species, and predicting to environments outside the range of the training data are demonstrated, and new capabilities discussed."

  Additional papers on Maxent available from Steven Phillips' (one of the Maxent developers) website http://www2.research.att.com/~phillips/density.php provide guidance on handling data bias.

  Conclusion: There are methods to handle biased data. The contractor should first determine what species occurrence data are available, what the biases are for existing data (perhaps in concert with species experts), and what steps are needed to correct those biases; corrective steps should then be implemented. As noted in Comment #1, biased data will also be a problem for expert-based models.

- **Comment 4**: Regarding paragraph 2 on page 4 which states: "In addition to the limitations noted above, the technical complexities and rigorous nature of developing ground-up, DRECP-specific statistically-based models were not considered compatible with the scope and schedule of the DRECP Preliminary Conservation Strategy. By way of example, the statistically-based species distribution model for the desert tortoise developed by the US Geological Survey (Nussear et al. 2009) utilized 15,311 data points and involved an extensive amount of species presence data standardization, environmental variable resampling, and model tuning. A similar model is being developed for the Mohave ground squirrel, which when completed will have taken about 18 months to develop."

Issue of technical complexity: Numerous ecological studies have successfully applied statistically-based model algorithms (Elith et al. 2006, Elith et al. 2011) including algorithms appropriate for situations where there are only records of a species' presence ("presence-only" algorithms), as well as algorithms that are appropriate for situations where there are both presence and absence records ("presence-absence" algorithms) for a species. Both statistical and expert opinion models require environmental data (e.g. appropriate GIS layers) and a clean or standardized set of species occurrence data indicating where the species has occurred. Once species occurrence and environmental data are available then it should not take one technician longer than one week to create model output for one species. Additional refinement may be necessary but that will be required regardless of the chosen modeling method. Also, while Nussear et al. (2009) produced a robust model with a dataset of 15,311 data points (which was later resampled to 6,350 grid-cell points), a smaller number of data points can be used to produce a robust model (see Comment #1 regarding sample size).

If the contractor is encountering technical difficulties because of lack of technical expertise with statistically-based modeling methods then they should obtain assistance from those more knowledgeable with those methods. Lack of technical expertise should not be used as rationale to not use important conservation tools such as statistically-based models (See Comment #1).

Conclusion: Statistically-based modeling methods are frequently used by ecologists and conservation practitioners in order to reach scientifically defensible conclusions. These methods were considered preferable to expert based models by the DRECP ISA (2010) and these approaches have been used successfully by other scientists working within the DRECP boundaries (i.e. Nussear et al. 2009). If the contractor finds these approaches to be overly complex, then the contractor should seek out the expertise and assistance of those more knowledgeable in this field. The contractor has not provided any evidence that they have even attempted to apply statistically-based modeling methods.

Literature Cited

DRECP Independent Science Advisors (DRECP ISA). 2010. Recommendations of Independent Science Advisors for The California Desert Renewable Energy Conservation Plan (DRECP). Prepared for the Renewable Energy Action Team.

Elith, J., Phillips, S., Hastie, T., Dudik, M., Chee, Y., Yates, C. 2011 A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43-57.

Elith, J., J.R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time *Annu. Rev. Ecol. Evol. Syst.* 2009. 40:677–97

Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. M. Overton, A.T. Peterson, S.J. Phillips, K. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S. Williams, M.S. Wisz, and N.E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.

Nussear, K.E., T.C. Esque, R.D. Inman, L. Gass, K.A. Thomas, C.S.A. Wallace, J.B. Blainey, D.M. Miller,and R.H. Webb. 2009. Modeling habitat of the desert tortoise (Gopherus agassizii) in the Mojave and parts of the Sonoran Deserts of California, Nevada, Utah, and Arizona: U.S. Geological Survey Open-File Report 2009-1102, 18 p.

Wisz, M.S., R.J. Hijmans, J. Li, A.T. Peterson, C.H. Graham, A. Guisan, and NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species models. *Diversity and Distributions.* 14: 763-773.