Does addressing imperfect detection improve predictive performance of species distribution models?

Christopher T. Rota¹,², Robert J. Fletcher¹ Jr., Jason M. Evans¹,³, and Richard L. Hutto⁴

¹University of Florida, Dept. of Wildlife Ecology and Conservation
²University of Missouri, Dept. of Fisheries and Wildlife Science
³University of Georgia, Carl Vincent Institute of Government
⁴University of Montana, Division of Biological Science
Accurate predictions of species distributions are essential.

Conservation and management of species

Predicting the effects of global climate change on species distributions
Species Distribution Models (SDMs)

- Logistic Regression
  Presence/Absence (detection/non-detection)
- Maximum Entropy
  Presence-only
- Failure to account for imperfect detection may bias SDMs
Accounting for imperfect detection may improve SDM predictions

- Occupancy models estimate occurrence and detection probability simultaneously

- Flexible approach to modeling species distributions
Objectives

• Contrast predictive performance of occupancy models with logistic regression and maximum entropy SDMs.

• Evaluate how predictive performance changes over a range of detectability.

• *By accounting for imperfect detection, we expect occupancy models to improve predictive performance relative to other SDM approaches.*
Field Methods and Transect Selection

- Training = 5,745 points
- Validation = 2,290 points
Species Selection

Difficult to Detect

Highly Detectable
Habitat and Detection Covariates

- Selected unique covariates for each species to reflect known habitat associations

- Modeled detectability as a function of survey-specific covariates

- Derived all habitat covariates from GIS based measures
Occupancy Models

• Hierarchical Bayesian Occupancy Model

• Observation model for detections on a ‘removal’ sampling protocol

• Model averaging based on posterior model probabilities (Kuo and Mallick 1998)
Logistic Regression Models

• Hierarchical Bayesian Logistic Regression

• Model averaging based on posterior model probabilities
Maximum Entropy

- Estimated distributions with Maxent machine learning algorithm (Phillips et. al 2006)

- Overlaid known presence locations on GIS layers describing habitat covariates
Comparing Predictive Performance

- Predicted observations to the validation data.
- Calculated Area Under the Curve (AUC) of a ROC plot.
- Corrected predicted probability of occurrence for detection with occupancy models.
Relative Predictive Performance

Relative AUC

Occupancy  LR  Maxent
Occupancy vs. Logistic Regression

The diagram illustrates the relative AUC (Area Under the Curve) against detection probability for various models. The x-axis represents the detection probability, while the y-axis shows the relative AUC. Points are labeled with acronyms (e.g., BHCO, PIWO, CBCH, etc.) and are color-coded to indicate model performance. Two regions are highlighted: Greater Occupancy Model Performance and Greater Logistic Regression Performance.
Occupancy vs. Maxent

![Graph showing relative AUC vs. detection probability with points labeled BHC0, PIWO, TOSO, CBCH, MGWA, VATH, SWTH, GCKI, and TOWA. The graph indicates greater occupancy model performance compared to greater Maxent performance.]
Logistic vs. Maxent

The graph compares the relative AUC (Area Under the Curve) against detection probability for different species. The x-axis represents detection probability, and the y-axis represents relative AUC. The points indicate species performance: those closer to the top left corner show greater logistic regression performance, while those in the bottom right corner indicate greater Maxent performance.

Species mentioned include CBCH, TOSO, MGWA, VATH, SWTH, PIWO, GCKI, BHO, and TOWA.
Discussion

- Improved predictive performance for difficult to detect species
- Minimal gain in predictive performance
Discussion

Minimal gain may be explained by:
1. Only two replicate surveys
2. Homogeneous estimates of detection probability across sites
Discussion

Logistic Regression performs best when species highly detectable.
Discussion

Maxent performs relatively well when species difficult to detect.
Conclusions

Advantages of occupancy modeling approach:
1. Perform well across a gradient of detectability
2. Clear interpretation of results
3. Flexible modeling approach
4. Can collect detectability data with little additional effort

• Ecologists must weigh advantages against added field effort and computational expense.
Acknowledgements

Data and GIS Support: Jock Young & Anna Noson

Hierarchical Modeling Advice: Bob Dorazio

Committee Members: Katie Sieving & John Hayes
Questions?