Model Evaluation: Used Habitat Calibration (UHC) Plots

John Fieberg, Professor

Department of Fisheries, Wildlife and Conservation Biology

Model Evaluation

Calibration: how well does the model describe the probability (or relative probability) of using different areas?

Model Evaluation

Calibration: how well does the model describe the probability (or relative probability) of using different areas?

- If the model predicts that 10% of forested sites will be occupied, are 10% of forested sites occupied?

Model Evaluation

Calibration: how well does the model describe the probability (or relative probability) of using different areas?

- If the model predicts that 10% of forested sites will be occupied, are 10% of forested sites occupied?
- If the model predicts sites within 1 km of water are twice as likely to be used as sites 10km from water, does this happen?

Model Evaluation

Calibration: how well does the model describe the probability (or relative probability) of using different areas?

- If the model predicts that 10% of forested sites will be occupied, are 10% of forested sites occupied?
- If the model predicts sites within 1 km of water are twice as likely to be used as sites 10km from water, does this happen?

Discrimination: how well does the model discriminate between ‘good’ and ‘bad’ habitat?

Model Evaluation

Calibration: how well does the model describe the probability (or relative probability) of using different areas?

- If the model predicts that 10% of forested sites will be occupied, are 10% of forested sites occupied?
- If the model predicts sites within 1 km of water are twice as likely to be used as sites 10km from water, does this happen?

Discrimination: how well does the model discriminate between ‘good’ and ‘bad’ habitat?

- Given two different locations, one close to water and one far from water, can the model identify which location is most likely to be used?

Model Evaluation

Calibration: how well does the model describe the probability (or relative probability) of using different areas?

- If the model predicts that 10% of forested sites will be occupied, are 10% of forested sites occupied?
- If the model predicts sites within 1 km of water are twice as likely to be used as sites 10km from water, does this happen?

Discrimination: how well does the model discriminate between ‘good’ and ‘bad’ habitat?

- Given two different locations, one close to water and one far from water, can the model identify which location is most likely to be used?

Often calibration and discrimination go hand-in-hand, but that need not be the case.

Model Evaluation

Methods largely borrowed from binary regression modeling literature

Model Evaluation

Methods largely borrowed from binary regression modeling literature

Discrimination: often measured using AUC (area under the receiver operating curve)

Model Evaluation

Methods largely borrowed from binary regression modeling literature

Discrimination: often measured using AUC (area under the receiver operating curve)

Calibration: usually explored with calibration plots

- Compare observed (y) and predicted ($\hat{y}$) response data
- Use data splitting, cross-validation, bootstrapping, or data from a different study site (i.e., out of sample data)

Model Evaluation

Methods largely borrowed from binary regression modeling literature

Discrimination: often measured using AUC (area under the receiver operating curve)

Calibration: usually explored with calibration plots

- Compare observed (y) and predicted ($\hat{y}$) response data
- Use data splitting, cross-validation, bootstrapping, or data from a different study site (i.e., out of sample data)

This talk will focus on calibration methods.

Calibration plots: Logistic Regression

1. Fit logistic regression model to *training* data $(x_i^{train}, y_i^{train})$:
   $logit(\pi_i^{train}) = \log \frac{(\pi_i^{train})}{1-\pi_i^{train})} = x_i^{train}\beta^{train}$

Calibration plots: Logistic Regression

1. Fit logistic regression model to *training* data $(x_i^{train}, y_i^{train})$:
   $logit(\pi_i^{train}) = \log \frac{(\pi_i^{train})}{1-\pi_i^{train})} = x_i^{train}\beta^{train}$
2. Form predictions for *test* data (x_i^{test}, y_i^{test}) using the fit from [1]:
   $\hat{\pi}_i^{test} = \frac{e^{x_i^{test}\hat{\beta}^{train}}}{1+e^{x_i^{test}\hat{\beta}^{train}}}$

Calibration plots: Logistic Regression

1. Fit logistic regression model to *training* data $(x_i^{train}, y_i^{train})$:
   $logit(\pi_i^{train}) = \log \frac{(\pi_i^{train})}{1-\pi_i^{train})} = x_i^{train}\beta^{train}$
2. Form predictions for *test* data (x_i^{test}, y_i^{test}) using the fit from [1]:
   $\hat{\pi}_i^{test} = \frac{e^{x_i^{test}\hat{\beta}^{train}}}{1+e^{x_i^{test}\hat{\beta}^{train}}}$
3. Calibration plot:
   - Option 1: Bin data (e.g., based on quantiles of $\hat{\pi}_i$). Plot the proportion of values where $y_i^{test} = 1$ in each bin versus mean $\hat{\pi}_i^{train}$ in each bin.

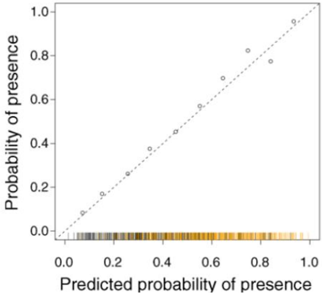
Calibration plots: Logistic Regression

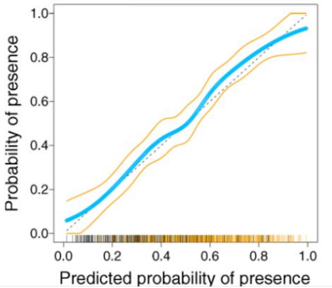
1. Fit logistic regression model to *training* data $(x_i^{train}, y_i^{train})$:
$logit(\pi_i^{train}) = \log \frac{(\pi_i^{train})}{1-\pi_i^{train})} = x_i^{train}\beta^{train}$
2. Form predictions for *test* data (x_i^{test}, y_i^{test}) using the fit from [1]:
$\hat{\pi}_i^{test} = \frac{e^{x_i^{test}\hat{\beta}^{train}}}{1+e^{x_i^{test}\hat{\beta}^{train}}}$
3. Calibration plot:
   - Option 1: Bin data (e.g., based on quantiles of $\hat{\pi}_i$). Plot the proportion of values where $y_i^{test} = 1$ in each bin versus mean $\hat{\pi}_i^{train}$ in each bin.
   - Option 2: Fit a new logistic regression model $logit(\pi_i^{test}) = b_0 + b_1(x_i^{test}\beta^{\hat{train}})$. $b_0 = 0, b_1 = 1$ indicates perfect calibration.

Calibration plots: Logistic Regression

1. Fit logistic regression model to *training* data $(x_i^{train}, y_i^{train})$:
   $logit(\pi_i^{train}) = \log \frac{(\pi_i^{train})}{1-\pi_i^{train})} = x_i^{train}\beta^{train}$
2. Form predictions for *test* data (x_i^{test}, y_i^{test}) using the fit from [1]:
   $\hat{\pi}_i^{test} = \frac{e^{x_i^{test}\hat{\beta}^{train}}}{1+e^{x_i^{test}\hat{\beta}^{train}}}$
3. Calibration plot:
   - Option 1: Bin data (e.g., based on quantiles of $\hat{\pi}_i$). Plot the proportion of values where $y_i^{test} = 1$ in each bin versus mean $\hat{\pi}_i^{train}$ in each bin.
   - Option 2: Fit a new logistic regression model $logit(\pi_i^{test}) = b_0 + b_1(x_i^{test}\beta^{\hat{train}})$. $b_0 = 0, b_1 = 1$ indicates perfect calibration.
   - Option 3: Fit a more flexible, non-linear model: $logit(\pi_i^{test}) = f(x_i^{test}\beta^{\hat{train}})$

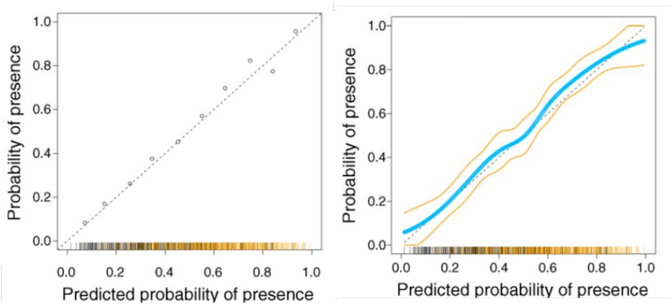
Calibration plots: Logistic Regression





Phillips, S. J., & Elith, J. (2010). POC plots: calibrating species distribution models with presence-only data. Ecology, 91(8), 2476-2484.

Calibration plots: Logistic Regression



Phillips, S. J., & Elith, J. (2010). POC plots: calibrating species distribution models with presence-only data. Ecology, 91(8), 2476-2484.

Adapted to presence-only data, transforming the y-axis since some of the 0's might actually be used.

The Boyce Cross-validation Method

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K. (2002). Evaluating resource selection functions. Ecol. Model., 157, 281–300.

The Boyce Cross-validation Method

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K. (2002). Evaluating resource selection functions. Ecol. Model., 157, 281–300.

1. Generate predicted suitability scores, $w(x_i^{test}) = \exp(x_i^{train}\beta)$ for the used and observed locations

The Boyce Cross-validation Method

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K. (2002). Evaluating resource selection functions. Ecol. Model., 157, 281–300.

1. Generate predicted suitability scores, $w(x_i^{test}) = \exp(x_i^{train}\beta)$ for the used and observed locations
2. Bin the suitability scores (e.g., into deciles)

The Boyce Cross-validation Method

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K. (2002). Evaluating resource selection functions. Ecol. Model., 157, 281–300.

1. Generate predicted suitability scores, $w(x_i^{test}) = \exp(x_i^{train}\beta)$ for the used and observed locations
2. Bin the suitability scores (e.g., into deciles)
3. Count the number (or proportion) of used locations within each bin

The Boyce Cross-validation Method

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K. (2002). Evaluating resource selection functions. Ecol. Model., 157, 281–300.

1. Generate predicted suitability scores, $w(x_i^{test}) = \exp(x_i^{train}\beta)$ for the used and observed locations
2. Bin the suitability scores (e.g., into deciles)
3. Count the number (or proportion) of used locations within each bin
4. Calculate Spearman correlation (bin rank, number of observations in a bin)

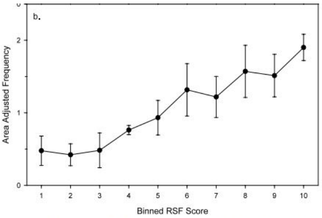


Fig. 4. Area-adjusted frequency of binned cross-validated use locations for fall (late-hyperphagia) RSF models in the Greater Yellowstone Ecosystem, USA. Frequency values for individual cross-validation sets ($n = 5$) are depicted with unique symbols (graph a). Mean ($\pm$S.D.) frequency values by bin are illustrated in graph b. A Spearman-rank correlation for mean frequency values by bins ($r_s = 0.972$, $P < 0.001$) indicates that the model predicted cross-validated use locations well.

Modified Boyce Method

Johnson, C.J., Nielsen, S.E., Merrill, E.H., McDonald, T.L. & Boyce, M.S. (2006). Resource selection functions based on use-availability data: Theoretical motivation and evaluation methods. J. Wildlife Manage., 70, 347–357.

Modified Boyce Method

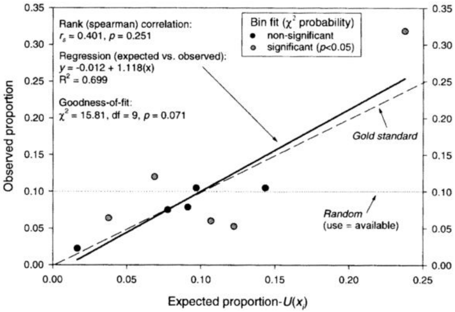
Johnson, C.J., Nielsen, S.E., Merrill, E.H., McDonald, T.L. & Boyce, M.S. (2006). Resource selection functions based on use-availability data: Theoretical motivation and evaluation methods. J. Wildlife Manage., 70, 347–357.

Use $\pi_i^{test} = \frac{\exp(x_i^{train}\beta^{train})}{\sum_{j=1}^{n_{test}} \exp(x_j^{train}\beta^{train})}$ to estimate the expected number of observations within each bin.

Modified Boyce Method

Johnson, C.J., Nielsen, S.E., Merrill, E.H., McDonald, T.L. & Boyce, M.S. (2006). Resource selection functions based on use-availability data: Theoretical motivation and evaluation methods. J. Wildlife Manage., 70, 347–357.

Use $\pi_i^{test} = \frac{\exp(x_i^{train}\beta^{train})}{\sum_{j=1}^{n_{test}}\exp(x_j^{train}\beta^{train})}$ to estimate the expected number of observations within each bin.



But what about?

- SSFs where availability changes with each used point?

But what about?

- SSFs where availability changes with each used point?
- When models are not well-calibrated? How do we gain insights into *why*?

Used-habitat calibration plots (UHC plots)

ECOGRAPHY

Research

Used-habitat calibration plots: a new procedure for validating species distribution, resource selection, and step-selection models

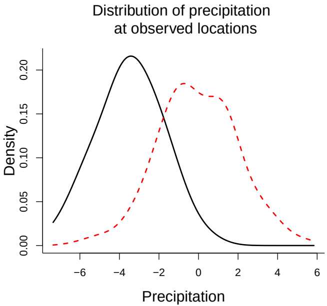
John R. Fieberg, James D. Forester, Garrett M. Street, Douglas H. Johnson, Althea A. ArchMiller and Jason Matthiopoulos

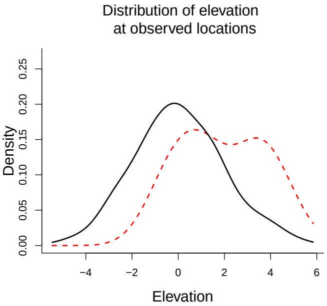
Focus on predicting the characteristics of the used locations in out-of-sample data

- Treats the environmental variables, x, as random (rather than the y's)
- Easily generalizes to step-selection functions
- Can compliment existing approaches for model evaluation

Producing an UHC Plot

Step 0: Split the data into test and training data sets.
Step 1: Summarize the distribution of the environmental variables at the available and **used** locations.





Red = available distribution **Black** = used distribution

Producing an UHC Plot

Step 2: Fit a model to the training data set

- Logistic regression (for resource-selection functions)
- Conditional logistic regression (for step-selection functions)

Store $\hat{\beta}$ and its uncertainty ($\widehat{cov}(\hat{\beta})$).

Producing an UHC Plot

Step 2: Fit a model to the training data set

- Logistic regression (for resource-selection functions)
- Conditional logistic regression (for step-selection functions)

Store $\hat{\beta}$ and its uncertainty ($\widehat{cov}(\hat{\beta})$).

Can capture uncertainty using a bootstrap or posterior distribution.

Step 3: Repeat Steps M Times

A. Draw random values of β (to represent our uncertainty), $\tilde{\beta}_i$

Step 3: Repeat Steps M Times

A. Draw random values of β (to represent our uncertainty), $\tilde{\beta}_i$
B. Estimate relative probability of selecting points in test data,
$\tilde{w}_i = \exp(x^{test}\tilde{\beta}_i)$

Step 3: Repeat Steps M Times

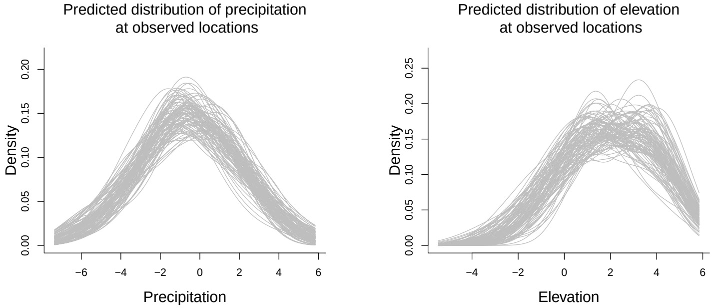
A. Draw random values of β (to represent our uncertainty), $\tilde{\beta}_i$
B. Estimate relative probability of selecting points in test data, $\tilde{w}_i = \exp(x^{test}\tilde{\beta}_i)$
C. Select n_u^{test} used locations from test data set with probability proportional to $\tilde{w}_i$

Step 3: Repeat Steps M Times

A. Draw random values of β (to represent our uncertainty), $\tilde{\beta}_i$
B. Estimate relative probability of selecting points in test data, $\tilde{w}_i = \exp(x^{test}\tilde{\beta}_i)$
C. Select n_u^{test} used locations from test data set with probability proportional to $\tilde{w}_i$
D. Summarize predicted distribution of x^{test} at chosen locations.

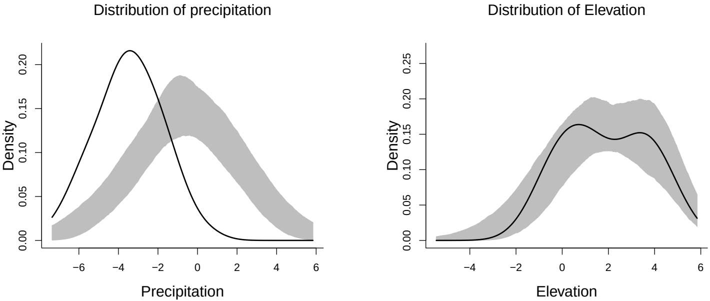
Step 3: Repeat Steps M Times

A. Draw random values of β (to represent our uncertainty), $\tilde{\beta}_i$
B. Estimate relative probability of selecting points in test data, $\tilde{w}_i = \exp(x^{test}\tilde{\beta}_i)$
C. Select n_u^{test} used locations from test data set with probability proportional to $\tilde{w}_i$
D. Summarize predicted distribution of x^{test} at chosen locations.



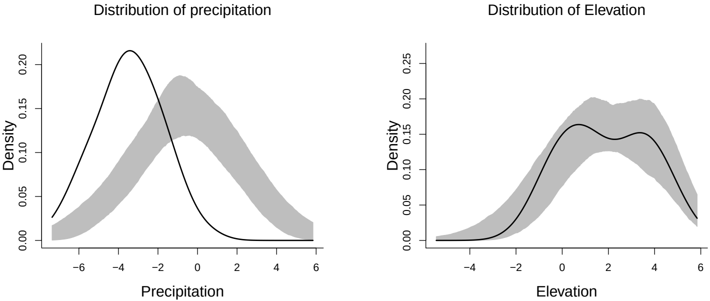
Producing an UHC Plot

Step 4: Compare observed (black) and predicted (gray) distributions



Producing an UHC Plot

Step 4: Compare observed (black) and predicted (gray) distributions

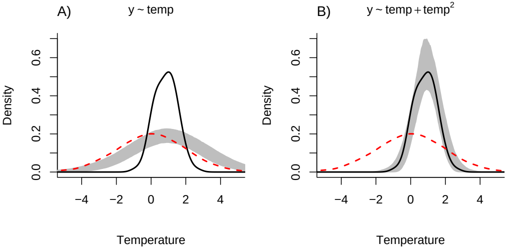


Similar to a posterior predictive check to see if the model can produce data like those that were observed.

Simulation Example: Non-linear relationship

Species distribution driven by temperature (x_3)

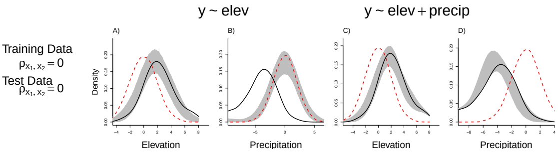
- Probability of use proportional to $\exp(2x_3 - x_3^2)$.
- Fit models: y$\sim$ temp (incorrect) and y$\sim$ temp + temp^2 (correct)



Red = available distribution **Black** = used distribution

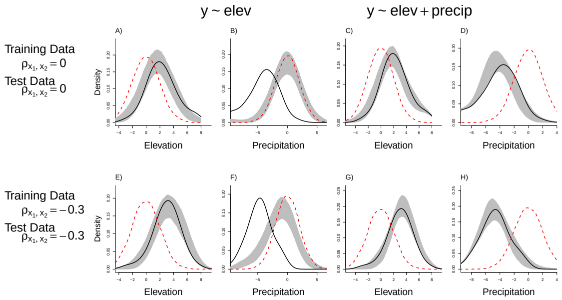
Simulation Example: Missing predictor

- Probability of use proportional to $\exp(0.5x_1 - x_2)$, with $(x_1, x_2) =$ (elevation, precipitation).
- Fit models: $y \sim$ elev (left two columns) and $y \sim$ elev + precip (right two columns)



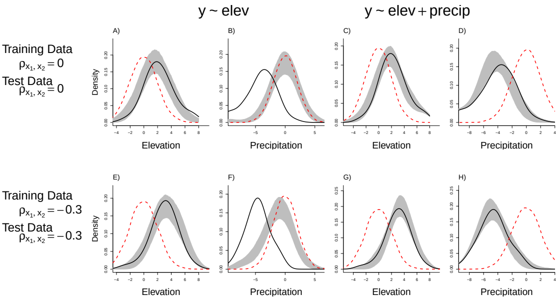
Simulation Example: Missing predictor

- Probability of use proportional to $\exp(0.5x_1 - x_2)$, with $(x_1, x_2) =$ (elevation, precipitation).
- Fit models: $y \sim$ elev (left two columns) and $y \sim$ elev + precip (right two columns)



Simulation Example: Missing predictor

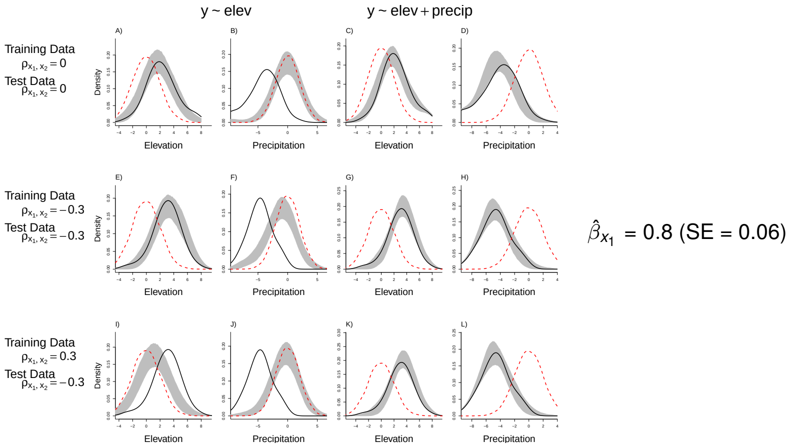
- Probability of use proportional to $\exp(0.5x_1 - x_2)$, with (x_1, x_2) = (elevation, precipitation).
- Fit models: $y \sim$ elev (left two columns) and $y \sim$ elev + precip (right two columns)



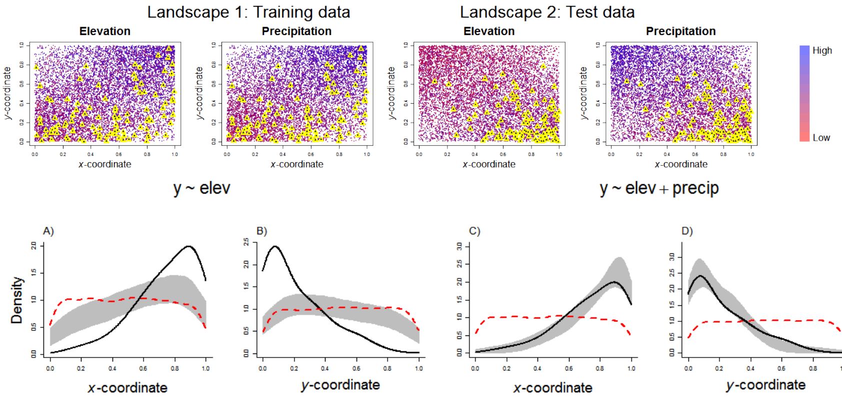
$\hat{\beta}_{x_1}$ = 0.8 (SE = 0.06)

Simulation Example: Missing predictor

- Probability of use proportional to $\exp(0.5x_1 - x_2)$, with $(x_1, x_2) =$ (elevation, precipitation).
- Fit models: $y \sim$ elev (left two columns) and $y \sim$ elev + precip (right two columns)



Simulation Example: Spatial Coordinates



Can use this approach to explore accuracy of predictions in space.

Summary

Used-habitat calibration (UHC) plots are simple, graphical methods that compare distributions of resources at:

- available locations
- observed locations (training data)
- locations predicted to be used (test data)

Easily adapted to any model that can *rank* observations in terms of predicted probability of use

Brian will demonstrate how to calculate uhc plots using `amt`.