Intrinsic sources of error in habitat models: Rounding error and variation in population densities

William B. Kristan and J. Michael Scott, Idaho Cooperative Fish and Wildlife Research Unit, University of Idaho, Moscow

INTRODUCTION

Habitat models predict the presence or absence of animals based on functional relationships between animal occurrences and habitat characteristics

Common techniques, such as logistic regression, predict probabilities of occurrence. These predictions are converted to 'presence' or 'absence' by rounding probabilities below 0.5 to 'absent' and above 0.5 as 'present'

Thus, up to 50% of some habitats that have a predicted 'presence' should actually not contain the animal (and will be considered 'commission errors'), and animals should be observed in up to 50% of some habitats that have a predicted 'absence' (and will be considered 'omission errors')

Rather than indicating a poor model, these errors represent intrinsic rounding error, and can mask the effectiveness of a model

Additionally, a logistic regression model fitted to data from one population can predict occupancy in another population poorly if the population densities differ, even if the same ecological relationship holds in both populations.

In this investigation, we explored the effects of three factors on prediction accuracy:
a) differences in animal density, b) the strength of the ecological relationship represented by the model, c) the structure of the landscape to which the models are applied

Density-dependent habitat use is common in animal populations – territorial behavior forces subordinate individuals to use sub-optimal habitat when densities are high, such that the range of habitats used is greater at high densities than at low densities

The strength of the ecological relationship between animals and habitat is represented here by the steepness of the curve.

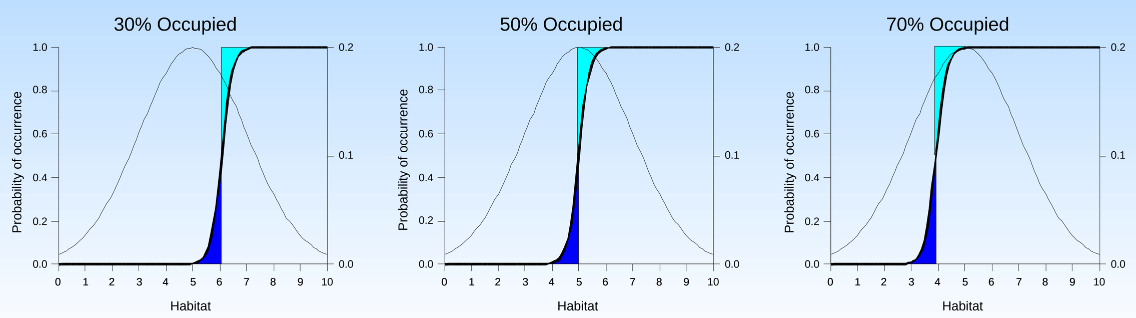
The steepness does not change as the percent of the habitat occupied changes, but the curve is shifted to the left, representing use of less-preferred habitat at higher population densities.

This source of error can be corrected by adjusting one parameter (the "intercept" in the logistic equation) while leaving the shape (the "slope") the same.

Model predictions depend on the population density

Same relationship at three different densities

Omission = commission = 2%



Predictions from models developed with 50% occupancy

Omission = 1%, commission = 20%

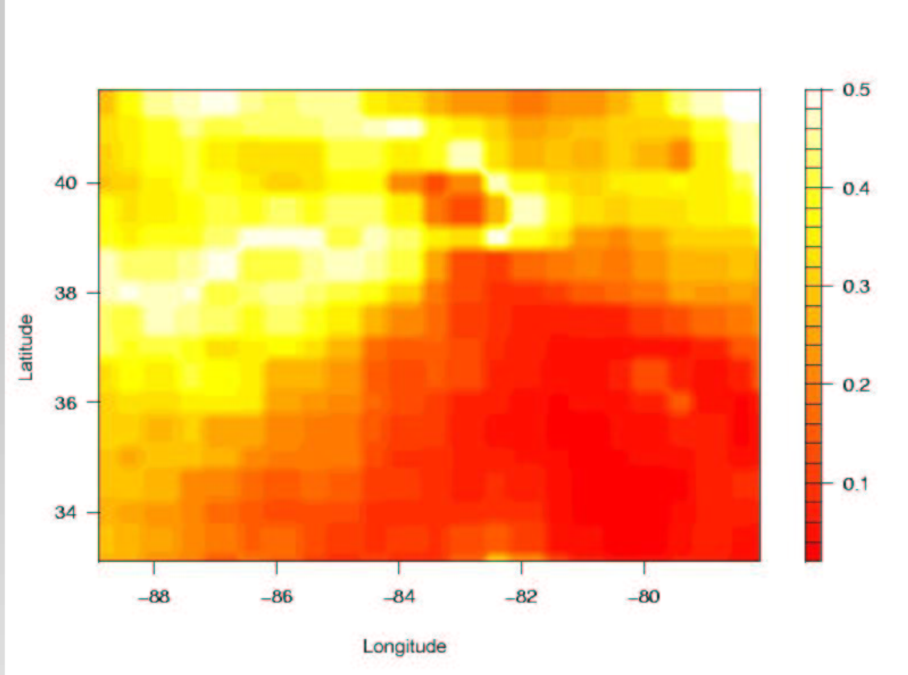
Omission = 20%, commission = 1%

Deriving a model from a population with 50% of the habitat occupied, then applying it to a population with 30% of the habitat occupied introduces a large number of predicted presences that have no animals

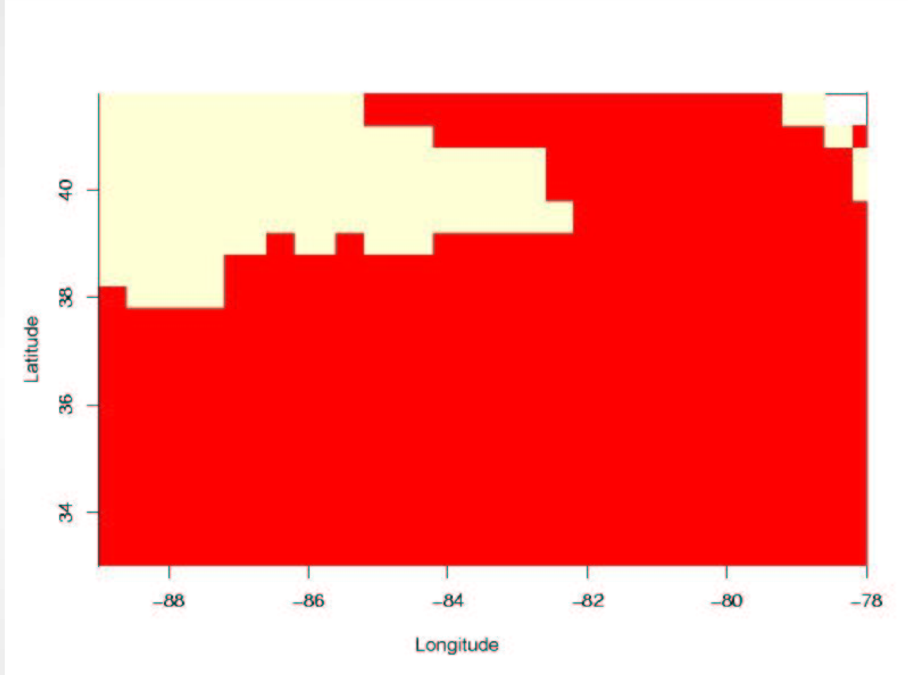
Deriving a model from a population with 50% of the habitat occupied, then applying it to a population with 70% of the habitat occupied introduces a large number of predicted absences that contain animals

A mapped example

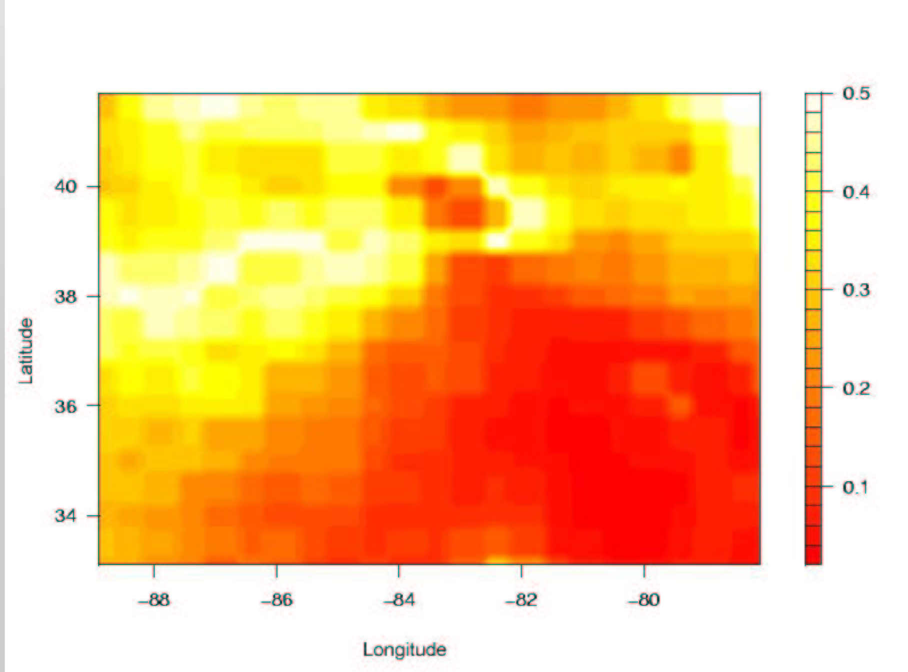
The predicted probability of occurrence of an animal



Predicted probabilities converted to presence (light) and absence (red)



Expected prediction error (|second image-first image|)



APPLICATIONS

1. Logistic models can be calibrated to the local population density prior to assessing model accuracy

2. The expected predictive accuracy of different models can be assessed and compared based only on the model and the map, and the best model for the particular application can be chosen

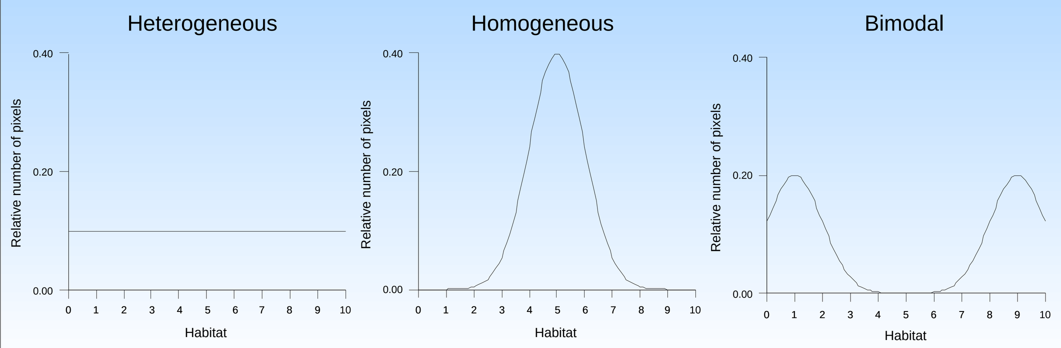
3. Model prediction errors can be compared to the expected error for a particular habitat, so that the source of errors can be better understood – that is, cases of models that poorly represent the habitat associations of an animal can be distinguished from a cases of landscapes having a large amount of habitat with an intermediate probability of use

Hypothetical distributions of the habitat variable
The heterogeneous landscape has all habitats equally represented
The homogeneous landscape has most of the habitat clustered around one value
Bimodal landscape is either at large or small values, but with little habitat near the middle

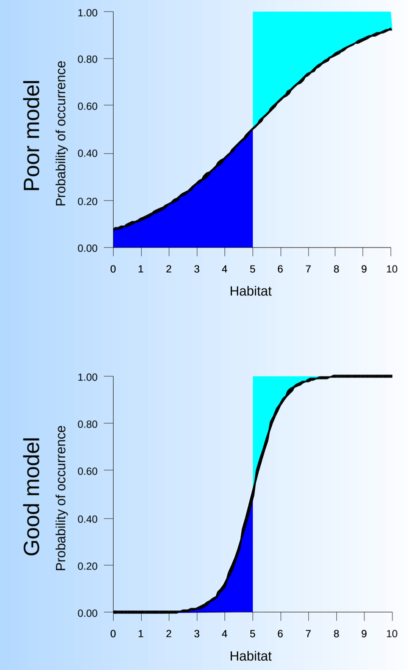
A model with a gradual transition between low and high values indicates a weak ecological relationship between the animal and the habitat variable – a poor model

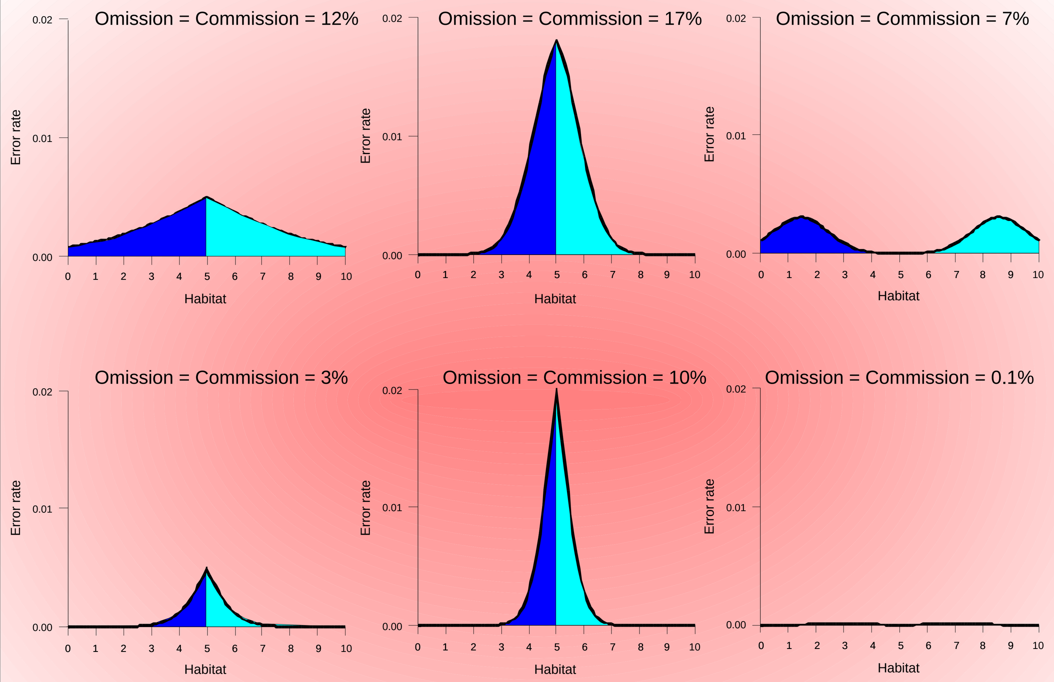
A model with a sharp transition between low and high values indicates a strong ecological relationship between the animal and the habitat variable – a good model

Structure of the habitat



Strength of the ecological relationship





Expected error from applying the models to these landscapes

The rounding error introduced by applying the model to a landscape depends both on the model (whether it is a 'good' or 'poor' model), and on the structure of the habitat.

A landscape with a large amount of habitat at the part of the curve with the greatest uncertainty have the highest error due to rounding error – for each model, the greatest amount of error occurs with homogeneous habitat

If these models represent two different habitat variables, the 'poor' model can predict a species better than the 'good' model, if little habitat falls where the model performs poorly (e.g. compare the error rate of the poor model applied to the bimodal habitat variable with the good model applied to the homogeneous habitat variable)

CONCLUSIONS

Rates of omission and commission error are uninformative without information about intrinsic sources of error.

The combination of the habitat model and landscape structure determine the magnitude of intrinsic error, which can be measured independently of animal occurrence.

Differences in population density between the population from which the model was derived and to which the model is applied can result in large error rates.

Assuming that the steepness of a logistic function represents the ecological relationship, and the position along the habitat axis is determined by density, it is simple to adjust models to the proportion of habitats occupied before predictions are made.

It is possible to use this approach to not only improve prediction accuracy, but to test for density-dependent habitat use in a species – thus, habitat model predictions can be used as an investigative tool as well as a planning tool.