Spatial stream network modeling of trout occurrence in the upper Snake River

Cody J. Feldman, Ernest R. Keeley, and Charles R. Peterson

Department of Biological Sciences, Idaho State University

Introduction

Goal

- The goal of this project is to compare modeling approaches that account for spatial autocorrelation with statistical models over a large riverine system to identify probability of occurrence of salmonid fish species including brook trout, rainbow trout, cutthroat trout, and brown trout.
- This project is part of Genes by Environment: Modeling, Mechanisms, and Mapping (GEM3) study, NSF Idaho EPSCoR project OIA-1757324.

Questions

- What available range-wide environmental variables are useful in predicting the distribution of salmonids?
- Can presence/absence models be improved by models that account for spatial autocorrelation?

Approach

- We acquired presence/absence records from Idaho Department of Fish and Game to assess remotely sensed environmental data and its usefulness for predicting salmonid presence/absence.
- We acquired presence/absence records from Idaho Department of Environmental Quality to test how well models performed.
- We used 11 environmental variables to create spatially explicit and non-spatial models to compare accuracy in predicting trout presence and absence.

Importance

- Widespread loss of habitat is a key factor leading to species declines in many ecosystems worldwide
- Trout, like many species, have been affected by human activities.
- It is important to understand the change in their current habitat range to illustrate how these losses will affect fish populations.

Study Area

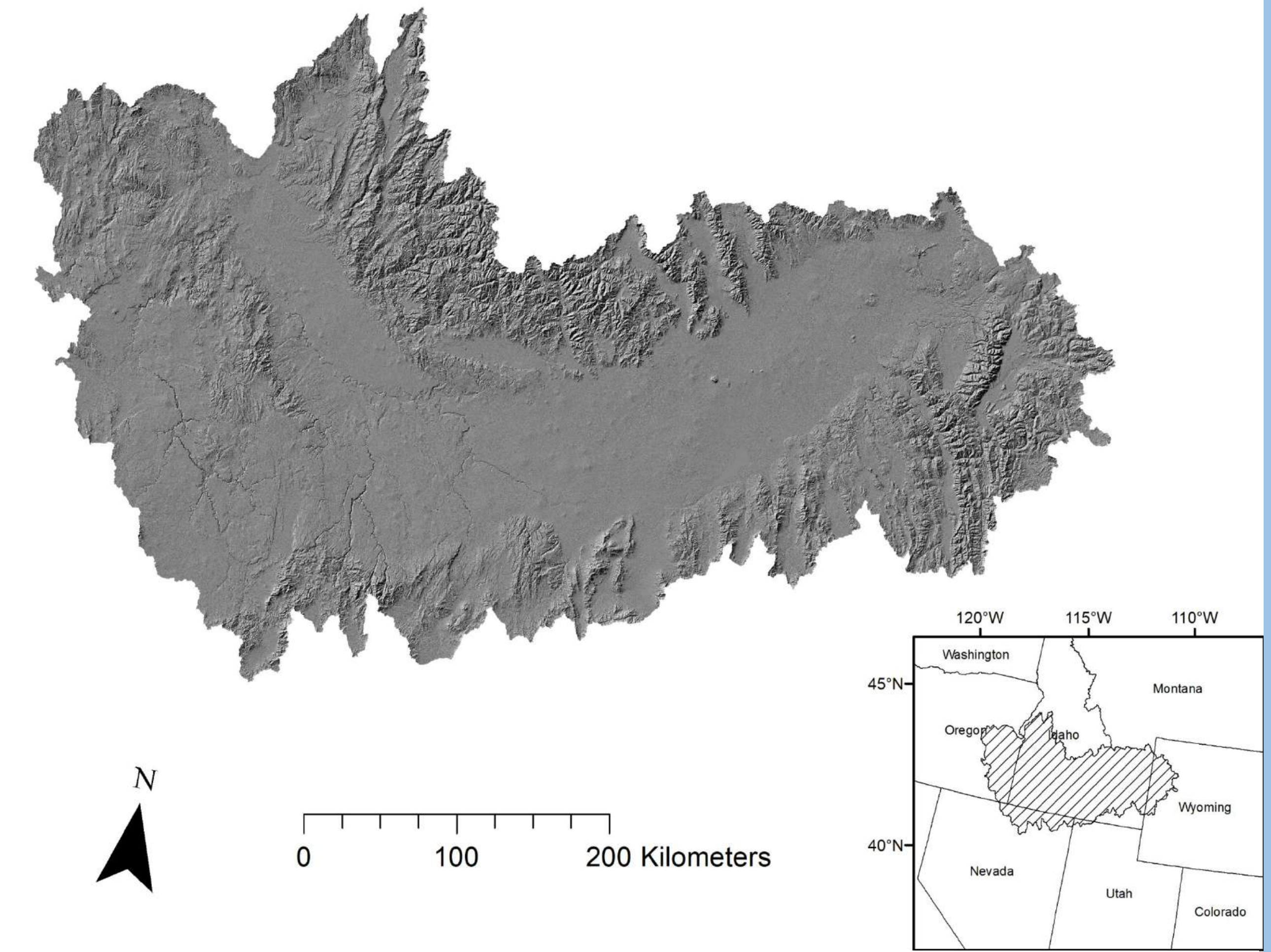


Figure 1. The Middle Snake-Boise, Upper Snake and Snake Headwater watersheds.

- The study area is comprised of 3 HUC 6 watersheds: Middle Snake-Boise, Upper Snake, and Snake Headwaters (figure 1).
- It has an area of 178,060 km².
- The watersheds range in elevation from 4,094 meters to 626 meters.

Methods

- We used records from 1944 sampling sites provided by the Idaho Department of Fish and Game (figure 2)
- Records were imported into ArcMap 10.5
- We identified 11 environmental variables (stream temperature, stream order, landcover, canopy cover, human footprint index, NDVI, precipitation, stream slope, isothermality, elevation, and water vapor pressure) through literature review that may identify salmonid habitat and is representable in a GIS (geographic information system) (figure 4)
- Using the records and environmental data we used an ArcMap toolbox Spatial Tools for the Analysis of River Systems (STARS) and a package for R statistical software Spatial Stream Network (SSN) to fit non-spatial and spatially explicit models to a stream network.
- We then used 3506 records from the Idaho Department of Environmental Quality (figure 3) to assess how models performed.

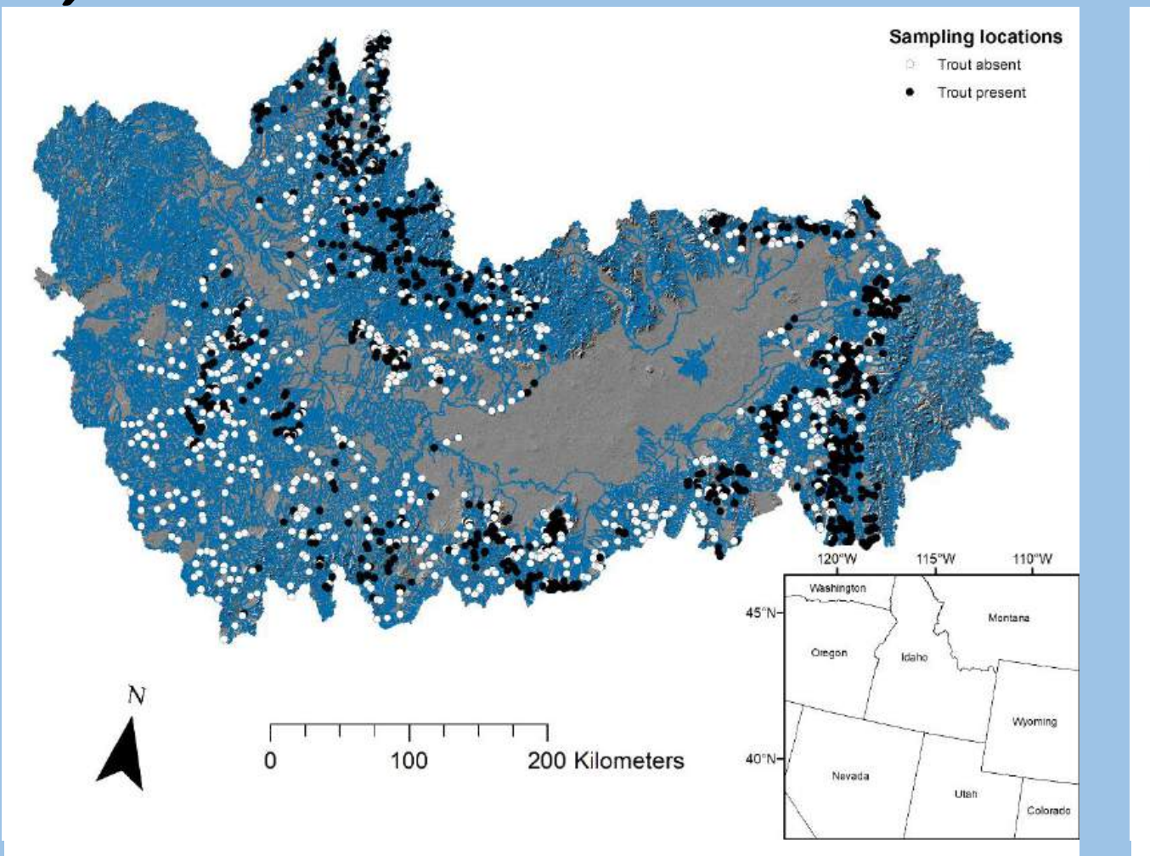


Figure 2. Distribution of sampling sites (n=1944) used for model creation.

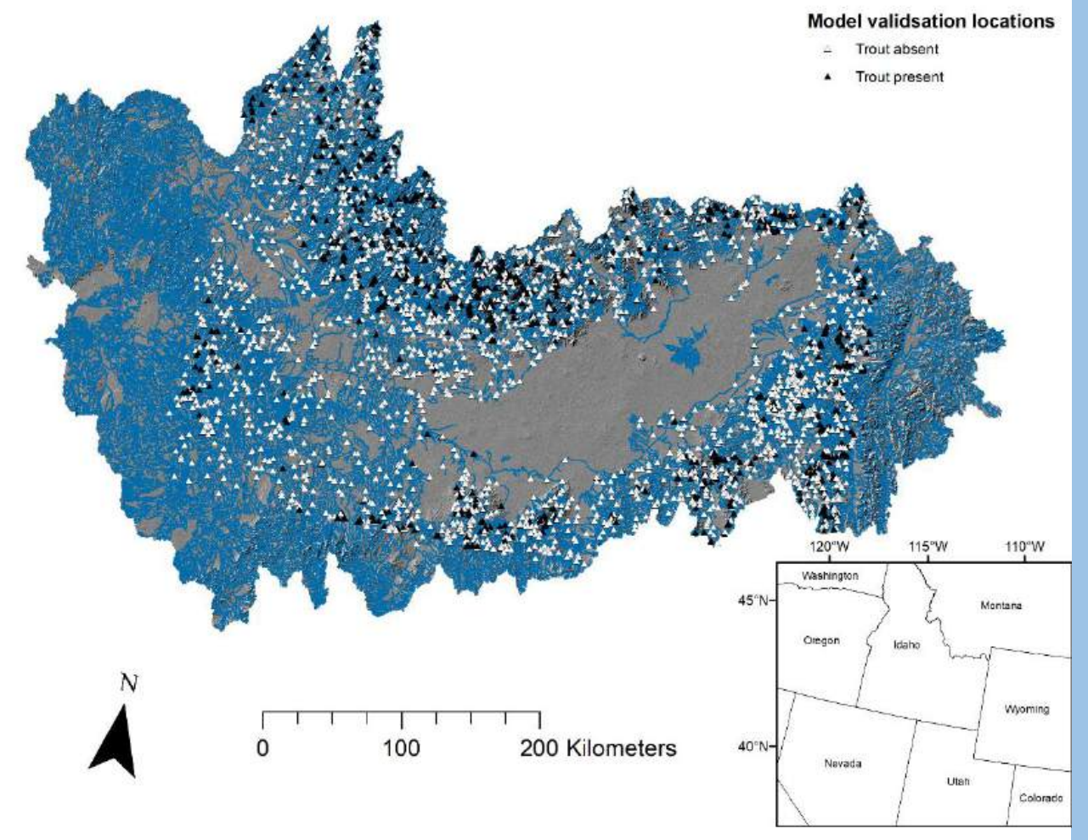


Figure 3. Distribution of sites (n=3506) used for model validation.

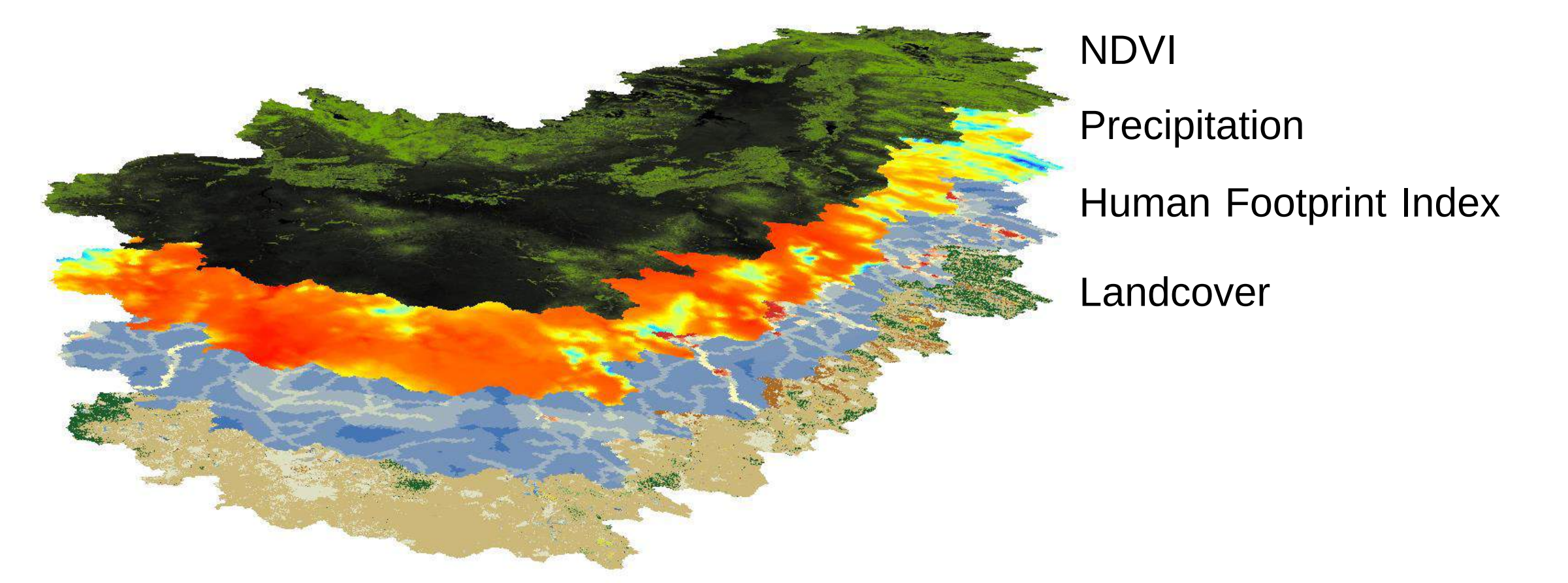


Figure 4. Four of 12 environmental variables used for model creation.

Results

- The multivariate logistic regression model (figure 5) had significant values for 8 of the environmental variables (stream order, landcover, canopy cover, human footprint index, water vapor pressure, NDVI, precipitation, and stream slope). This model produced an area under the curve (AUC) value of 0.651 (figure 6a). An error matrix was created to illustrate type I and II errors (figure 6b) with a true positive rate of 0.694 and a true negative rate of 0.489. Overall accuracy of the non-spatial model was 0.636. Few of the presence and absence points were present on the 90-100% and 0-10% probabilities, respectively (figure 6c). There majority of stream kilometers fell within the 10-50% range of probabilities (figure 6d)
- The Null model (figure 7) slightly improved AUC values (figure 8a), true positive rates, true negative rates, and accuracy (figure 8b). Presence and absence points no longer fell within the 0-10% and 90-100% probabilities (figure 8c) but stream kilometers (figure 8d) are distributed more evenly throughout the predicted probabilities.
- The spatially explicit multivariate model (figure 9) used the same environmental variables as the non-spatial model (stream order, landcover, canopy cover, human footprint index, water vapor pressure, NDVI, precipitation, and stream slope). This model further improves on AUC values (figure 10a), true positive rate, true negative rate, and accuracy (figure 10b). The presence and absence points (figure 10c) shows decreases falling within the 40-50% probabilities and an expansion to the extreme ends of the probabilities. The stream kilometers (figure 10d) show a similar pattern as the Null model but with an expansion of the stream kilometers that fall within the 0-10% probabilities.

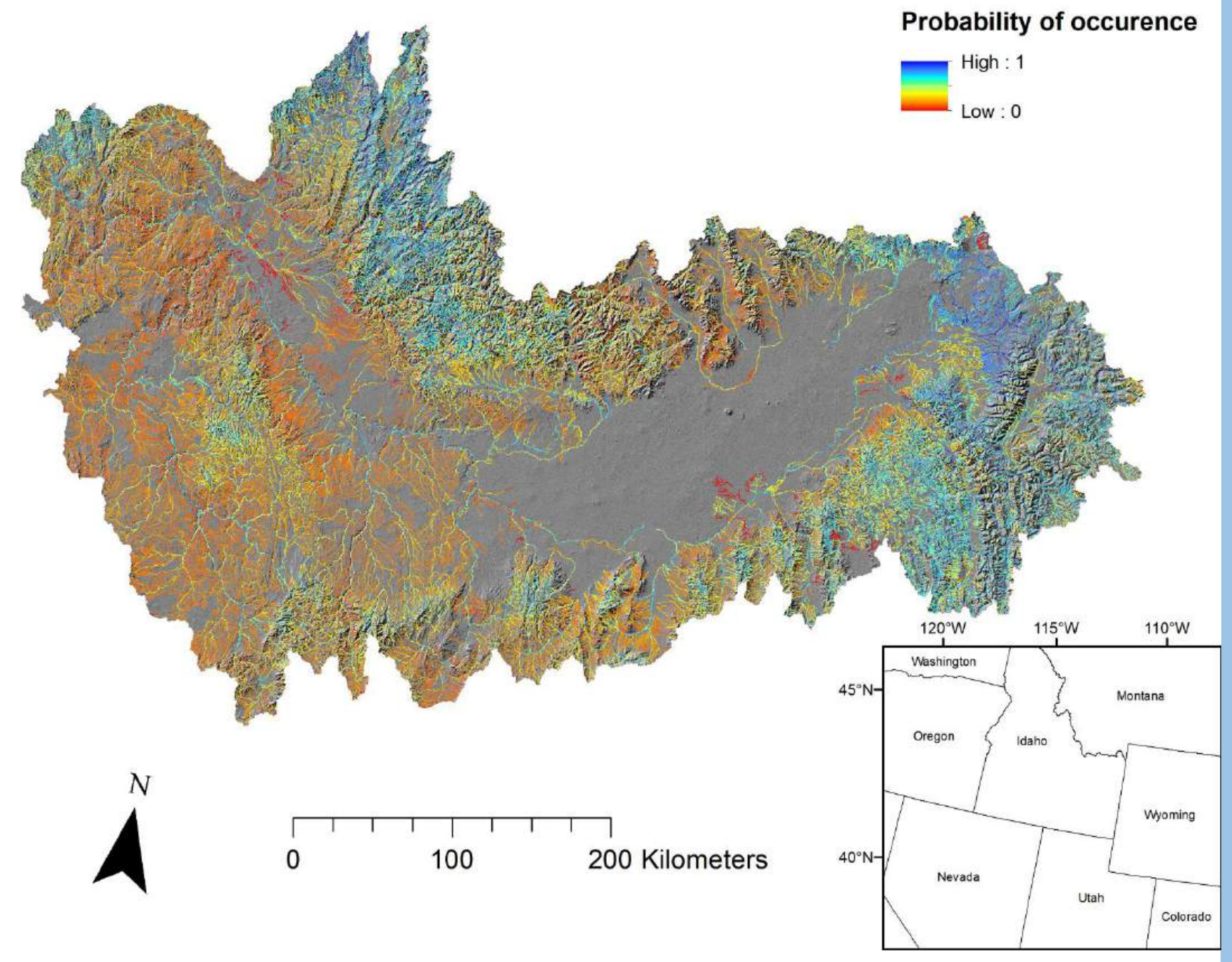


Figure 5. Probability of occurrence of a multivariate logistic regression model to evaluate trout presence/absence

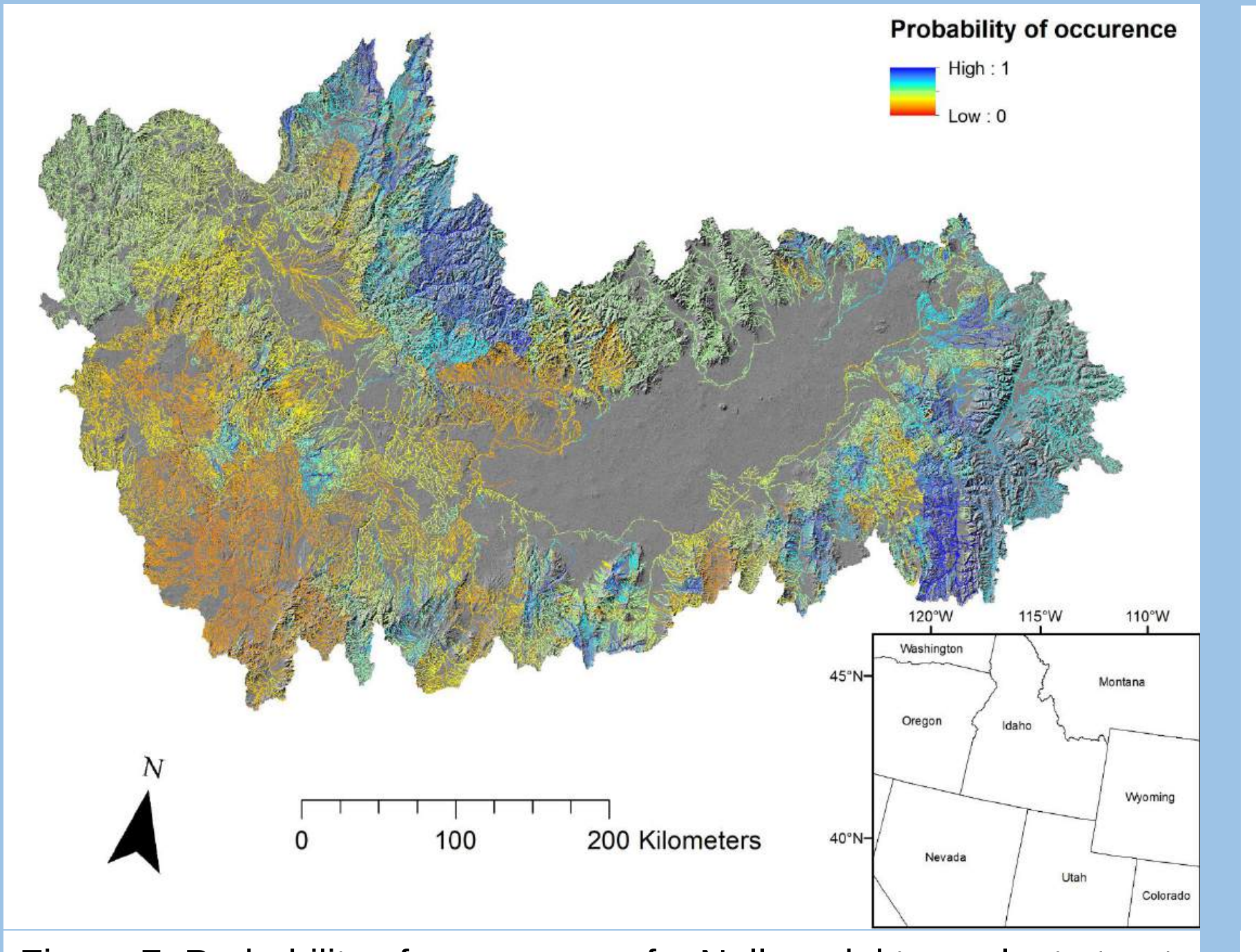


Figure 7. Probability of occurrence of a Null model to evaluate trout presence/absence

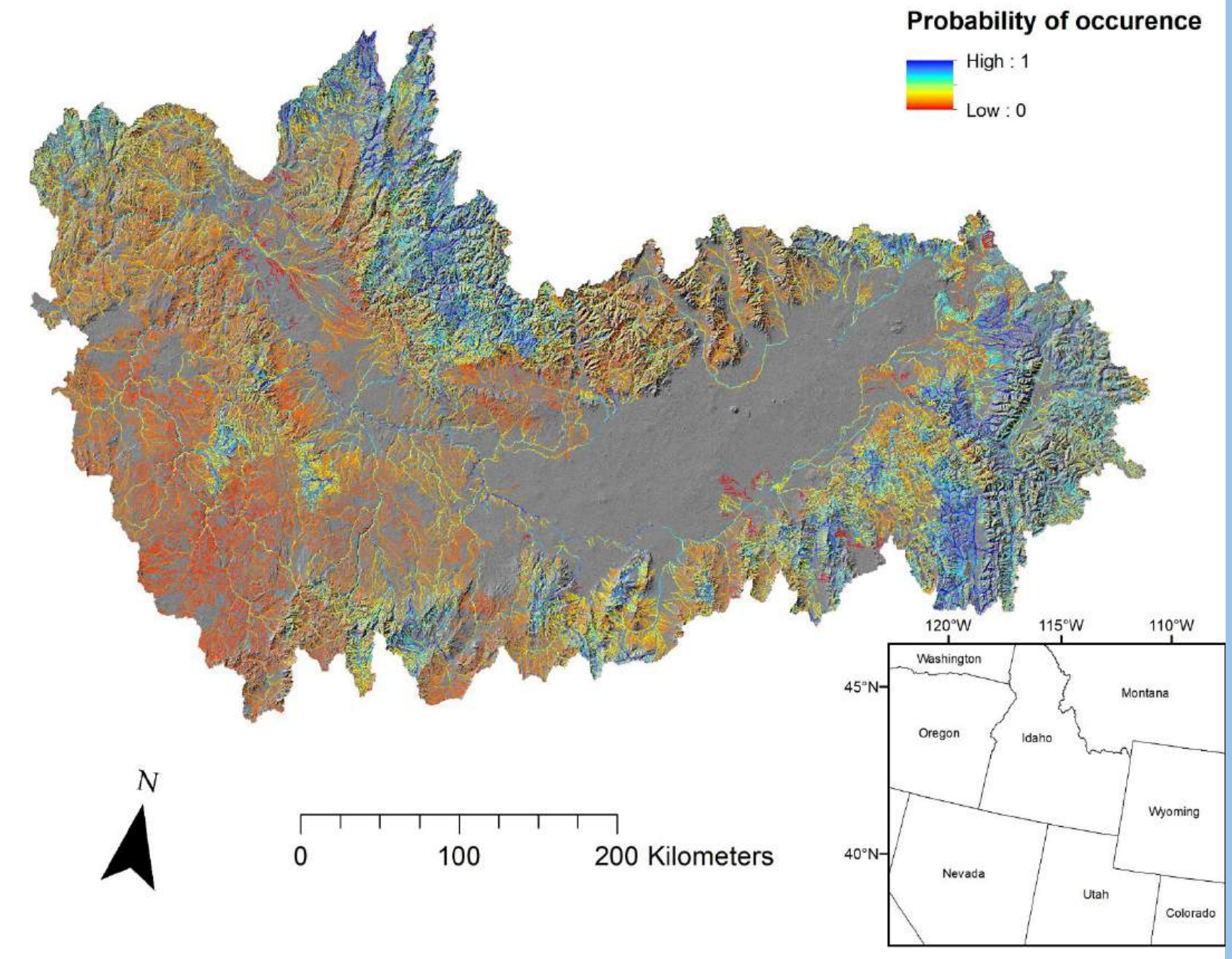


Figure 9. Probability of occurrence of a spatial multivariate logistic regression model to evaluate trout presence/absence

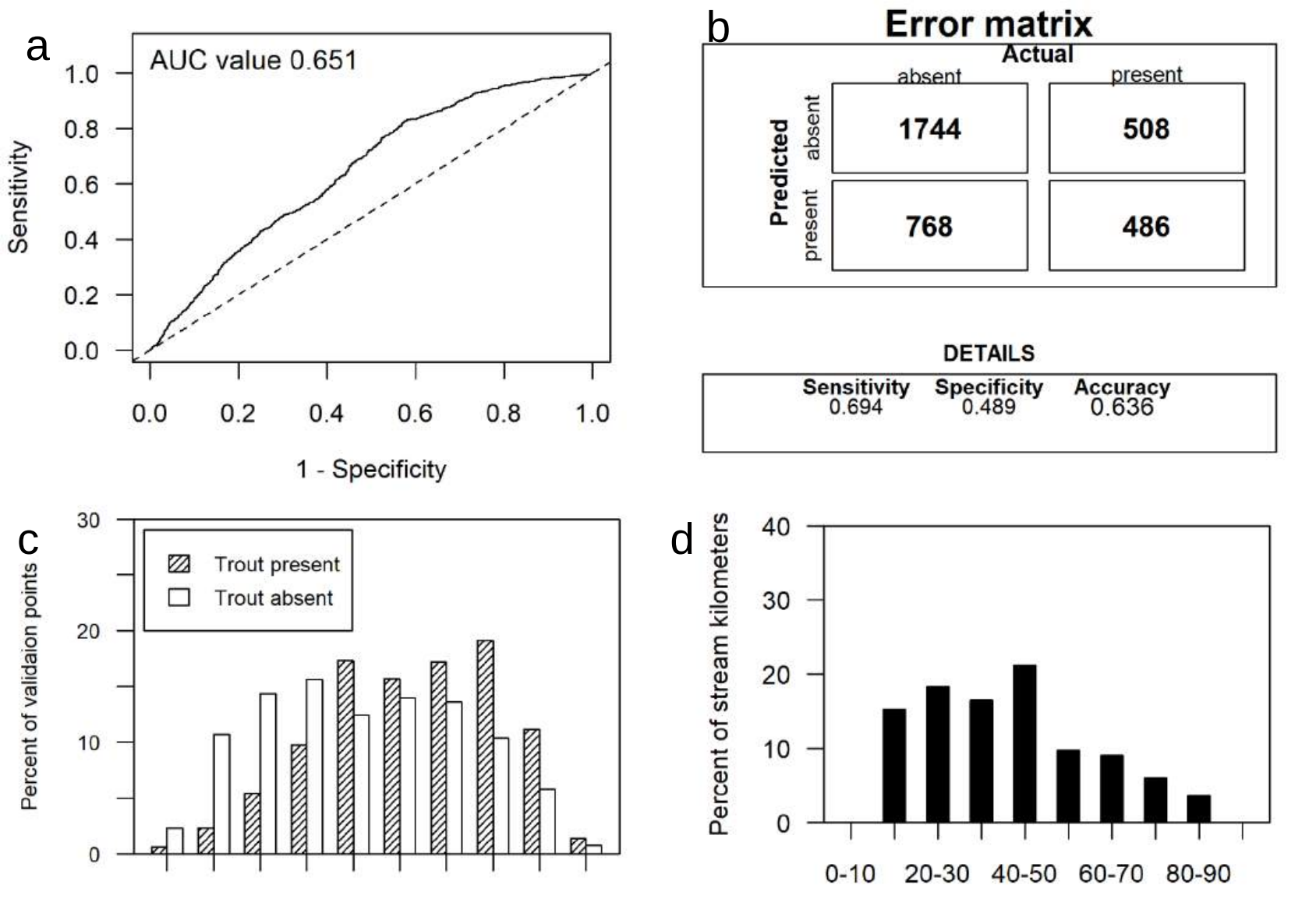


Figure 6. Predictive assessment of a multivariate logistic regression model

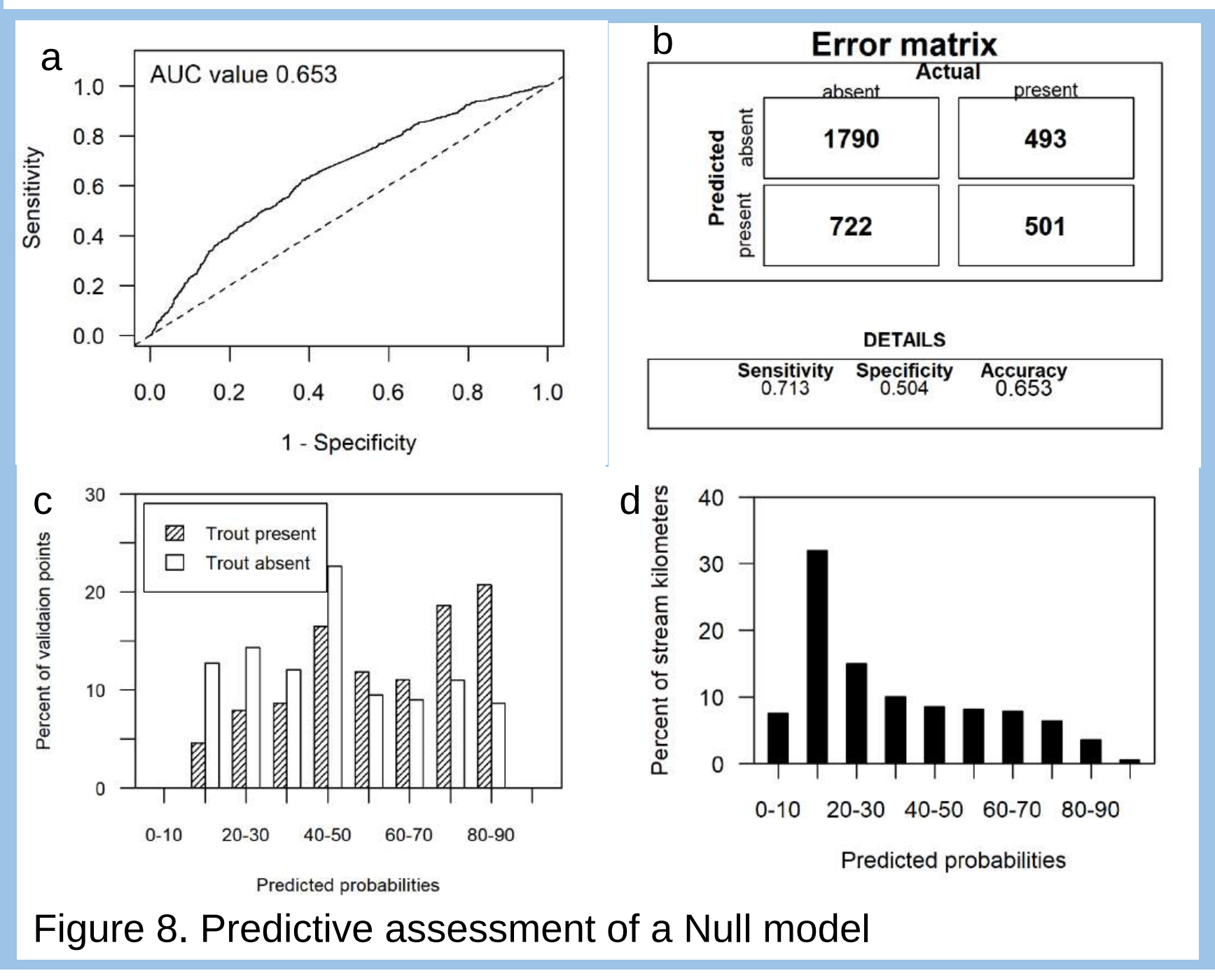


Figure 8. Predictive assessment of a Null model

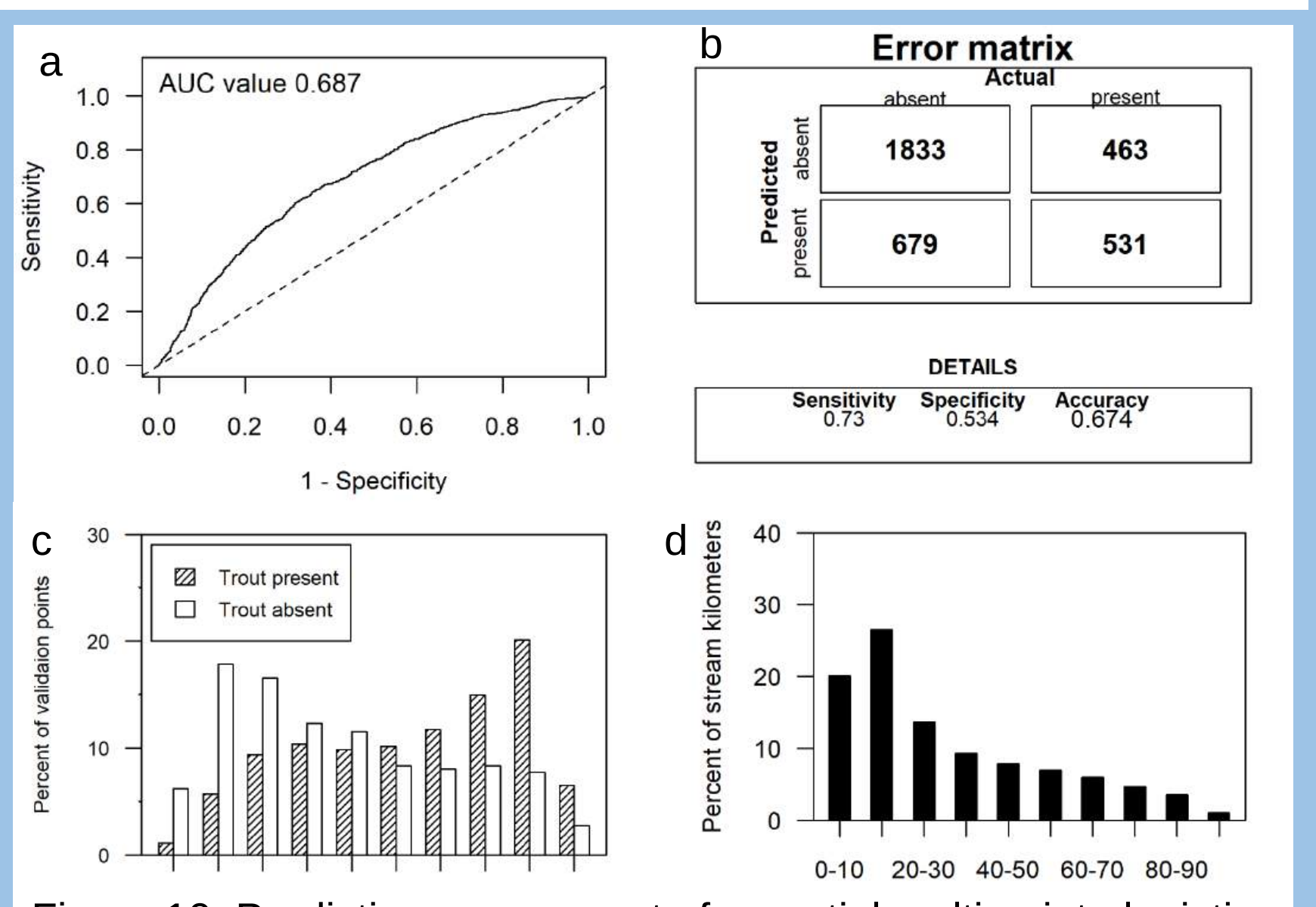
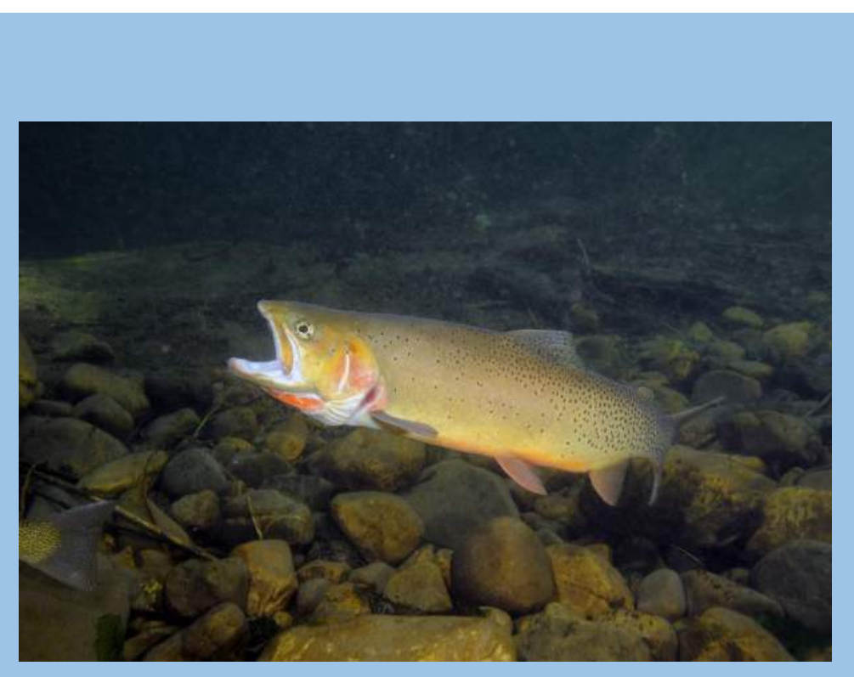
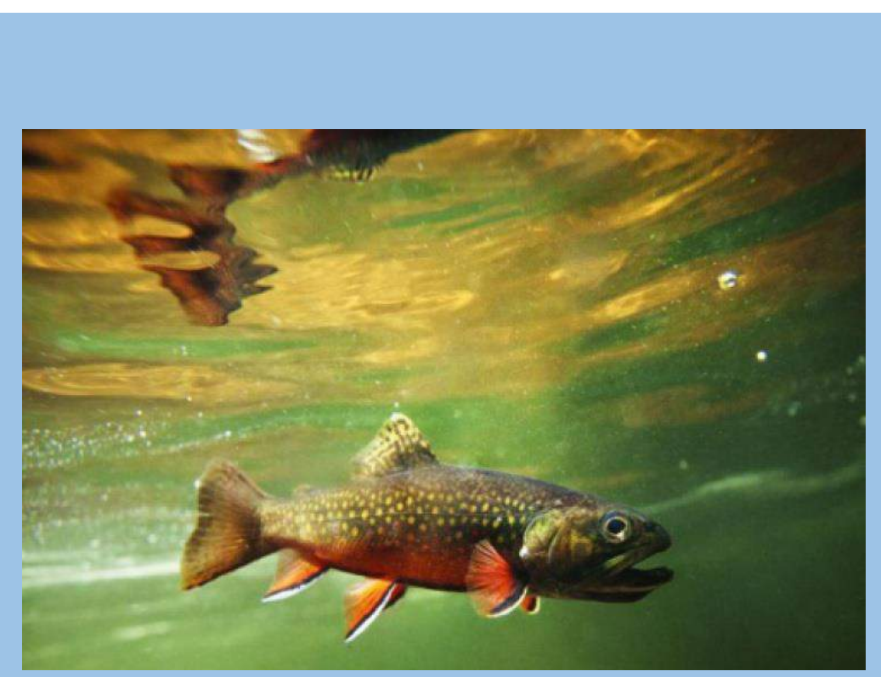


Figure 10. Predictive assessment of a spatial multivariate logistic regression model





Summary

Spatial autocorrelation of sampling locations is an important aspect of modeling species distributions. These models illustrate that modeling stream species may benefit from including spatial autocorrelation with environmental variables allowing conservationists and managers to prepare for future climate scenarios and assess current populations. It can also give managers valuable information when attempting to sample a large species distribution range when identifying the persistence of species. This technique can be used as a supplement to field sampling to create cost effective management of stream species.