Advances in Ecological Niche Modeling

David R.B. Stockwell
San Diego Supercomputer Center
Email:davids@sdsc.edu
Web:http://biodi.sdsc.edu
Basic ENM process

- Occurrence and random background points overlain on variables
- Test accuracy on points
- Model applied to variables
- Frequency histogram of occurrence and background points
- Variable values
- Probability model
Classical 1D parabolic response

'Cerulean Warbler'
Why ENM?

Resolution
Use correlative variables of high resolution to improve resolution of scarce occurrence data.
GARP – Genetic Algorithm for Rule-set Production

1. Bioclimatic Envelope
IF
   Dev=[1,2] AND StC=[0,1] AND SdC=[3,5] AND
   StQ=[1,3] AND FlN=[2,3] AND Slp=[1,1] AND
   Ero=[3,3]
THEN ExM= 3

2. Logistic Regression
IF
   - Dev*0.10 - StC*0.10 + SdC*0.09`+ StQ*0.06
   - FlN*0.19 + Slp*0.40
THEN ExM= 1

3. GARP rule
IF
   Dev=[1,2] AND SdC=[0,0] AND StQ=[0,3]
THEN ExM= 0

4. Atomic rule
IF
   Dev= 0 AND StC= 1 AND SdC= 2`AND StQ= 1 AND
   FlN= 1
   AND Slp= 3 AND Ero= 1
THEN ExM= 1

Why use multiple models?

Robustness

• Use of the consensus of multiple models compensates for problems in one model and provides adequate results on most occasions

• Does not necessarily provide the highest accuracy in a specific case
Predicted range of the Asian Longhorn Beetle in the USA (Anoplophora glabripennis) by A. Townsend Peterson (KU)
Leo Joseph (Academy of Sciences, Phil.)

Migration of Swainson’s Flycatcher

Myiarchus swainsoni across South America
Growth in GARP Citations

Williams PH, Margules CR, Hilbert DW Data requirements and data sources for biodiversity priority area selection, J BIO SCIENC E 27 (4): 327-338 Suppl. 2 JUL 2002
35. Lim BK, Peterson AT, Engstrom MD Robustness of ecological niche modeling algorithms for mammals in Guyana, BIODIVERS CONSERV 11 (7): 1237-1246 JUL 2002
42. Bobbin J, Becknagel P Knowledge discovery for prediction and explanation of blue-greens algal dynamics in lakes by evolutionary algorithms, ECOL MODEL 146 (1-3): 253-262 Sp. Iss. SI DEC 1 2001
43. Mackey BG, Lindemayer DB Towards a hierarchical framework for modelling the spatial distribution of animals, J BIOGEOGR 28 (9): 1147-1166 SEP 2001
45. Peterson AT Predicting species' geographic distributions based on ecological niche modeling, CONDOR 103 (3): 599-605 AUG 2001
53. Joseph L, Stockwell D Temperature-based models of the migration of Swainson’s Flycatcher (Myiarchus swainsoni) across South America: A new use for museum specimens of migratory birds, P ACAD SCI PHILA 150: 293-300 APR 14 2000
54. Peterson AT, Soberon J, Sanchez-Cordero V Conservation of ecological niches in evolutionary time, SCIENCE 285 (5431): 1265-1267 AUG 1999
55. Peterson AT, Cohoon KP Sensitivity of distributional prediction algorithms to geographic data completeness, ECOL MODEL 117 (1): 159-164 APR 1 1999
63. Stockwell DRB, LBS - BAYESIAN LEARNING SYSTEM FOR RAPID EXPERT SYSTEM DEVELOPMENT, EXPERT SYST APPL 6 (2): 137-147 APR-JUN 1993 STOCKWELL DRB, NOBLE IR PREDICTION OF SETS OF RULES FROM ANIMAL DISTRIBUTION DATA – A ROBUST AND INFORMATIVE METHOD OF DATA-ANALYSIS, MATH COMPUT SIMULAT 33 (5-6): 385-390
Why continue collecting?

Accuracy

• Most methods perform well with enough data

• Adequate occurrence points is a major limitation to accuracy

• 10 points produces 90% of maximum accuracy

• Of all museum specimens, 49% have any georeferences, 8% have > 10 georeferences
Data Sources: Museum Data http://speciesanalyst.net

- **Institution** | **Server** | **Database** | **Status**
- KUNHM           | habanero.nhm.ukans.edu:210 | KUBirds | OK
- KUNHM           | habanero.nhm.ukans.edu:210 | KUMammals | OK
- KUNHM           | habanero.nhm.ukans.edu:210 | KBSPlants | OK
- UNAM            | fcbiologia.fciencias.unam.mx:210 | Mamife | OK
- UNAM            | fcbiologia.fciencias.unam.mx:210 | MexBirds | OK
Lifemapper - www.lifemapper.org

- a screensaver grid computing project
- develop a fauna and flora using the world's museum data
- installed base of over 30,000 screensavers on personal computers.

Accuracy by number correlates

Accuracy

Channels

CW-in
CW-ex
RV-ex
RV-in
SF-ex
SF-in
Why fewer correlates?

Explanation

• Identify those factors that maximize accuracy
• Parsimonious model - with the right variables, and non-linear response can be as few as one or two
Some sources of correlative data

**Terrestrial >500**  Marine >500

- Global Ecosystems database (1deg - 1km) Topographic (DEM), Atmospheric, Climatic and Meteorologic, Hydrologic, Oceanographic, Ecosystems and Biogeochemical Dynamics, Geological and Geophysical Data - 10GB
- Landsat 1km - %cover of treecover, evergreen, deciduous, broadleaf - 4GB
- Marine - productivity, annual temperatures and deviations, salinity, at various depths
- Satellite - MODIS 12 level 3-4 land and 15 level 3-4 ocean products 1km to 250m tiled, 1TB to 4TB per annum
- Digital Elevation - min, max, median elevation, slope, aspect, rugosity, hydrological variables - 30m - 1TB
- BIOCLIM variables
SEEK EcoGrid
http://seek.ecoinformatics.org

LTER Network (24)  Natural History Collections (>> 100)
Organization of Biological Field Stations (180)
UC Natural Reserve System (36)
Partnership for Interdisciplinary Studies of Coastal Oceans (4)
Multi-agency Rocky Intertidal Network (60)
Cerulean Warbler

WhyWhere?

Answers the question "Where is it and why?" on a global scale.
WhyWhere? vs. GARP $\uparrow 14\%$

![Graph comparing WhyWhere? and GARP with accuracy vs. number of training data](image)
WhyWhere? Parallel Prediction Algorithm

**Archive:** 1. Point Data, 2. Cropped and Sized Environmental Correlates

**Iteration 1 test**
- 0.6 layer1
- 0.75 layer2
- ...
- 0.91 layer(m) best
- ...
- 0.89 layer(n)

**Iteration 2 test**
- 0.92 m.layer1
- 0.75 m.layer2
- ...
- 0.95 m.layer(o) best
- ...
- 0.9 m.layer(n)

**Iteration 3 test**
- 0.92 o.m.layer1
- 0.75 o.m.layer2
- ...
- 0.96 o.m.layer(p) best
- ...
- 0.9 o.m.layer(n)

Output model color palette with probabilities, Internal and External Accuracy of best combinations
Víctor Sánchez-Cordero, Sahotra Sarkar, David Stockwell and Howei Liu

Competition limits the southern distribution of bobcats *Lynx rufus*
Socio-Scientific Summary

- Increased Resolution ↔ Correlative Models
- Increased Robustness ↔ Consensus Models
- Increased Accuracy ↔ Increased Occurrence Data ↔ Natural History Museums
- Increased Explanation ↔ Increased Correlative Data ↔ NASA
Advance in ENM

1. Empirical and Theoretical Statistical Studies

2. Algorithmic Software Developments

3. Information Infrastructure Development

4. New Science Applications

http://biodi.sdsc.edu/ww_home.html