Working with what we know – presence-only / ecological niche models in marine mammal science

Kristin Kaschner, Colin MacLeod, Laura Mandleberg, Ross Comptom

Sea Around Us Project, Fisheries Centre, University of British Columbia, Canada & FTZ Büsum, Kiel Christian-Albrechts-University, Germany
School of Biological Sciences (Zoology), University of Aberdeen, UK
Outline

• What are presence-only models?
• Why do we use them?
• Which ones are there?
• Do they work?
• What can we do with them?
What are they?

- Predict ecological niches
- Use only presence data
Why do we use them?

- Data paucity
- Absence data issues
- Niche modeling vs. distribution
Data paucity

Hawaii
Data paucity

WHY?
OBIS-SEAMAP (http://seamap.env.duke.edu/species)
- compilation & storage of marine mammal occurrence data
- out of 115 species, geo-referenced / effort corrected data
  - available/accessible for ~ 50%
  - representative coverage = ~ 2 %
Absence Data Issues

Model evaluation
Presence-absence confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Recorded presence</th>
<th>Recorded absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted presence</td>
<td>a (true presence)</td>
<td>b (false presence)</td>
</tr>
<tr>
<td>Predicted absence</td>
<td>c (false absence)</td>
<td>d (true absence)</td>
</tr>
</tbody>
</table>
Absence Data Issues

Model evaluation
Presence-absence confusion matrix

<table>
<thead>
<tr>
<th>Predicted presence</th>
<th>Recorded presence</th>
<th>Recorded absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>a (true presence)</td>
<td>b (false presence)</td>
<td>c (false absence)</td>
</tr>
</tbody>
</table>

Omission error / Model overfitting
Commission error / Model overprediction
Absence Data Issues

Model evaluation

Presence-absence confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Recorded presence</th>
<th>Recorded absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted presence</td>
<td>a (true presence)</td>
<td>b (false presence)</td>
</tr>
<tr>
<td>Predicted absence</td>
<td>c (false absence)</td>
<td>d (true or perceived absence)</td>
</tr>
</tbody>
</table>

WHY?
Absence Data Issues

- Site with environmental value X
- Presence-only data

Species present?  Site visited?  Species detected?

M. Nakamura, CONABIO, 2005
GBIF Ecological Niche Modelling Workshop, KU
Absence Data Issues

Species present? Site visited? Species detected?

Site with environmental value X

Presence-absence data

M. Nakamura, CONABIO, 2005
GBIF Ecological Niche Modelling Workshop, KU
Absence Data Issues

Species present? Site visited? Species detected?

Site with environmental value X

True absence data

False absence data

M. Nakamura, CONABIO, 2005
GBIF Ecological Niche Modelling Workshop, KU
Absence Data Issues

WHY?

Harbour porpoise density

No animals in deeper waters!

Scheidat, Gilles et al, (unpublished data)
Absence Data Issues

WHY?

Harbour porpoise density

No animals in deeper waters?
True absences???

Aerial survey effort

Scheidat, Gilles et al, (unpublished data)
Absence Data Issues

Scheidat, Gilles et al, (unpublished data)
Absence Data Issues

Scheidat, Gilles et al, (unpublished data)
Absence Data Issues

Scheidat, Gilles et al, (unpublished data)
Absence Data Issues

WHY?

Spring

Summer

Fall

Mean # sightings

Effort category [km2]

True absences???

Scheidat, Gilles et al, (unpublished data)
### Absence Data Issues

**Example: 100 trials**

<table>
<thead>
<tr>
<th>Species #1</th>
<th>Species #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable A</td>
<td>Variable A</td>
</tr>
<tr>
<td>Variable B</td>
<td>Variable B</td>
</tr>
<tr>
<td>Variable C</td>
<td>Variable C</td>
</tr>
<tr>
<td>Presence</td>
<td>Presence</td>
</tr>
<tr>
<td>0.20</td>
<td>0.32</td>
</tr>
<tr>
<td>0.80</td>
<td>0.16</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bias</td>
<td>Bias</td>
</tr>
<tr>
<td>0.8</td>
<td>0.5</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>Obs. rate</td>
<td>Obs. rate</td>
</tr>
<tr>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Observed</td>
<td>Observed</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

M. Nakamura, CONABIO, 2005
GBIF Ecological Niche Modelling Workshop, KU
Ecological Niche vs Distribution

WHY?

Model Complexity

Prediction Error

Low Variance

High Variance

Low

High

Training sample

Test sample

Hastie et al. (2001)
WHY?

Ecological Niche vs Distribution

Model Complexity vs Prediction Error

Low Variance → High Variance

Low → High

Ecological Niche Model

Distribution Model

Test sample

Training sample

Hastie et al. (2001)
Ecological Niche vs Distribution

Geographic space (2 dimensions)

Ecological space (n dimensions / hypervolume)

J. Soberon, CONABIO, 2005
GBIF Ecological Niche Modelling Workshop, KU
Which ones are out there?

Envelope models
- BioClim
- DOMAIN
- Fuzzy bioclimatic envelope model
- RES*

Machine-learning
- Garp**
- Maxent**
- ENFA (Biomapper)*

* Models have been applied to and tested for marine mammals
** Preliminary applications to marine mammals
Which ones are out there?

Web-based applications

- AquaMaps*
- KGS-Mapper**
- WhyWhere?

* Models have been applied to and tested for marine mammals
** Preliminary applications to marine mammals
Ecological Niche Models

WHICH?

Geographic space

Ecological / Environmental space

J. Soberon
BioClim / Climate Envelope Range

- Based on presence cells
- Very simple & intuitive
- No interactions between variables
- Unweighted variables
- Binary predictions
- No extrapolations
- Tends to overpredict

Relative Environmental Suitability Model
(Fuzzy Bioclimatic Envelope Model)

- Based on relative occurrences
- Very simple, transparent & intuitive
- Expert knowledge based (no point data required)
- No interactions between variables
- Unweighted variables
- Continuous output
- Tends to overpredict

Assigned habitat usage categories: Depth, SST, Ice edge

Kaschner et al. 2006, MEPS / www.seaaroundus.org
Relative Environmental Suitability Model
(Fuzzy Bioclimatic Envelope Model)

Assigned habitat usage categories: Depth, SST, Ice edge

- Based on relative occurrences
- Very simple, transparent & intuitive
- Expert knowledge based (no point data required)
- No interactions between variables
- Unweighted variables
- Continuous output
- Tends to overpredict

Kaschner et al. 2006, MEPS / www.seaaroundus.org
DOMAIN

- Based on presence cells
- Cluster algorithm in environmental space
- No interactions between variables
- Unweighted variables
- Non-binary predictions
- Tends to overfit

Freeware: http://www.cifor.cgiar.org/docs/_ref/research_tools/domain/
GARP
Genetic Algorithm of Rule-set Prediction

- Based on presence cells
- Machine learning / automated model optimization (not transparent)
- Generates pseudo-absence data
- Optimization using training / test data sets
- Interactions between weighted variables
- Non-binary predictions
- Tends to overpredict?

Freeware: [http://www.lifemapper.org/desktopgarp/](http://www.lifemapper.org/desktopgarp/)
GARP
Genetic Algorithm of Rule-set Prediction

- Based on presence cells
- Machine learning / automated model optimization (not transparent)
- Generates pseudo-absence data
- Optimization using training / test data sets
- Interactions between weighted variables
- Non-binary predictions
- Tends to overpredict?

Minke whale (IWC whaling data)

Freeware: http://www.lifemapper.org/desktopgarp/
Maxent
Maximum entropy

- Based on presence cells
- Machine learning / automated model optimization (not transparent)
- Optimization using training / test data sets
- Interactions between weighted variables
- Non-binary predictions
- Tends to overfit?

Phillips et al. 2006, *Ecological Modelling*
Freeware: http://www.cs.princeton.edu/~schapire/maxent/

Minke whale (IWC whaling data)
Maxent
Maximum entropy

• Based on presence cells
• Machine learning / automated model optimization (not transparent)
• Optimization using training / test data sets
• Interactions between weighted variables
• Non-binary predictions
• Tends to overfit?

Phillips et al. 2006, *Ecological Modelling*
Freeware: http://www.cs.princeton.edu/~schapire/maxent/
ENFA
Ecological Niche Factor Analysis

- Based on presence cells
- Compares species occurrence to all available habitat
- Multivariate (Interactions between weighted variables)
- Non-binary predictions (HIS)

Biomapper freeware: http://www2.unil.ch/biomapper/
ENFA
Ecological Niche Factor Analysis

- Based on presence cells
- Compares species occurrence to all available habitat
- Multivariate (Interactions between weighted variables)
- Non-binary predictions (HIS)

Biomapper freeware: http://www2.unil.ch/biomapper/
ENFA
Ecological Niche Factor Analysis

Northern bottlenose whale
(NW Atlantic & Total N Atlantic)

MacLeod, 2005, PhD Thesis, U of Aberdeen
AquaMaps (Coming soon....)

Kaschner et al, in prep, www.fishbase.de
AquaMaps (Coming soon....)

Kaschner et al, in prep, www.fishbase.de
AquaMaps (Coming soon....)

Kaschner et al, in prep, www.fishbase.de
Do they work?
Model Evaluations

- Test statistics
- Cross-validation
- External testing
- Comparison with other models
**Test statistics**

<table>
<thead>
<tr>
<th></th>
<th>Recorded presence</th>
<th>Recorded absence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predicted presence</strong></td>
<td>a (true presence)</td>
<td>b (false presence)</td>
</tr>
<tr>
<td><strong>Predicted absence</strong></td>
<td>c (false absence)</td>
<td>d (true absence)</td>
</tr>
</tbody>
</table>

**Sensitivity:** Proportion of observed presences correctly predicted

\[ \frac{a}{a + c} \]

R. Pearson, AMNH, 2005
GBIF Ecological Niche Modelling Workshop, KU
## Test statistics

<table>
<thead>
<tr>
<th></th>
<th>Recorded presence</th>
<th>Recorded absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted presence</td>
<td>a (true presence)</td>
<td>b (false presence)</td>
</tr>
<tr>
<td>Predicted absence</td>
<td>c (false absence)</td>
<td>d (true absence)</td>
</tr>
</tbody>
</table>

**Sensitivity:** Proportion of observed presences correctly predicted

\[
a/(a + c)
\]

**Specificity:** Proportion of observed (or assumed) absences correctly predicted

\[
d/(b + d)
\]

---

R. Pearson, AMNH, 2005
GBIF Ecological Niche Modelling Workshop, KU
Test statistics:
Receiver Operator Curve

DO THEY WORK?

AUC = 0.5 = random

Area Under Curve (AUC)
**Test statistics**

<table>
<thead>
<tr>
<th></th>
<th>Recorded presence</th>
<th>Recorded absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted presence</td>
<td>a (true presence)</td>
<td>b (false presence)</td>
</tr>
<tr>
<td>Predicted absence</td>
<td>c (false absence)</td>
<td>d (true absence)</td>
</tr>
</tbody>
</table>

Cohen’s Kappa:

\[
k = \frac{\left( (a + d) - \left( \frac{((a + c)(a + b) + (b + d)(c + d))}{n} \right) \right)}{\left( n - \left( \frac{((a + c)(a + b) + (b + d)(c + d))}{n} \right) \right)}
\]

R. Pearson, AMNH, 2005
GBIF Ecological Niche Modelling Workshop, KU
Cross-validation

Araújo et al. 2005 *Gl. Ch. Biol.*

- Same data set
- Jack-knife procedure

Northern bottlenose whale (NW Atlantic)

Environmental envelope


- Different data set
- Permutation

External Testing

100%

Evaluation

Calibration

New region

New resolution

New time

Projection

Average SPUE per cell

RES category

Southern bottlenose whale

Kaschner et al, 2006, MEPS

<table>
<thead>
<tr>
<th>Common name</th>
<th>Spearman's non-parametric rank correlation analysis of RES vs mean SPUE</th>
<th>Comparison with random data sets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rho</td>
<td>p</td>
</tr>
<tr>
<td>Northern fur seal</td>
<td>0.54</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Harbour porpoise</td>
<td>0.59</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Sperm whale</td>
<td>0.66</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Antarctic minke whale</td>
<td>0.71</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Killer whale</td>
<td>0.56</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Fin whale</td>
<td>0.53</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Blue whale</td>
<td>0.48</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Humpback whale</td>
<td>0.20</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Hourglass dolphin</td>
<td>0.68</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Southern bottlenose whale</td>
<td>0.83</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Dwarf minke whale</td>
<td>-0.77</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
Model Comparison

PCA  ENFA

GARP  GLM

Harbour porpoise

Presence
Absence

Mandelbaum, 2005, MSc Thesis, U of Aberdeen
Model Comparison

DO THEY WORK?

Mandelbaum, 2005, MSc Thesis, U of Aberdeen

- PCA: AUC = 0.746
- ENFA: AUC = 0.745
- GARP: AUC = 0.773
- GLM: AUC = 0.828

Significantly different
Do they work?

Limitations

- Presence cells vs. encounter rates
- Effort / Sampling biases
Presence cells vs Encounter Rate

Maxent prediction (IWC whaling data)

Minke whale

Kaschner et al, in prep
Presence cells vs Encounter Rate

Maxent prediction (IWC whaling data)

Minke whale

SOWER sightings
Whaling data

Kaschner et al, in prep
Presence cells vs Encounter Rate

Maxent prediction (IWC whaling data)

Minke whale

Kaschner et al, in prep
Sampling Biases

Minke whale  Blue whale  Humpback whale

Kaschner et al, 2006, MEPS
Sampling Biases

Minke whale

Blue whale

Humpback whale

Kaschner et al., 2006, MEPS
DO THEY WORK?

Sampling Biases

Minke whale

Blue whale

Humpback whale

Kaschner et al, 2006, MEPS
What can we do with them?
Potential Applications

- Biodiversity Mapping
- Management / Research Prioritization
  - Risk mitigation
  - Marine mammal-fisheries interactions
  - Marine Protected Areas
  - Climate change
What can we do with them?

Potential Applications

- Biodiversity Mapping
- Management / Research Prioritization
  - Risk mitigation
  - Marine mammal-fisheries interactions
  - Marine Protected Areas
  - Climate change
Applications

- Management / Research Prioritization
  - Risk mitigation
  - Marine mammal fisheries interactions
  - Marine Protected Areas
  - Climate change
Applications

- Management / Research Prioritization
  - Risk mitigation
  - Marine mammal fisheries interactions
  - Marine Protected Areas
  - Climate change

MacLeod, 2005, PhD Thesis, U of Aberdeen

- M. densirostris
- All Mesoplodon Spp.
Applications

• Management / Research Prioritization
  – Risk mitigation
  – Marine mammal-fisheries interactions
  – Marine Protected Areas
  – Climate change

Kaschner et al, accepted, CJFAS
Applications

- Management / Research Prioritization
  - Risk mitigation
  - Marine mammal fisheries interactions
  - Marine Protected Areas
- Climate change

Kaschner, 2006, *Ecology of Seamounts*
Applications

• Management / Research Prioritization
  – Risk mitigation
  – Marine mammal fisheries interactions
  – Marine Protected Areas
  – Climate change

Predicted relative changes in distribution size by 2020

Small toothed whales
Conclusions

Presence-only / Ecological niche models

- require less data / maximize available data
- less affected by false absences
- more general, useful to investigate large scale patterns & ecological interactions
- time & cost efficient starting points
- can supplement small scale studies and help to focus research and management efforts
Acknowledgements

- Ed Gregr, MMRU, UBC, Vancouver
- Anita Gilles & Roger Mundry, FTZ Buesum
- Instructors & Participants of the GBIF Ecological Niche Modelling Workshop, 5-9 Dec 2005, Lawrence, Kansas