FUTURE OF MARINE ANIMAL POPULATIONS

Species distribution modelling in the marine environment: opportunities and dangers

Derek Tittensor

11th October 2009

Quebec City

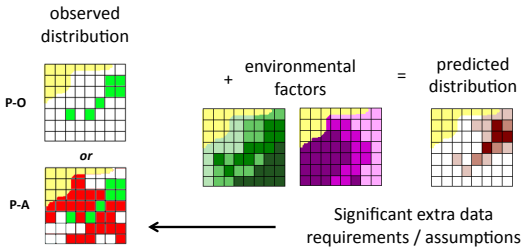
Outline

1. Introduction
2. Examples of presence-only marine models
3. Methods
4. When should I use a presence-only model?
5. Challenges & Dangers
6. Conclusions

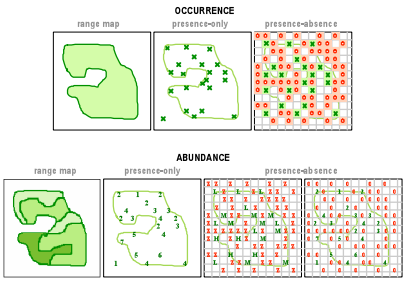
1. Introduction

Types of model

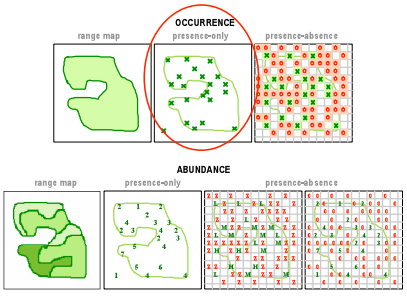
• Can be broadly broken down into presence-only and presence-absence, by data requirements



Types of distribution data



Types of distribution data



Why use presence only models?

Ideal scenario

- Good spatial coverage
- Reliable absence data
- Comparable level of effort between cells (locations)

Why use presence only models?

Ideal scenario

- Good spatial coverage
- Reliable absence data
- Comparable level of effort between cells

*Reality**

- Sparse data
- Often potential false absences
- Frequently not standardised effort
- Very difficult to prove absence

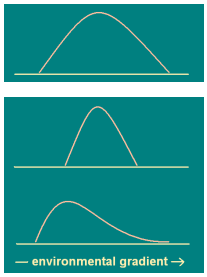
* At least in the marine realm

Niche concepts

fundamental niche:
potential to survive, grow, reproduce
- physiological tolerance (abiotic)
- resources (biotic & abiotic)

realised niche:
actual survival, growth, reproduction
- competitors, predators, parasites & pathogens (biotic)

➢ non-normal response curves
➢ occurrence ≠ optimal conditions
➢ potentially several niche configurations



Example modelling methods

Envelope Models
- BIOCLIM, DOMAIN, Mahalanobis distance, RES/ AquaMaps

Canonical Methods
- ENFA, discriminant analysis

Regression Techniques
- GLM, GAM, generalized dissimilarity models, (boosted) regression trees, MARS

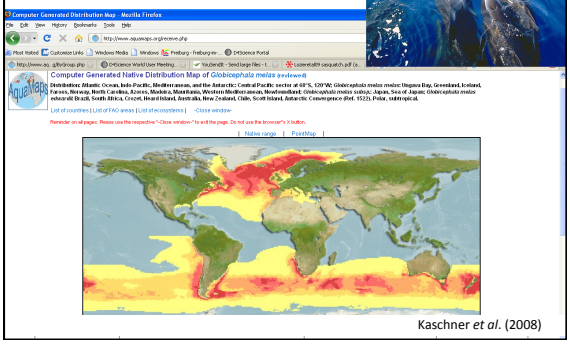
Machine learning methods
- GARP, artificial neural networks, MAXENT

Terrestrial vs. marine

- Species distribution modelling is somewhat less frequent in the marine realm
- 84 of 995 (< 8.5%) of SDM papers from 1991 to 2008 were 'marine' (Macpherson, pers. comm.)
- Why is this?
  - Sampling more challenging, and data requirements more difficult to meet?

2. Examples of presence-only marine models

Aquamaps



Kaschner *et al*. (2008)

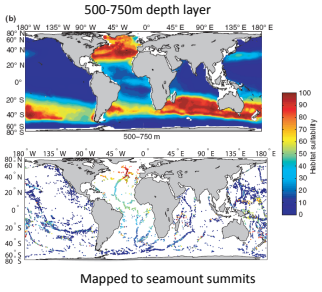
Aquamaps

- Automatically generated maps for >9,000 marine species
- Maps can be reviewed and verified by experts
- Based on (supplemented) environmental envelopes (modified RES model)
- Developed for particularly data-poor situations

Kaschner *et al*. (2008)

Cold-water stony corals on seamounts

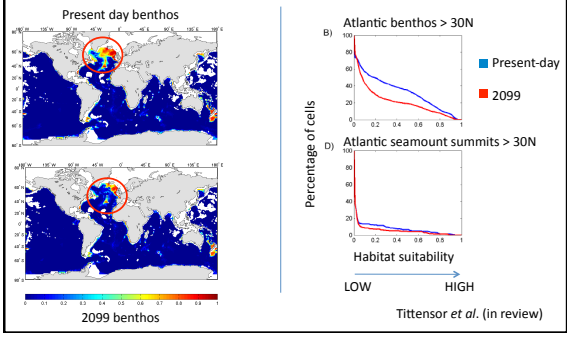
- Maxent and ENFA
- Modelled habitat suitability for deep-sea stony corals on seamounts
- Three-dimensional approach



Mapped to seamount summits

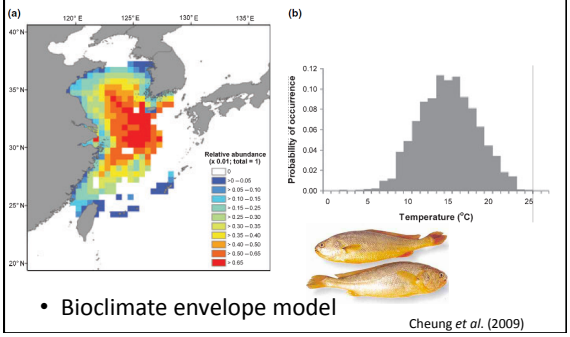
Tittensor *et al*. (2009)

Effects of ocean acidification on cold-water corals: seamount refugia



Tittensor *et al*. (in review)

Impacts of climate change on marine biodiversity



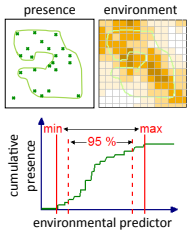
- Bioclimate envelope model

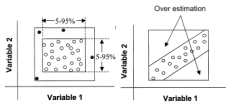
Cheung *et al*. (2009)

3. Methods

BIOCLIM

e.g. Farber & Kadmon 2003 Ecological Modelling 160: 115-130

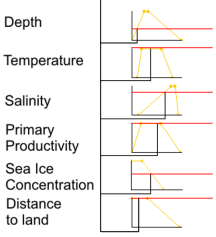




- requires only presence
- rectilinear envelope
  - ➢ does not cope well with collinearity or interactions among predictors
- http://www.diva-gis.org/

RES / AquaMaps

$P_{Cell} = \sqrt[6]{(0.4 \times 1.0 \times 0.8 \times 1.0 \times 0.5 \times 1.0)} = 0.737$



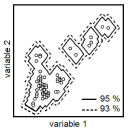
$P_{Depth} = 0.4$
$P_{Temp} = 1.0$
$P_{Salinity} = 0.8$
$P_{PProd} = 1.0$
$P_{IceCon} = 0.5$
$P_{LandDist} = 1.0$

- Trapezoidal response curve
- Developed specifically for marine environment and data

RES: Kaschner *et al.* (2006)
Aquamaps: Kaschner *et al.* (2008)
Slide: J.Ready

DOMAIN

e.g. Carpenter *et al.* 1993
Biodiversity & Cons. 2: 667-680



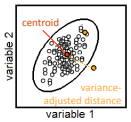
- requires only presence
- point-similarity envelope

  Gower metric:

  $$d_{AB} = \frac{1}{p}\sum_{k=1}^{p}\left(\frac{|A_k - B_k|}{range_k}\right)$$
- http://www.diva-gis.org/

Mahalanobis distance

e.g. Corsi *et al.* 1999
Conservation Biology 13: 150-159

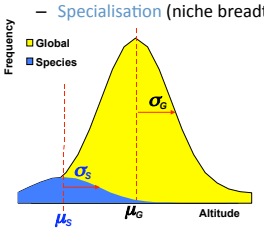


- requires only presence
- elliptic envelope

  Mahalanobis distance:
  $D^2 = (x - m)^T C^{-1} (x - m)$
  NB: assumes normality
- R base: *mahalanobis()*

Environmental Niche Factor Analysis (ENFA)

- Species niche is a subset of the global environment.
- Species set of EGV differs from global set by:
  - Marginality (deviation from the global mean)
  - Specialisation (niche breadth)

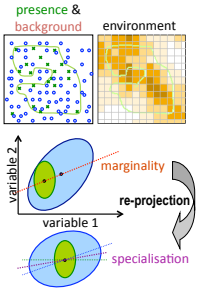


$$\text{Marginality} = \frac{|\mu_G - \mu_S|}{1.96\sigma_G}$$

$$\text{Specialisation} = \frac{\sigma_G}{\sigma_S}$$

Environmental Niche Factor Analysis (ENFA)

e.g. Brotons *et al.* 2004 Ecography 27: 437-448



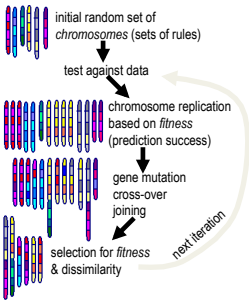
- compares conditions at presence localities to 'global' conditions
- computes an alternative set of axes (factors):

  Marginality
  maximizes difference in means

  Specialisation
  maximizes variance ratio
- NB: assumes normality
- www.unil.ch/biomapper/

Genetic Algorithm for Rule-set Prediction (GARP)

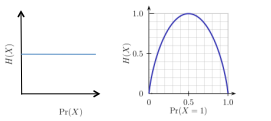
e.g. Peterson *et al.* 1999 Science 285: 1265-1267



- uses data on presence & (pseudo)absence
- genes = rules/functions chromosome = rule set
- 3 rule types: atomic (threshold), range & logit
- non-parametric
- binary output
- predictors' contributions to solution unknown
- www.lifemapper.org/desktopgarp/

Maximum entropy models (MAXENT)

$$H(\hat{\delta}) = -\sum_{x \in X} \hat{\delta}(x) \ln \hat{\delta}(x)$$



Distn. of max. entropy (unconstrained)

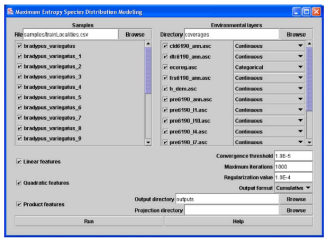
Distn. of max. entropy (coin toss)

- Based on Shannon's entropy
- Presence and background data
- Identifies statistical distribution that best fits observed data while minimizing constraints (maximizing entropy)
- Maximum likelihood approach with optimal solution

Phillips et al. (2006)
Phillips et al. (2008)

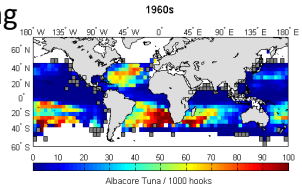
Maxent (continued)

- 5 constraint types:
  - Linear
  - Quadratic
  - Product
  - Threshold
  - Binary
- Thus can fit a wide variety of distribution functions
- Good at handling correlated predictor variables

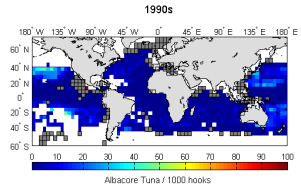


3. When should I use a presence-only model?

Japanese longlining data



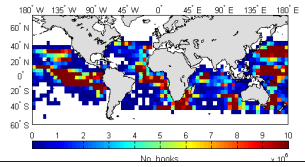
Albacore tuna
1960s & 1990s

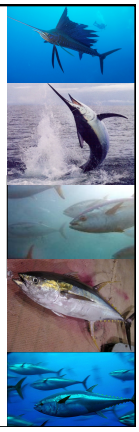


E.g. Myers & Worm (2003)
Worm et al. (2005)

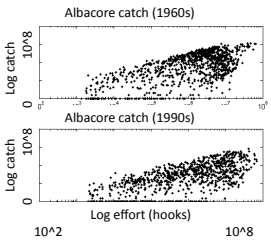
Japanese longlining data

- >20 billion hooks from 1950 to 1999!
- Yet still some 5 deg cells in which species were not caught, but are listed as present by FAO

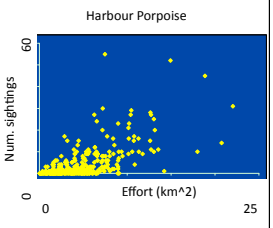




The relationship between effort and presence



Tittensor & Worm (in prep)



Scheidat, Gilles et al, (unpublished data)

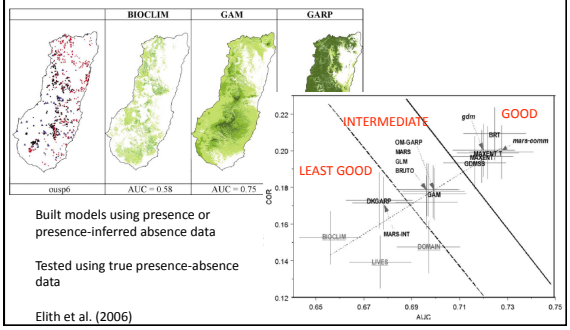
When should you use a presence-only model?

- If you have *reliable* absence data, it is better to use a P/A model (Elith *et al*. 2006)
- However, it is better to use a presence-only model rather than a P/A model with problematic absence data
- Otherwise you can be inaccurately representing species niche

Presence-only model validation

- How to test model performance?
- Field is evolving extremely rapidly
- Threshold-independent metrics have recently been developed (e.g. AUC, Phillips *et al*. 2006)
- Cross-validation important to prevent over-fitting

How do presence-only models perform in comparison to P/A models?



Built models using presence or presence-inferred absence data

Tested using true presence-absence data

Elith et al. (2006)

AquaMaps / RES

Species	AMG		AMEG		GAM		GLM		MAX		OMG	
	ROC-AUC	Rho	ROC-AUC	Rho	ROC-AUC	Rho	ROC-AUC	Rho	ROC-AUC	Rho	ROC-AUC	Rho
HYPLA	NO	NO	NO	YES	NO	YES	NO	NO	NO	NO	NO	NO
CAEQU	YES	NO	YES	NO	NO	YES	NO	NO	NO	N/A	NO	NO
PSERU	NO	NO	NO	YES	NO	NO	NO	NO	NO	YES	NO	N/A
CLHAR	YES	YES	YES	YES	NO	YES	YES	YES	YES	YES	NO	NO
BABAQ*	NO	NO	NO	NO	NO	NO	YES	YES	NO	NO	NO	NO
BOBOL*	NO	NO	NO	NO	NO	N/A	NO	NO	NO	YES	NO	N/A
TRTRA*	NO	NO	NO	NO	NO	YES	NO	NO	NO	NO	NO	N/A
SQMEG	YES	YES	YES	YES	NO	N/A	NO	NO	YES	YES	NO	NO
ZEFAB	YES	NO	YES	YES	NO	N/A	NO	NO	YES	YES	NO	NO
PHPHO	NO	NO	NO	NO	NO	YES	NO	YES	NO	NO	NO	NO
SQACA	NO	YES	NO	NO	NO	YES	NO	YES	NO	NO	NO	N/A
BAPHY	NO	YES	NO	YES	NO	YES	NO	YES	NO	NO	NO	NO
YES/Possible	4/12	4/12	4/12	6/12	0/12	7/10	2/12	5/12	3/12	5/11	0/12	0/8
YES/Possible Excluding *	4/9	4/9	4/9	6/9	0/9	6/8	1/9	4/9	3/9	4/8	0/9	0/7

- Compares well with existing methods
- Inclusion of expert knowledge tends to improve predictions

Ready et al. (accepted)

Which method should I use?

- Depends on your problem
- I am most familiar with ENFA and Maxent and both complement one another – ENFA is easily interpretable, Maxent tends to perform better under cross-validation
- I would thus advise implementing multiple methods to get a robust understanding of species-environment relationships.

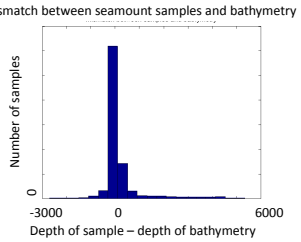
5. Challenges and dangers

Presence-only challenges

- Spatial autocorrelation not yet able to be resolved in presence-only models (Dormann *et al*. 2007)

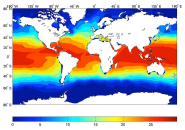
Marine-specific challenges

- Inherently three-dimensional environment
- Best approach depends on organism – benthic, pelagic, mid-water?
- Model the volume in which a species lives, not just the 'area'

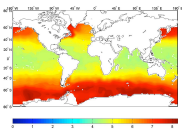


More (marine) challenges

- Highly correlated environmental variables can present model identifiability issues



Sea surface temperature



Sea surface dissolved O_2

General modelling challenges

- Potential scale mismatches between drivers and observations
- Biotic interactions
- KEY ASSUMPTION: Samples cover the range of environmental space occupied by a species

Dangers

- Over-fitting
- Software packages are terrifically easy to use (which means they are often applied with insufficient thought)
- Modelling potential vs. realized niche, and understanding that difference.

Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling

J. D. Lozier[1]*, P. Aniello[2] and M. J. Hickerson[3] (2009)

- Modelling without using biological & ecological knowledge of organism(s) under study is foolish

6. Conclusions

In conclusion

- Presence-only methods are very useful when absence data are unreliable
- Ideal for data compiled from non-standardised or effort-corrected sources (e.g. museum collections, multiple surveys with different methodologies)

In conclusion

- Performance compares well to presence-absence models
- Many opportunities as studies in the marine environment are limited.
- Field is evolving very rapidly, so important to keep an eye on the literature.

Thank you

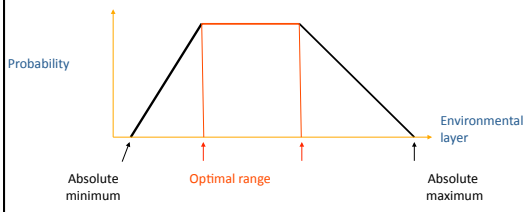
- Kristin Kaschner
- Jana MacPherson
- Boris Worm
- UNEP
- FMAP
- LenFest foundation

- I have key papers and software in a range modelling library on my flash drive for anyone who wants

METHOD

AquaMaps: Environmental envelopes

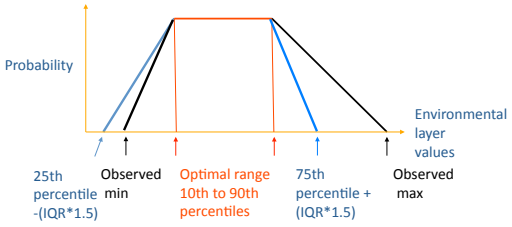
Depth: derived from published information about habitat usage



METHOD

AquaMaps: Environmental envelopes

SST, Salinity, PP, Sea ice concentration: derived from point data



IQR = interquartile range

Slide: courtesy J.Ready

METHOD

AquaMaps: Expert input