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?

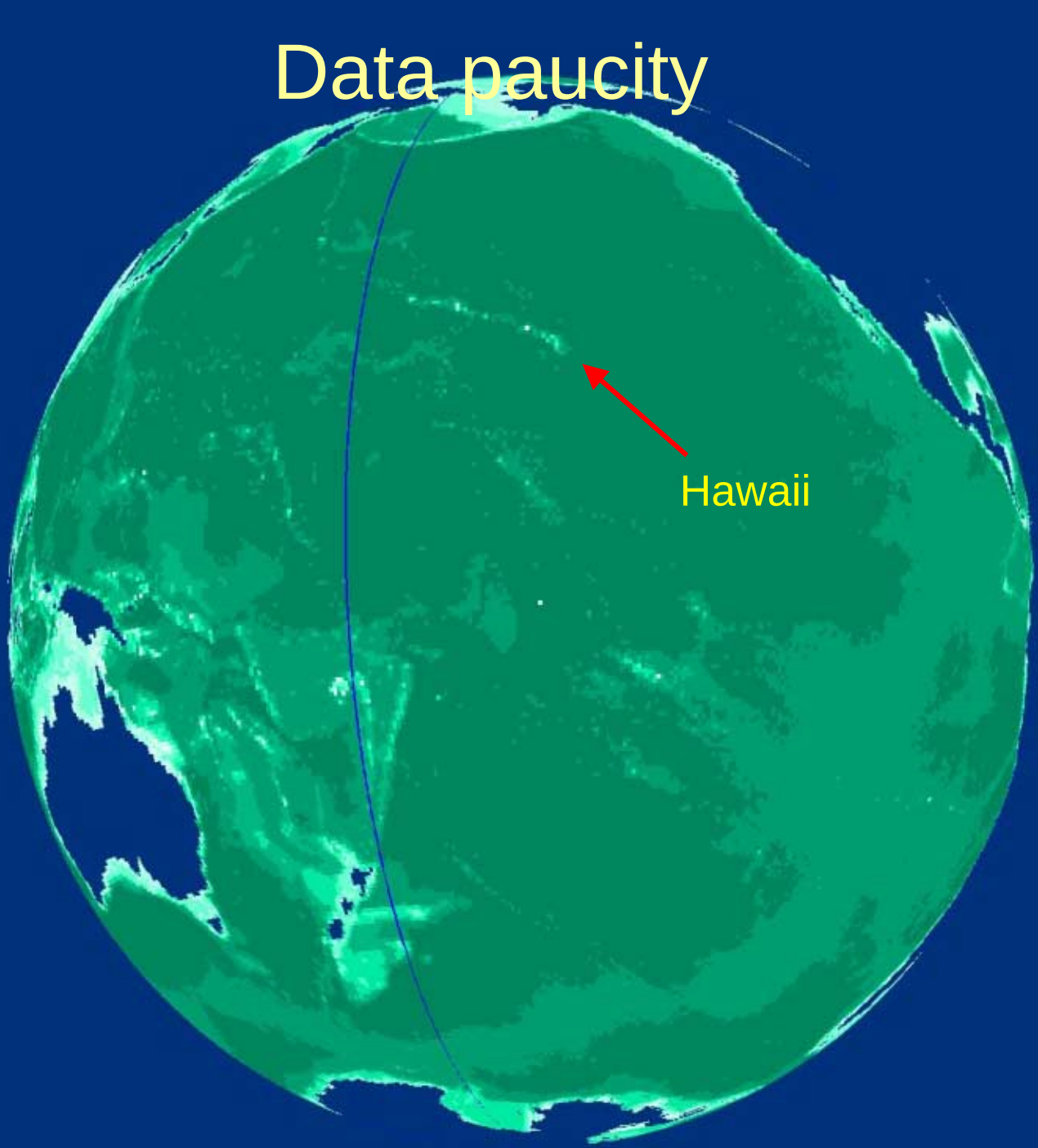
What are they?

- Predict ecological niches
- Use only presence data

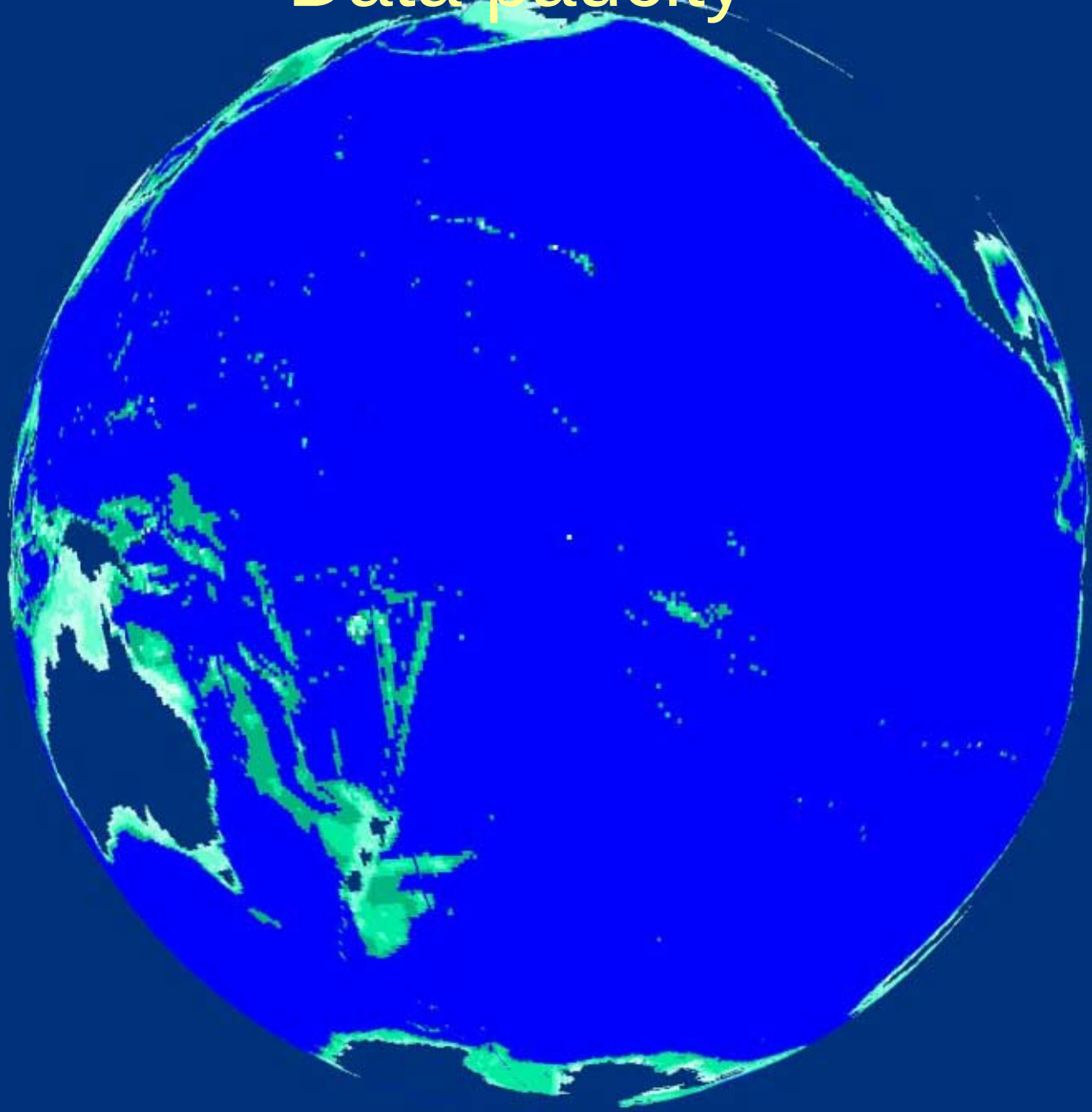
Why do we use them?

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

Data paucity



Data paucity



Data paucity

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 %

WHY?

Absence Data Issues

Model evaluation
Presence-absence confusion matrix

	Recorded presence	Recorded absence
Predicted presence	a (true presence)	b (false presence)
Predicted absence	c (false absence)	d (true absence)

WHY?

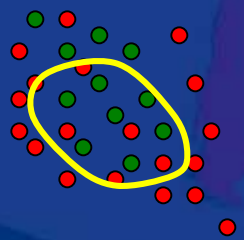
Absence Data Issues

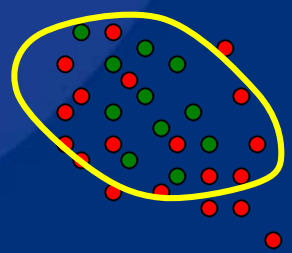
Model evaluation
Presence-absence confusion matrix

	Recorded presence	Recorded absence
Predicted presence	a (true presence)	b (false presence)
Predicted absence	c (false absence)	d (true absence)

Omission error / Model overfitting

Commission error / Model overprediction





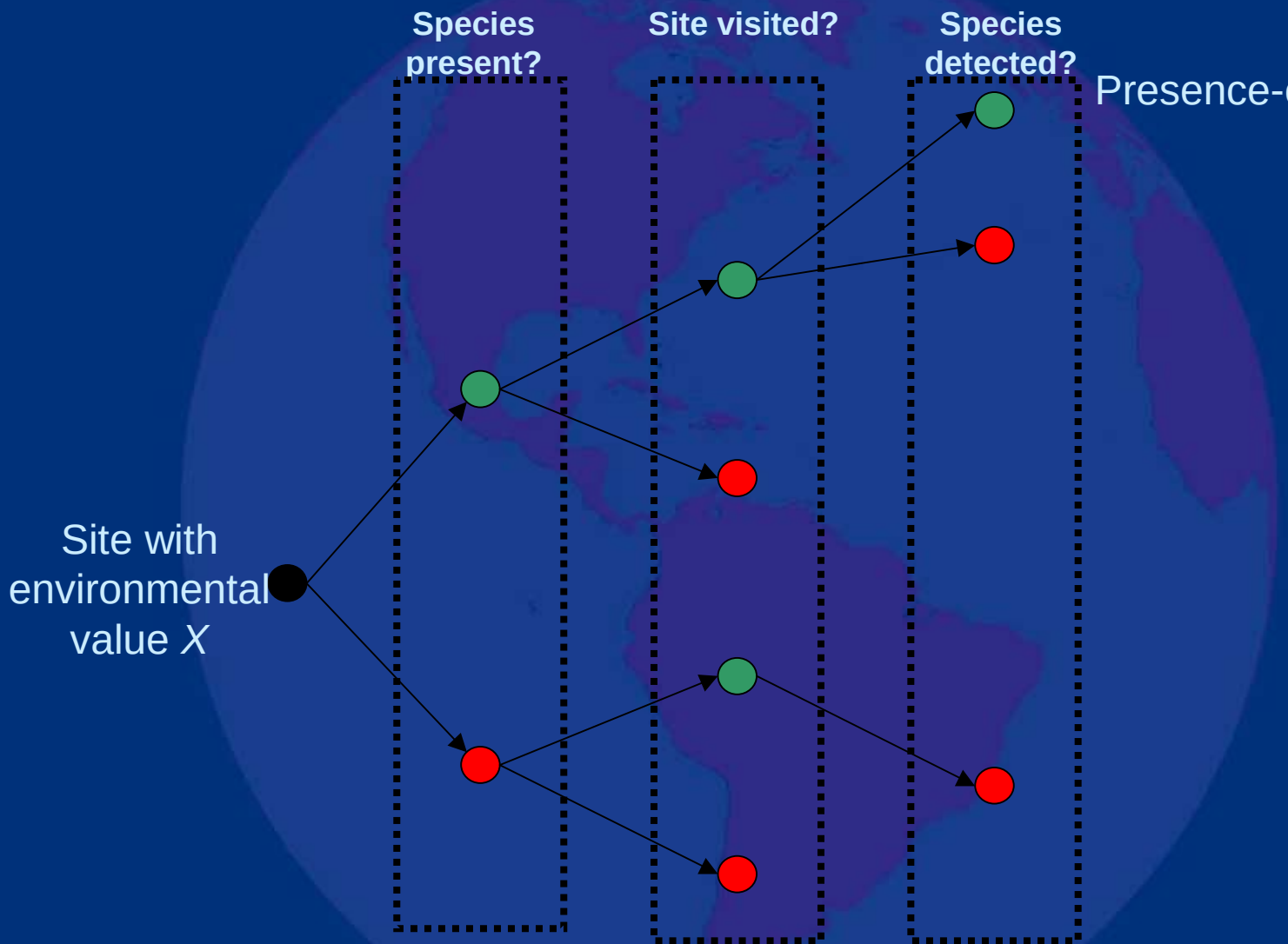
WHY?

Absence Data Issues

Model evaluation
Presence-absence confusion matrix

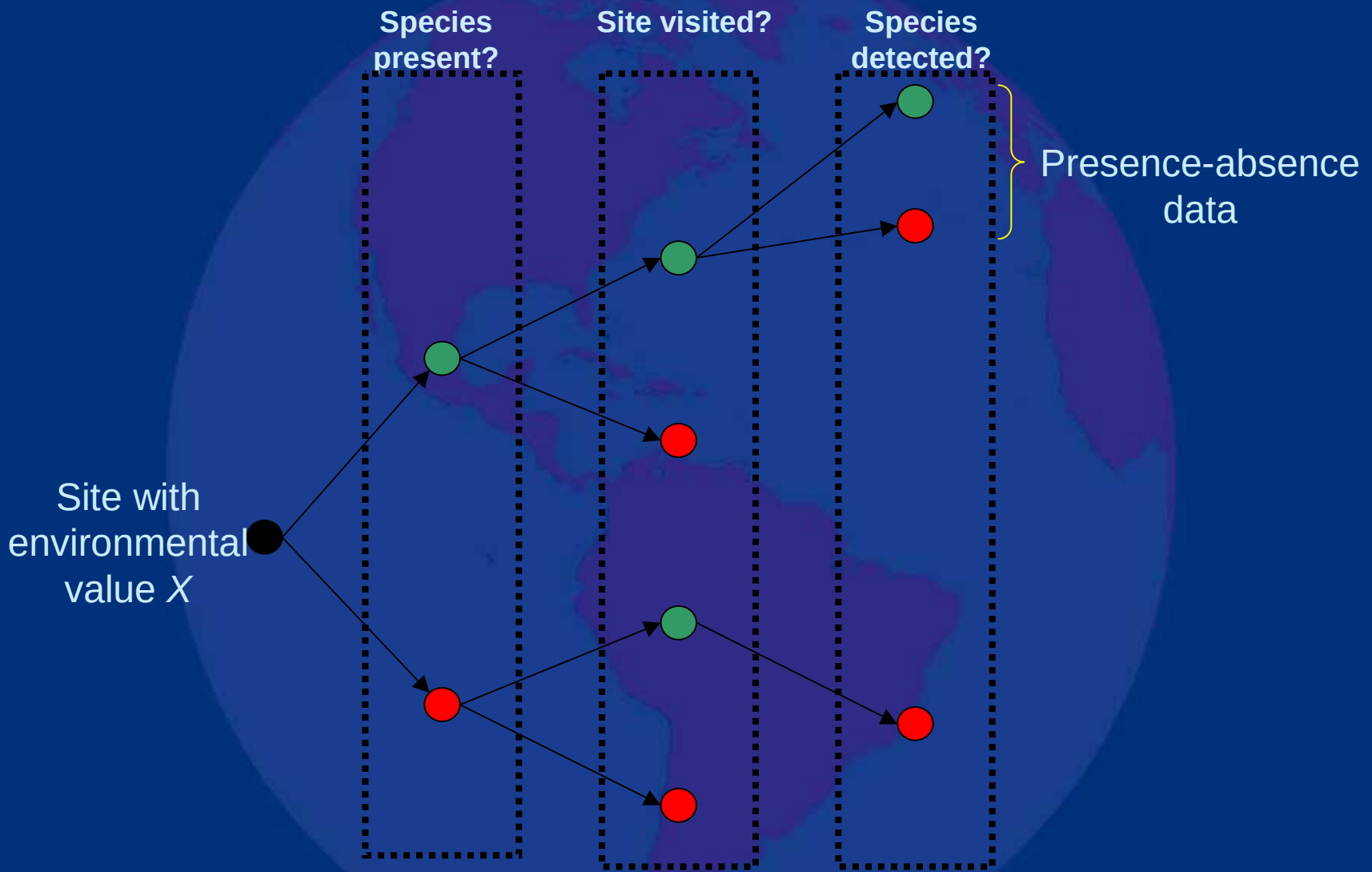
	Recorded presence	Recorded absence
Predicted presence	a (true presence)	b (false presence)
Predicted absence	c (false absence)	d (true or perceived absence????)

Absence Data Issues



M. Nakamura, CONABIO, 2005

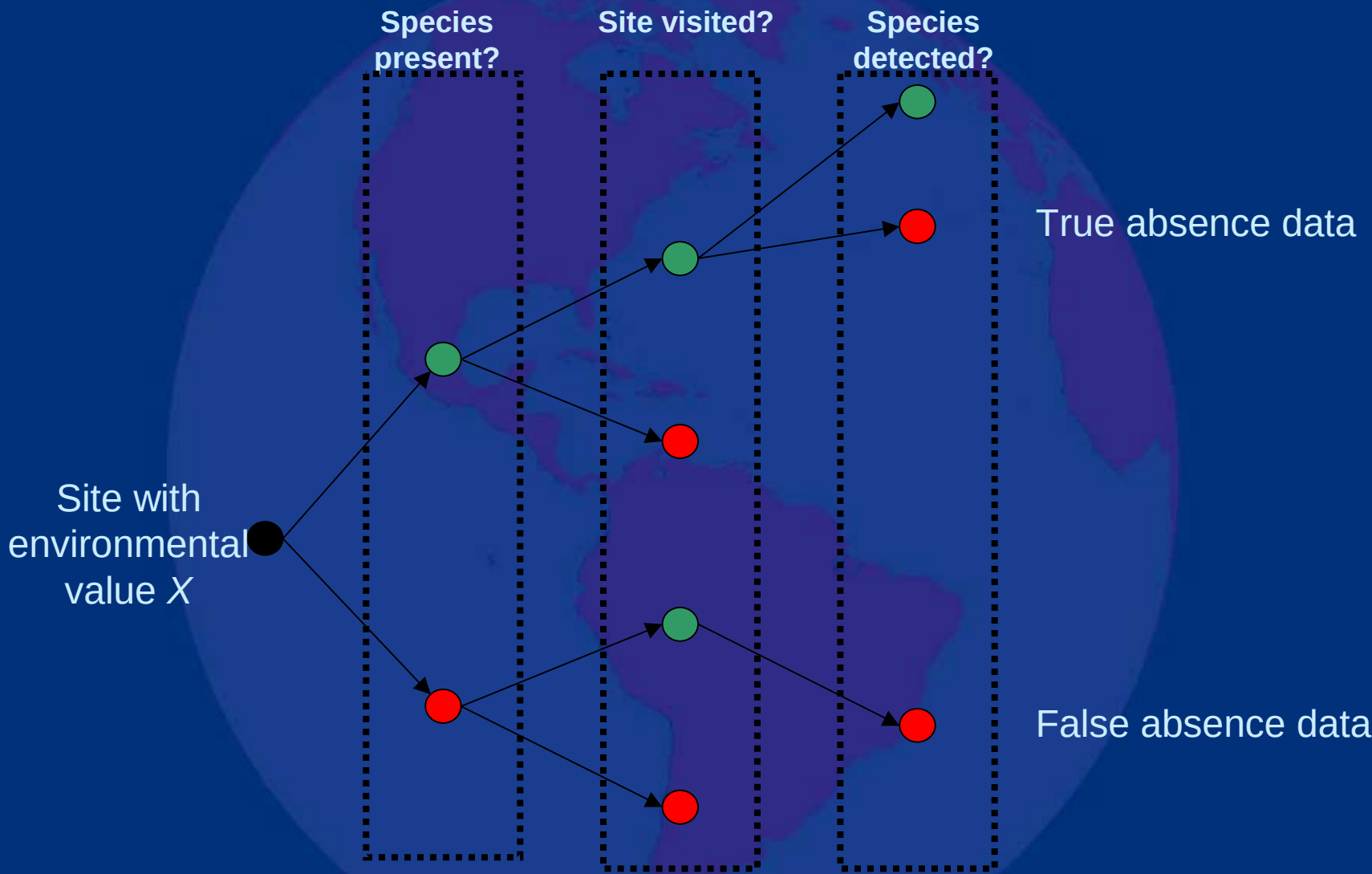
Absence Data Issues



M. Nakamura, CONABIO, 2005

WHY?

Absence Data Issues

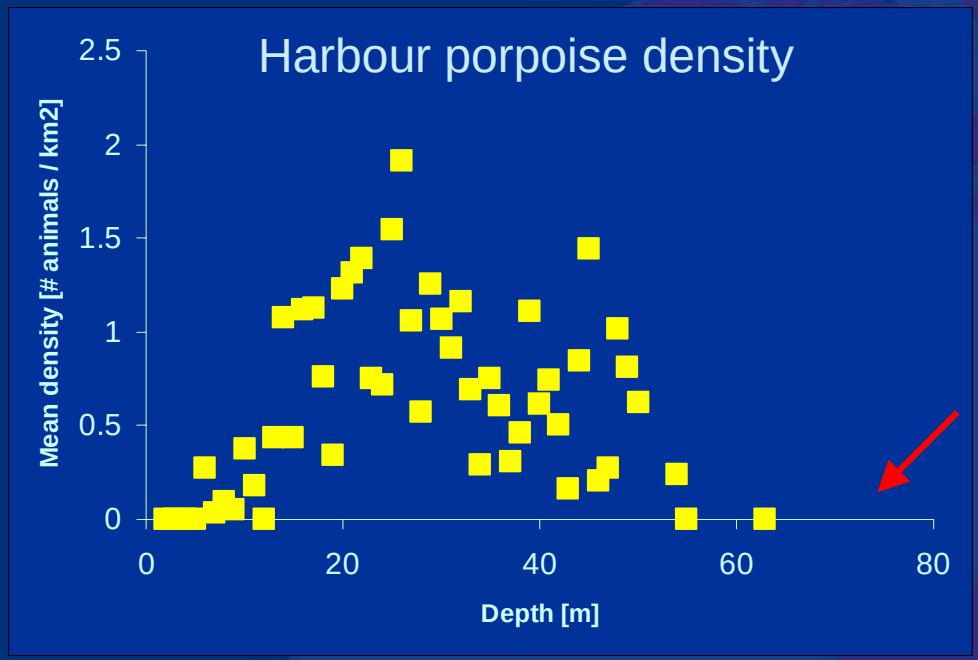


M. Nakamura, CONABIO, 2005

GBIF Ecological Niche Modelling Workshop, KU

WHY?

Absence Data Issues

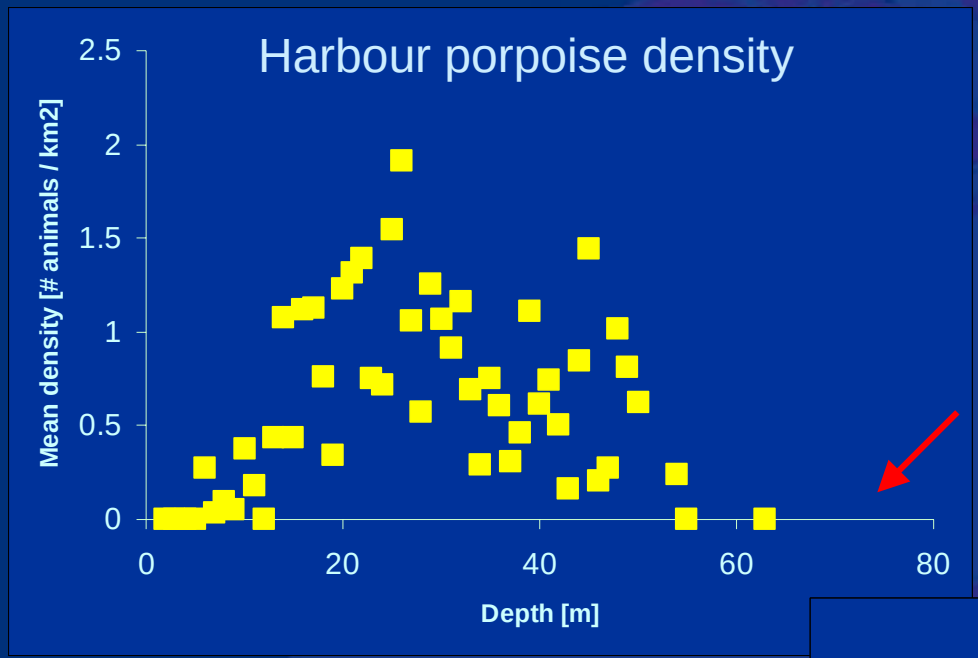


No animals in deeper waters!

Scheidat, Gilles et al, (unpublished data)

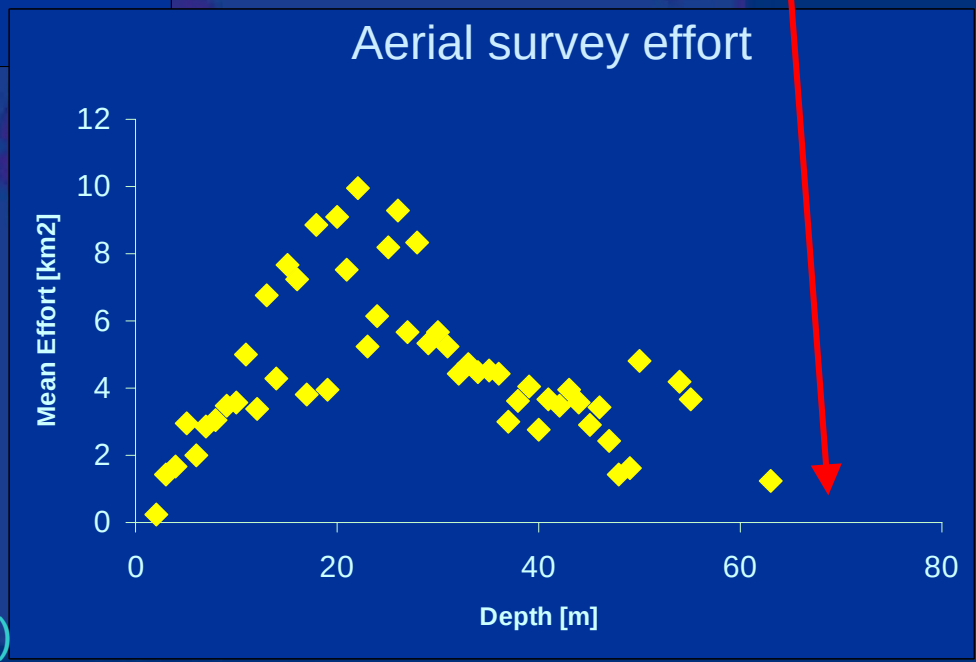
WHY?

Absence Data Issues



No animals in deeper waters?

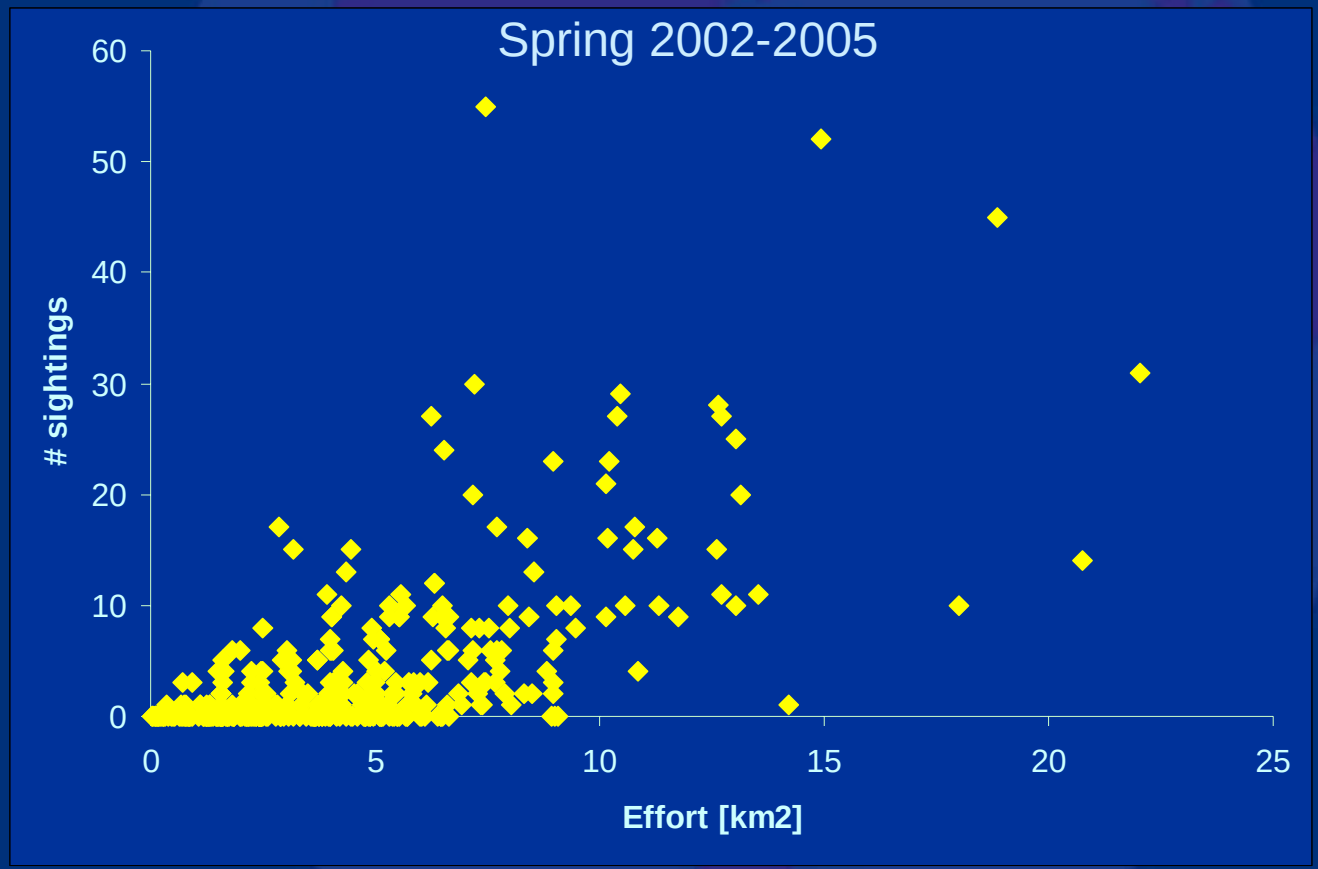
True absences???



Scheidat, Gilles et al, (unpublished data)

WHY?

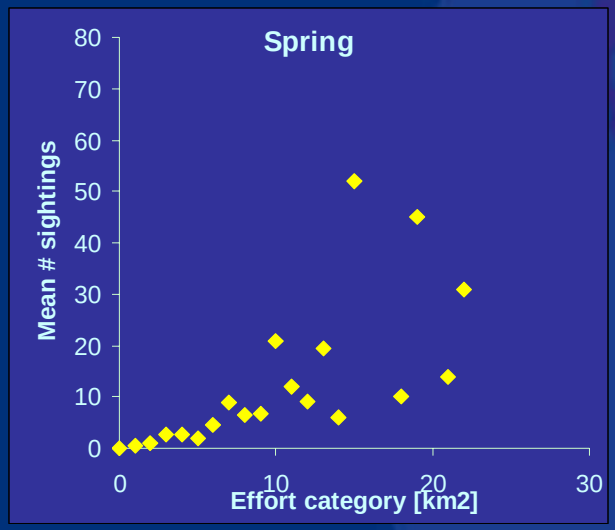
Absence Data Issues

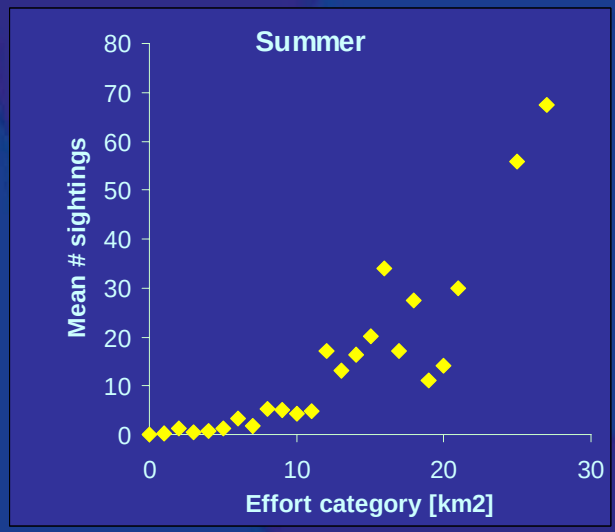


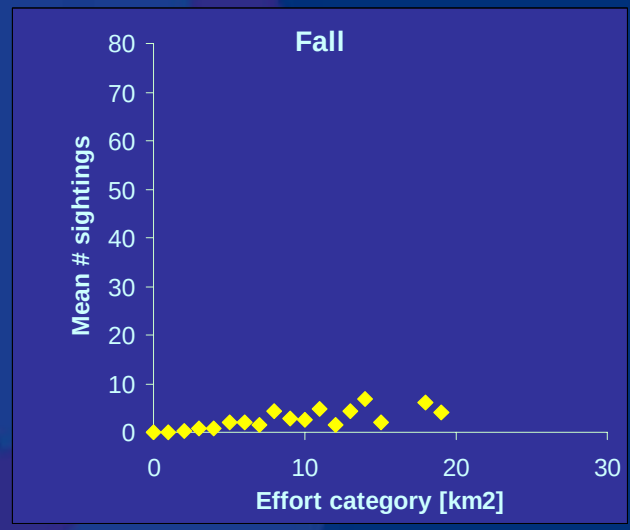
Scheidat, Gilles et al, (unpublished data)

WHY?

Absence Data Issues



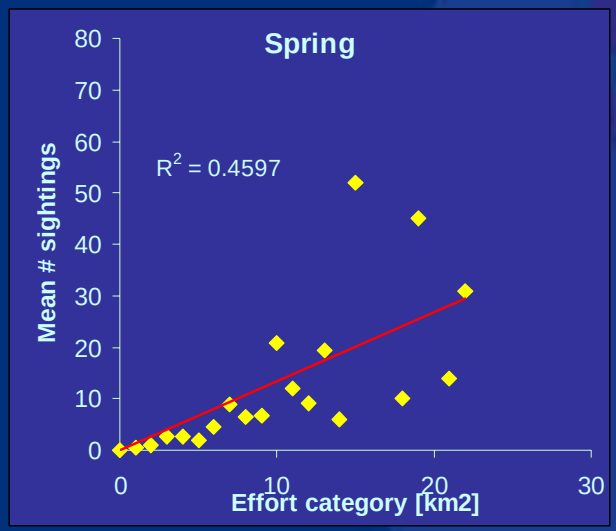


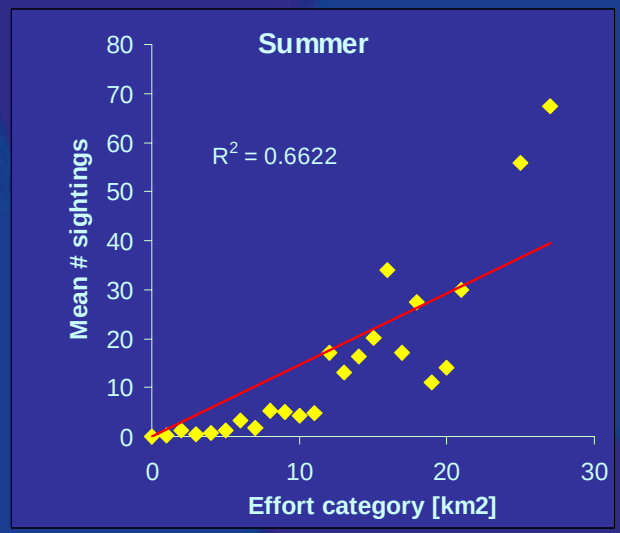


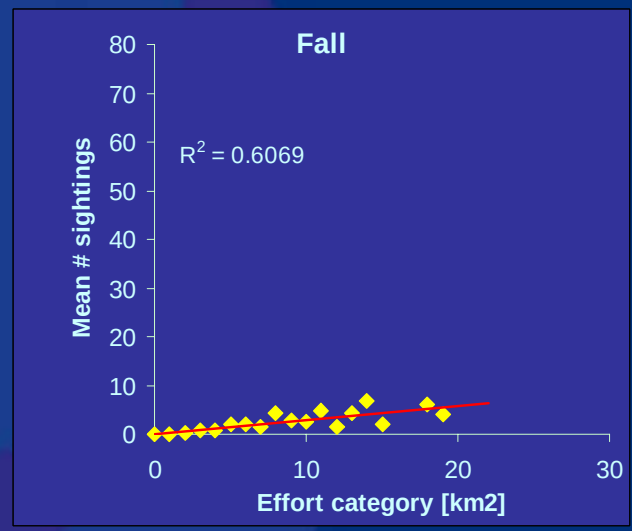
Scheidat, Gilles et al, (unpublished data)

WHY?

Absence Data Issues



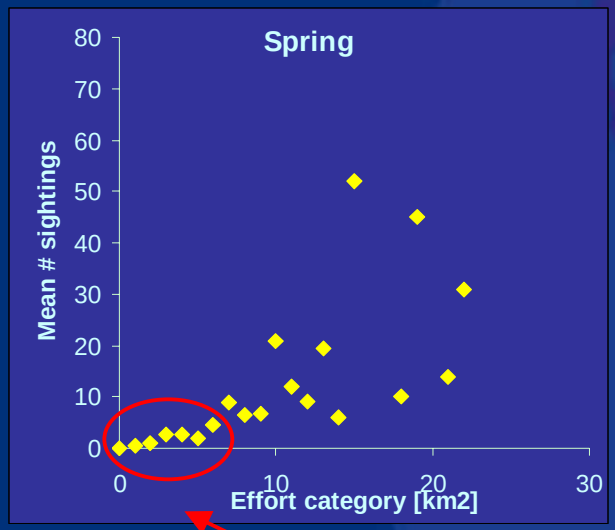


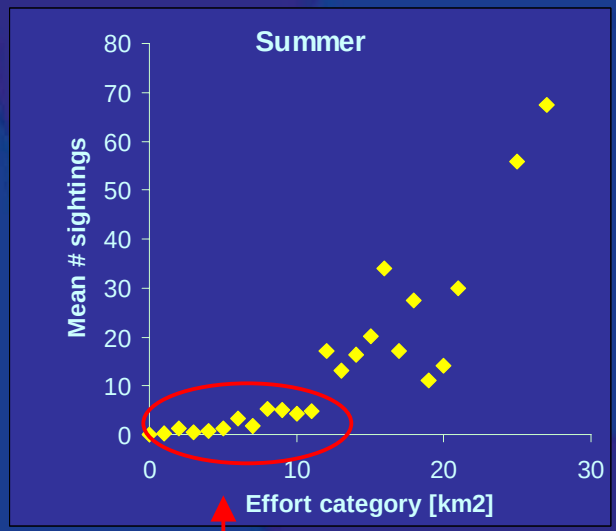


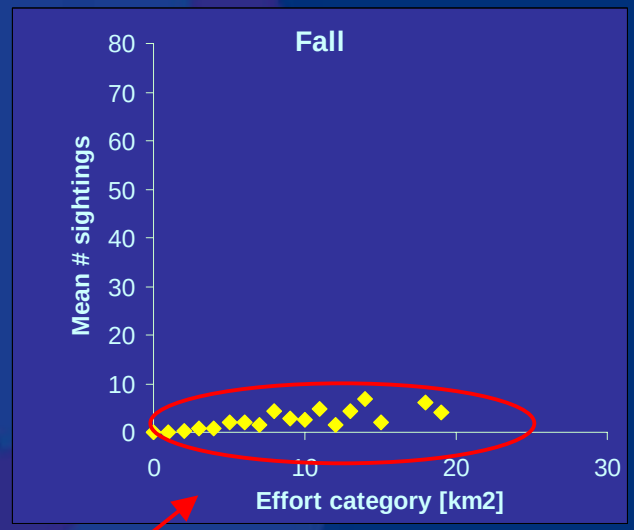
Scheidat, Gilles et al, (unpublished data)

WHY?

Absence Data Issues







True absences???

Scheidat, Gilles et al, (unpublished data)

WHY?

Absence Data Issues

Example: 100 trials

Species #1

	Variable A	Variable B	Variable C
Presence	.20	.80	0
	×	×	×
Bias	.8	.1	.1
	=	=	=
Obs. rate	.16	.08	0
Observed	16	8	0

Species #2

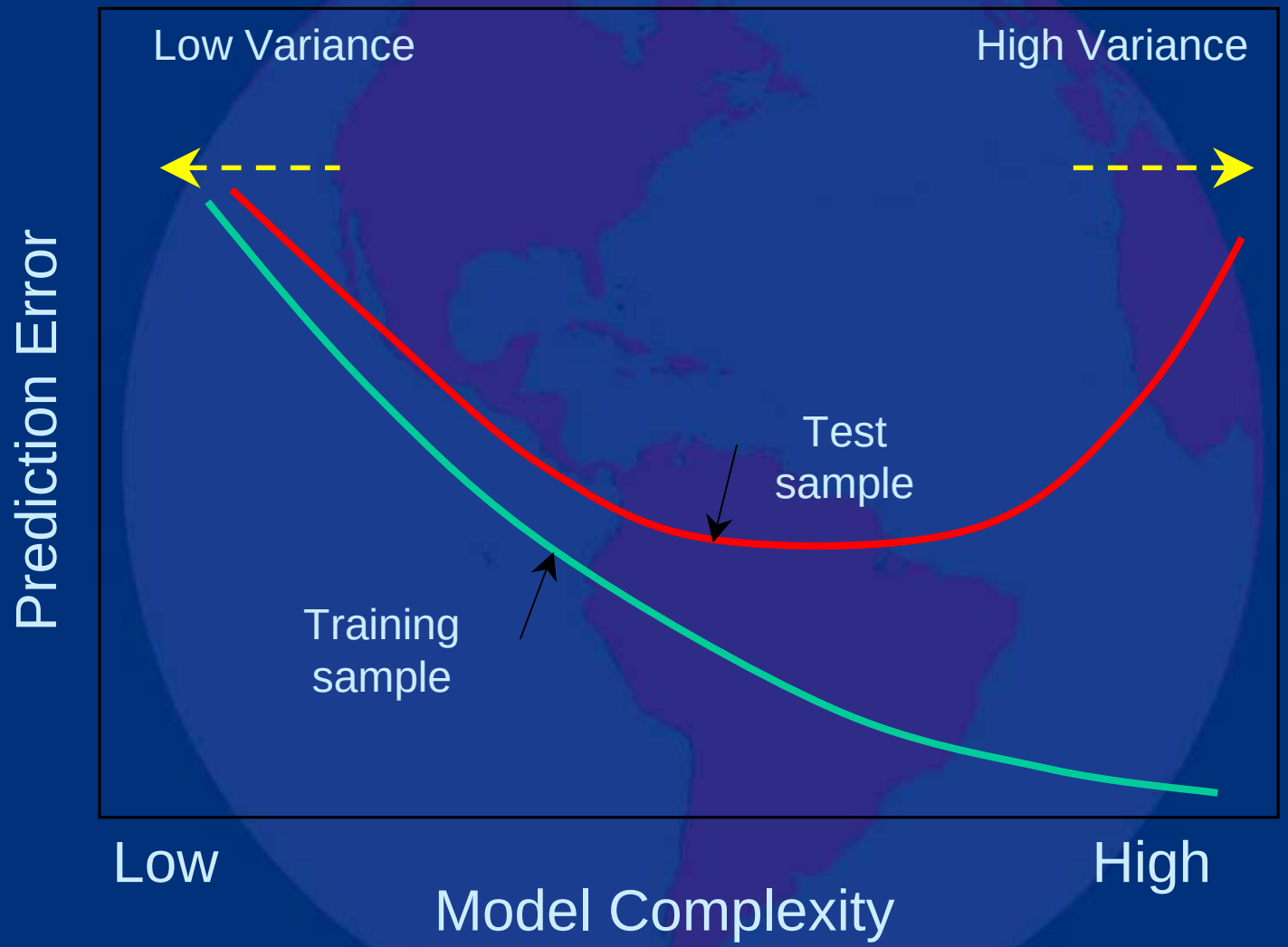
	Variable A	Variable B	Variable C
Presence	.32	.16	.52
	×	×	×
Bias	.5	.5	0
	=	=	=
Obs. rate	.16	.08	0
Observed	16	8	0

M. Nakamura, CONABIO, 2005

GBIF Ecological Niche Modelling Workshop, KU

WHY?

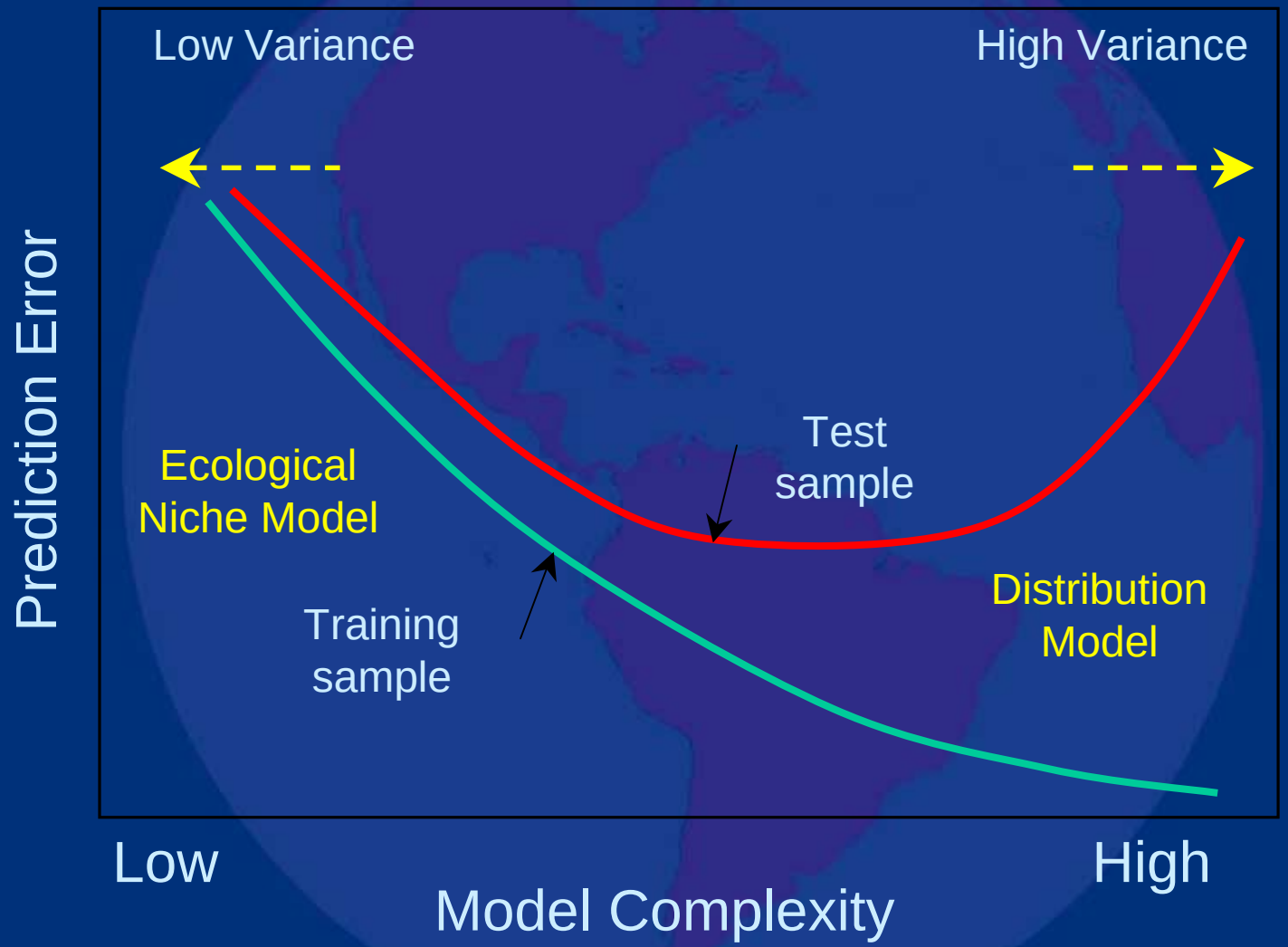
Ecological Niche vs Distribution



Hastie *et al*. (2001)

WHY?

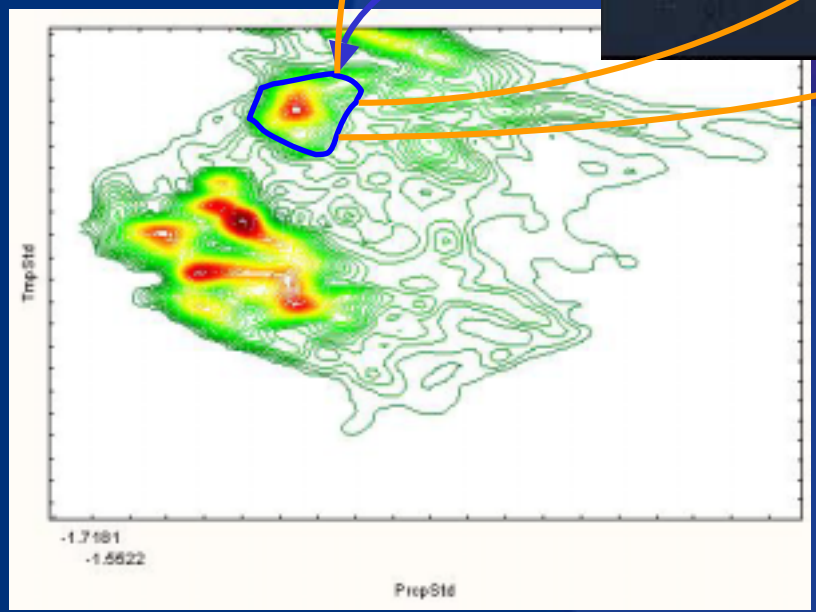
Ecological Niche vs Distribution



Hastie *et al*. (2001)

WHY?

Ecological Niche vs Distribution



Geographic space
(2 dimensions)

Ecological space
(n dimensions / hypervolume)

J. Soberon, CONABIO, 2005
GBIF Ecological Niche Modelling Workshop, KU

WHICH?

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?

Which ones are out there?

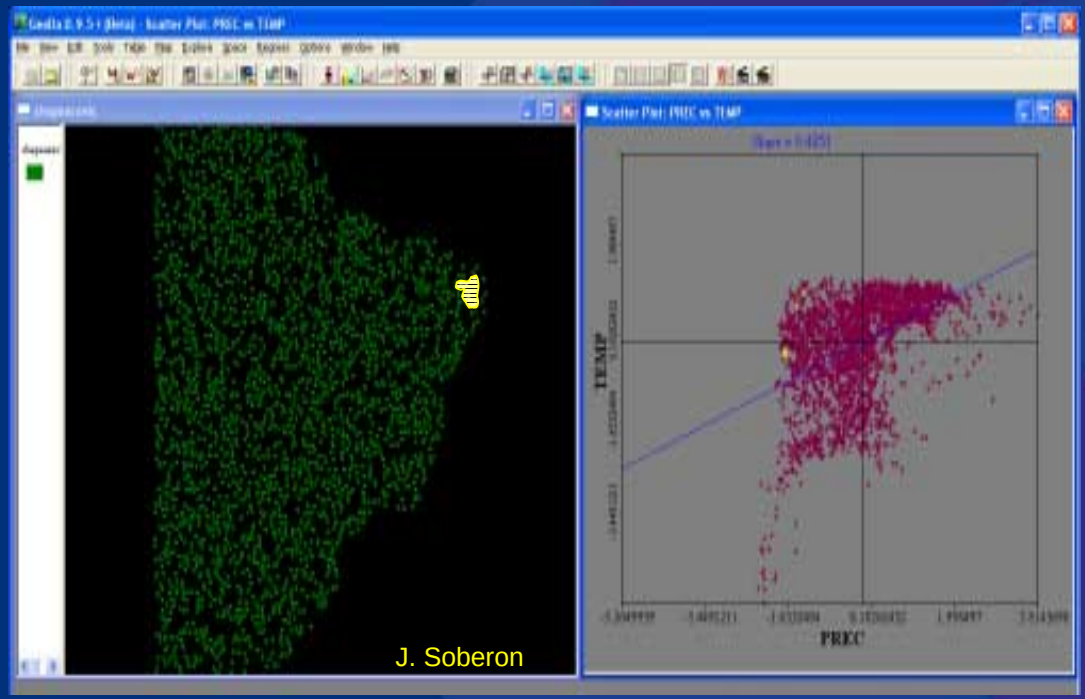
Web-based applications

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

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

WHICH?

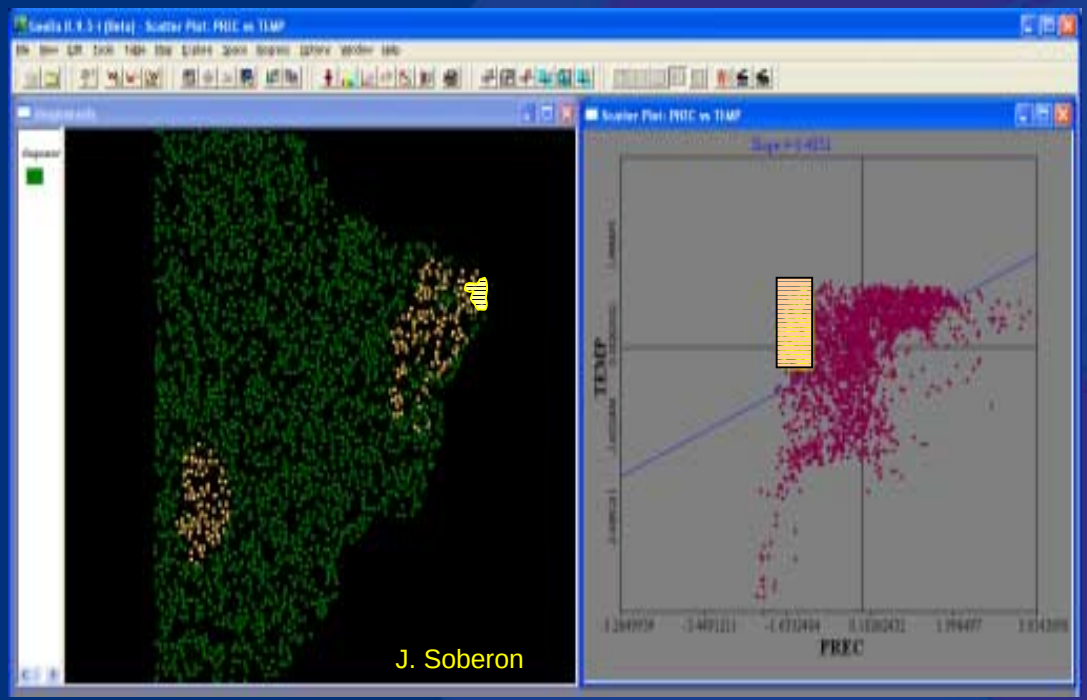
Ecological Niche Models



Geographic space

Ecological / Environmental space

WHICH? BioClim / Climate Envelope Range



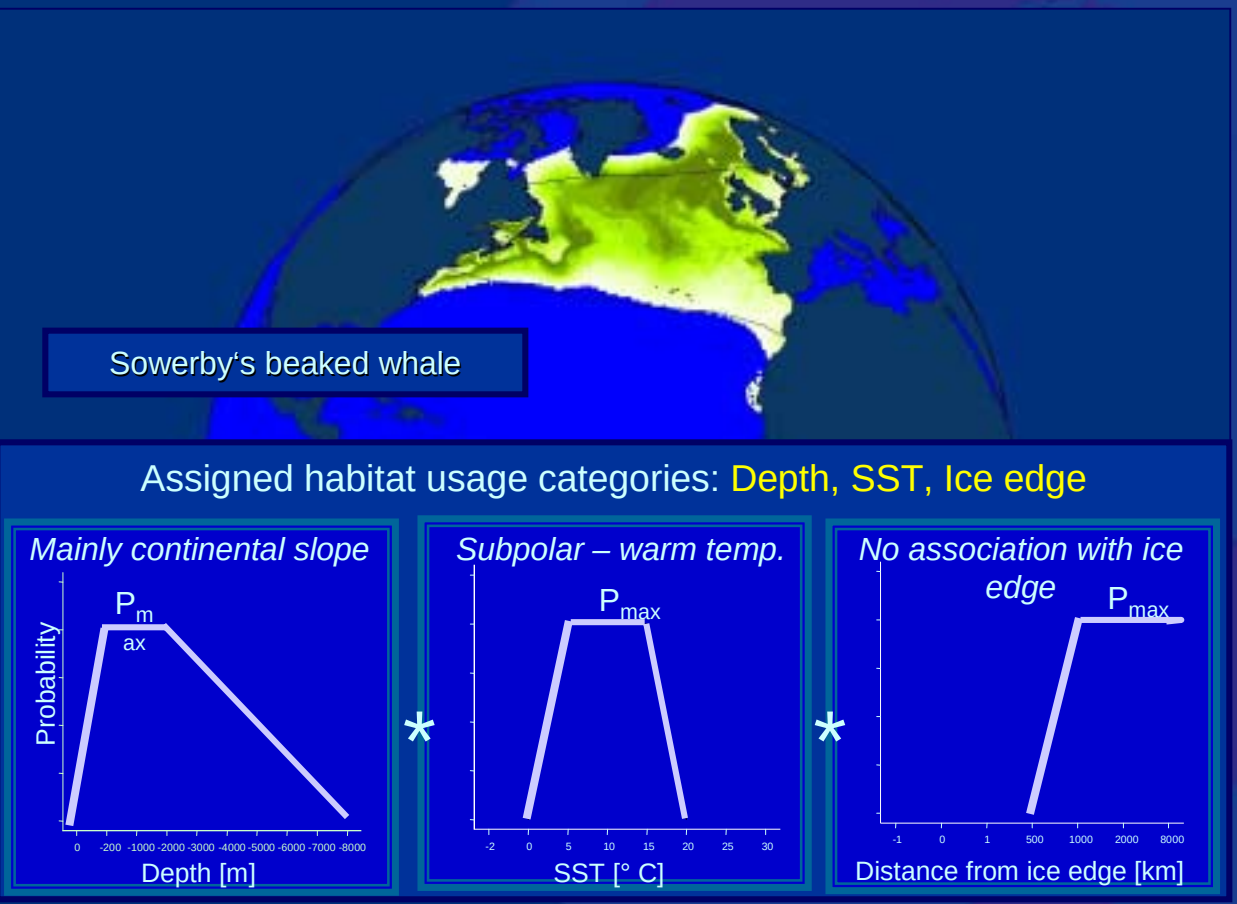
- Based on presence cells
- Very simple & intuitive
- No interactions between variables
- Unweighted variables
- Binary predictions
- No extrapolations
- Tends to overpredict

Lindenmayer et al. 1991 *J. Biogeog.* 18: 371-383.
Arcscript: http://arcscripts.esri.com/details.asp?dbid=13745

WHICH?

RES

Relative Environmental Suitability Model (Fuzzy Bioclimatic Envelope Model)



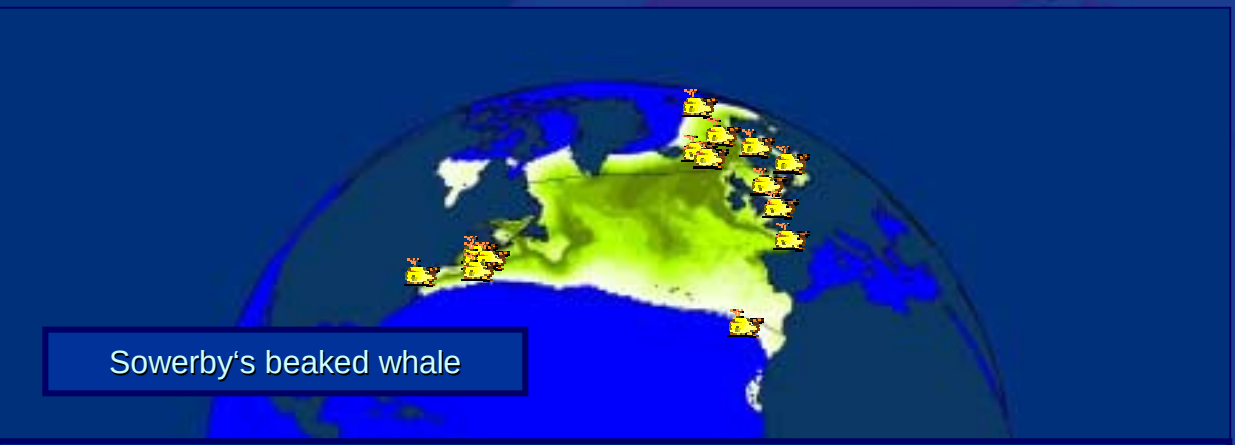
- Based on relative occurrences
- Very simple, transparent & intuitive
- Expert knowledge based (no point data required)
- No interactions between variables
- Unweighted variables
- Continuous output
- Tends to overpredict

Kaschner et al. 2006, *MEPS* / www.seaaroundus.org
Skov & Svenning, 2004, *Ecography*, 27:366-380

WHICH?

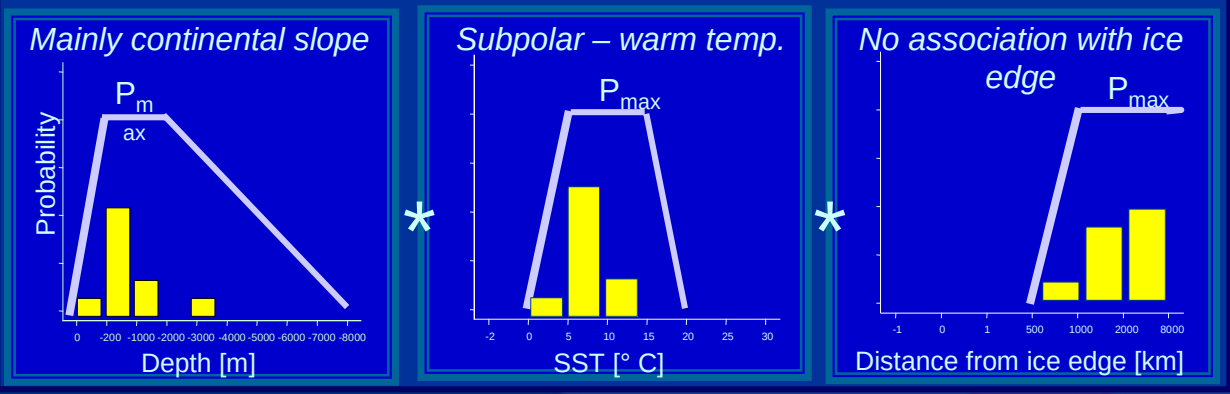
RES

Relative Environmental Suitability Model (Fuzzy Bioclimatic Envelope Model)



Sowerby‘s beaked whale

Assigned habitat usage categories: Depth, SST, Ice edge

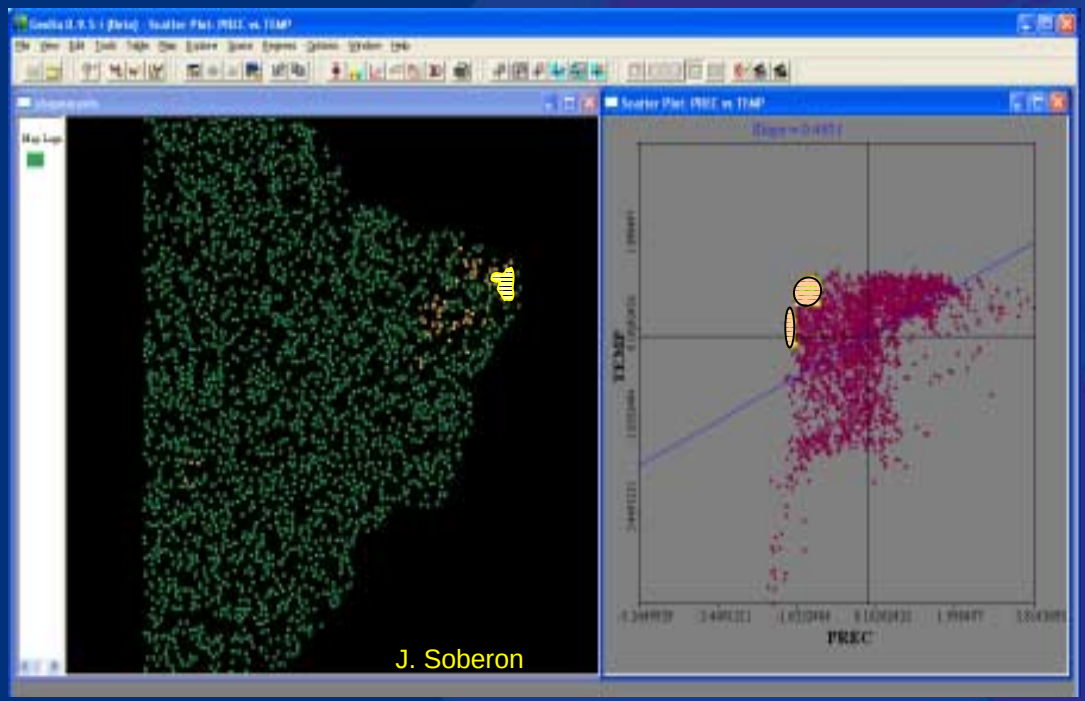


- Based on relative occurrences
- Very simple, transparent & intuitive
- Expert knowledge based (no point data required)
- No interactions between variables
- Unweighted variables
- Continuous output
- Tends to overpredict

Kaschner et al. 2006, *MEPS* / www.seaaroundus.org
Skov & Svenning, 2004, *Ecography*, 27:366-380

WHICH?

DOMAIN



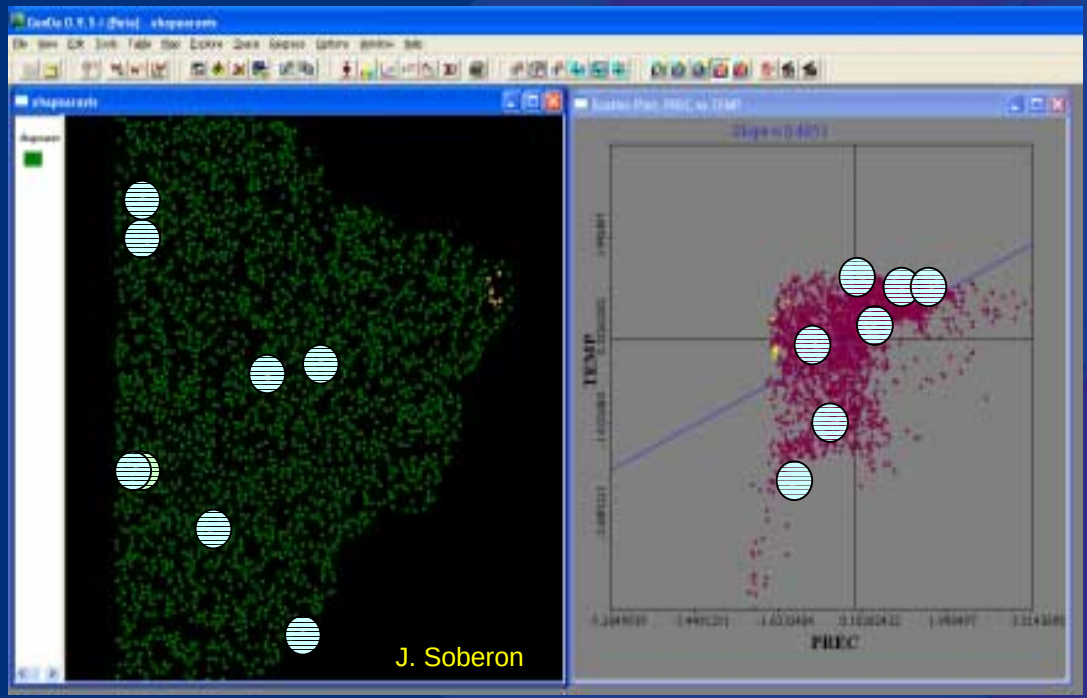
- Based on presence cells
- Cluster algorithm in environmental space
- No interactions between variables
- Unweighted variables
- Non-binary predictions
- Tends to overfit

Carpenter et al. 1993 *Biodiv. Conservation* 2: 667-680.
Freeware: http://www.cifor.cgiar.org/docs/_ref/research_tools/domain/

WHICH?

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?

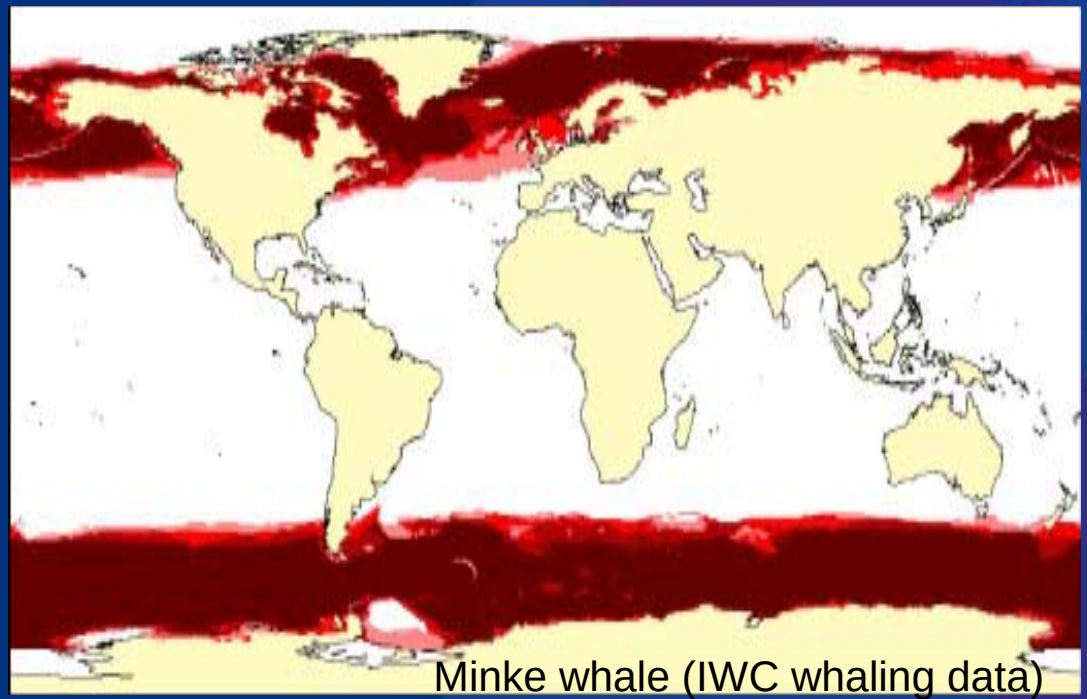
Stockwell & Noble 1992, *Math. & Comp. in Simulation* 33:385-390

Freeware: http://www.lifemapper.org/desktopgarp/

WHICH?

GARP

Genetic Algorithm of Rule-set Prediction



Minke whale (IWC whaling data)

- 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?

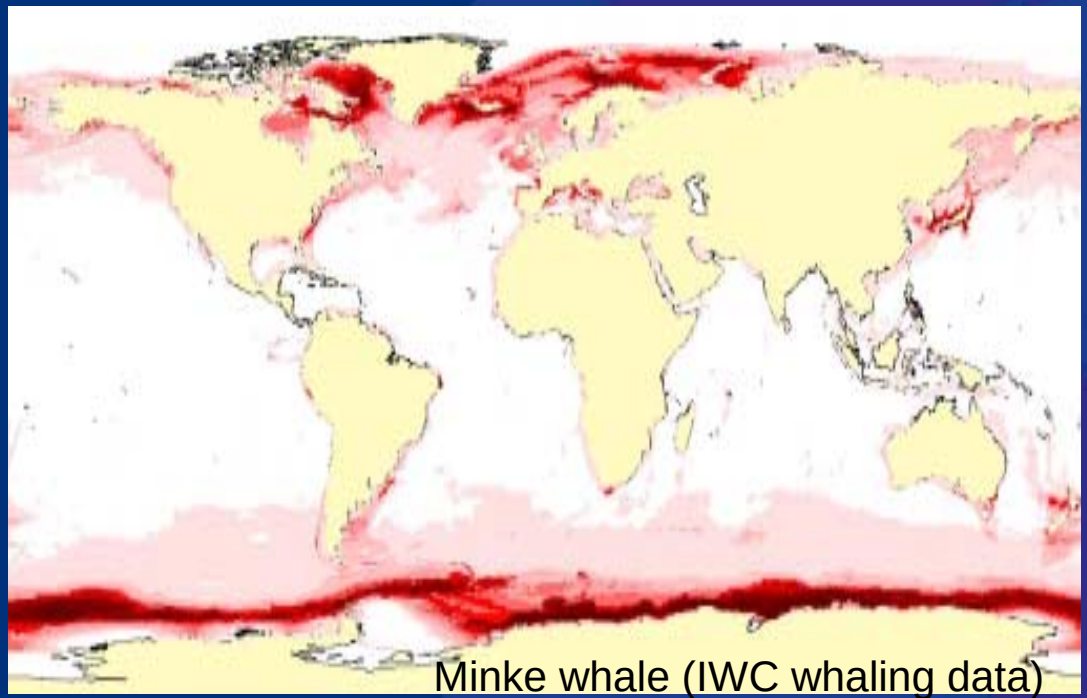
Stockwell & Noble 1992, *Math. & Comp. in Simulation* 33:385-390

Freeware: http://www.lifemapper.org/desktopgarp/

WHICH?

Maxent

Maximum entropy



Minke whale (IWC whaling data)

- 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?

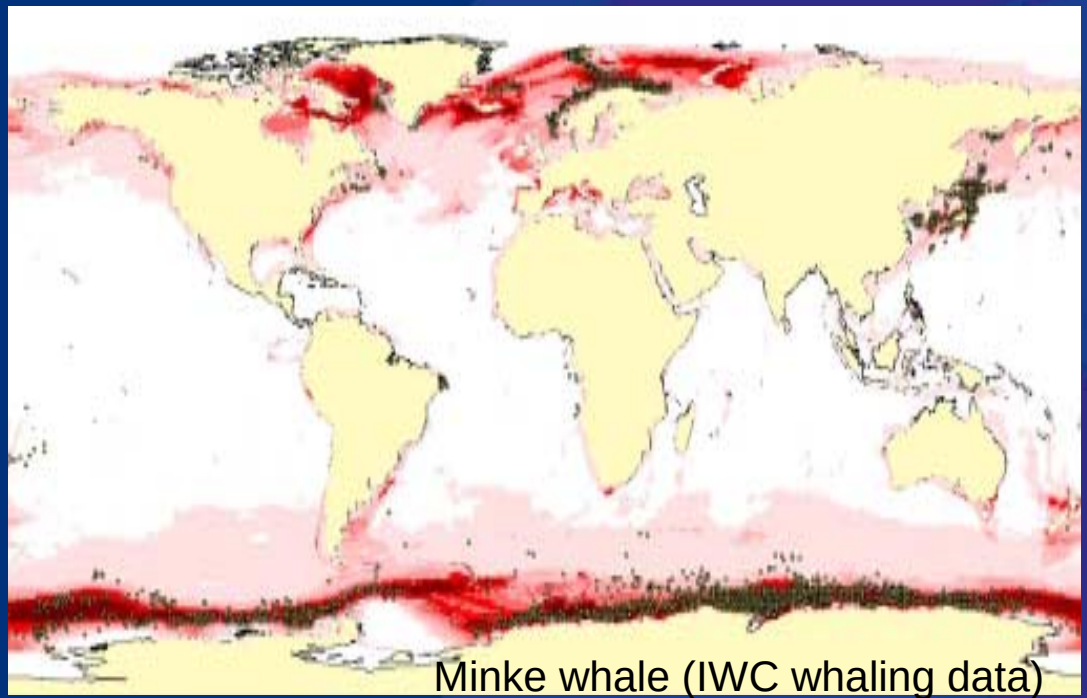
Phillips et al. 2006, *Ecological Modelling*

Freeware: http://www.cs.princeton.edu/~schapire/maxent/

WHICH?

Maxent

Maximum entropy



Minke whale (IWC whaling data)

- 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?

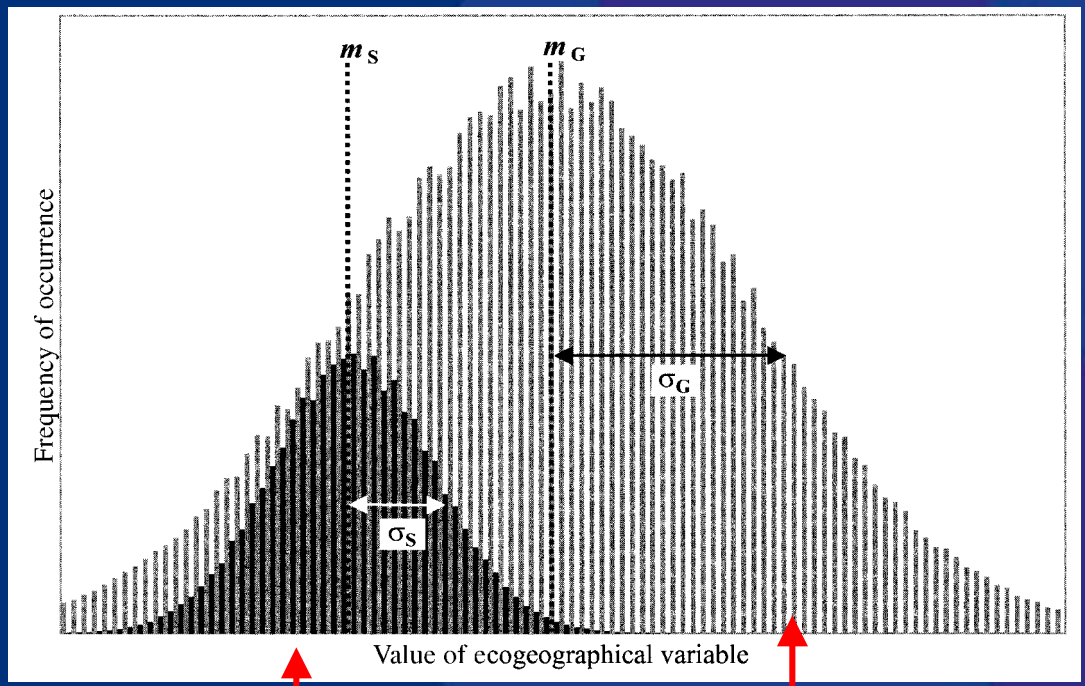
Phillips et al. 2006, *Ecological Modelling*

Freeware: http://www.cs.princeton.edu/~schapire/maxent/

WHICH?

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)

Species occurrence cells

Available habitat

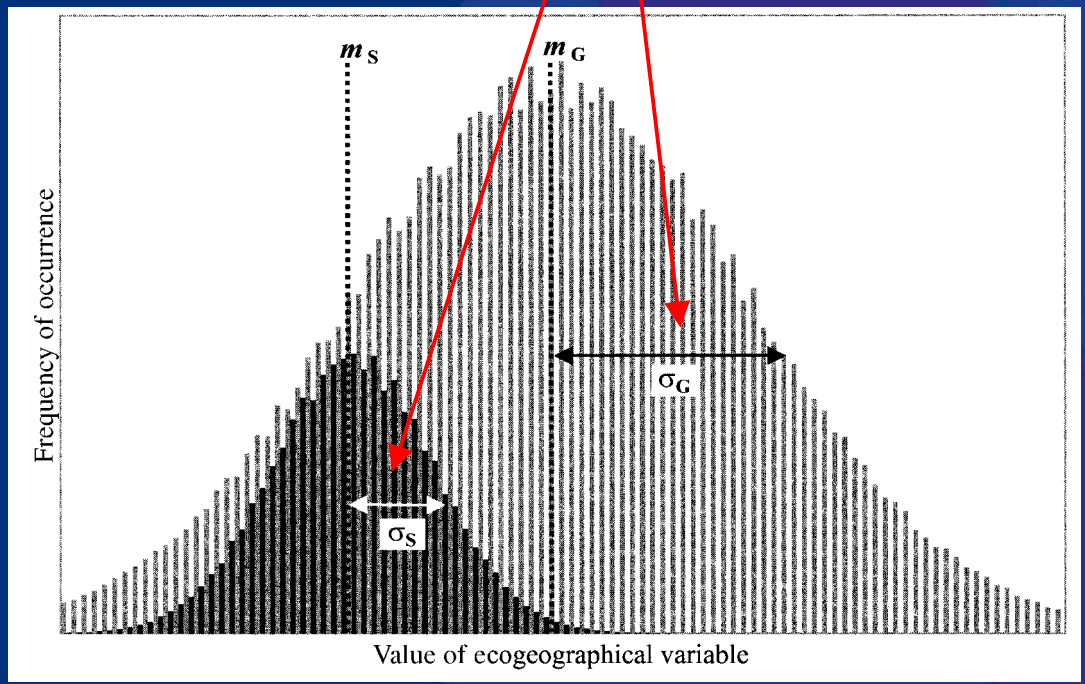
Hirzel et al. 2002, *Ecology* 83: 2027-2036.
Biomapper freeware: http://www2.unil.ch/biomapper/

WHICH?

ENFA

Ecological Niche Factor Analysis

Specialization



Marginality

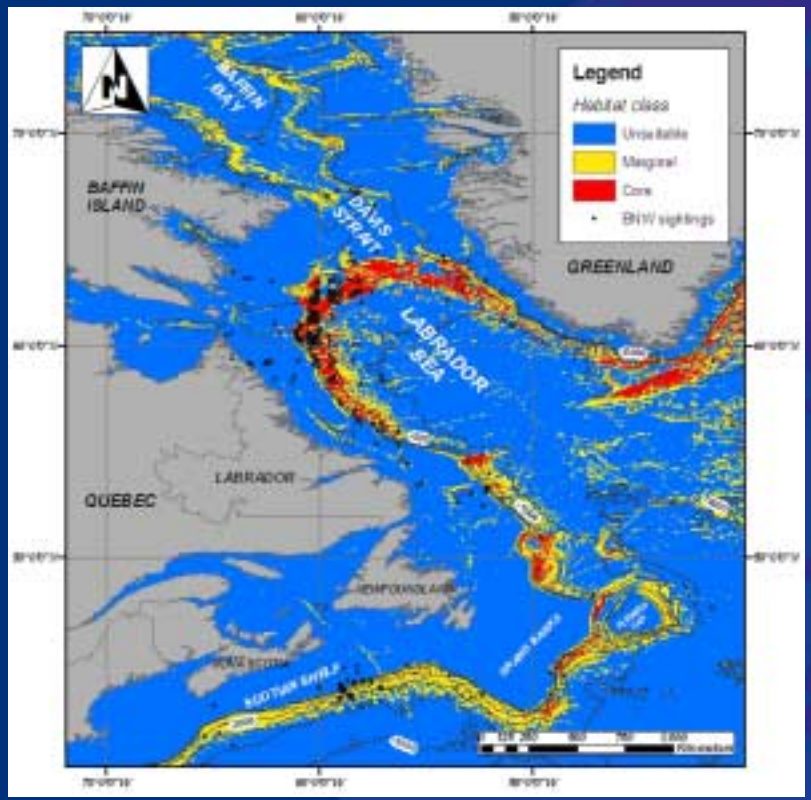
- Based on presence cells
- Compares species occurrence to all available habitat
- Multivariate (Interactions between weighted variables)
- Non-binary predictions (HIS)

Hirzel et al. 2002, *Ecology* 83: 2027-2036.
Biomapper freeware: http://www2.unil.ch/biomapper/

WHICH?

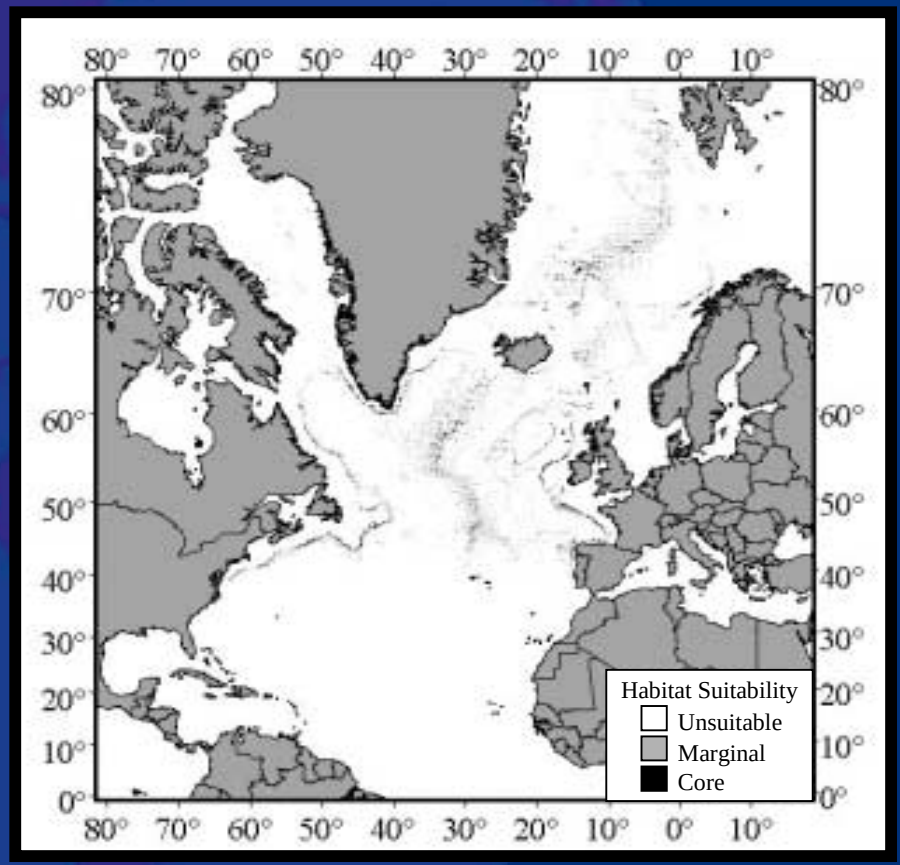
ENFA

Ecological Niche Factor Analysis



Northern bottlenose whale

(NW Atlantic & Total N Atlantic)



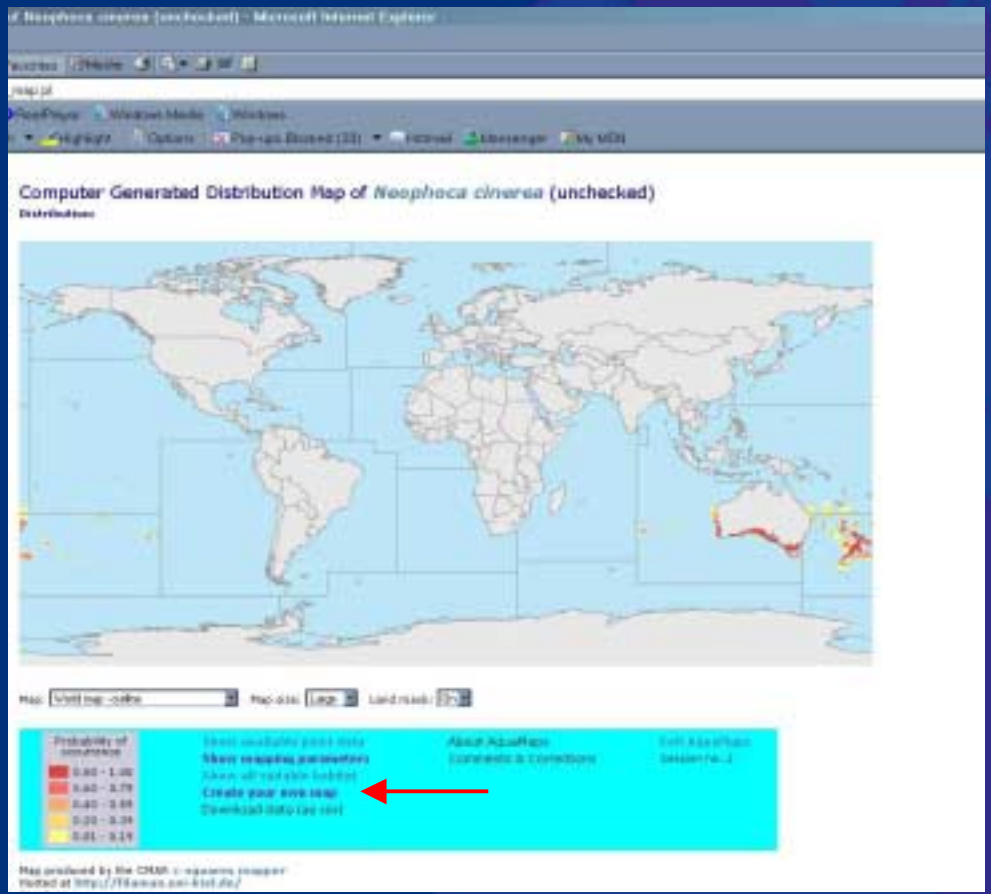
Compton, 2004, MRes Thesis, U of Plymouth

MacLeod, 2005, PhD Thesis, U of Aberdeen

AquaMaps (Coming soon....)

Kaschner et al, in prep,
www.fishbase.de

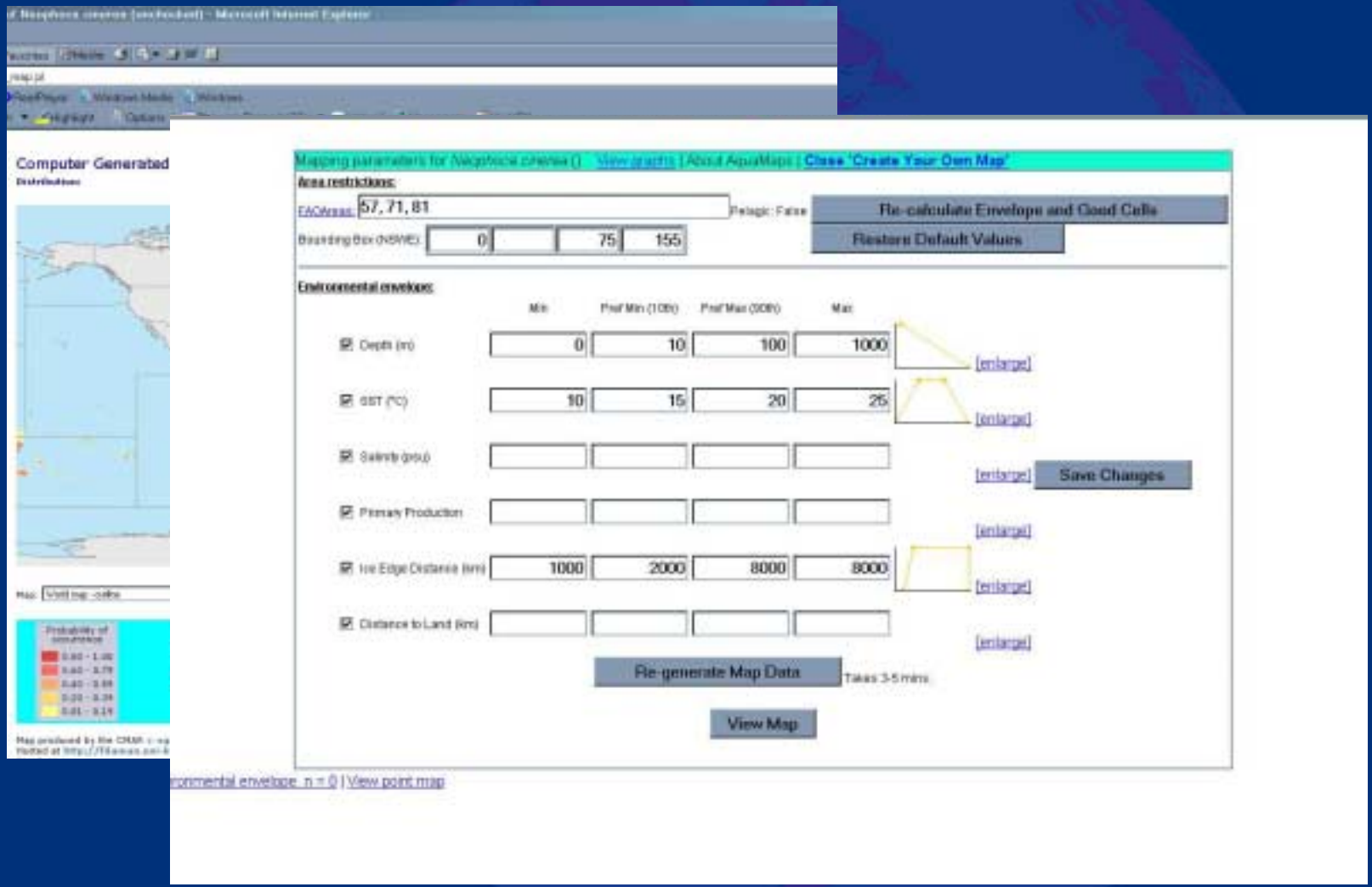
AquaMaps (Coming soon....)



Kaschner et al, in prep,
www.fishbase.de

WHICH?

AquaMaps (Coming soon....)



Kaschner et al, in prep,
www.fishbase.de

DO THEY WORK?

Do they work? Model Evaluations

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

DO THEY WORK?

Test statistics

	Recorded presence	Recorded absence
Predicted presence	a (true presence)	b (false presence)
Predicted absence	c (false absence)	d (true absence)

Sensitivity: Proportion of observed presences correctly predicted

$a/(a + c)$

Test statistics

	Recorded presence	Recorded absence
Predicted presence	a (true presence)	b (false presence)
Predicted absence	c (false absence)	d (true absence)

Sensitivity: Proportion of observed presences correctly predicted

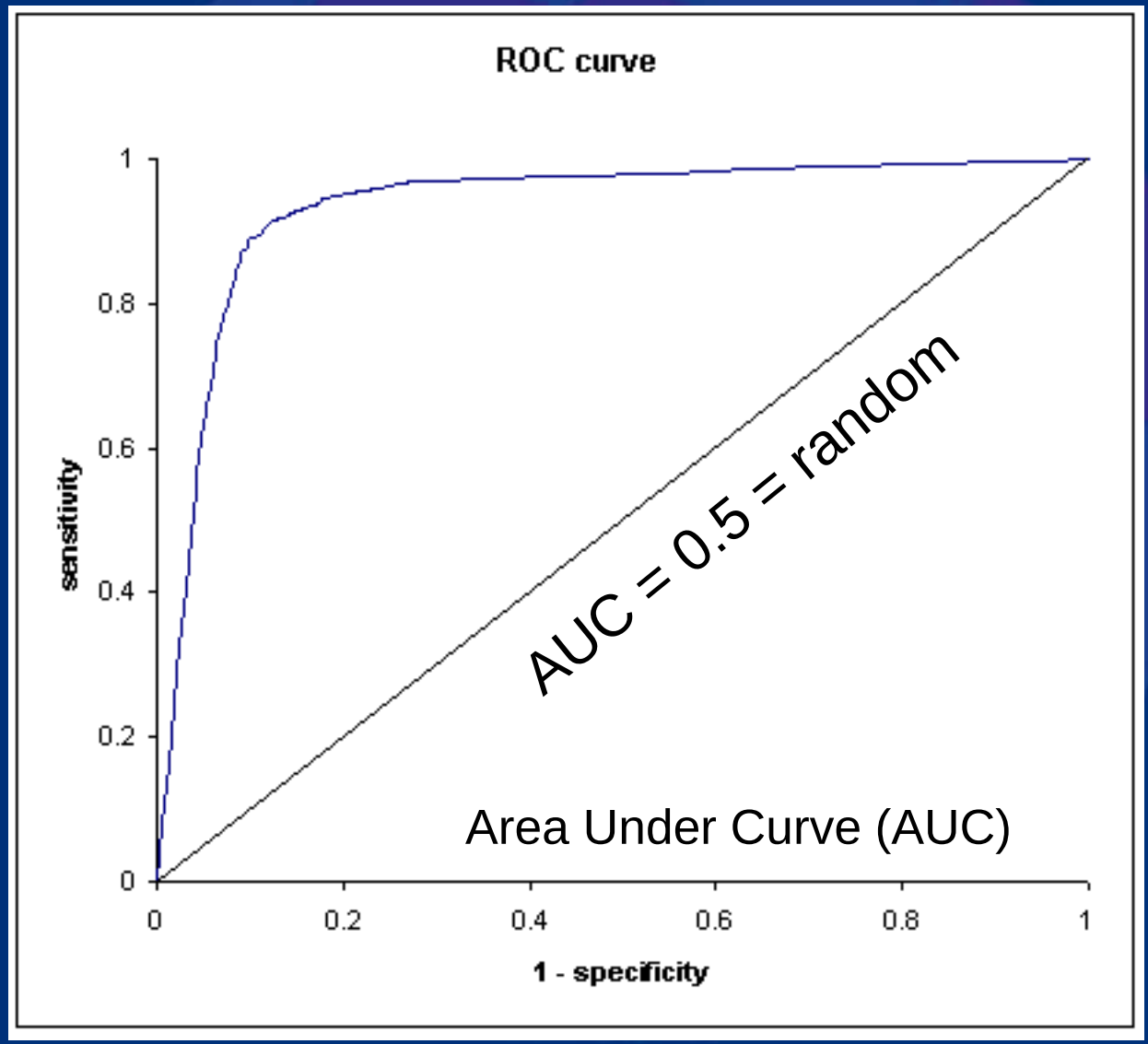
$$a/(a + c)$$

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

$$d/(b + d)$$

DO THEY WORK?

Test statistics: Receiver Operator Curve



DO THEY WORK?

Test statistics

	Recorded presence	Recorded absence
Predicted presence	a (true presence)	b (false presence)
Predicted absence	c (false absence)	d (true absence)

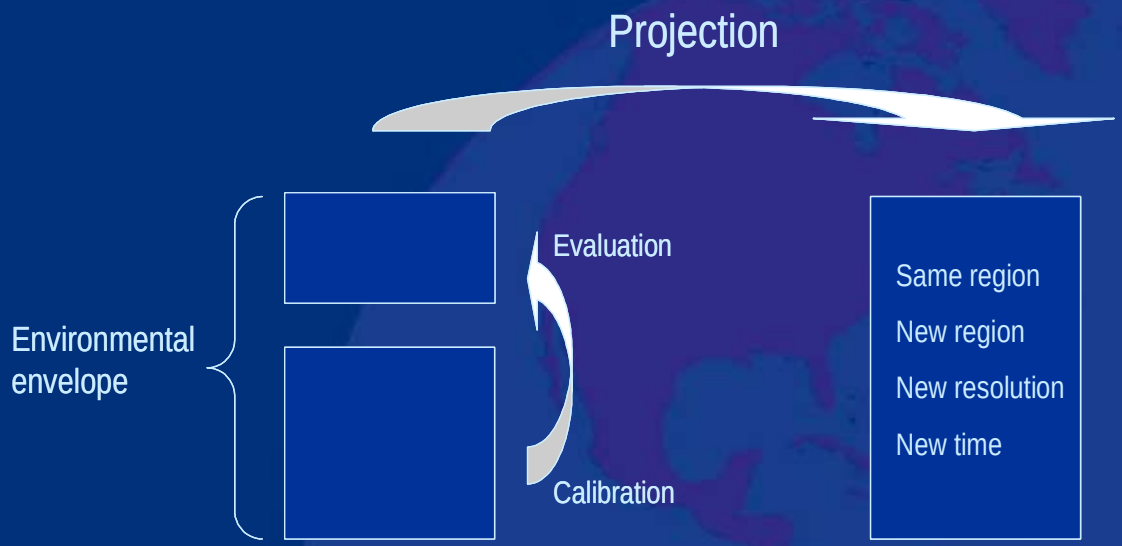
Cohen's Kappa:

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

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

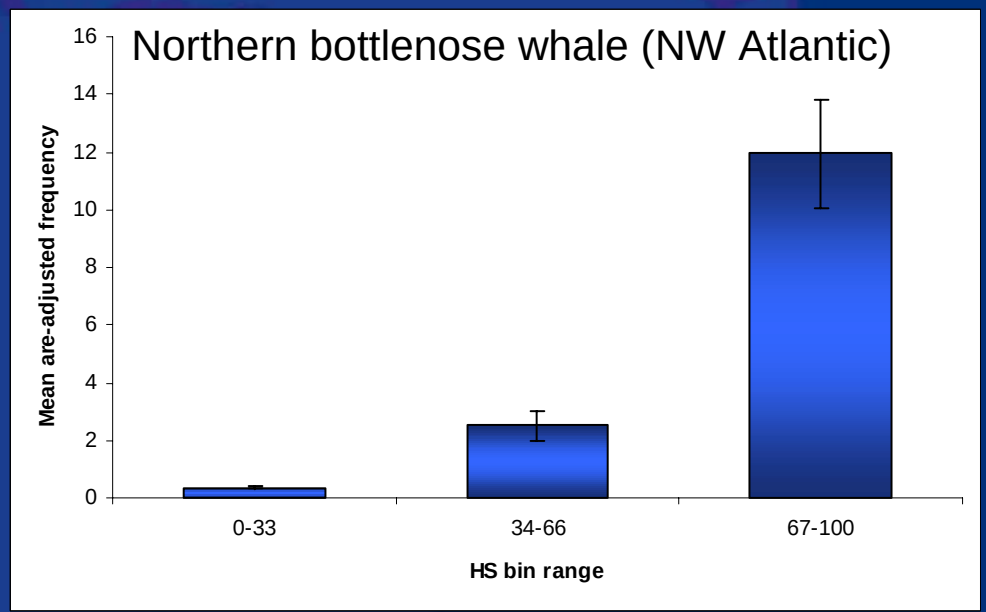
DO THEY WORK?

Cross-validation



Araújo et al. 2005 *Gl. Ch. Biol.*

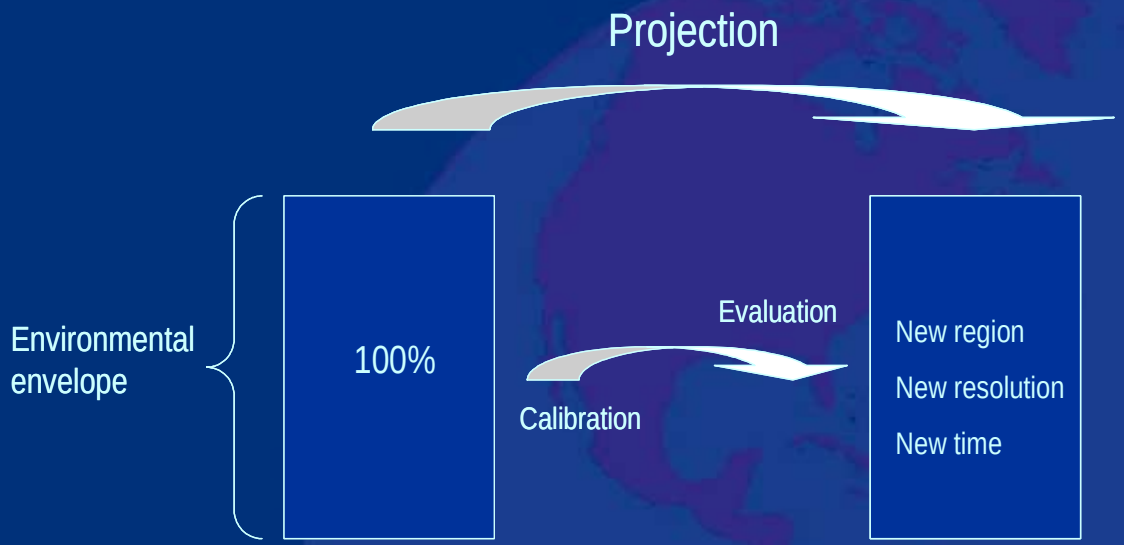
- Same data set
- Jack-knife procedure



Compton, 2004, MRes Thesis, U of Plymouth

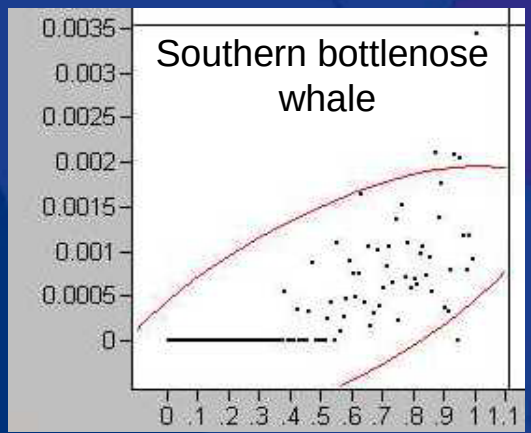
DO THEY WORK?

External Testing



Araújo et al. 2005 *Gl. Ch. Biol.*

- Different data set
- Permutation



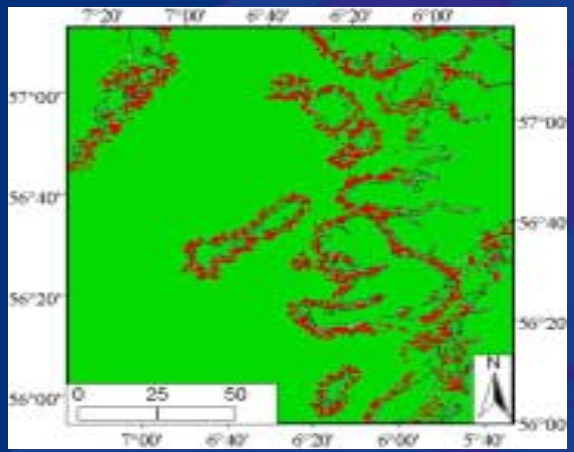
Common name	Spearman's non-parametric rank correlation analysis of RES vs mean SPUE		Comparison with random data sets
	rho	p	Simulated p-value
Northern fur seal	0.54	< 0.0001	0
Harbour porpoise	0.59	< 0.0001	0
Sperm whale	0.66	< 0.0001	0
Antarctic minke whale	0.71	< 0.0001	0
Killer whale	0.56	< 0.0001	0.54
Fin whale	0.53	< 0.0001	0
Blue whale	0.48	< 0.0001	0.268
Humpback whale	0.20	<0.05	0.006
Hourglass dolphin	0.68	< 0.0001	0
Southern bottlenose whale	0.83	< 0.0001	0
Dwarf minke whale	-0.77	< 0.0001	0

Kaschner et al, 2006, *MEPS*

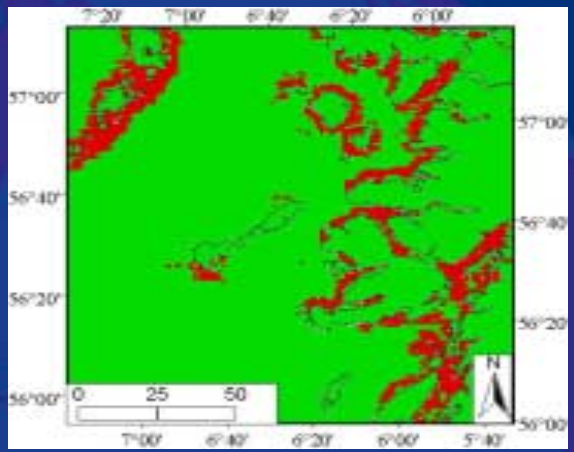
DO THEY WORK?

Model Comparison

PCA

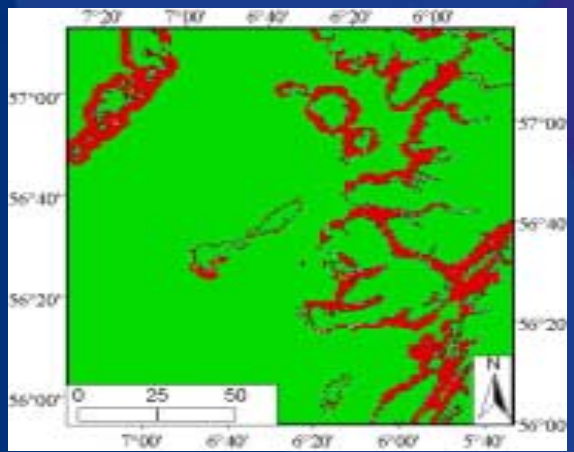


ENFA

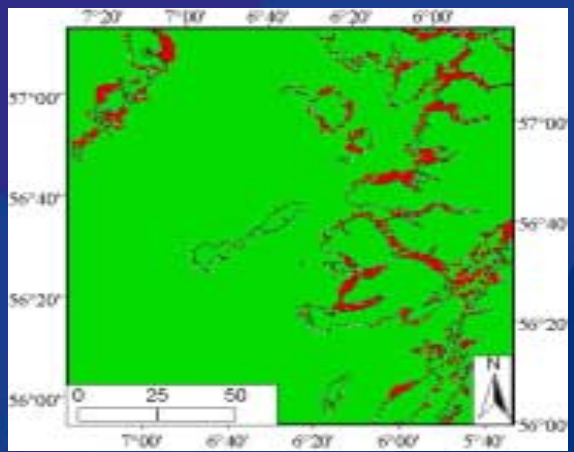


Harbour porpoise

GARP



GLM



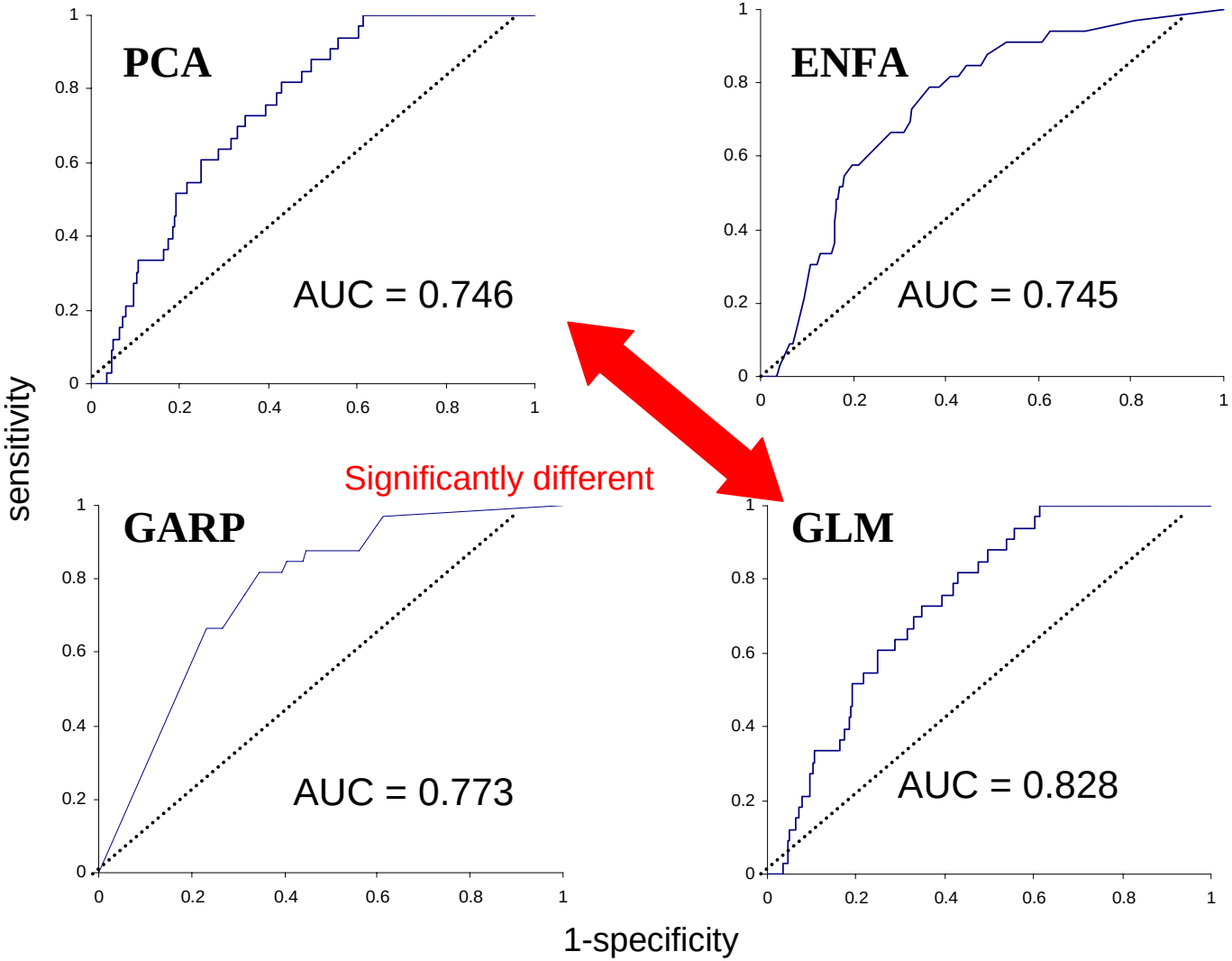
Presence

Absence

Mandelbaum, 2005, MSc Thesis, U of Aberdeen

DO THEY WORK?

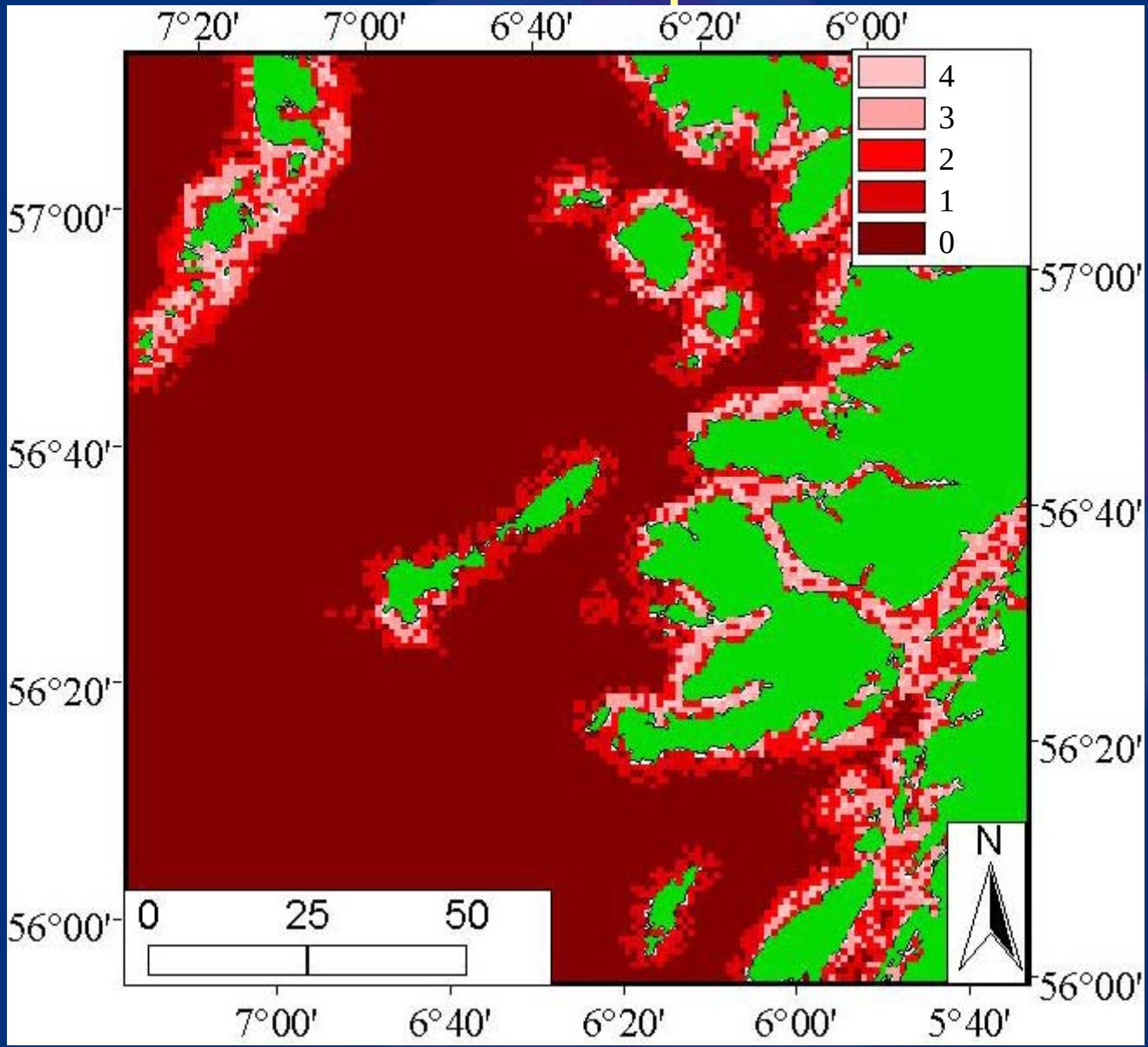
Model Comparison



Mandelbaum, 2005, MSc Thesis, U of Aberdeen

DO THEY WORK?

Model Comparison



Combined model

Mandelbaum, 2005, MSc Thesis, U of Aberdeen

DO THEY WORK?

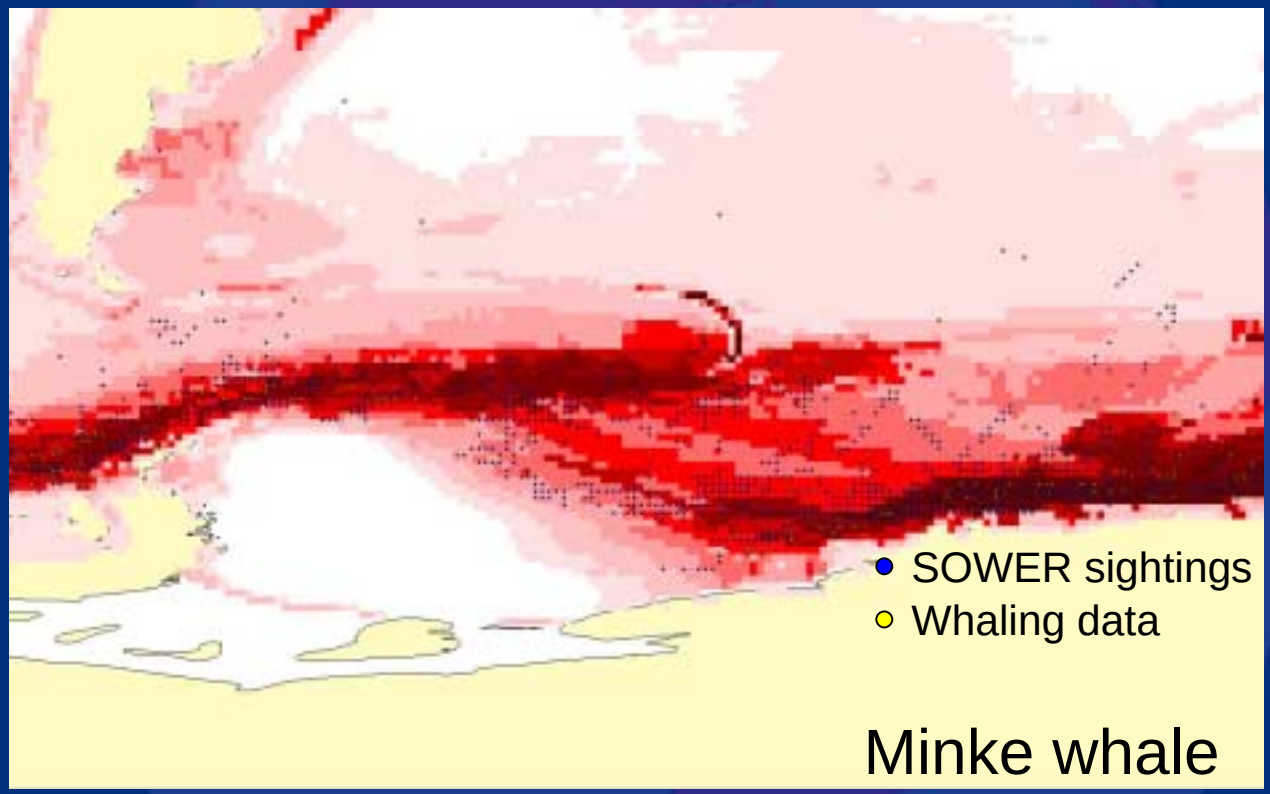
Do they work? Limitations

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

DO THEY WORK?

Presence cells vs Encounter Rate

Maxent prediction (IWC whaling data)

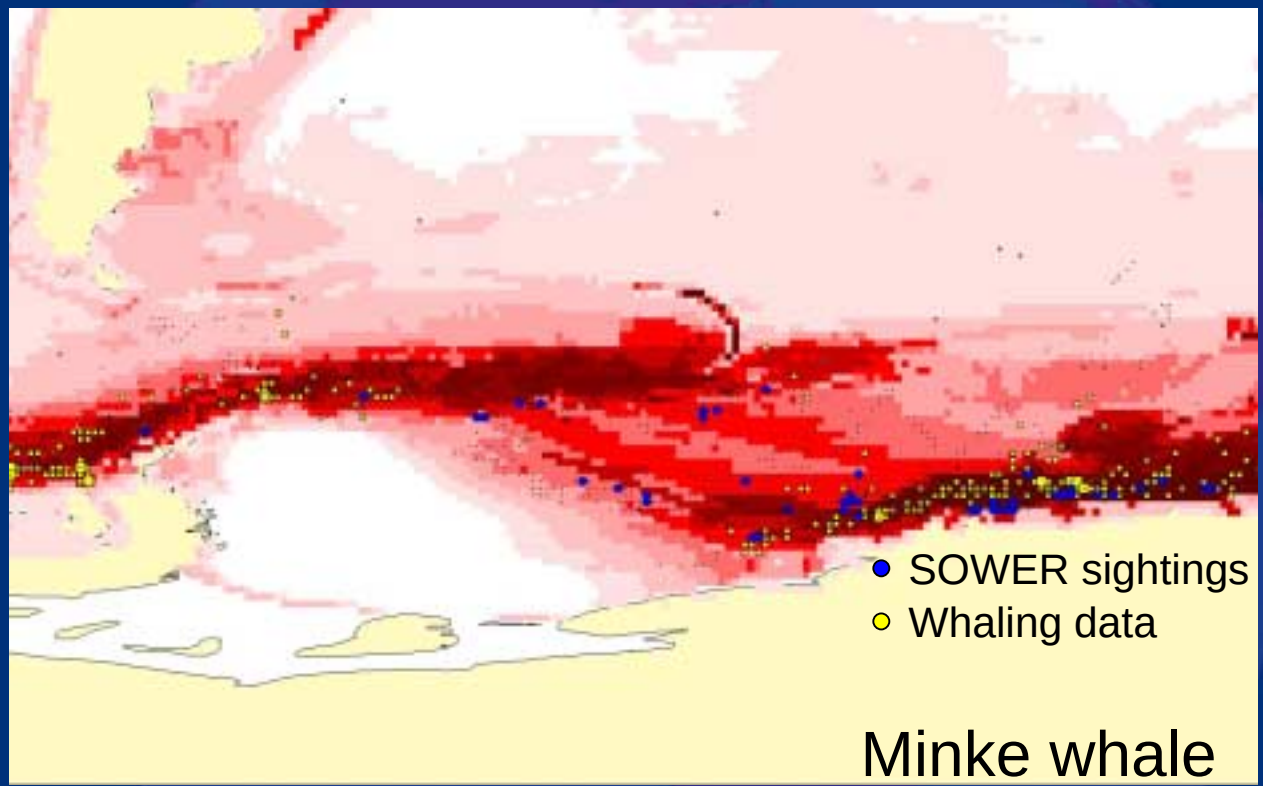


Kaschner et al, in prep

DO THEY WORK?

Presence cells vs Encounter Rate

Maxent prediction (IWC whaling data)

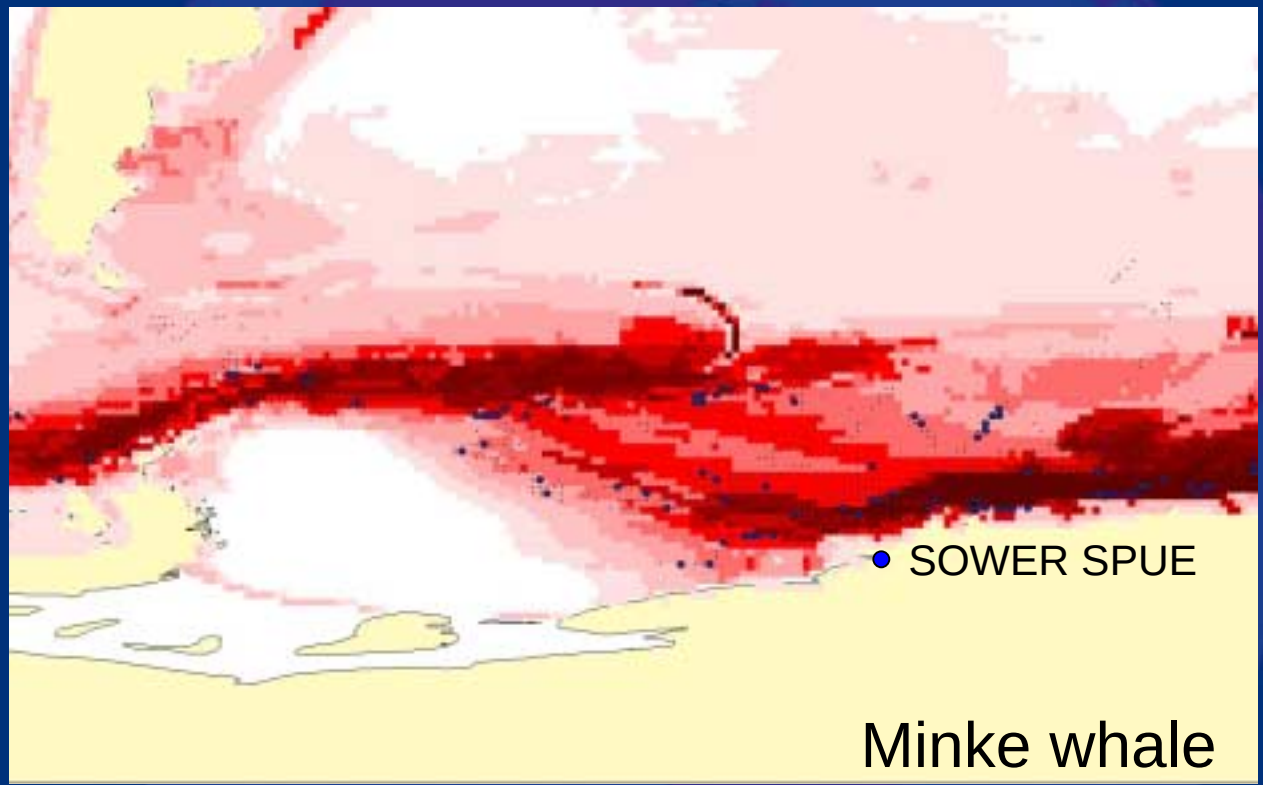


Kaschner et al, in prep

DO THEY WORK?

Presence cells vs Encounter Rate

Maxent prediction (IWC whaling data)

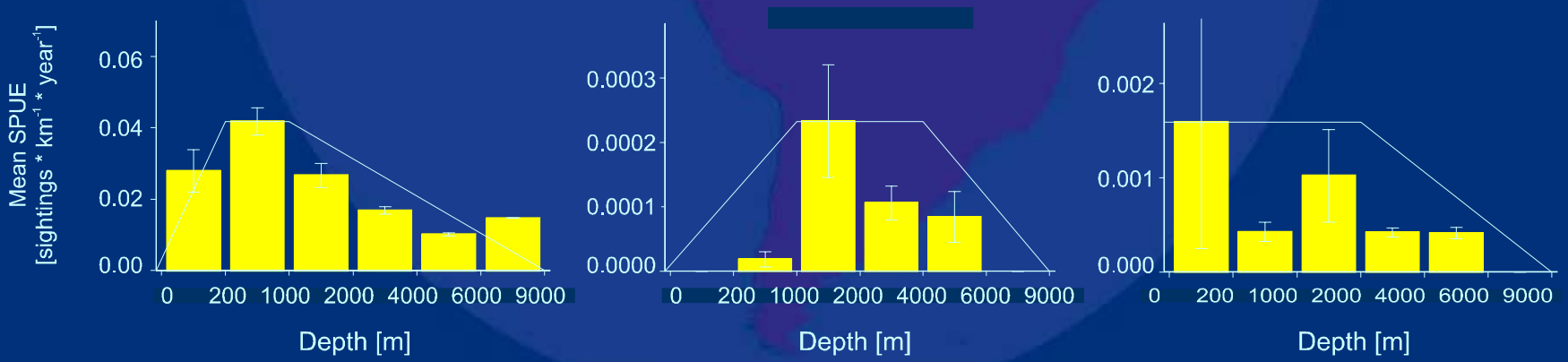


Kaschner et al, in prep

DO THEY WORK?

Sampling Biases

Minke whale | Blue whale | Humpback whale



Kaschner et al, 2006, *MEPS*

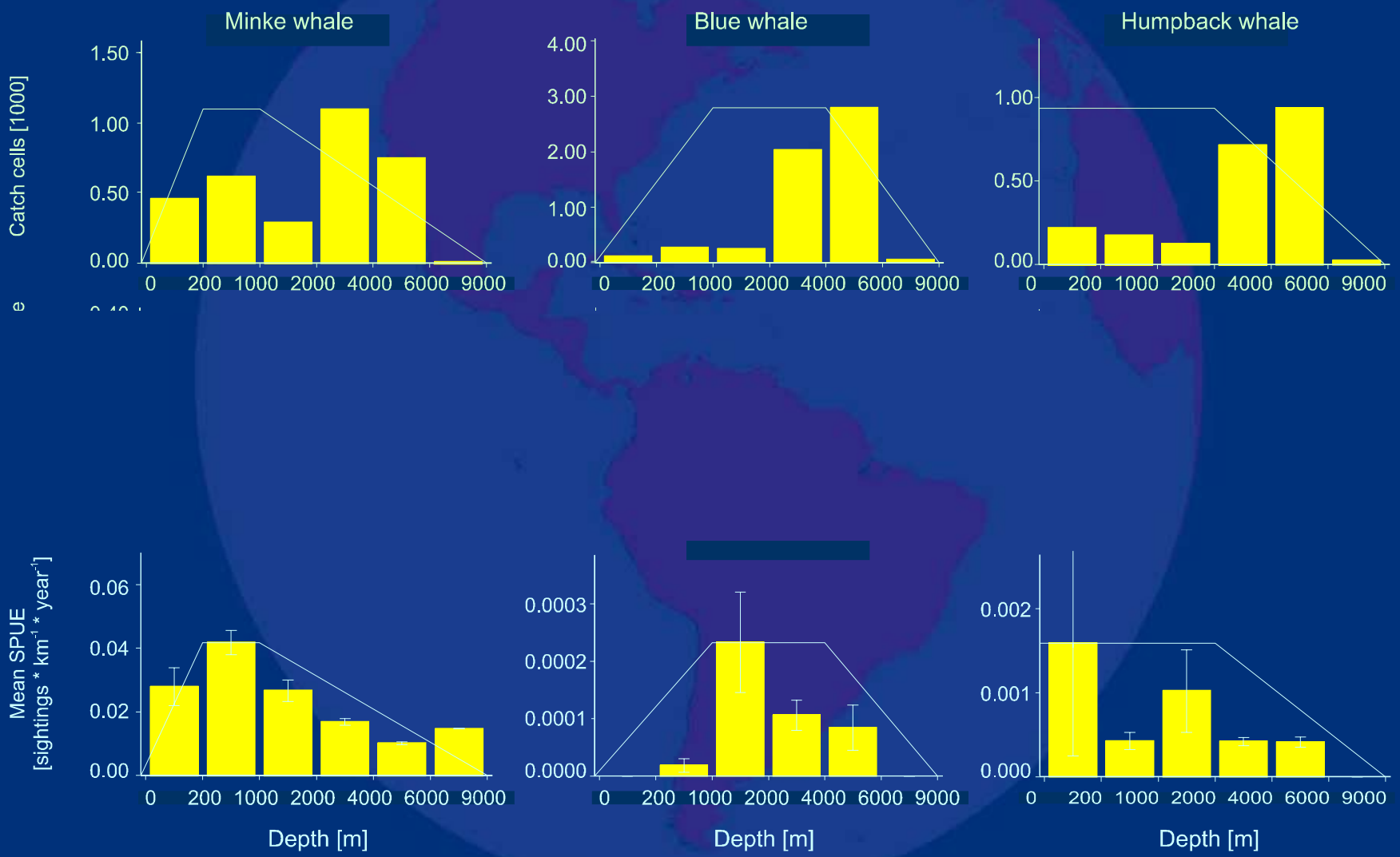
DO THEY WORK?

Sampling Biases

Minke whale | Blue whale | Humpback whale

A

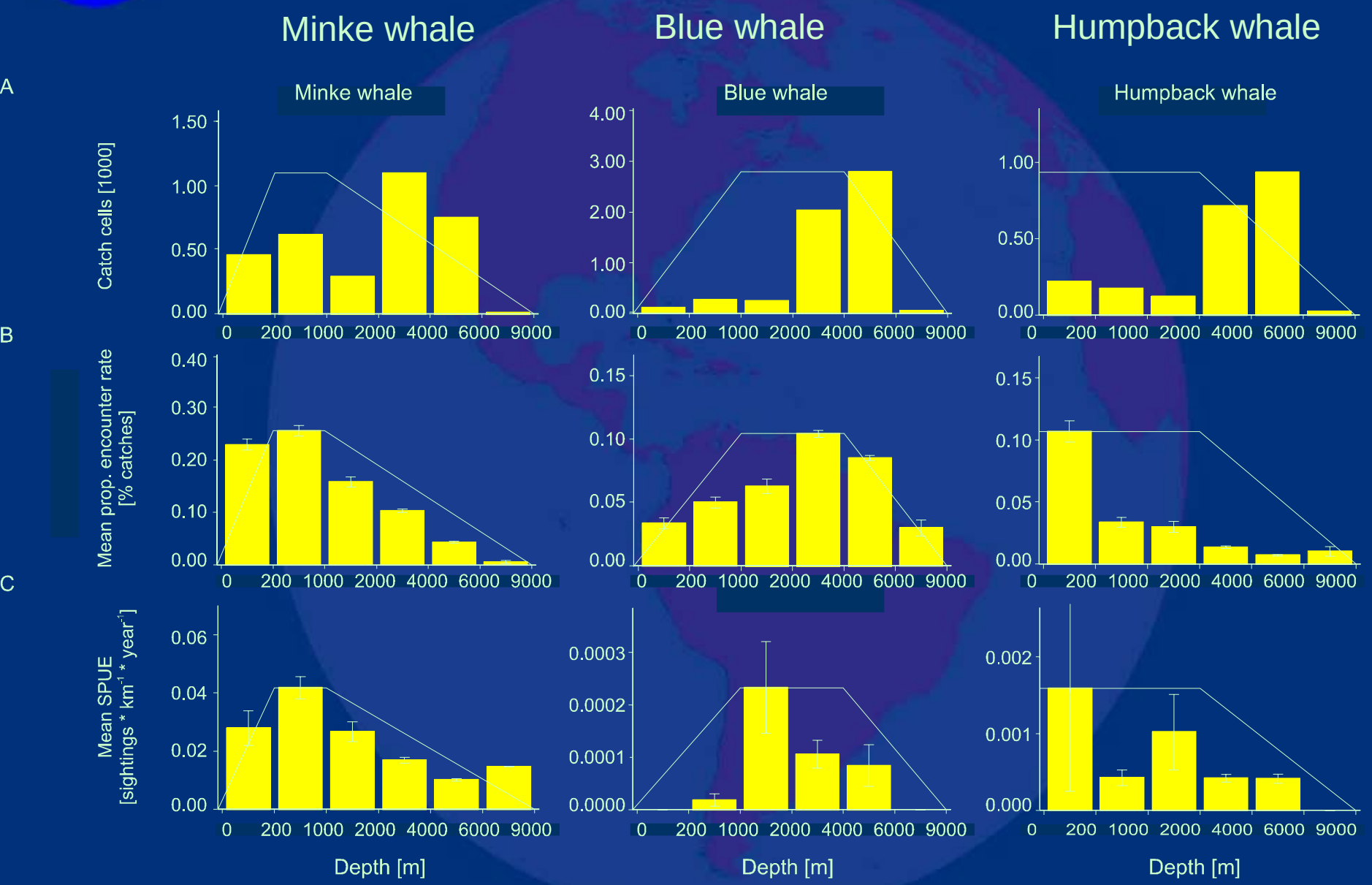
B



Kaschner et al, 2006, *MEPS*

DO THEY WORK?

Sampling Biases



Kaschner et al, 2006, *MEPS*

What can we do with them? Potential Applications

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

What can we do with them? Potential Applications

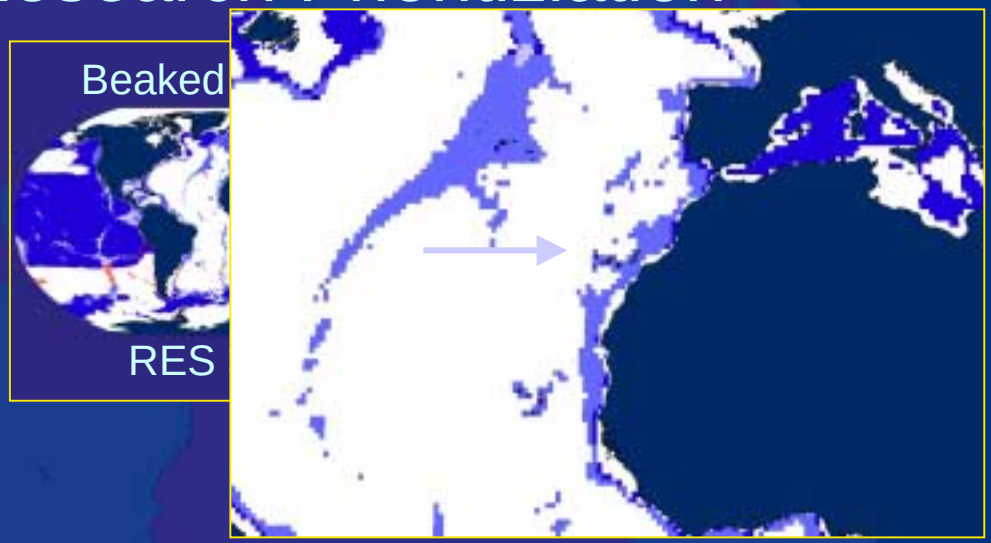
- Biodiversity Mapping

Kaschner & Worm
Wednesday, 15:15

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

Applications

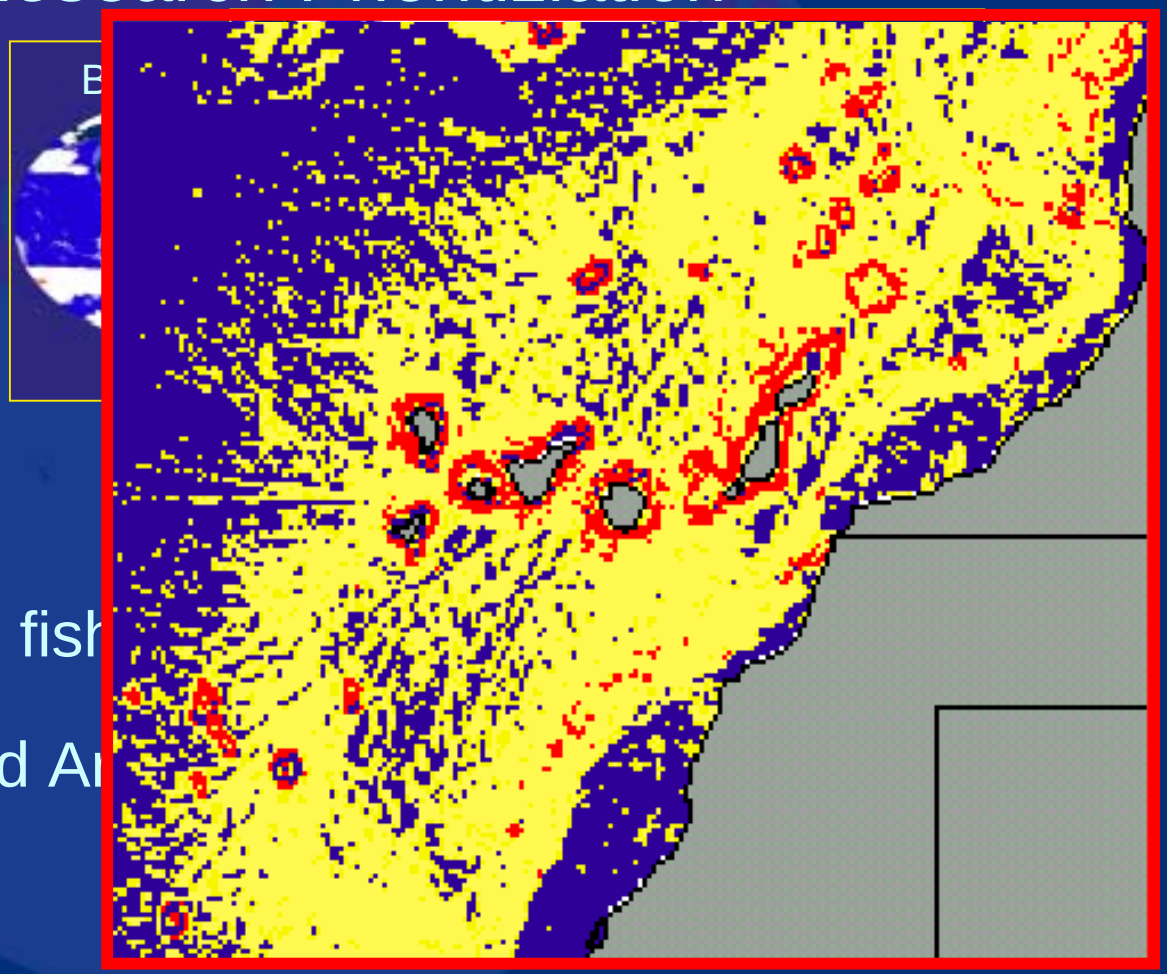
- Management / Research Prioritiziation
  - Risk mitigation



  - Marine mammal fisheries interactions
  - Marine Protected Areas
  - Climate change

Applications

- Management / Research Prioritiziation
  - Risk mitigation
  - Marine mammal fish
  - Marine Protected Ar
  - Climate change



M. densirostris

All *Mesoplodon* Spp.

MacLeod, 2005, PhD Thesis, U of Aberdeen

Applications

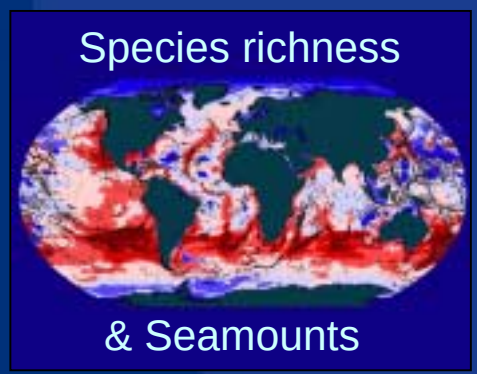
- Management / Research Prioritiziation
  - Risk mitigation
  - Marine mammal-fisheries interactions

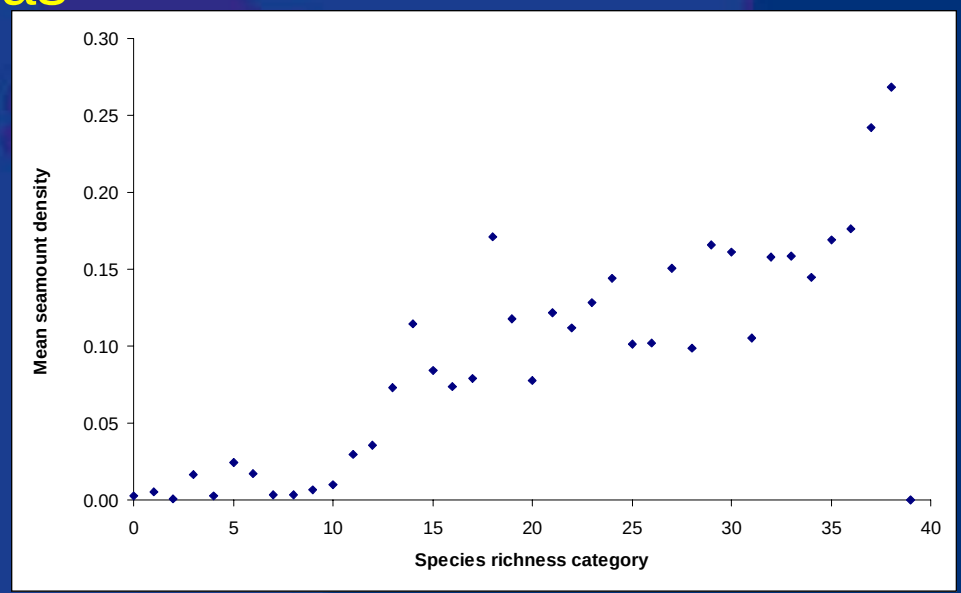
  - Marine Protected Areas
  - Climate change

Kaschner et al, accepted, *CJFAS*

Applications

- Management / Research Prioritiziation
  - Risk mitigation
  - Marine mammal fisheries interactions
  - Marine Protected Areas



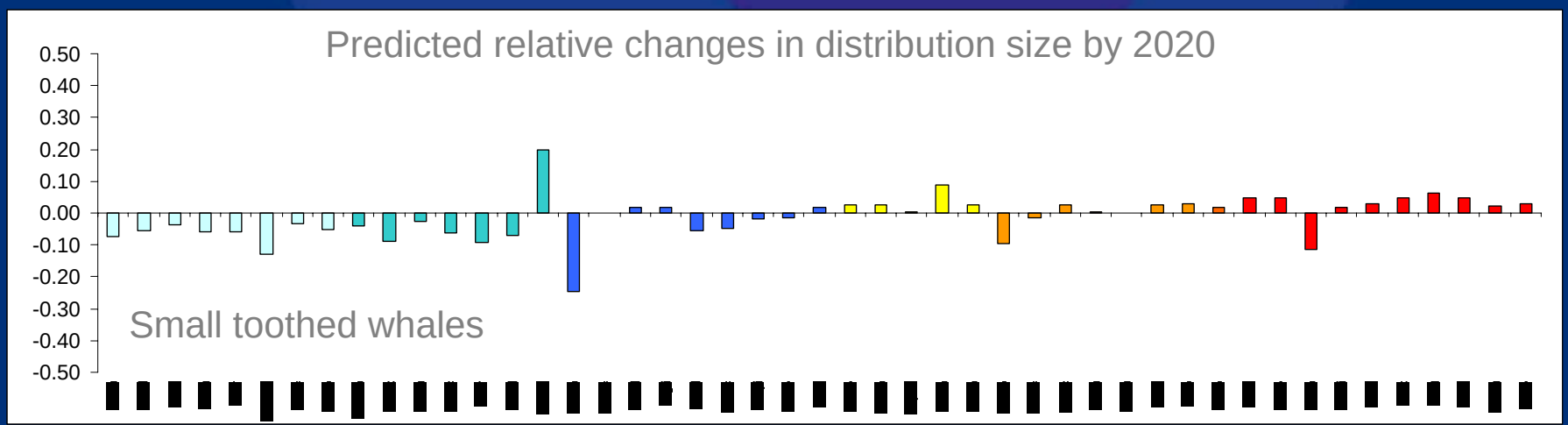


Kaschner, 2006, *Ecology of Seamounts*

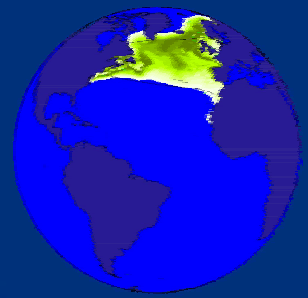
  - Climate change

Applications

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



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