Application of habitat models to highly mobile marine animals – Cetaceans in the North Pacific as case studies

Hiroto Murase *National Research Institute of Far Seas Fisheries*

Toshihide Kitakado *Tokyo University of Marine Science and Technology*

Yu Kanaji *National Research Institute of Far Seas Fisheries*

Hiroko Sasaki *Hokkaido University/National Institute of Polar Research*

Yoko Mitani *Hokkaido University*

Koji Matsuoka *The Institute of Cetacean Research*

Makoto Okazaki *National Research Institute of Fisheries Science*

Naohisa Kanda *The Institute of Cetacean Research*

What is a Habitat Model?

Definition:
A model that describes a spatial distribution of a biological organism as a function of environmental factors at a certain time

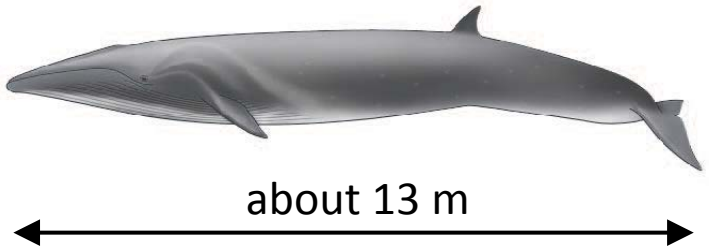
→ Observed relationship like a snapshot

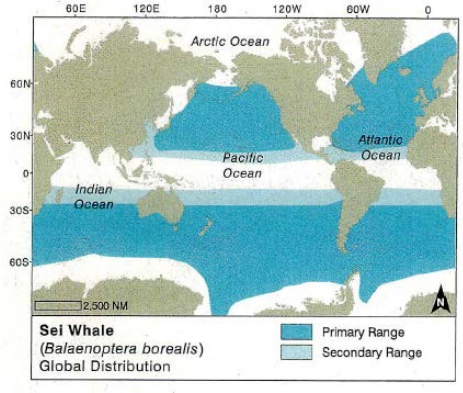
Habitat Model = Static (Statistical) Model

Caution!:
Habitat model can not capture the process (like movie) of the spatial distribution

→ Handled by Mechanistic Model

Case study :
Sei whale in the western North Pacific





(Jefferson et al. 2008)

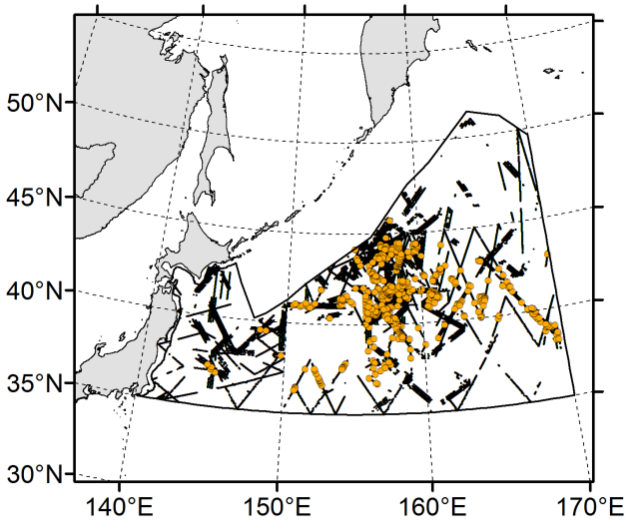
- Sei whale :
  - Medium sized baleen whales
  - Distributed mid-latitude, temperate zone
  - Feed on zooplankton (mainly copepods) and small pelagic fish

Data: Response variable

- Sighting data obtained in the western North Pacific in July from 2000 to 2007

Presence and absence of sei whales are used as the response variable

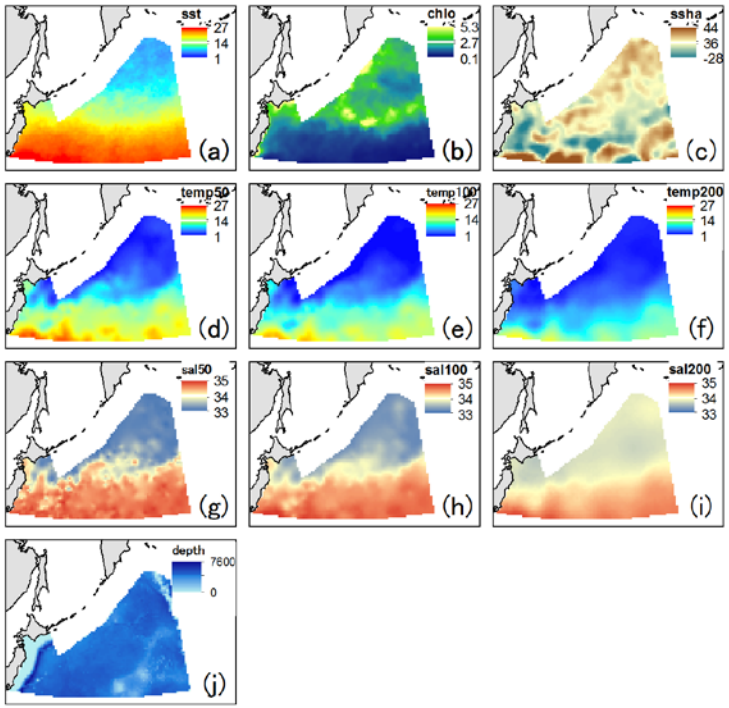
*Abundance (number of animals) or biomass is assumed presence of animals



* The sighting surveys were conducted as a part of The Japanese Whale Research Program under Special Permit in the western North Pacific Phase-II (JARPN II)

Data: Explanatory variables

- Sea surface environmental factors (SST, Chl-a and SSHa) from satellite
- Subsurface temperature and salinity from CTD and Argo floats
- Depth from ETOPO2
- Mean values in July from 2000 to 2007 are used
- Data are aggregated into 30 km grid cell (approximately equal to the resolution of SSHa data)



Statistical models

Methods	Response variable		
	Presence / absence	Presence only	Presence / Background
Regression	GLM (Generalized Linear Model) GAM (Generalized Additive Model)	-	-
Profile	-	BIOCLIM (Bioclimatic Analysis and Prediction System) DOMAIN	ENFA (Ecological Niche Factor Analysis)
Machine learning	BRT (Boosted Regression Tree) RF (Random Forest)	-	MaxEnt (Maximum Entropy)

Collinearity

Definition:

- Existence of correlated explanatory variables

If exist:

- Violating an assumption (= independence of explanatory variables) of standard statistical models such as regression models
- Estimated parameters may be unstable
- Statistical inference may be biased

How to identify:

- Using indices such as variance inflation factor (VIF)

How to handle:

- Removing explanatory variables based on values of the indices
- Dealing within models (e.g. Principal component regression (PCR))

Collinearity in the case study

- Assuming no collinearity is exist if values of VIF are less than 3

* The value is conventionally used but there is no consensus on it.

All variables

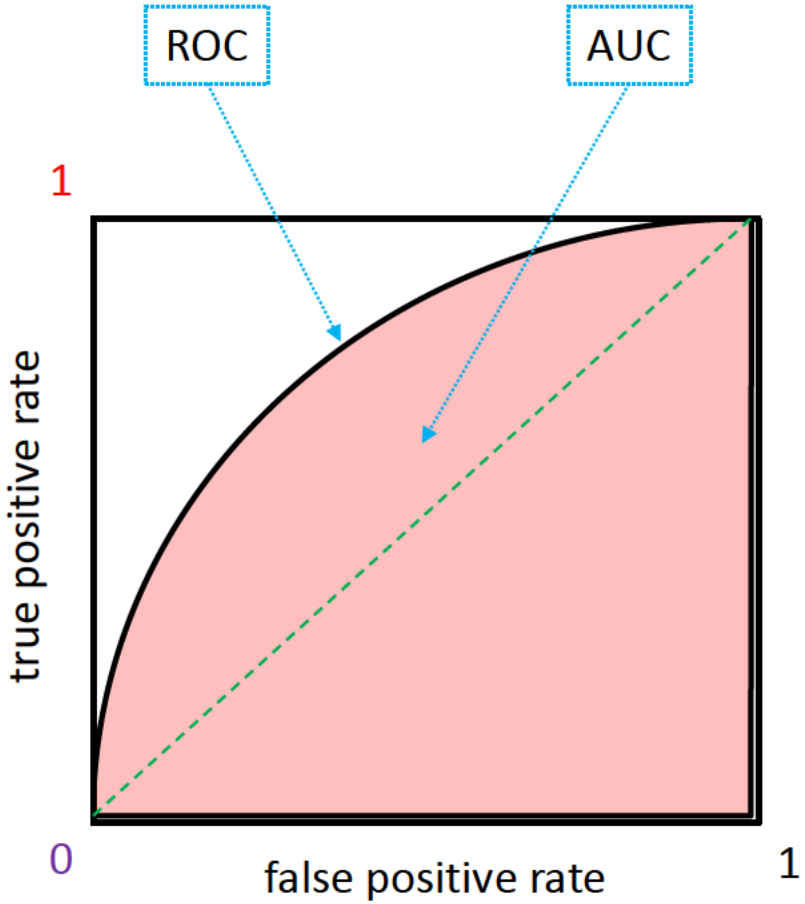
Variables	VIF
Depth	1.5
SST	12.2
SSHa	1.2
Chl-a	2.6
Temp. 50m	139.1
Temp. 100m	422.3
Temp. 200m	164.8
Sal. 50m	29.0
Sal. 100m	95.4
Sal. 200m	44.4

Selected variables

Variables	VIF
Depth	1.1
SST	1.8
SSHa	1.0
Chl-a	1.9

Model evaluation (AUC of ROC)

- Commonly used to evaluate habitat models which estimate occurrence



ROC: Receiver Operating Characteristic curve
AUC: Area under curve

AUC = 1

Estimation is perfect

AUC = 0.5

Estimation is at random

AUC = 0

Estimation is totally imperfect

Test and training data

Training data

- Utilized to develop habitat models

Test data

- Utilized to measure prediction success

* Generally, subsets of training data are used in habitat modeling.

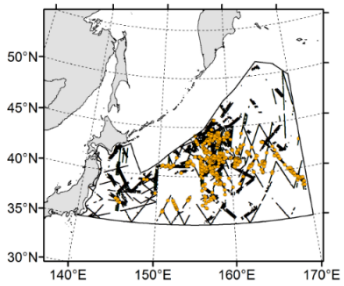
In the case study:

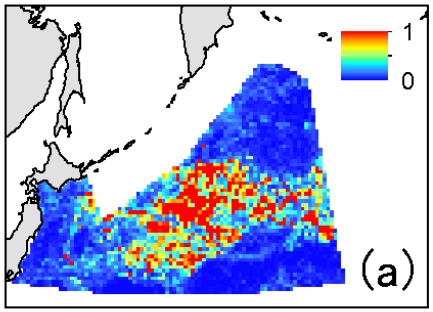
Training data are used to calculate AUC to compare the results among habitat models as both presence and absence data are available for modeling.

AUC in the case study

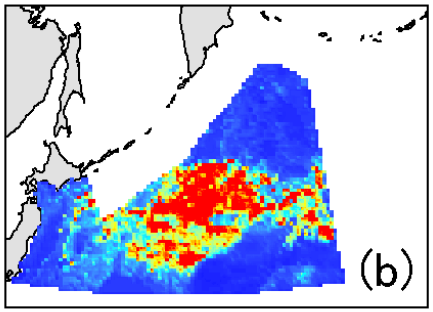
Methods	Response variable		
	Presence / absence	Presence only	Presence / Background
Regression	GLM **#6: AUC = 0.69** GAM **#3: AUC = 0.83**	-	-
Profile	-	BIOCLIM **#4: AUC = 0.75** DOMAIN **#5: AUC = 0.74**	ENFA **#4: AUC = 0.75**
Machine learning	BRT **#2: AUC = 0.94** RF **#1: AUC = 1.00**	-	MaxEnt **#3: AUC=0.83**

Estimated spatial distributions

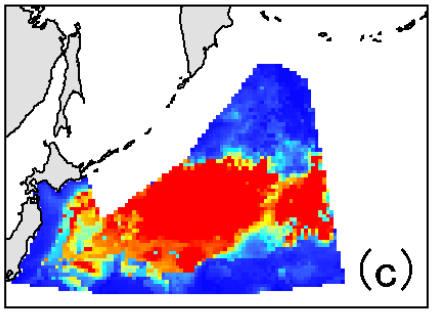




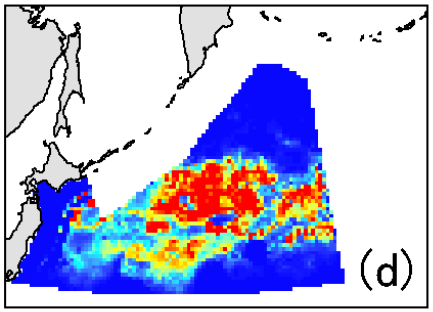
RF
AUC = 1.00



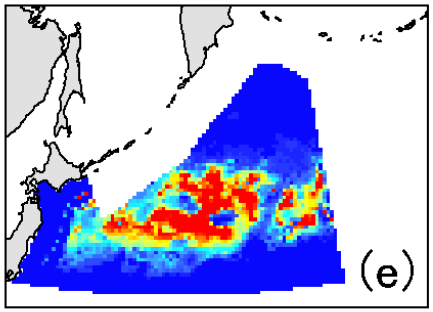
BRT
AUC = 0.94



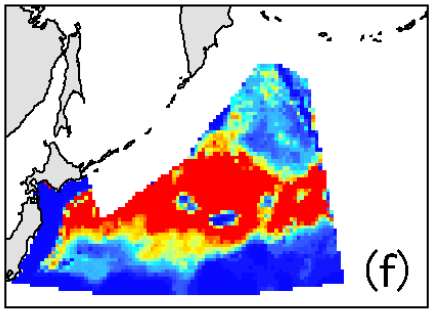
MAXENT
AUC = 0.83



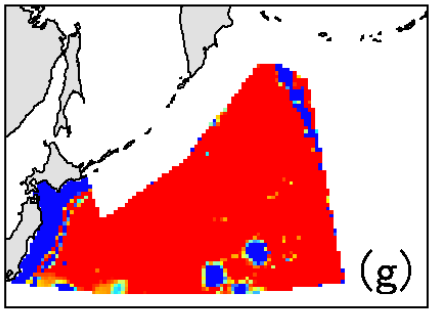
GAM
AUC = 0.83



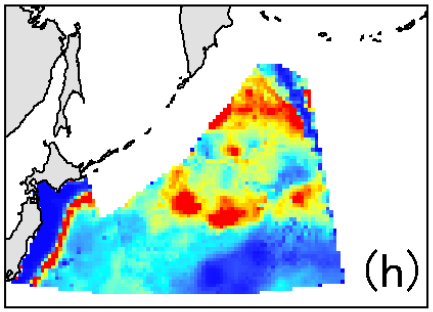
BIOCLIM
AUC = 0.75



ENFA
AUC = 0.75



DOMAIN
AUC = 0.74



GLM
AUC = 0.69

Ecological inference: Contribution of explanatory variables (in rank)

Variables	RF	BRT	MaxEnt	GAM	ENFA	GLM
Depth	2	2	3	1	2	1
SST	1	1	1	2	3	3
SSHa	3	3	4	4	4	4
Chl-a	4	4	2	3	1	2

*The results are for illustrative purpose as methods to determine the ranks are yet to be considered for some models

Habitat models: pros and cons

Pros:

- Relatively easy to construct as the modeling methods are well developed and access to the environmental data are getting easy
- Good starting point to understand effect of environmental factors which determine spatial distributions of organisms
- Easy to get overall spatial distribution maps

Cons:

- Difficult to assess the reliability of the estimates
- Difficult to understand the reasons why environmental factors affect the spatial distribution of target species as the habitat models are based on observation at a certain time

* Marine animals such as whales are highly mobile in both vertical and horizontal directions and they encounters a variety of environmental conditions while they move.

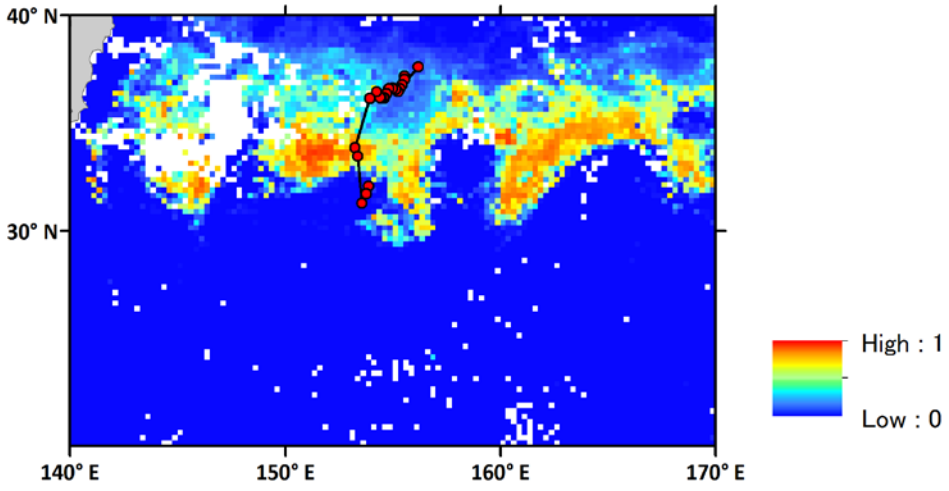
Horizontal movement

Satellite tracking of Bryde's whales as examples:

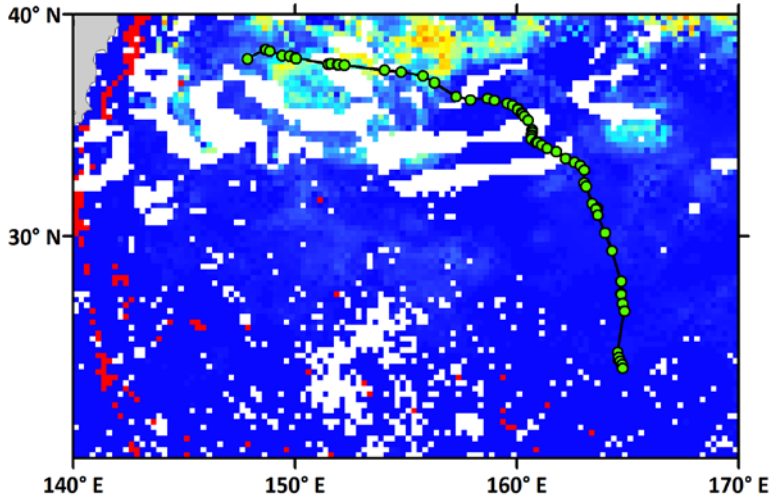
- Whales are capable to move long distance within a few weeks
- They encountered a variety of environmental condition while they move

Inference from sighting and satellite data might be different in terms of their habitat

Summer, 2006 (917 km in 13 days)



Summer, 2008 (2,649 km in 20 days)



Satellite tracking data: Nishiwaki et al. 2009 Habitat modeling: Sasaki et al. (unpublished data)

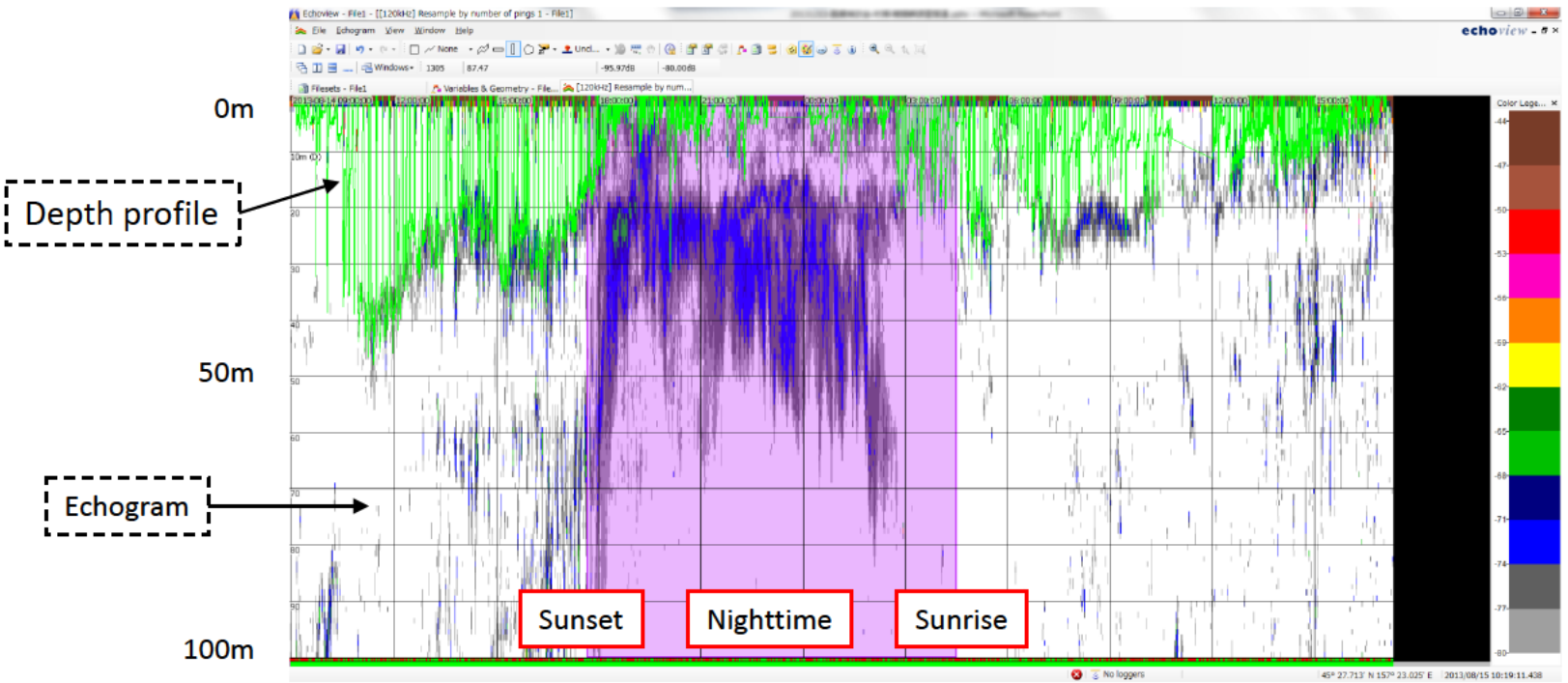
Vertical movement

Vertical movement of a sei whale recorded by an acoustic pinger:

- It showed diurnal vertical behavior
- It could be related to availability of prey as well as environment conditions

*Detailed analysis is undergoing

14-15 August 2013 (32 hours of observation)



Murase et al. . (unpublished data)

Inference from animal movement

- Animals encounter a variety of environmental conditions as they move

⇦ Habitat models might overlook such process as they are static models

- Required environmental conditions might be different for different behavior states (e.g. feeding, migration and breeding)

⇦ Habitat models generally overlook behavior states

* Spatiotemporal scales of habitat models are generally too coarse to capture these two points

Future direction: Combination with mechanistic models

To understand spatial distribution of marine animals fully:

- Development of mechanistic models such as spatially explicit ecosystem models could be one solution

  Advantage: Can reveal process of spatial distribution of animals

  Disadvantage: Required a lot of data sets more than habitat models

- Habitat models and spatially explicit ecosystem models are not mutually exclusive:
  - Outcome of habitat models can be used as input parameters
  - Estimated spatial distributions from two different types of models can be used for comparisons to get our better understating