Large-scale habitat-based models of cetacean density

An overview of collaborative research conducted by

Megan Ferguson, Elizabeth Becker, Jessica Redfern, Karin Forney, Jay Barlow, Paul Fiedler, Nacho Vilchis, and Lisa Ballance

NOAA, Southwest Fisheries Science Center

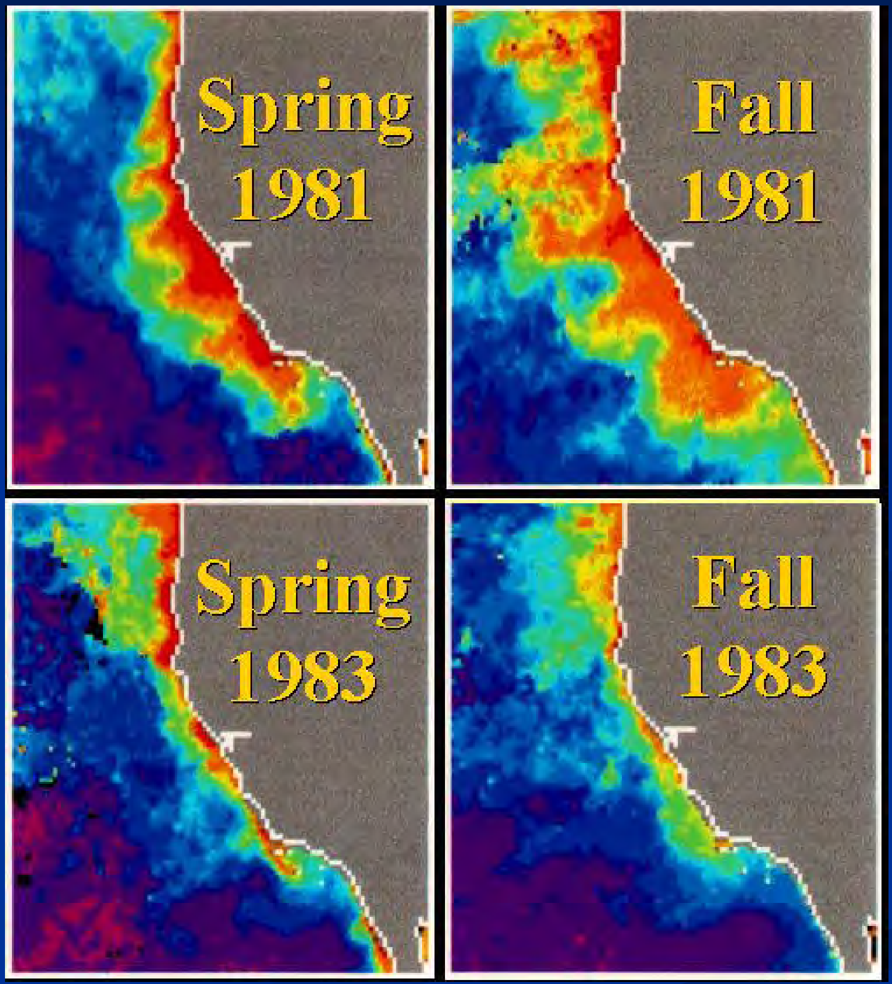
Presentation by Karin Forney
Ecological Modeling Workshop
Marine Mammal Conference
26 November 2011, Tampa FL

Why develop habitat models for cetaceans?

Curiosity?

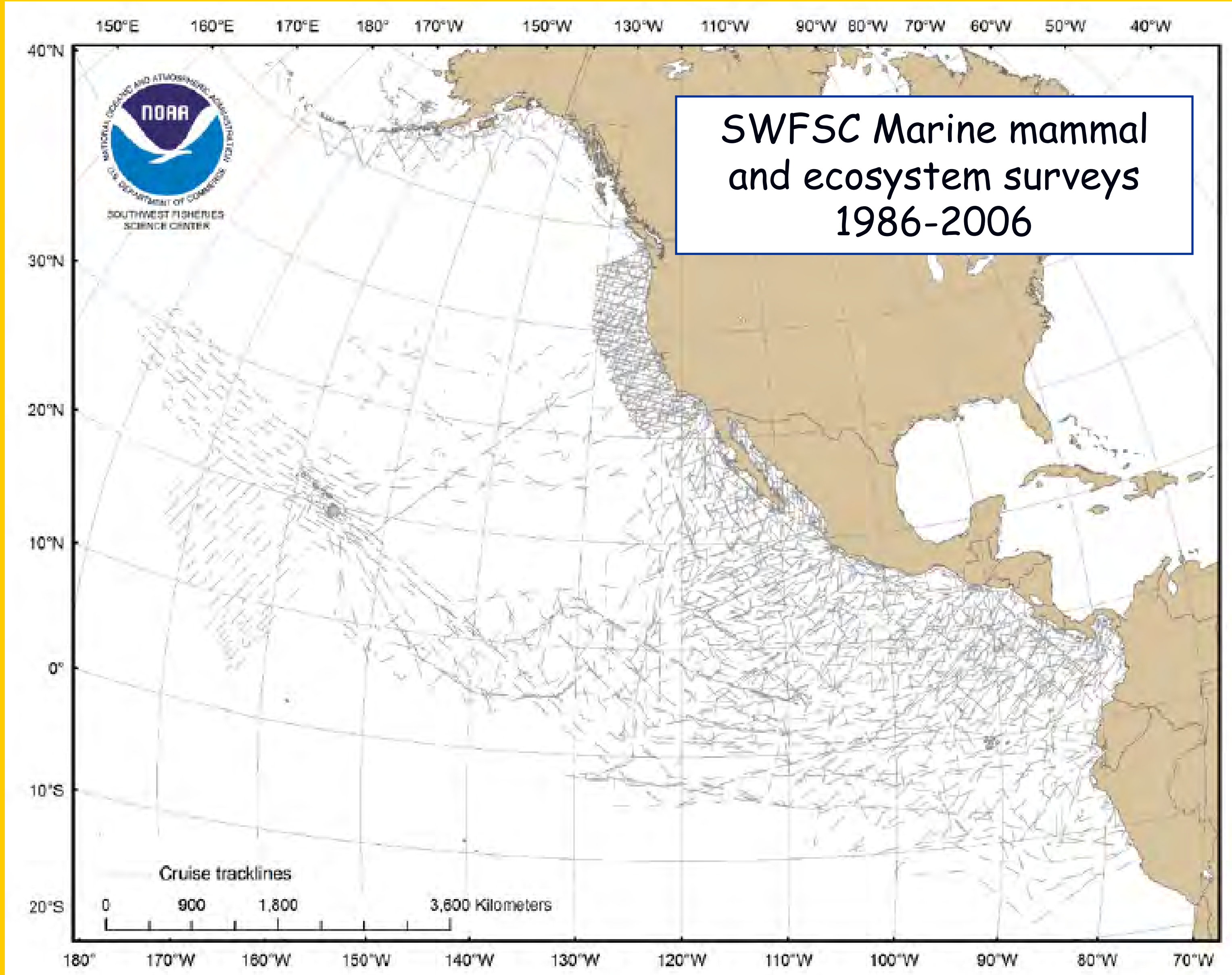
- Marine environments are dynamic
- Develop ecological insights and hypotheses
- Identify spatiotemporal patterns for management
- Improve estimates of abundance and trends

Ok – so how?



CZCS Surface chlorophyll concentration; Courtesy G. Mitchell, SIO

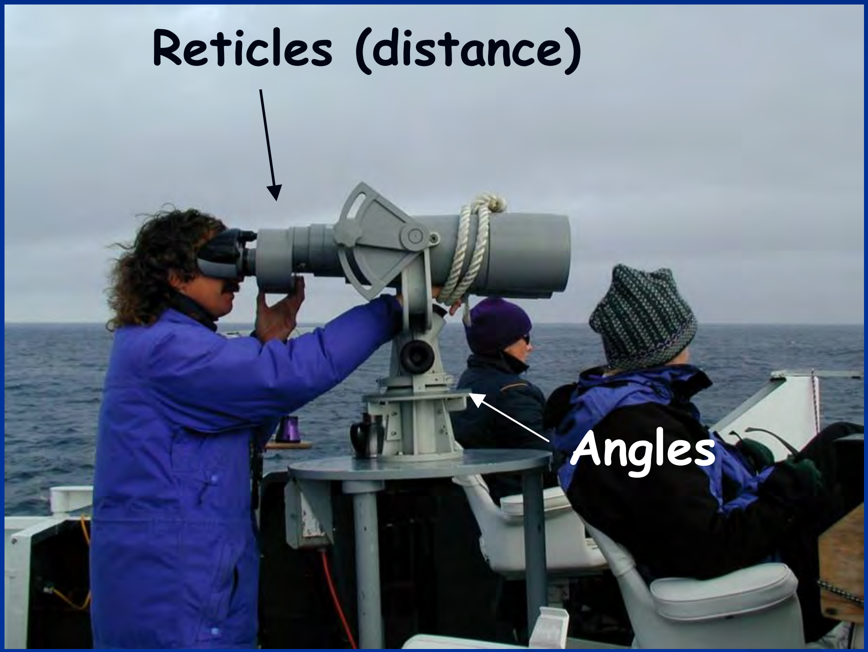
First, we need data...

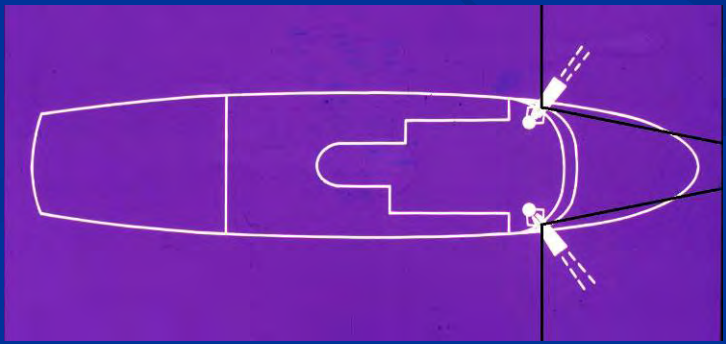


SWFSC Marine mammal and ecosystem surveys 1986-2006

Marine Mammal and Ecosystem Surveys, 1986-2006
Large-vessel, line-transect surveys

3 Observers:
- two 25x "big eye" binoculars
- one 7x binocular & unaided eye



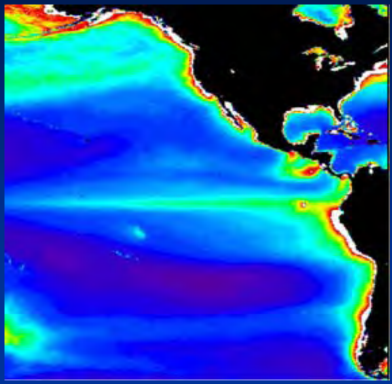


Marine Mammal and Ecosystem Surveys, 1986-2006
In situ ecosystem sampling

- XBTs & 1000-m CTDs
- Seabirds strip transect surveys
- Net tows
- SST, salinity, chlorophyll
- Acoustic backscatter



TECHNICAL APPROACH



Marine Mammal Survey Data

Habitat Data

Statistical models of marine mammal density relative to habitat variables

Many considerations…

- Identify modeling objectives
- Process survey data for model development
- Determine scale and types of predictor data
  - Remotely sensed vs. *in situ*
  - Spatial and temporal scales
  - Interpolation methods
- Select modeling framework
- Establish criteria for model selection and validation
- Characterize uncertainty

→ Provide examples from our projects

Identify modeling objectives

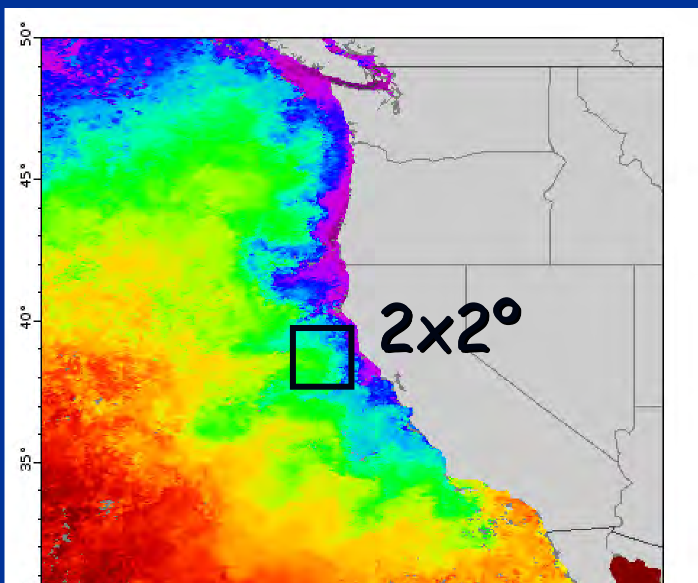
General types of models:

- Mechanistic/trophic: identify trophic linkages between cetaceans, prey and oceanographic variables
  - → Croll et al. 2005 *MEPS*; Baumgartner et al. 2003, *MEPS*
- Explanatory models: explain variability within a data set to improve estimation of abundance
  - → Hedley and Buckland 2004, *J Agri Biol & Env Stat*
- Predictive models: Identify (persistent) relationships between species and habitat variables to allow fine-scale prediction of densities within a study area
  - → Ferguson et al. 2006, *Ecological Modeling*
  - Barlow et al. 2009, *NOAA Tech Memo NMFS-SWFSC-444*
  - Forney et al., in press *ESR Special Issue*

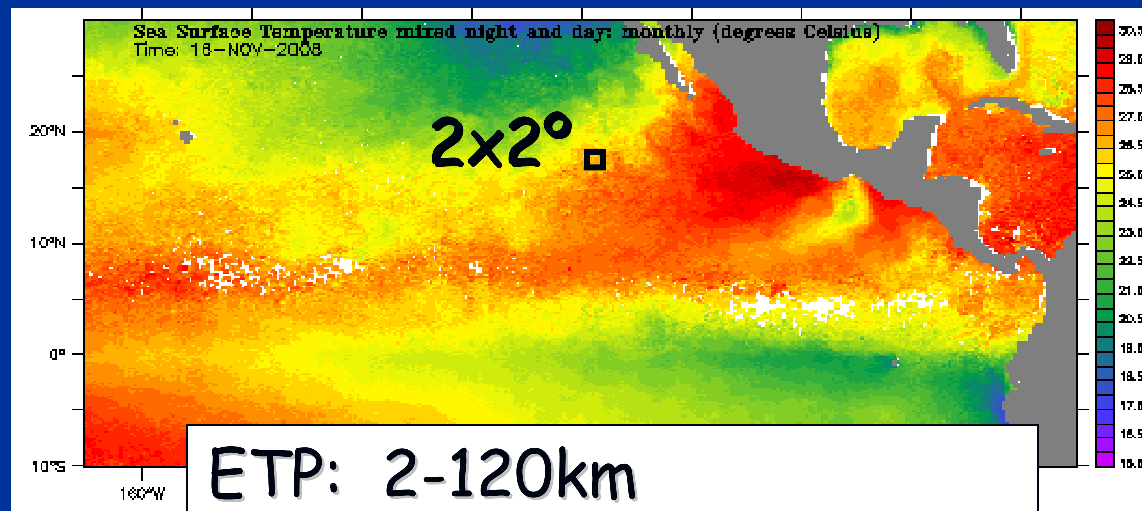
Process survey data for model development

Determine sampling unit (e.g. 10-km segments, 1x1° boxes,…)

- Depends on data
- Should relate to scale of ecological patterns
- May be tradeoff to minimize zeros in data



CA Current: 2-5 km
(Forney 2000, *Cons Biol*,
Becker et al. 2010, *MEPS*)

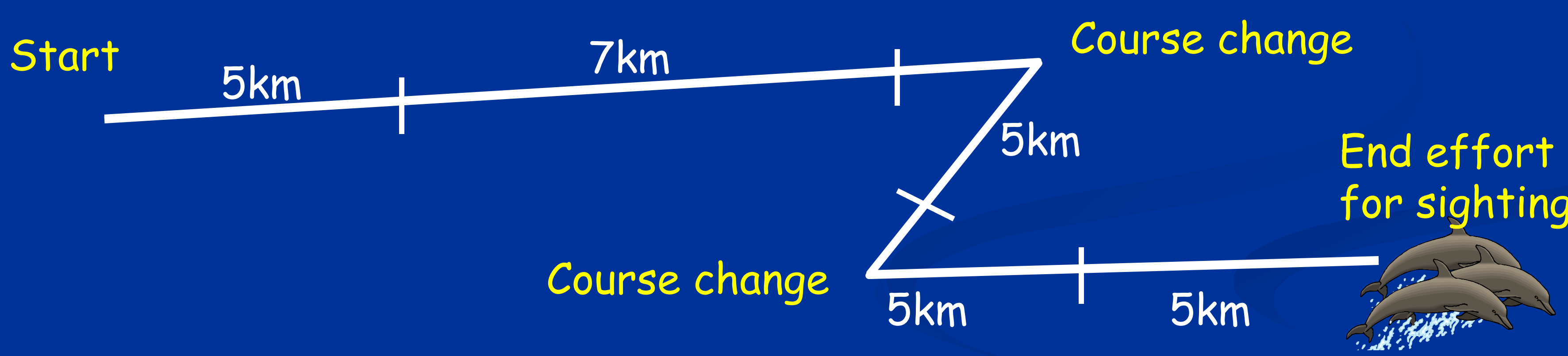


ETP: 2-120km
(Ferguson et al. 2006, *Ecol Appl*
Redfern et al. 2008, *MEPS*)

Process survey data for model development

Example: Creating 5-km segments along the survey track:

On-effort segment: total length = 27km; sighting at end



The extra 2km is randomly added to one of the 5km segments

Determine scale and types of input data
Underway environmental data

<u>Examples:</u>

- Thermosalinograph (temperature and salinity)
- Flow-through fluorometer (chlorophyll)
- Acoustic backscatter (zooplankton and nekton)
- Optical plankton counter
- CUFES (continuous underway fish egg sampler)
  Checkley et al. 1997, Fish. Ocean.

➢ Can readily average data within each sampling unit

➢ Matched in time and space to sighting data

Determine scale and types of input data
Station Data

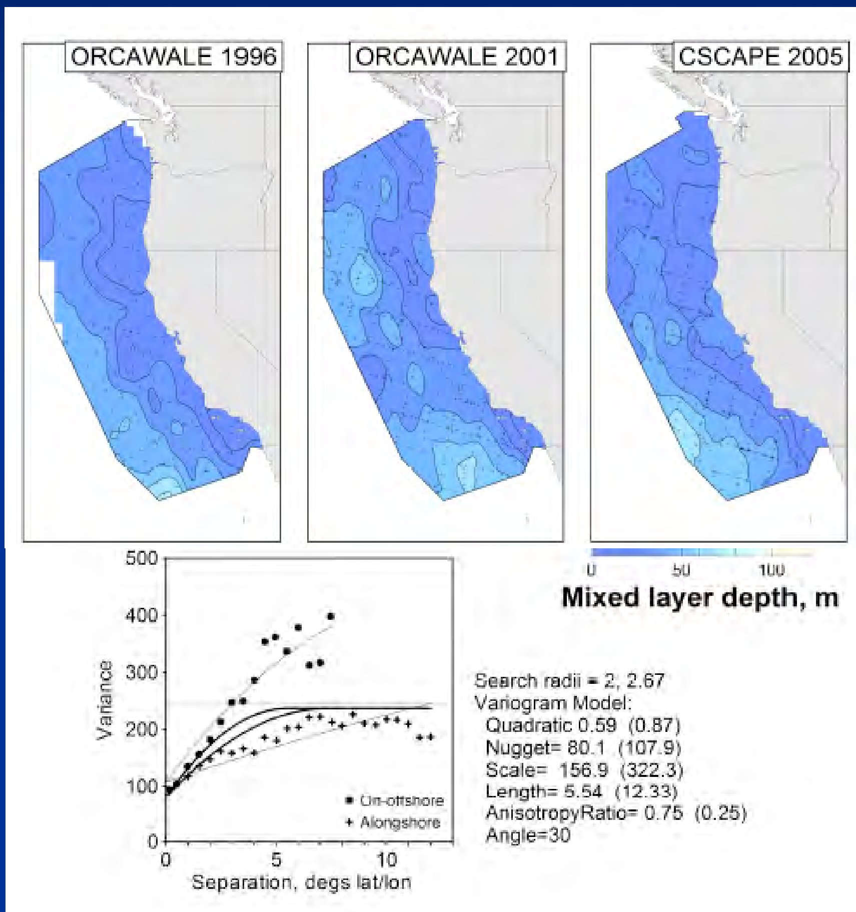
Examples:

- Conductivity-Temperature-Depth (CTD) water column profiles (temperature, salinity, mixed layer depth)
- Chlorophyll samples (surface and or with CTD)
- Net tows (zooplankton volume)

➢ These variables are often linked more closely to the trophic ecology of cetaceans

➢ Stations may be coarser than model sampling unit, requiring interpolation or averaging

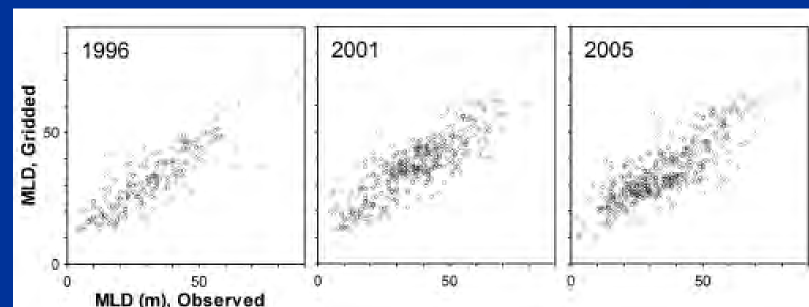
Determine scale and types of input data
Station Data – may require interpolation



Analysis by Paul Fiedler (see Barlow et al. 2009, *NOAA Tech Memo*)

<u>Examples:</u>

- Kriging
- Inverse Distance Weighting
- Local Polynomial



→ Spline interpolation used to create finer-scale interpolated fields, from which values for each segment were extracted using SURFER©, Golden Software Inc)

Determine scale and types of input data
Remotely sensed data

Examples:

- Sea surface temperature (SST) and STD(SST)
- Chlorophyll (e.g SeaWiFS 

)
- Sea surface height
- Derived products (Primary productivity, frontal probability, etc)

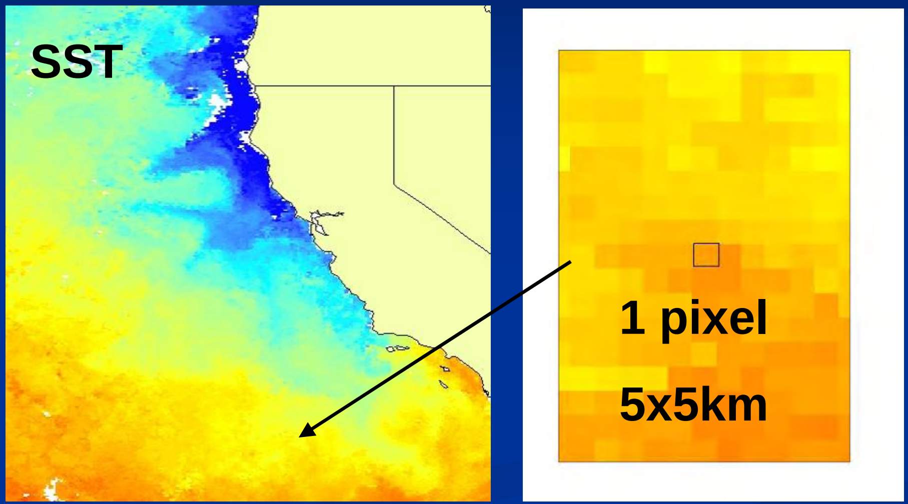
Becker et al. 2010, *MEPS*

- Compared models with *in situ* vs. remotely sensed SST variables for 10 cetacean species in California Current
- Models similar; remotely sensed predictors performed better when STD(SST) important.

Determine scale and types of input data

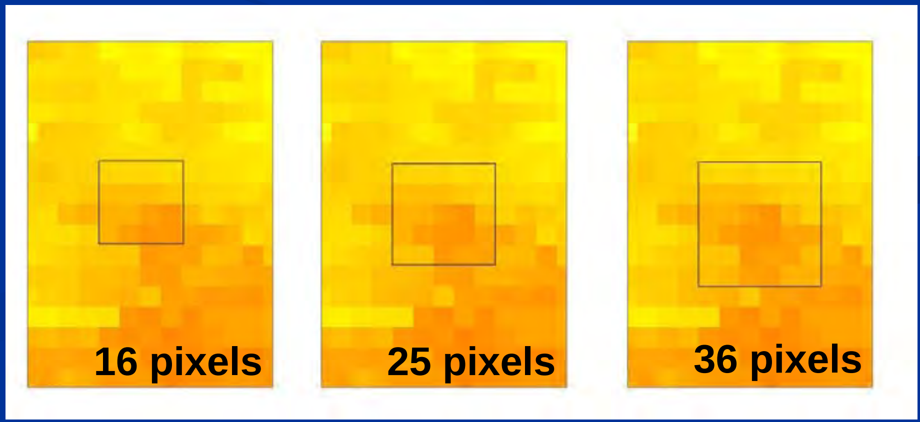
Remotely sensed data – temporal and spatial scales

- Data sets at varying spatial scales (5km, 9km, 25km)
- Cloud cover often requires 8-day or 30-day composites
- Species- and habitat-specific optimum resolution



Becker et al. 2010, *MEPS*

- Compared models that used various spatial scales (mean and STD across multiple pixels)
- Larger scales tended to perform better



Select modeling framework (FINALLY!)

Barlow et al. 2009 *NOAA Tech Memo NMFS-SWFSC-444*

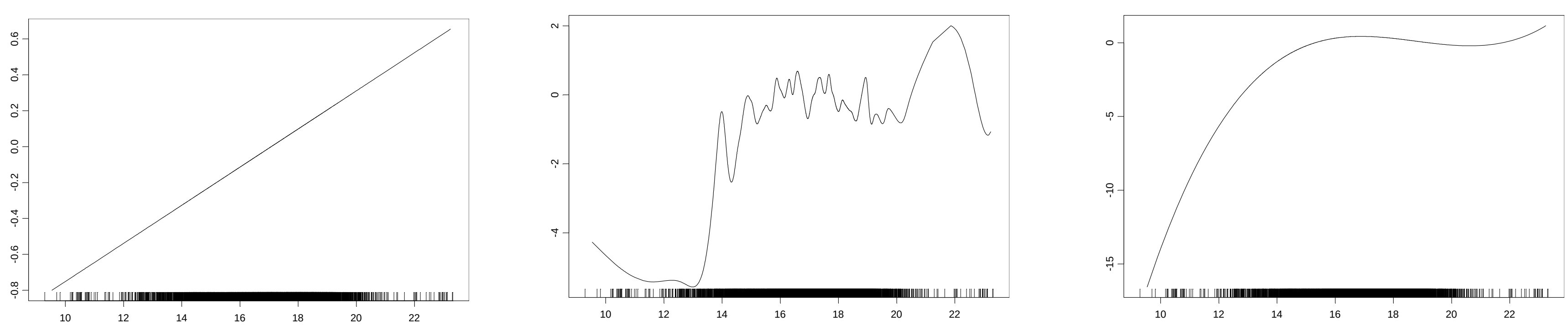
A variety of statistical model types were considered:

- Classification and Regression Trees (CART)
- Generalized Linear Models (GLM) and Generalized Additive Models (GAM) with 5 smoothing spline types
- 4 Algorithms
  - S-plus: gam
  - R packages: 'gam', 'mgcv', 'glm.nb'
- 8 criteria compared
  - predictors selected
  - predictor degrees of freedom
  - predictor functional forms
  - % explained deviance
  - AIC
  - Spatial plots of predictions
  - ASPE (response residuals)
  - ASPE (Anscombe)

Generalized Additive Models (GAMs)

$$link(\mu_i) = \alpha + \sum_{i=1}^{n} f_i(x_i)$$

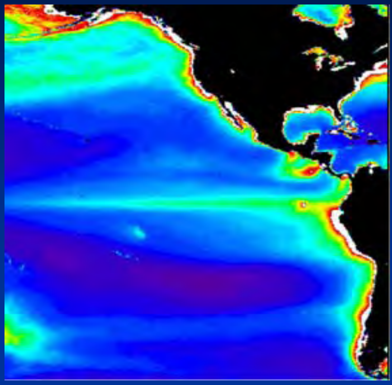
Each function, *f(x)*, can be a non-linear spline fit with variable degrees of freedom chosen to optimize the fit



linear 45 df 3 df

TECHNICAL APPROACH

Marine Mammal Survey Data



Habitat Data

Statistical models of marine mammal density

$$Density = \frac{n \cdot s}{L \cdot 2 \cdot w \cdot g(0)}$$

Line-transect framework

(Buckland et al. 2003)

n = # groups

L = length of transect

s = group size

w = effective strip ½-width

g(0) = probability of detection on transect line

TECHNICAL APPROACH – Generalized Additive Model

(Ferguson et al. 2006, *Ecol Appl*)

Encounter Rate (n/L):

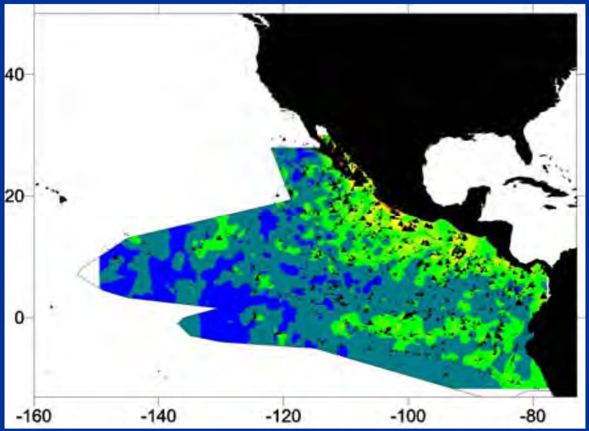
n ~ quasi-Poisson

$$\ln(n) = \text{offset}(L) + f(SST) + f(MLD) + f(\text{sea state}) + \ldots$$

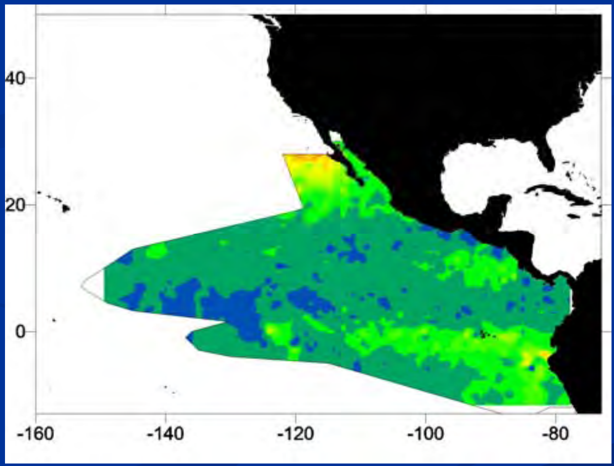
Group Size (s):

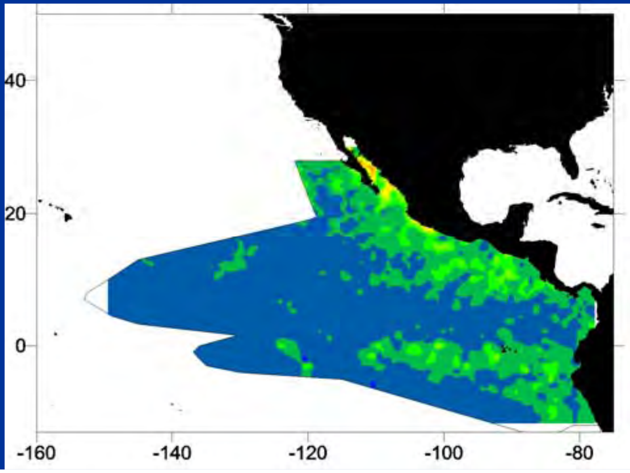
s ~ log-Normal

$$\ln(s) = f(SST) + f(depth) + f(MLD) + f(chl) + f(\text{sea state}) + \ldots$$



$$D = \frac{n \cdot s}{L \cdot 2 \cdot w \cdot g(0)}$$





Combined Density Model (D)

Model selection and validation

STEP 1 - Model Selection:

Identify model that best explains the observed patterns of variation

Goodness of fit measures, e.g.:

- R^2; explained variance/deviance
- AIC or similar criteria (each parameter is penalized)
- Visual inspection
- Beware of p-values!

This is not necessarily the best predictive model:

✓Insufficient variation
✓Model over-specification
✓Sample size limitations

STEP 2 – Model Validation:

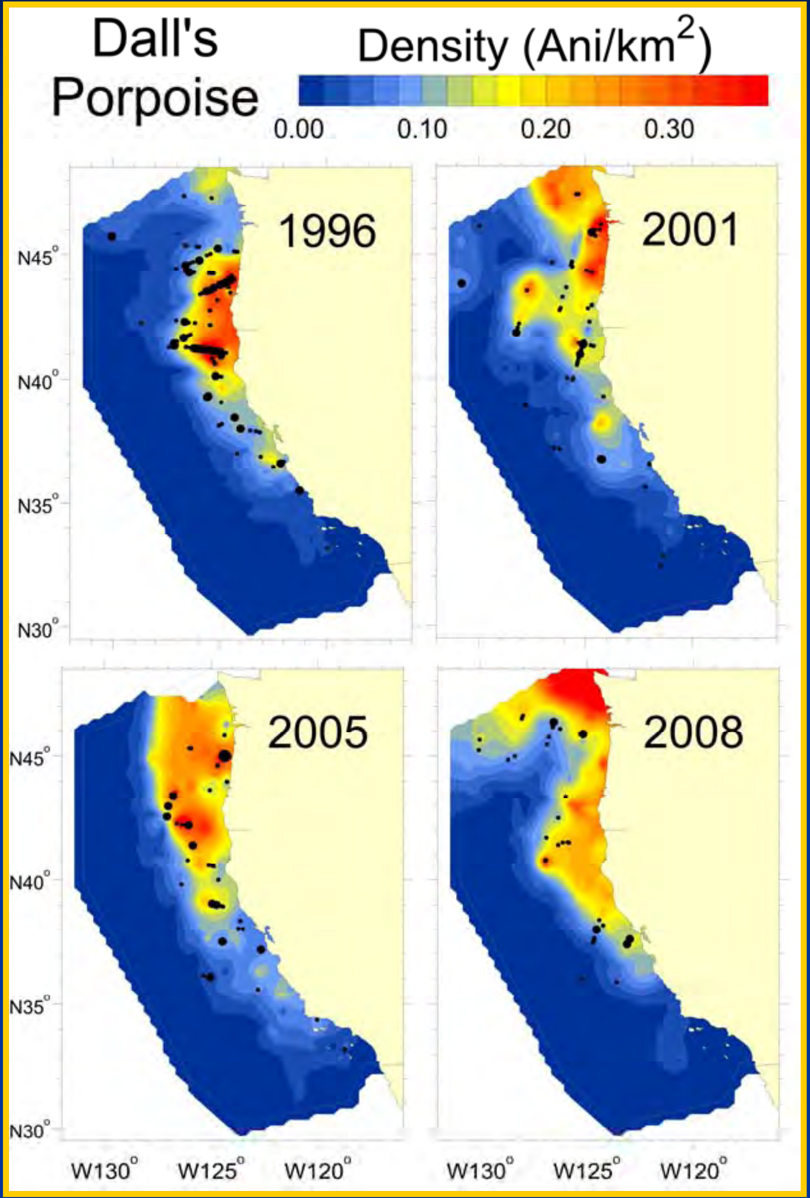
Evaluate predictive power on a novel data set

Validation measures, e.g.:

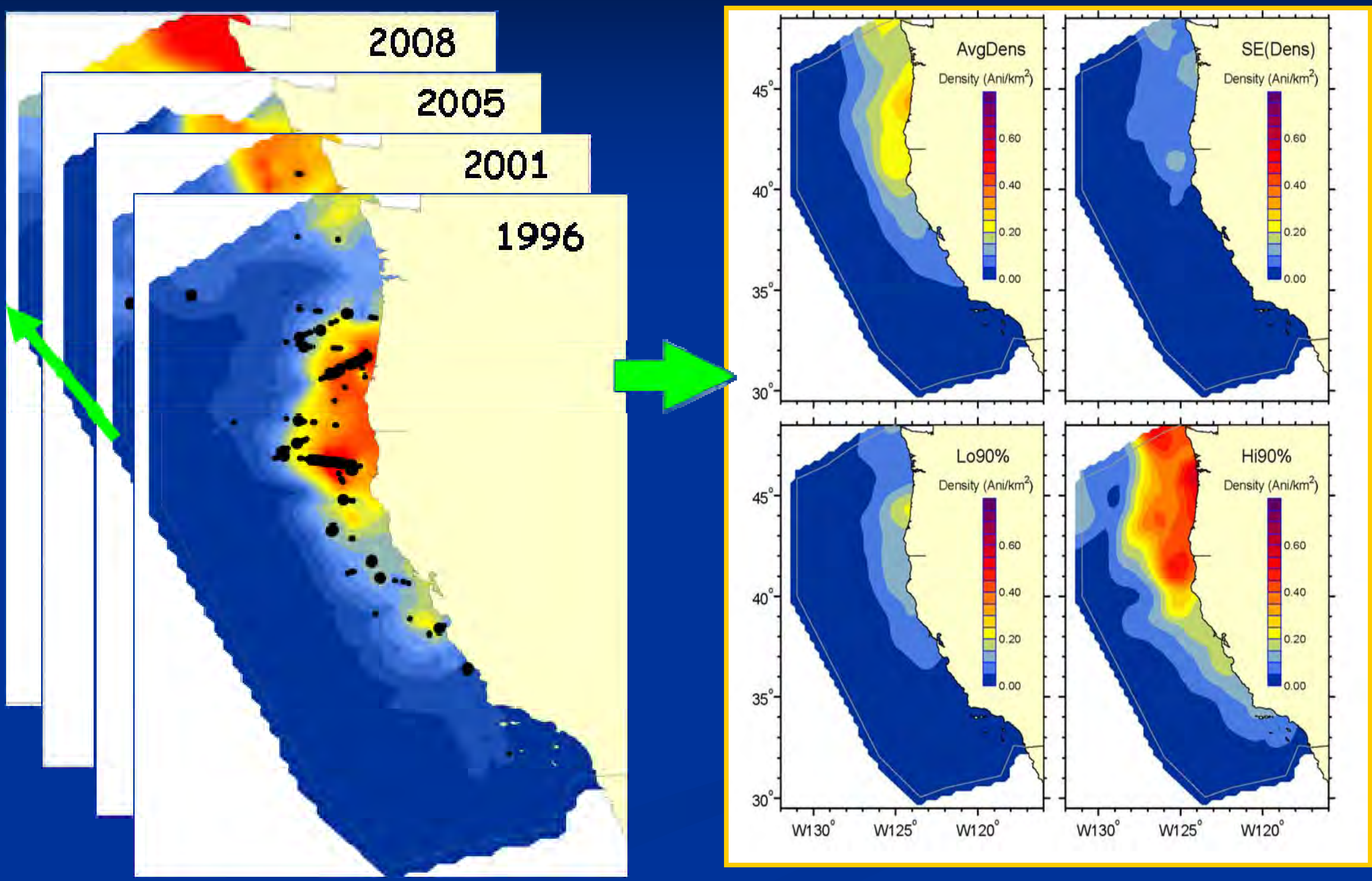
- Squared prediction error (ASPE or PRESS)
- Rank correlation tests
- Visual inspection of model prediction vs. new data

For example....

Characterize uncertainty



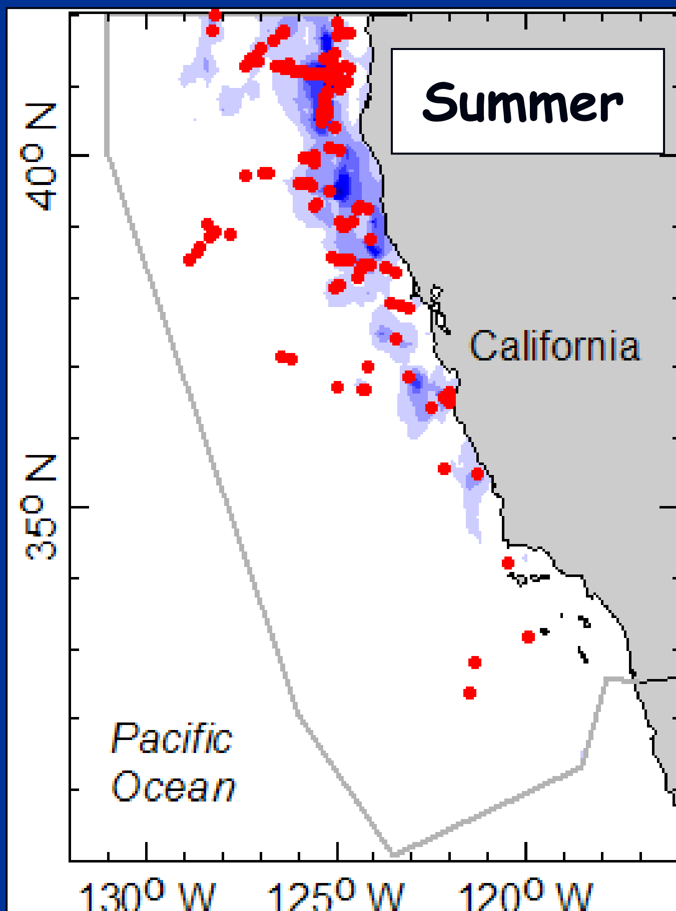
Characterize uncertainty

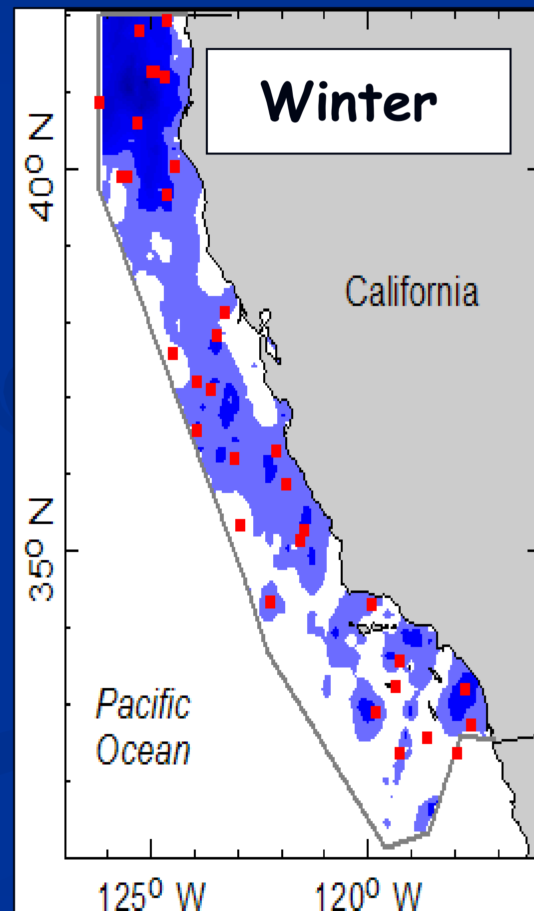


Examine seasonal performance

(Becker 2007, PhD Dissertation, UC Santa Barbara)

Models captured seasonal distribution changes for some species (e.g. Dall's porpoise, *Phocoenoides dalli*)





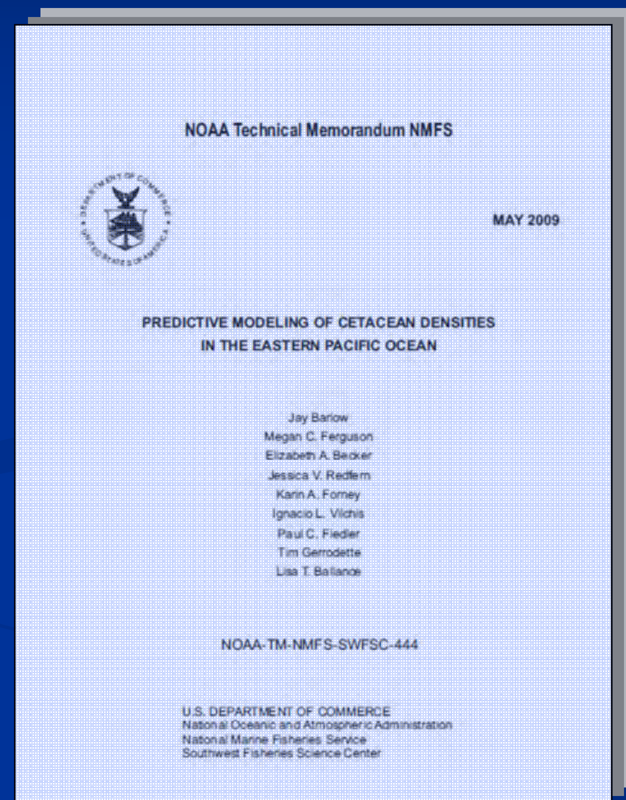
Conclusions

- Huge collaborative effort involving biologists (quantitative and field), oceanographers, etc.
- Many statistical and data considerations
- Many valid approaches – pick what is 'best'
- Model validation is key:

  "All models are wrong, but some are useful" (Box 1979)
- Future directions:
  - NOWCAST/FORECAST capabilities (see Becker presentation next, and Tue 08:30)
  - Area-searched offset instead of distance-searched (see Forney presentation Friday 13:30)

Modeling literature cited

- Barlow et al. 2009 (the nitty gritty)
  NOAA Tech Memo NMFS-SWFSC-444
- Forney 2000, *Conservation Biology*
- Ferguson et al 2006, *Ecological Applications*
- Redfern et al. 2006, *MEPS* (modeling review)
- Becker 2007, *PhD Diss., UC Santa Barbara*
- Redfern et al. 2008, *MEPS*
- Becker et al. 2010, *MEPS*
- Becker et al. (in press) *ESR Special Issue*
- Forney et al. (in press) *ESR Special Issue*


NOAA Technical Memorandum NMFS

MAY 2009

PREDICTIVE MODELING OF CETACEAN DENSITIES
IN THE EASTERN PACIFIC OCEAN

Jay Barlow
Megan C. Ferguson
Elizabeth A. Becker
Jessica V. Redfern
Karin A. Forney
Ignacio L. Vilchis
Paul C. Fiedler
Tim Gerrodette
Lisa T. Ballance

NOAA-TM-NMFS-SWFSC-444

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center