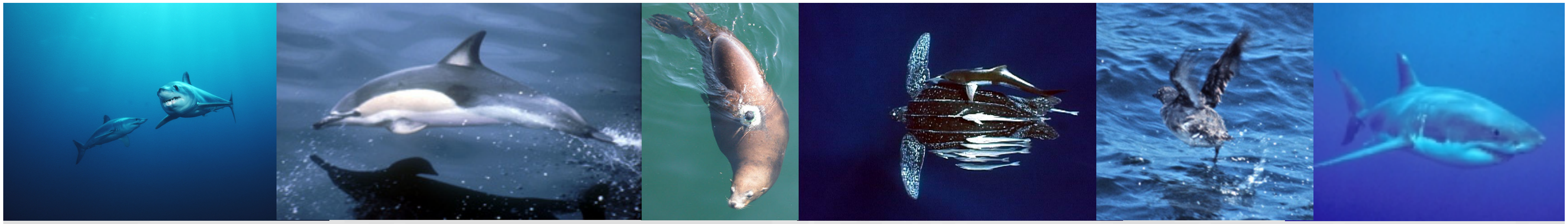


From oceanography to top predator habitat in TOPP:
Synthesis of tag data and modeling

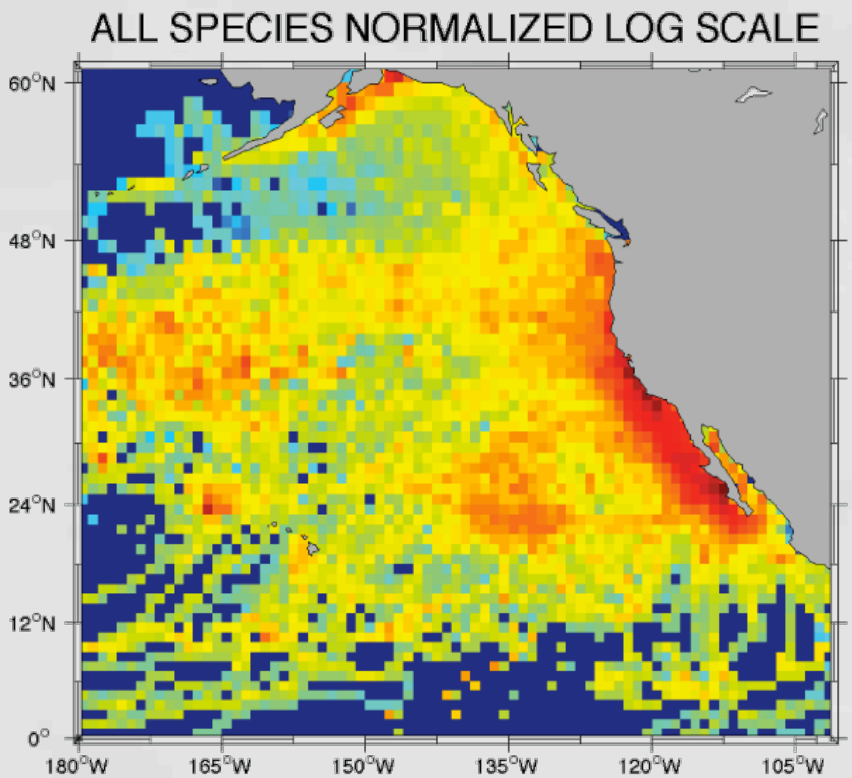
Elliott Hazen, UH-JIMAR / NOAA-SWCSC-PFEL

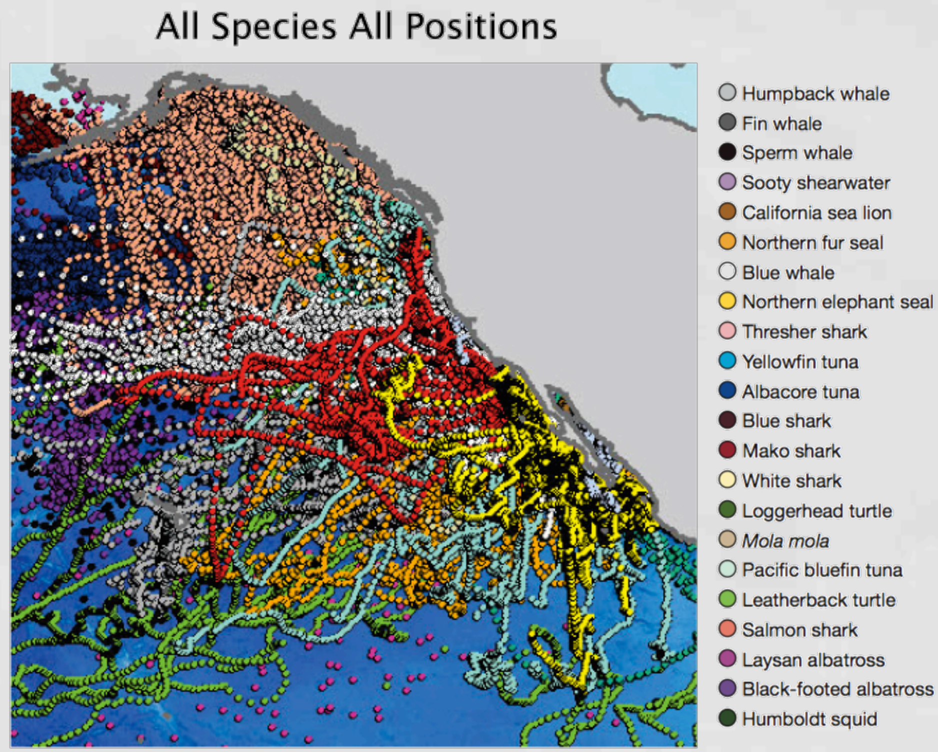
B. A. Block[1], I. D. Jonsen[2], S. J. Jorgensen[1], A. J. Winship[2], S. A. Shaffer[3], S. J. Bograd[4], E. L. Hazen[4], D. G. Foley[4], G. A. Breed[2,5], A.-L. Harrison[5], J. E. Ganong[1], A. Swithenbank[1], M. Castleton[1], H. Dewar[6], B. R. Mate[7], G. L. Shillinger[1], K. M. Schaefer[8], S. R. Benson[9], M. J. Weise[5], R. W. Henry[5] & D. P. Costa[5]

Outline of Presentation

- Introduction to the data
- Exploratory analyses & environmental sampling
- Choosing & running statistical models
- Where to go from here?

TOPP dataset





- From 2000-2010: 23 species; 4,300 tags; >1.1 Million profiles
- Tracking, conservation, ocean observation

Goals and Questions

- How do TOPP species use the ocean?
- Are there predictable variables that correspond to important habitat?
- How do seasonal patterns in the variables correspond to seasonal changes in top predator habitat use?

Dataset challenges

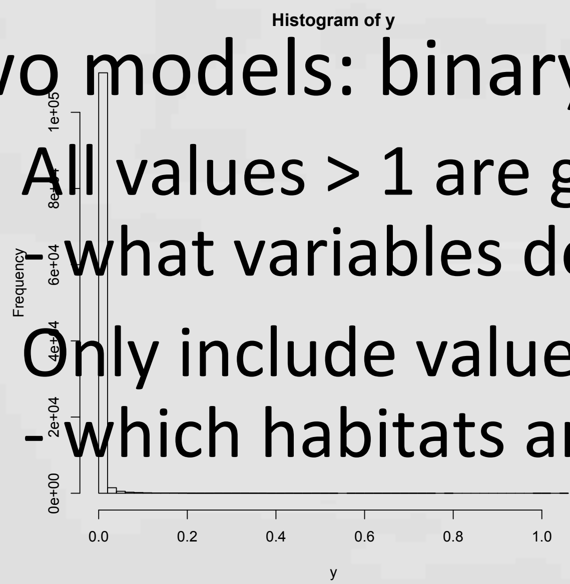
- Tag data are highly correlated spatially and temporally
  - Include spatial autocorrelation explicitly – correlation structure in the error term
  - Bin data at scales broader than correlation (e.g. 1° & ~90 days)
- Long time series can span multiple satellites e.g. SEAWiFS and MODIS for ocean color
- > 1.1 million points

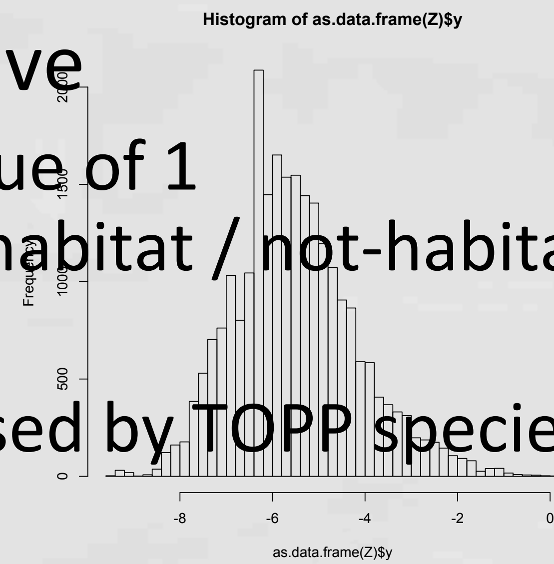
Relative use maps

- How do you convert tag data from multiple species & tag types into useful products?
  - State space models to account for multiple tag types and corresponding positional error
  - Weight track by length - longer tracks get higher weighting to minimize tag bias
  - Normalize by species – 1000 elephant seal tracks = 100 blue whale tracks
  - Relative use ~ f(# of species, # of individuals)

TOPP modeling framework

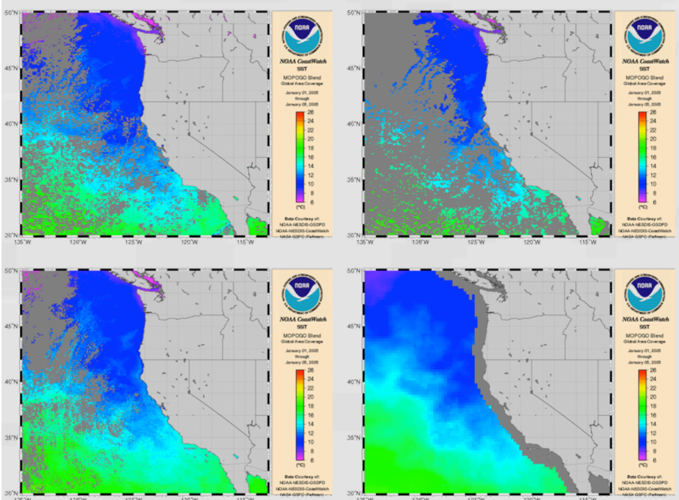
- R for statistics (mgcv, ape, spdep, ncf, AED)
- 5 environmental variables available for entire tracking dataset and spatial extent
- Binned data 1° & quarterly from 2000-2010
- Two models: binary & positive
  - All values > 1 are given a value of 1
    - what variables determine habitat / not-habitat
  - Only include values > 1
    - which habitats are more used by TOPP species

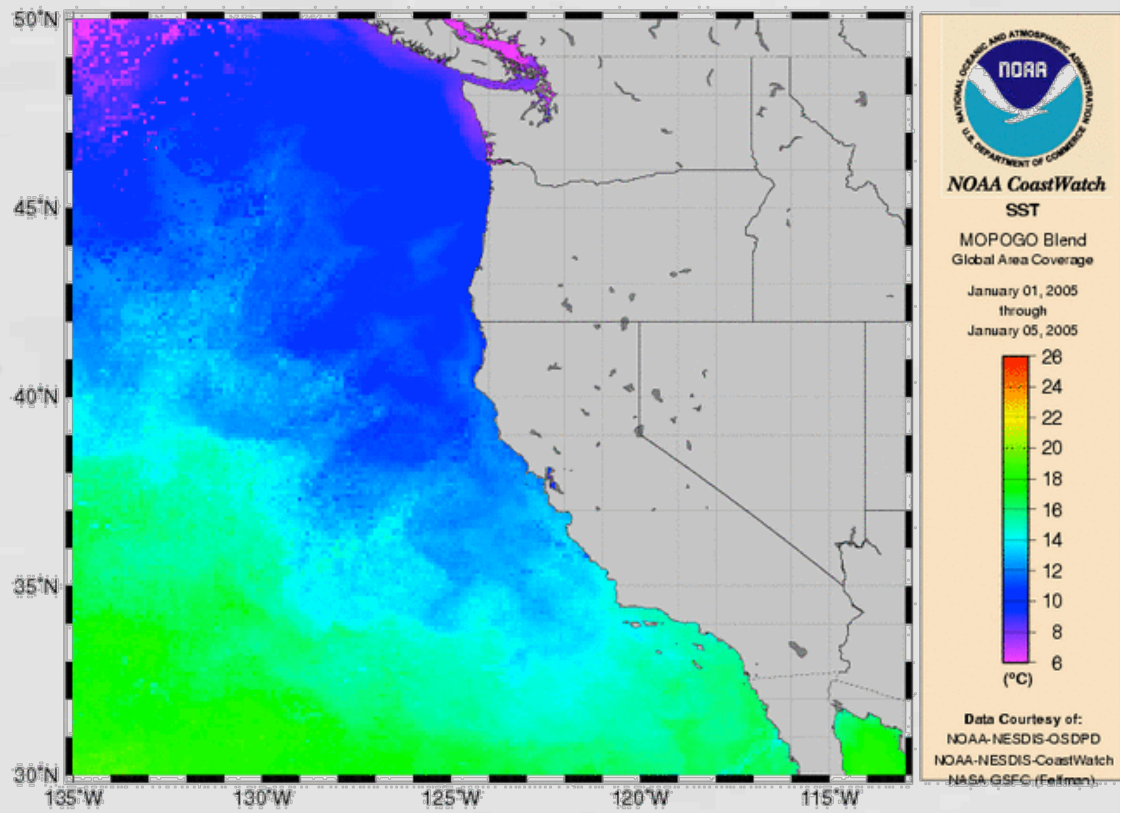




Environmental data

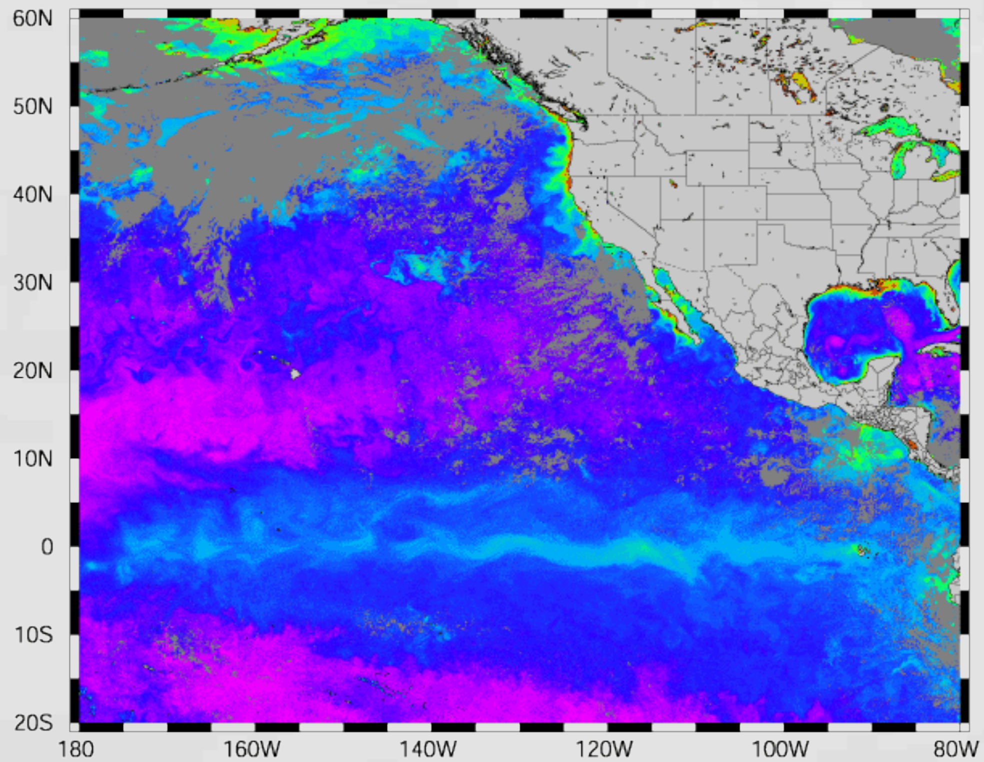
- Blended SST – AVHRR & GOES





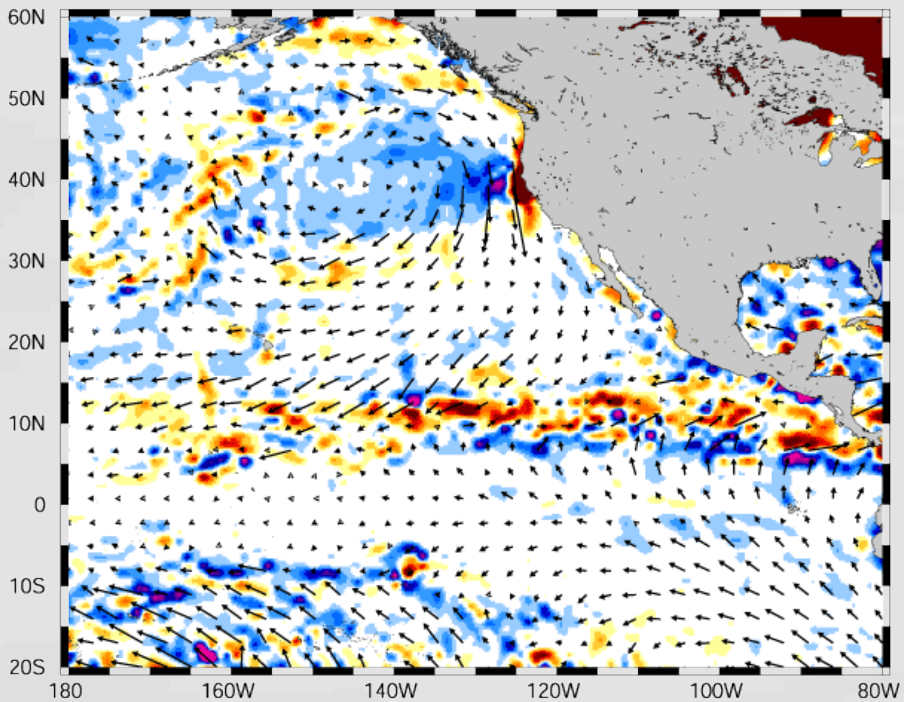
Environmental data

- Blended SST – AVHRR & GOES
- Chl – SeaWiFS (1997-2009) & MODIS (2002-)



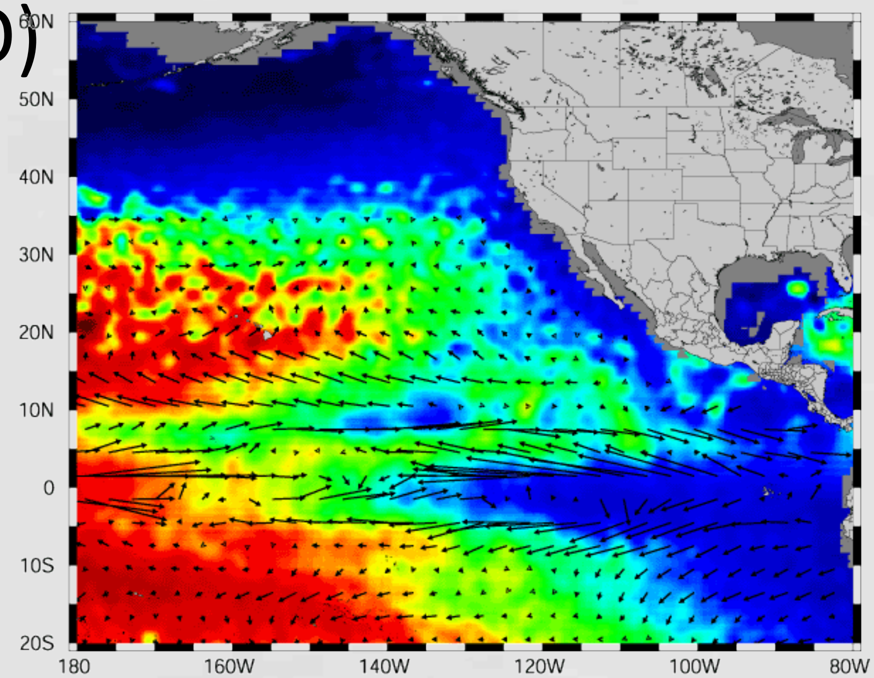
Environmental data

- Blended SST – AVHRR & GOES
- Chl – SeaWiFS (1997-2010) & MODIS (2002-)
- Wind stress curl (QuikSCAT, 1999-2009)



Environmental data

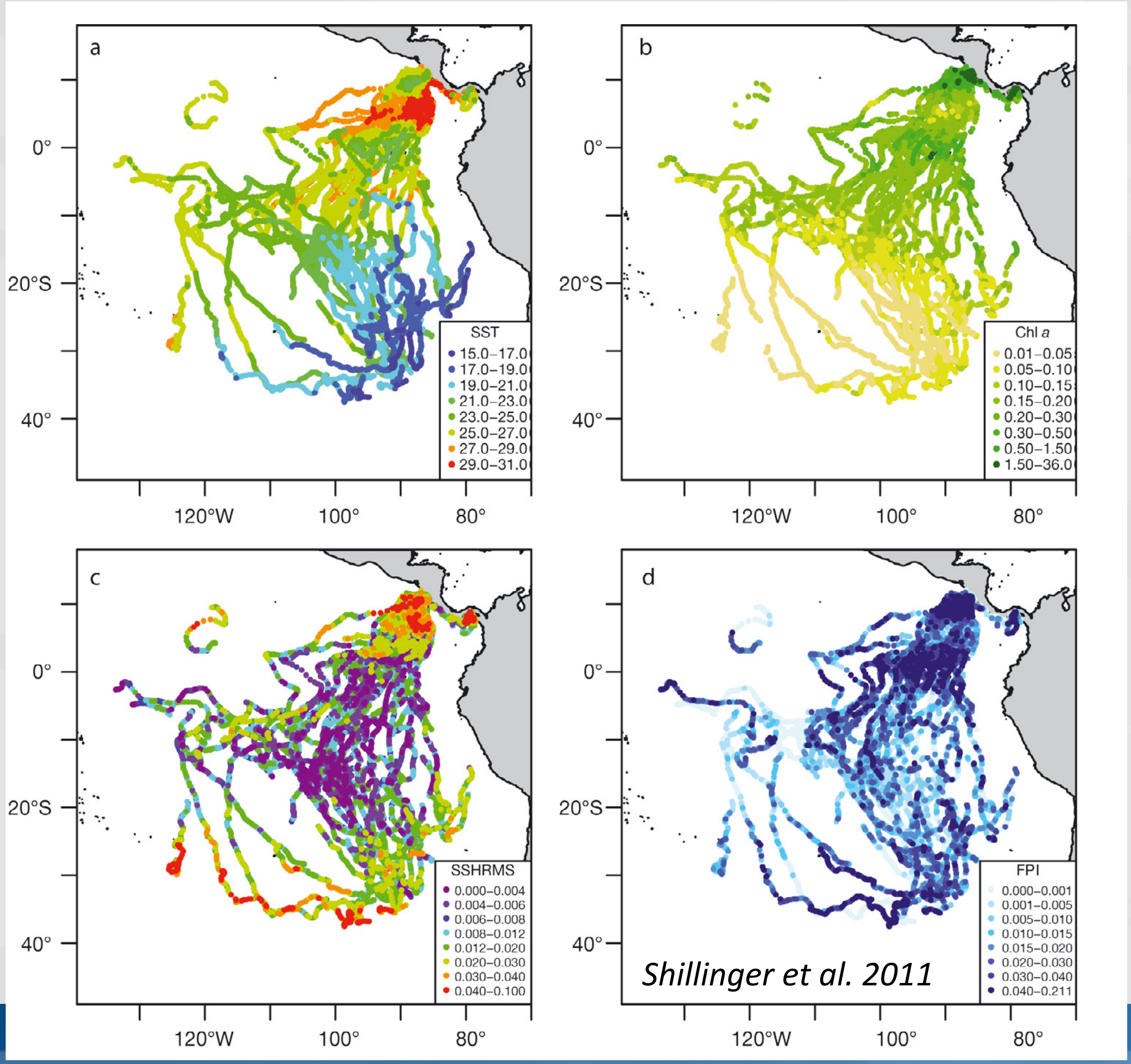
- Blended SST – AVHRR & GOES
- Chl – SeaWiFS (1997-2010) & MODIS (2002-)
- Wind stress curl (QuikSCAT)
- SSHa and SSHrms (AVISO)



Sampling RS environmental data

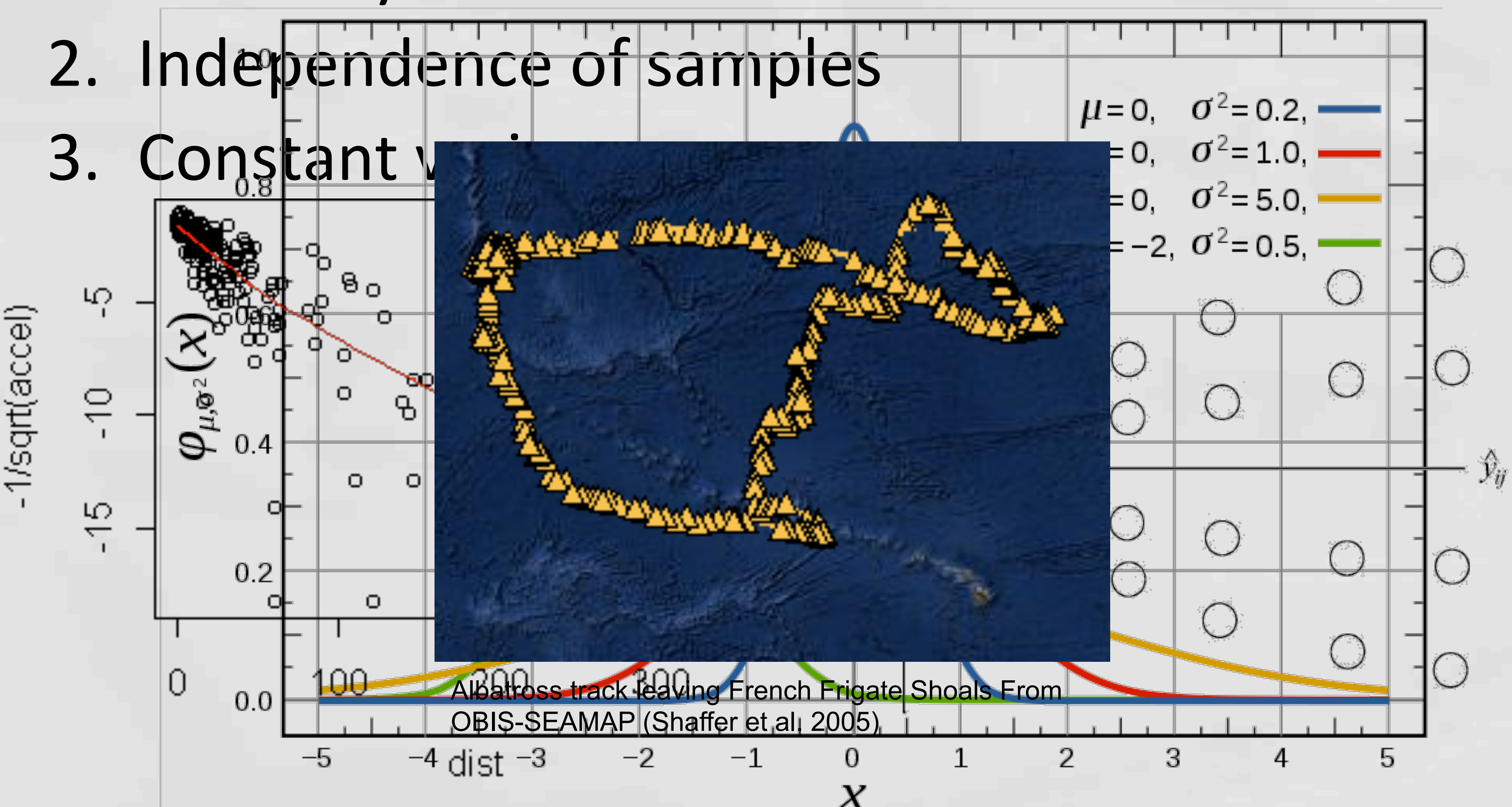
- Near Real Time
  - Live access server
    las.pfel.noaa.gov/oceanwatch.html
  - New and improved browser
    coastwatch.pfel.noaa.gov/coastwatch/CWBrowser.jsp
  - In ARCGIS – Environmental Data Connector
    http://www.pfeg.noaa.gov/products/edc/
- For tagging or other similar point data
  - Xtractomatic – in R and Matlab
    http://coastwatch.pfel.noaa.gov/xtracto/

Sampling RS environmental data



Common modeling assumptions

1. Normally distributed data
2. Independence of samples
3. Constant v



Albatross track leaving French Frigate Shoals From OBIS-SEAMAP (Shaffer et al. 2005)

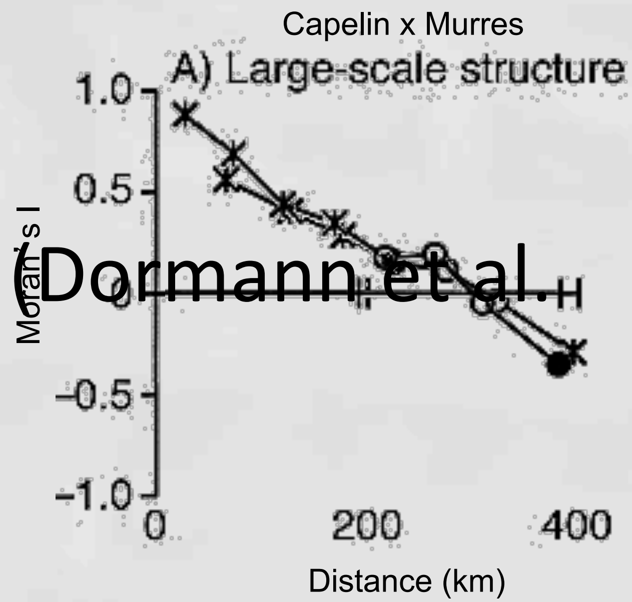
Spatial autocorrelation

- Why do we care?
- Spatial autocorrelation can bias model signif. (see Segurado et al. 2006)
- Spatial autocorrelation is a natural component of spatial data
  - For model residuals, there may be unmeasured spatial variables affecting data

Albatross track leaving French Frigate Shoals
From OBIS-SEAMAP (Shaffer et al. 2005)

Spatial autocorrelation

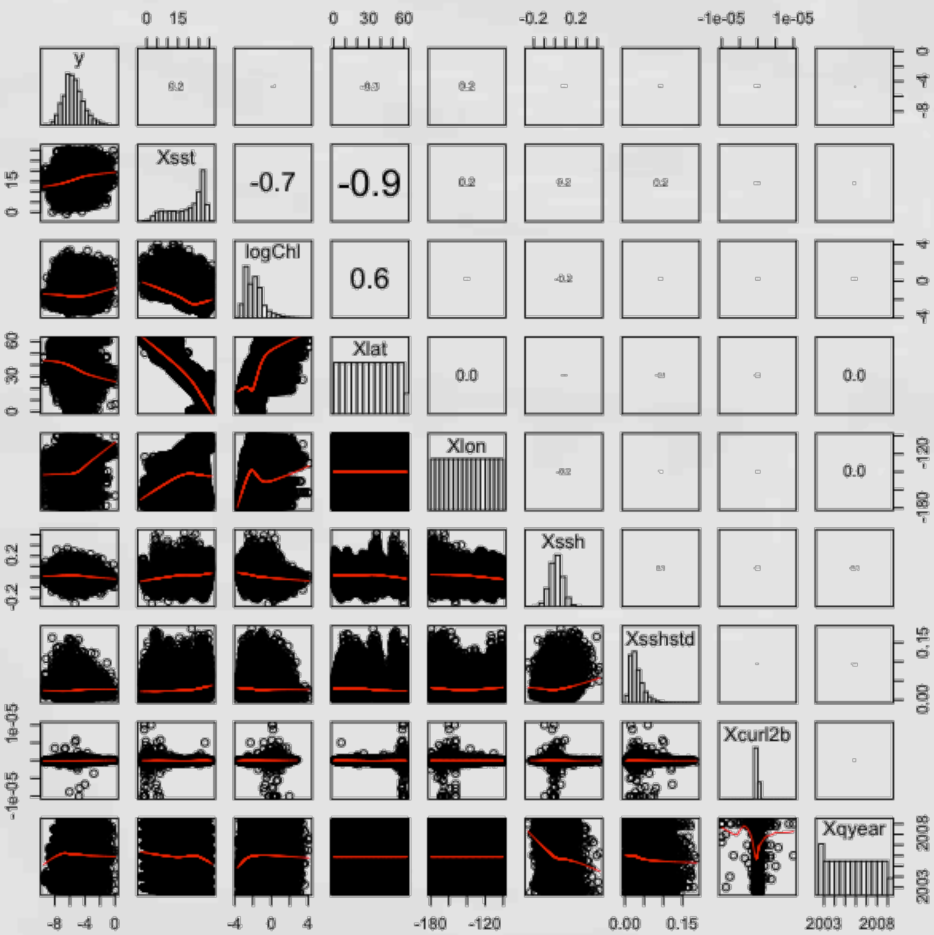
- Ways to test for it
  - Geary's C (0 to 2)
  - Moran's I (-1 to 1)
- Many methods to model it (Dormann et al. 2007)
  - Autocovariate regression & spatial eigenvectors
  - Generalized least squares (GLS), GLMMs, GEEs
  - Partial Mantel tests (Legendre and Legendre 1998)



From Fauchald et al. 2000

TOPP - Exploratory Analysis

- Examined cross-correlations & normalcy (Zuur et al. 2009)



- Chl was log transformed
- other variables were not improved via transformation

Types of statistical models

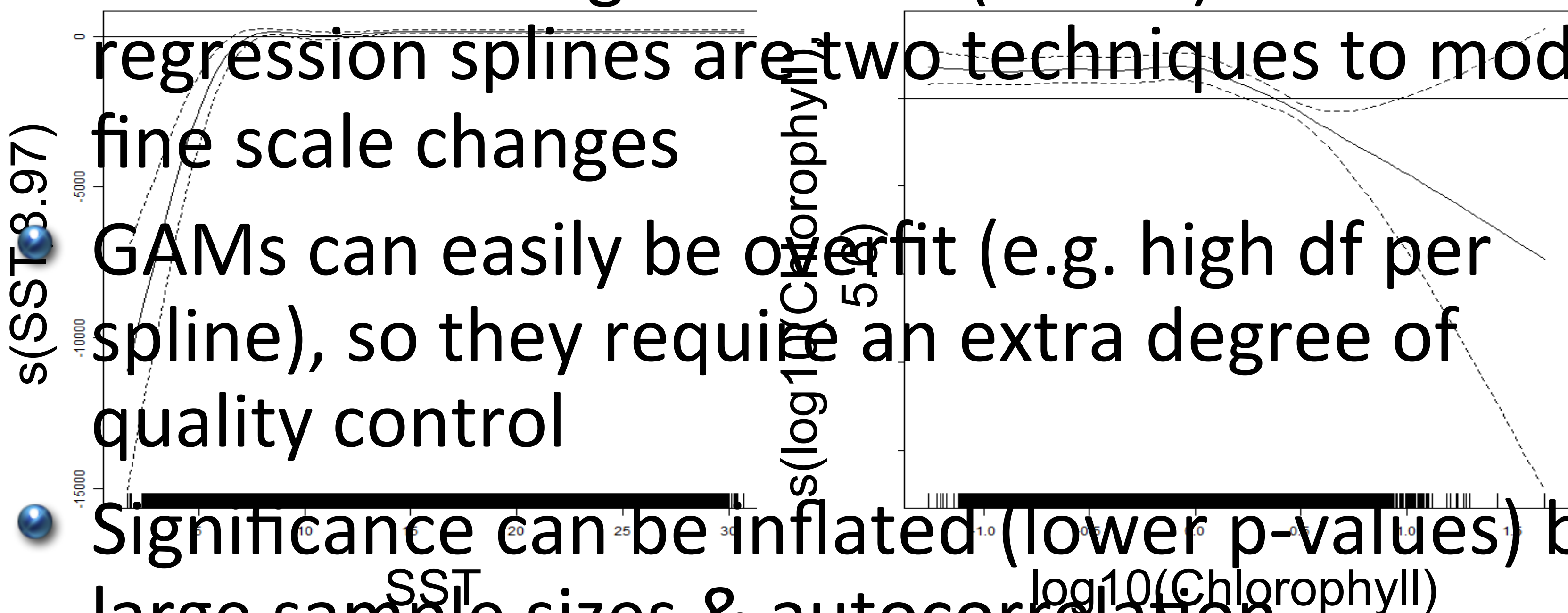
- There are many, and constantly changing / growing
- Correlation/Regression techniques – GLMs, GAMs (Austin 2002), Mixed models (Wood 2006), regression trees & random forests (Breiman 2001)
- Ordination –Multivariate dimensional scaling, e.g. CCAs (Guisan et al. 1999)
- Maximum Entropy models – species distributions "closest to uniform" (Phillips et al. 2006)
- Recent reviews of modeling approaches (Redfern et al. 2006, Elith et al. 2006, Dormann et al. 2007, Aarts et al. 2008, Elith et al. 2009)

Generalized Additive Models

- GAMs can use a combination of parametric and non-parametric functions:

  $(y \sim A+f(x_1)+f(x_2)\ldots+f(x_n)+\varepsilon)$

- Local smoothing functions (LOESS) and regression splines are two techniques to model fine scale changes



- GAMs can easily be overfit (e.g. high df per spline), so they require an extra degree of quality control
- Significance can be inflated (lower p-values) by large sample sizes & autocorrelation

TOPP GAMM Models

- Full models
  - RelUse ~ SST + log(chl) + SSH + SSHrms + Curl + Lat + Lon, family = gaussian, random = quarter+year
  - PresAbs ~ SST + log(chl) + SSH + SSHrms + Curl + Lat + Lon, family = binomial, random = quarter+year
- Model selection:

Akaike's Information Criterion

$$g(\mu) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

"Full" model

presence ~ s(SST)+s(prey.density)

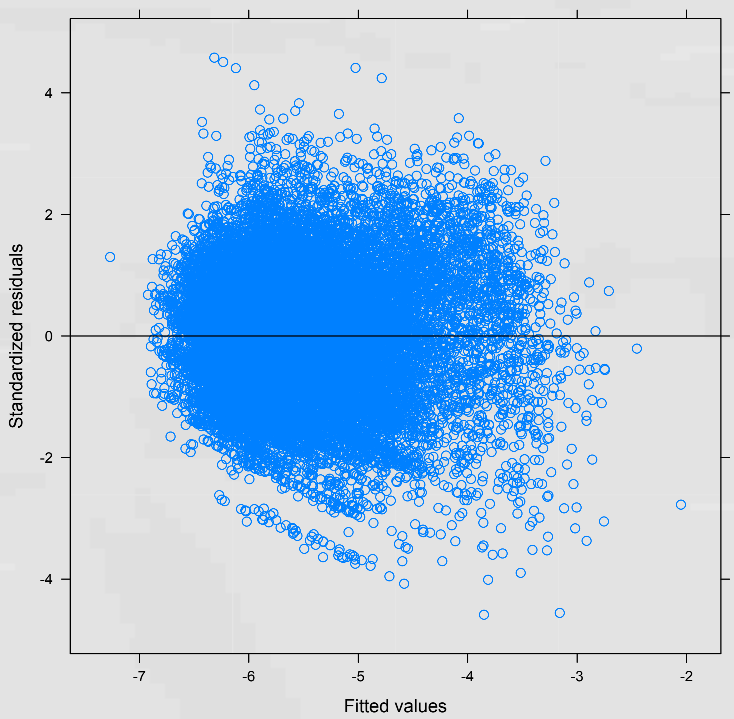
"Final" model

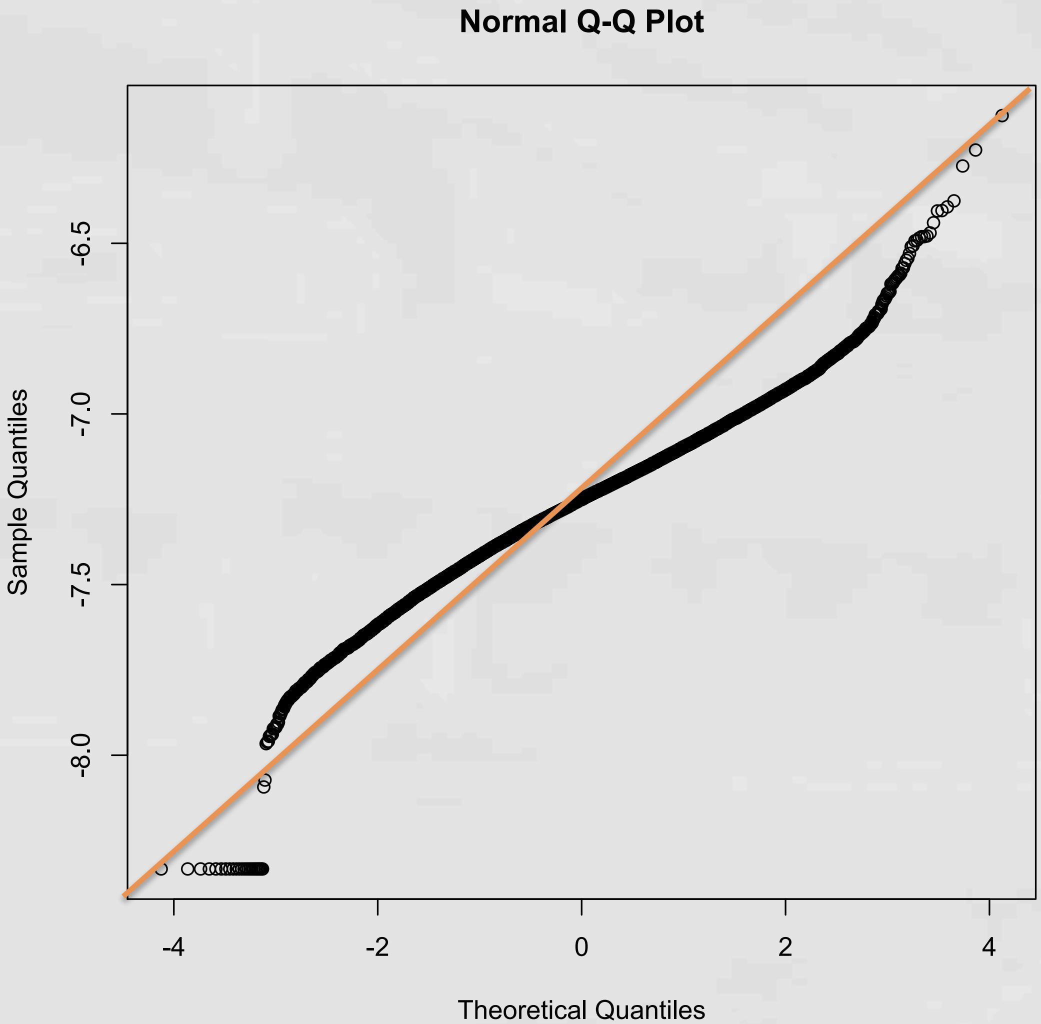
AIC = 2*k*-2ln(*L*)

n – sample size
k – number of parameters
L – Likelihood function

TOPP GAMM residuals

- Normalcy tests





Final GAMM model results

Supplementary Table 7: Summary statistics of fitted GAMMs. Morans I statistic was calculated from the residuals of the fitted models and can range from -1 to 1 with values > 0 indicating spatial clustering and values < 0 indicating dispersion. Morans I statistic p-values indicate that there was residual but non-significant spatial clustering in the presence/absence model and in the relative density model.

	Presence/absence	Relative density
Proportion of deviance explained	0.303	0.232
AIC	598375	74780.45
Moran's I statistic	0.013	0.050
Moran's I p-value	0.053	0.062

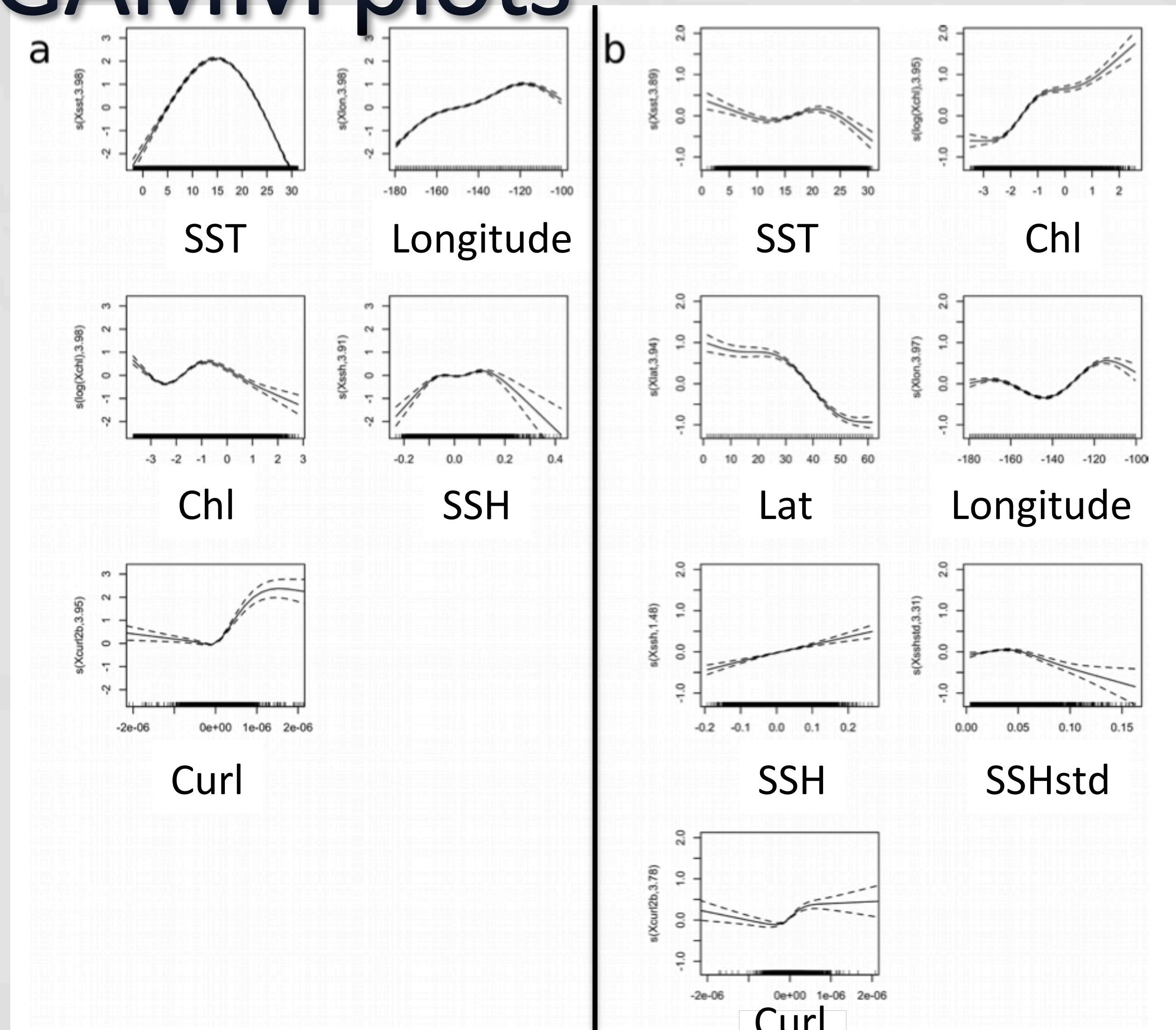
Final GAMM parameters

Supplementary Table 8: Model variables in final fitted presence/absence and relative density GAMMs.

Explanatory variable	Estimated DF	F	p-value
Presence/absence			
s(Xsst)	3.986	2506.99	< 0.001
s(log(Xchl))	3.945	1077.67	< 0.001
s(Xlon)	3.979	131.35	< 0.001
s(Xssh)	3.867	20.3	< 0.001
s(Xcurl2b)	3.935	58.6	< 0.001
Relative density			
s(Xsst)	3.90	31.01	< 0.001
s(log(Xchl))	3.95	226.84	< 0.001
s(Xlat)	3.94	122.57	< 0.001
s(Xlon)	3.97	147.62	< 0.001
s(Xssh)	1.48	72.25	< 0.001
s(Xsshstd)	3.31	20.11	< 0.001
s(Xcurl2b)	3.78	26.63	< 0.001

Final GAMM plots

Presence / Absence

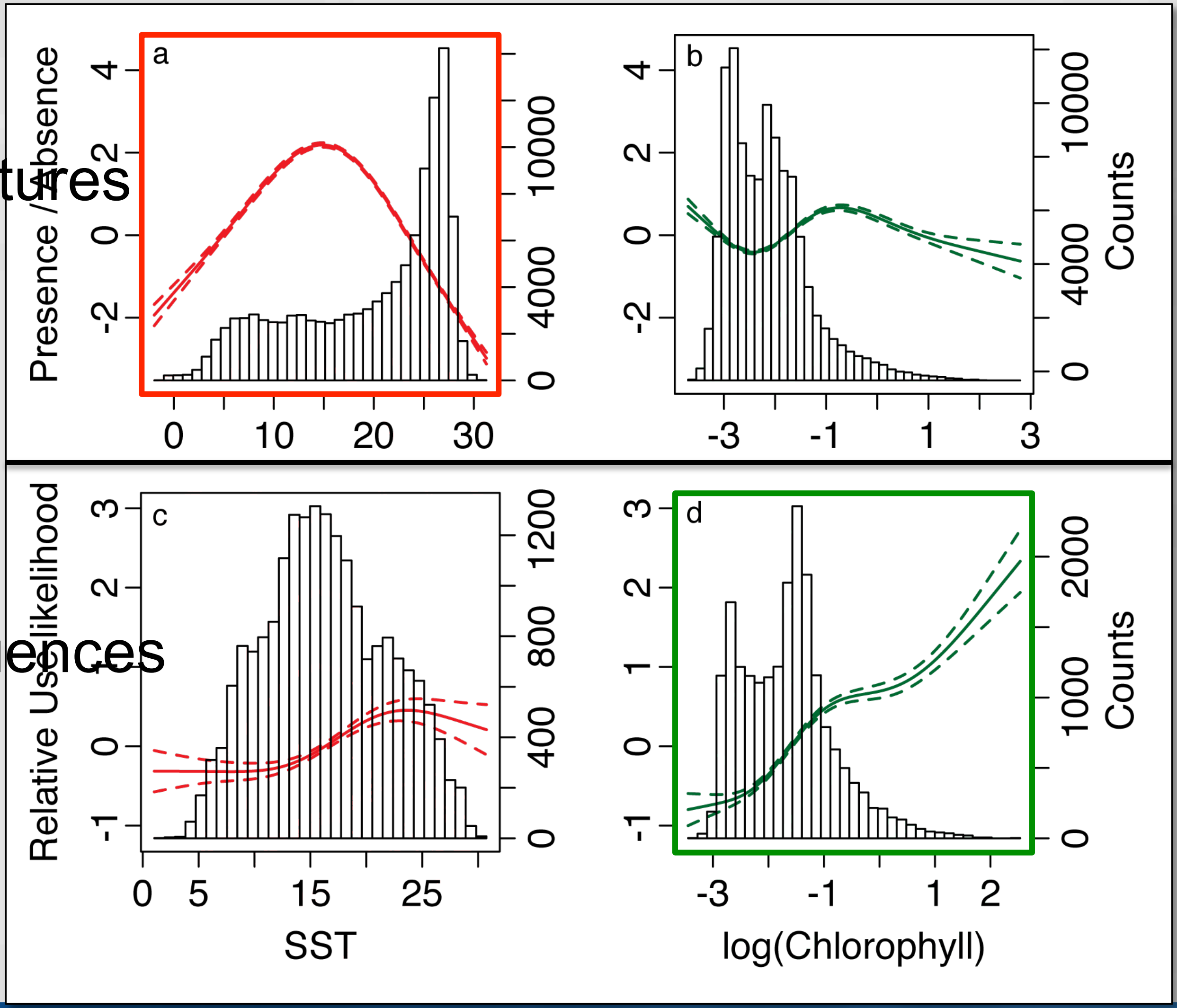


Relative Density

TOPP GAMM model

- SST structures habitat

- Chl-a influences use



Aknowledgements

Alfred P. Sloan Foundation 1934

The David & Lucile Packard Foundation

Gordon and Betty Moore Foundation

NOAA

Sea Grant California

Department of the Navy ONR Science & Technology

NOPP

NSF

Census of Marine Life

NASA