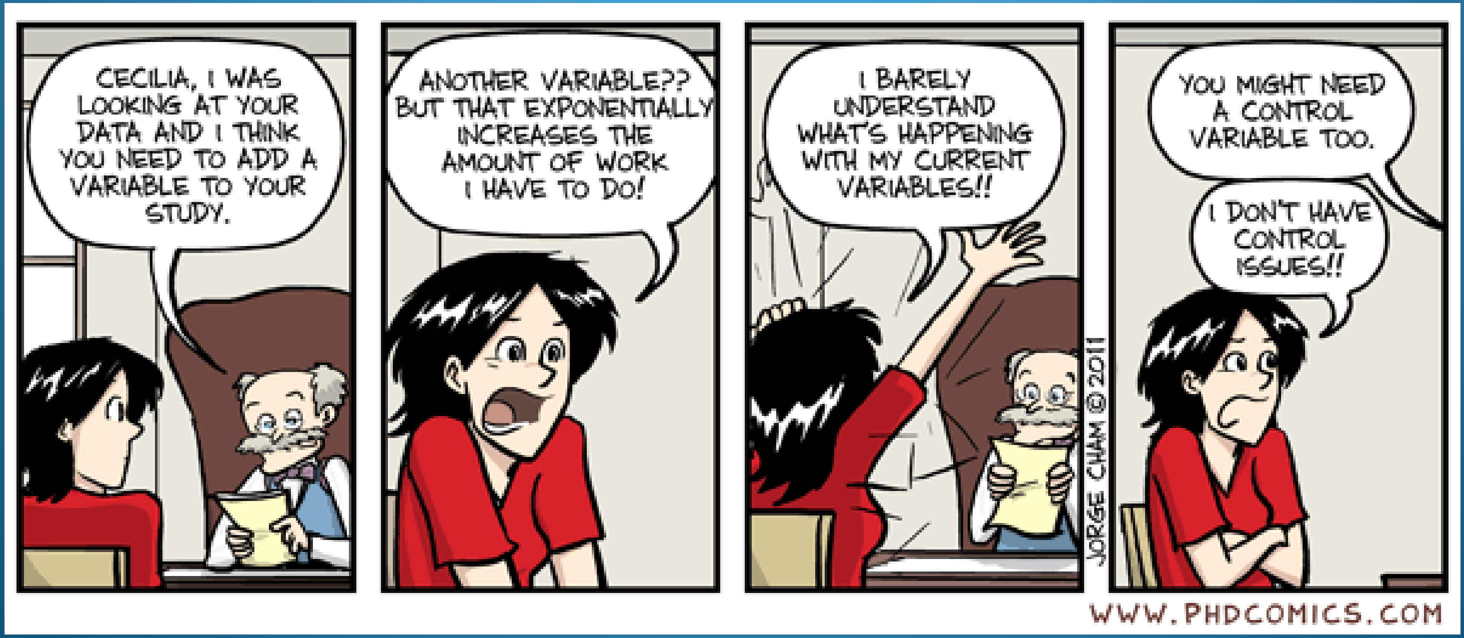
Adding covariates to density surface models

Jason Roberts, October 2015



Topics for this session

- Why use covariates other than x and y?
- What other covariates are there?
- Dynamic spatial covariates: how hard can it be?
- Covariates for our sperm whale model

Many images in this presentation are used without attribution. Please accept my apologies and belated thanks if I have used your image without permission.

Why use covariates other than x and y?

Three common motivations:

1. Desire for ecologically relevant covariates
   - Tie model to ecological theory (but correlation ≠ causation!)
   - Proximal variables → better correlations → better models

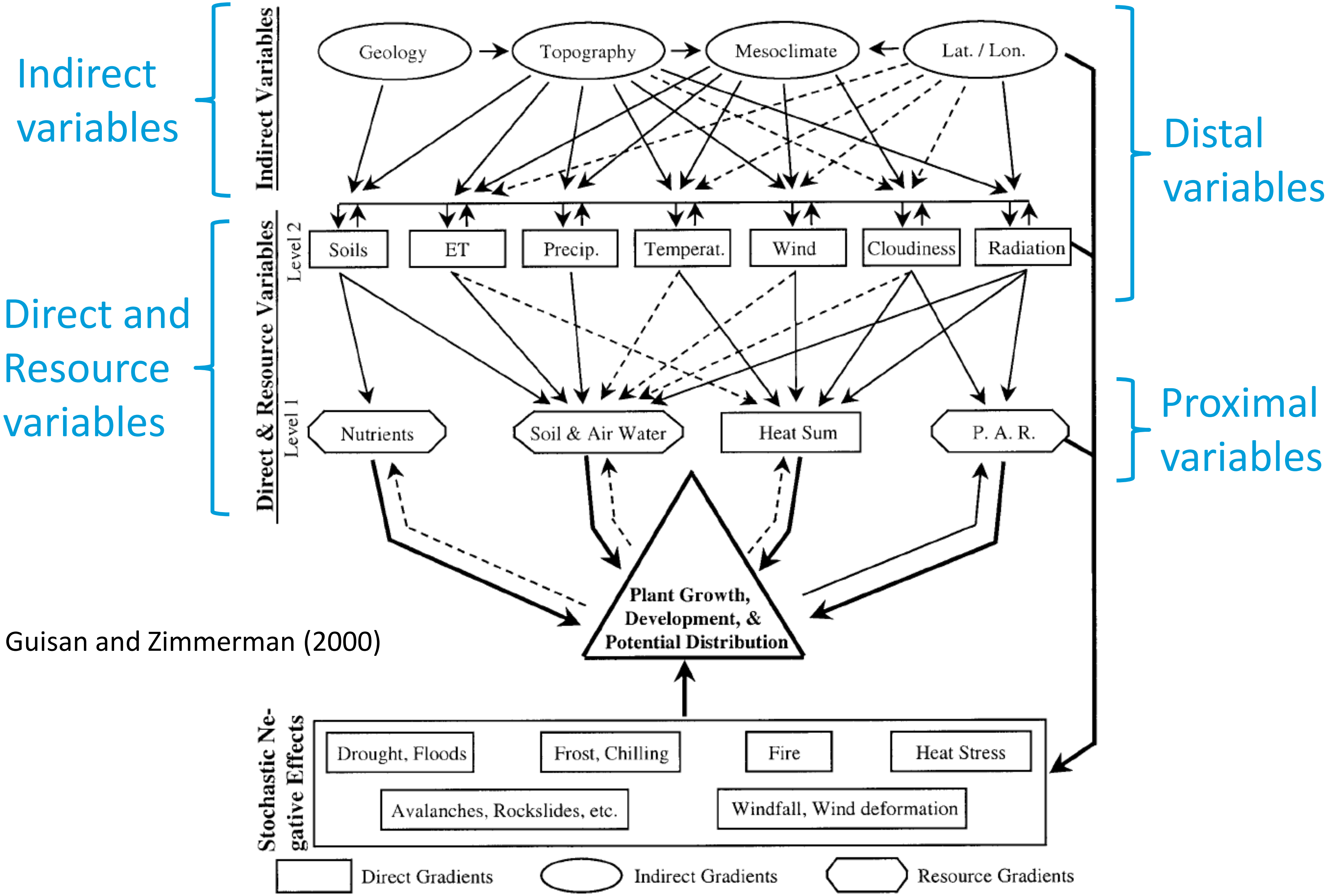
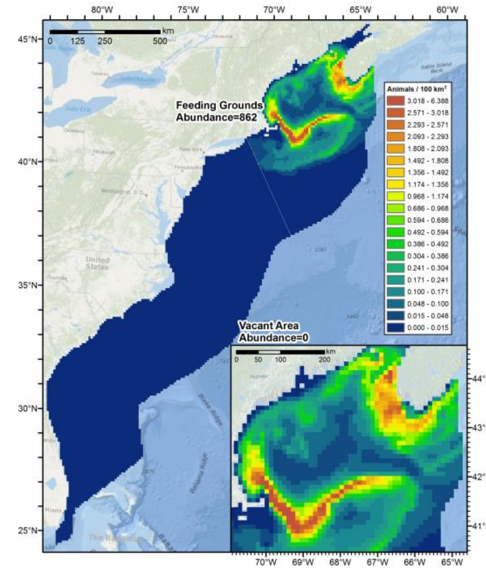


Fig. 3. Example of a conceptual model of relationships between resources, direct and indirect environmental gradients (see e.g. Austin and Smith, 1989), and their influence on growth, performance, and geographical distribution of vascular plants and vegetation.

Habitat-based density models for the U.S. Atlantic and Gulf of Mexico



Marine Geospatial Ecology Lab
Duke University

NASA

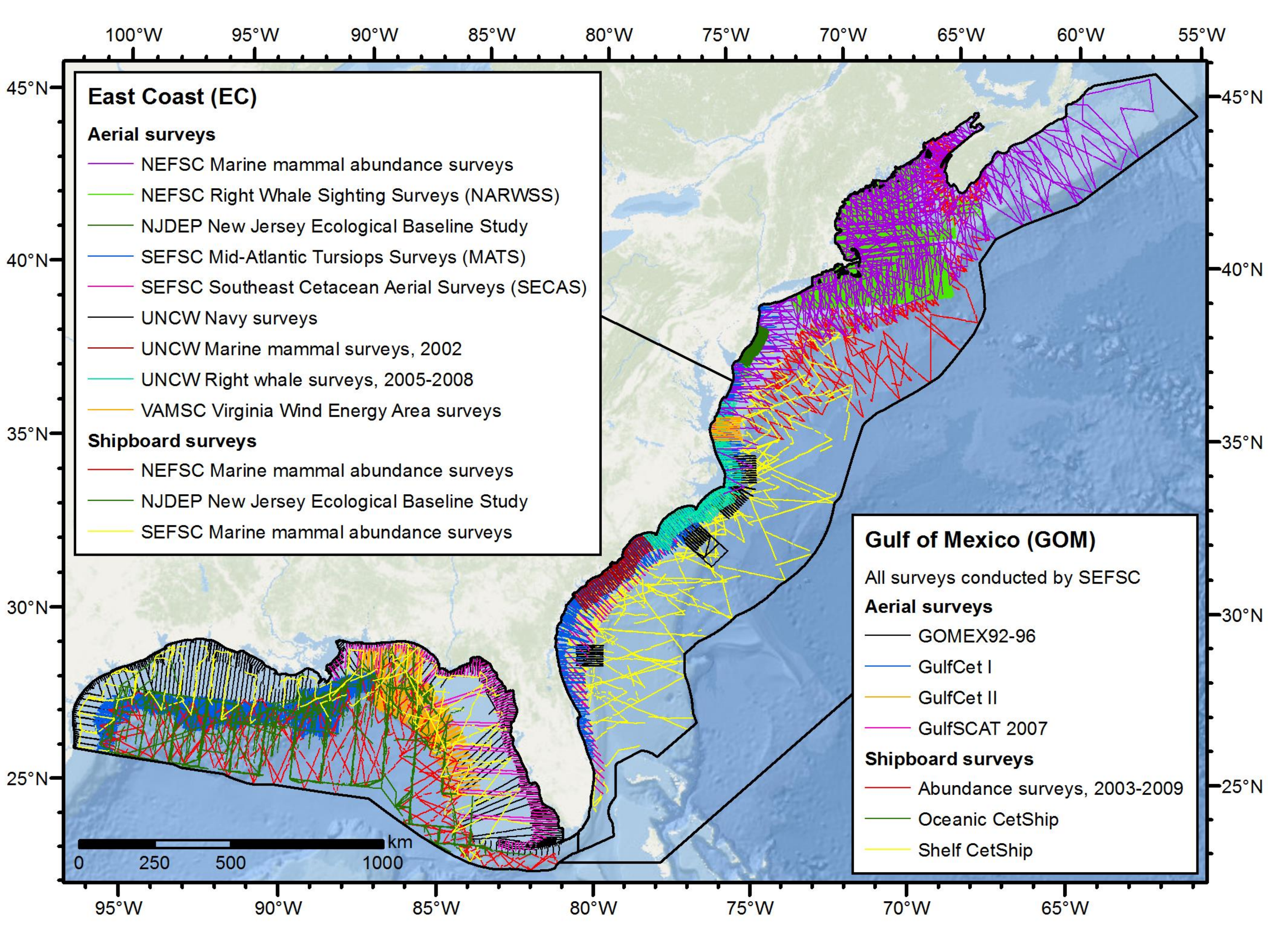
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NOAA
U.S. DEPARTMENT OF COMMERCE

Virginia Coastal Zone
MANAGEMENT PROGRAM

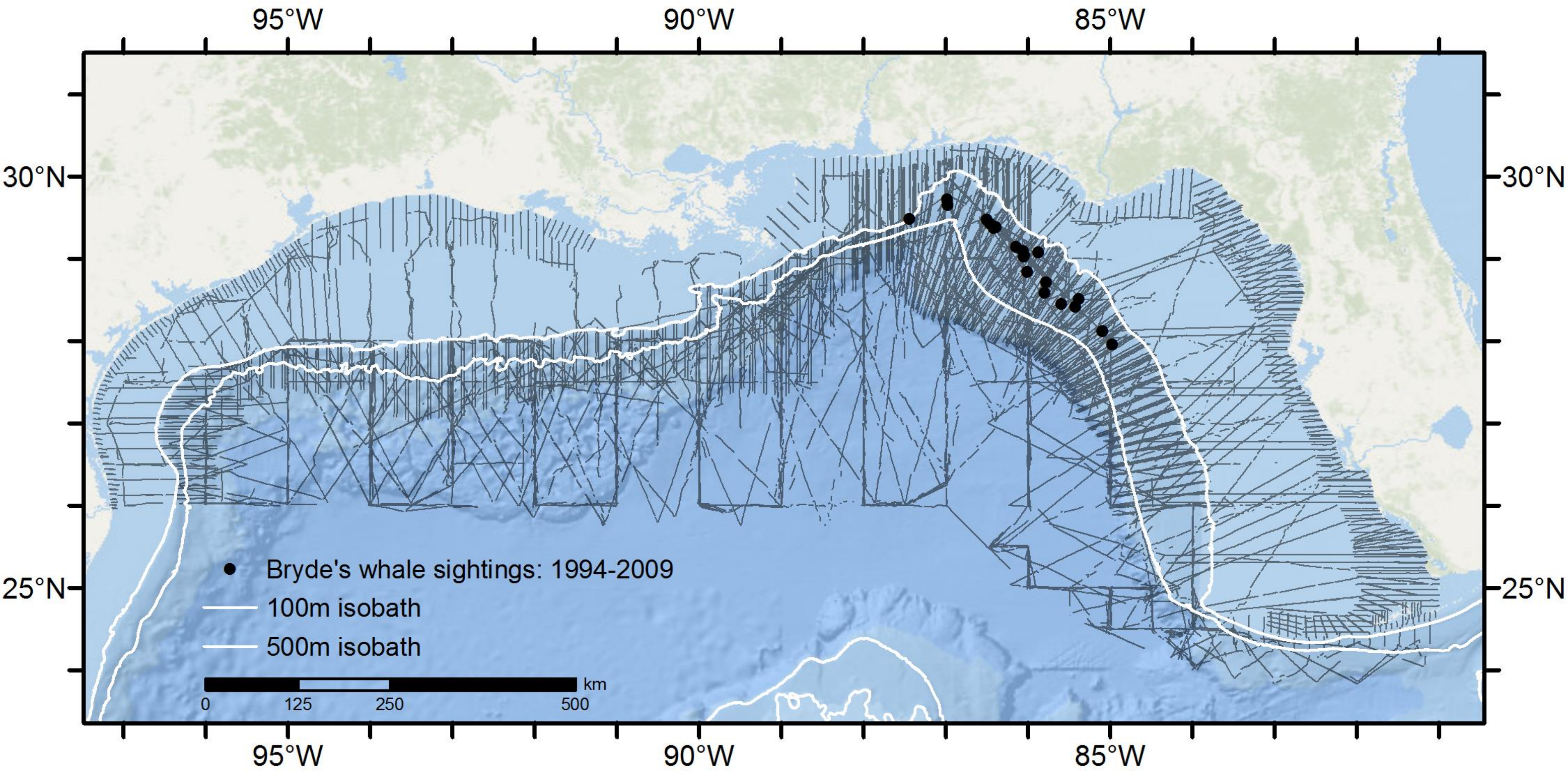
UNCW

NAVFAC
Naval Facilities Engineering Command

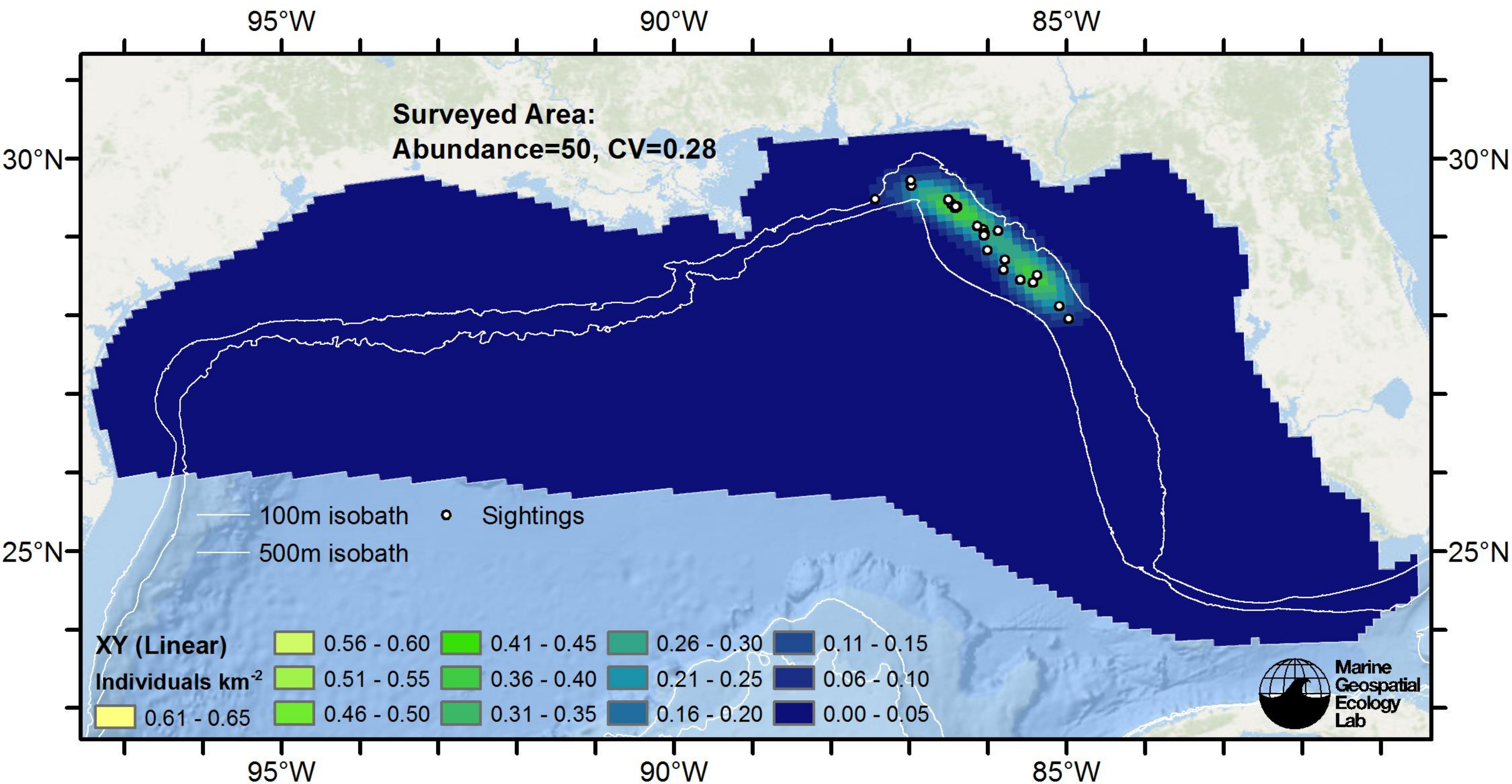
Photo: Whit Welles



100°W
95°W
90°W
85°W
80°W
75°W
70°W
65°W
60°W
55°W
45°N
40°N
35°N
30°N
25°N
East Coast (EC)
Aerial surveys
NEFSC Marine mammal abundance surveys
NEFSC Right Whale Sighting Surveys (NARWSS)
NJDEP New Jersey Ecological Baseline Study
SEFSC Mid-Atlantic Tursiops Surveys (MATS)
SEFSC Southeast Cetacean Aerial Surveys (SECAS)
UNCW Navy surveys
UNCW Marine mammal surveys, 2002
UNCW Right whale surveys, 2005-2008
VAMSC Virginia Wind Energy Area surveys
Shipboard surveys
NEFSC Marine mammal abundance surveys
NJDEP New Jersey Ecological Baseline Study
SEFSC Marine mammal abundance surveys
Gulf of Mexico (GOM)
All surveys conducted by SEFSC
Aerial surveys
GOMEX92-96
GulfCet I
GulfCet II
GulfSCAT 2007
Shipboard surveys
Abundance surveys, 2003-2009
Oceanic CetShip
Shelf CetShip
km
0
250
500
1000



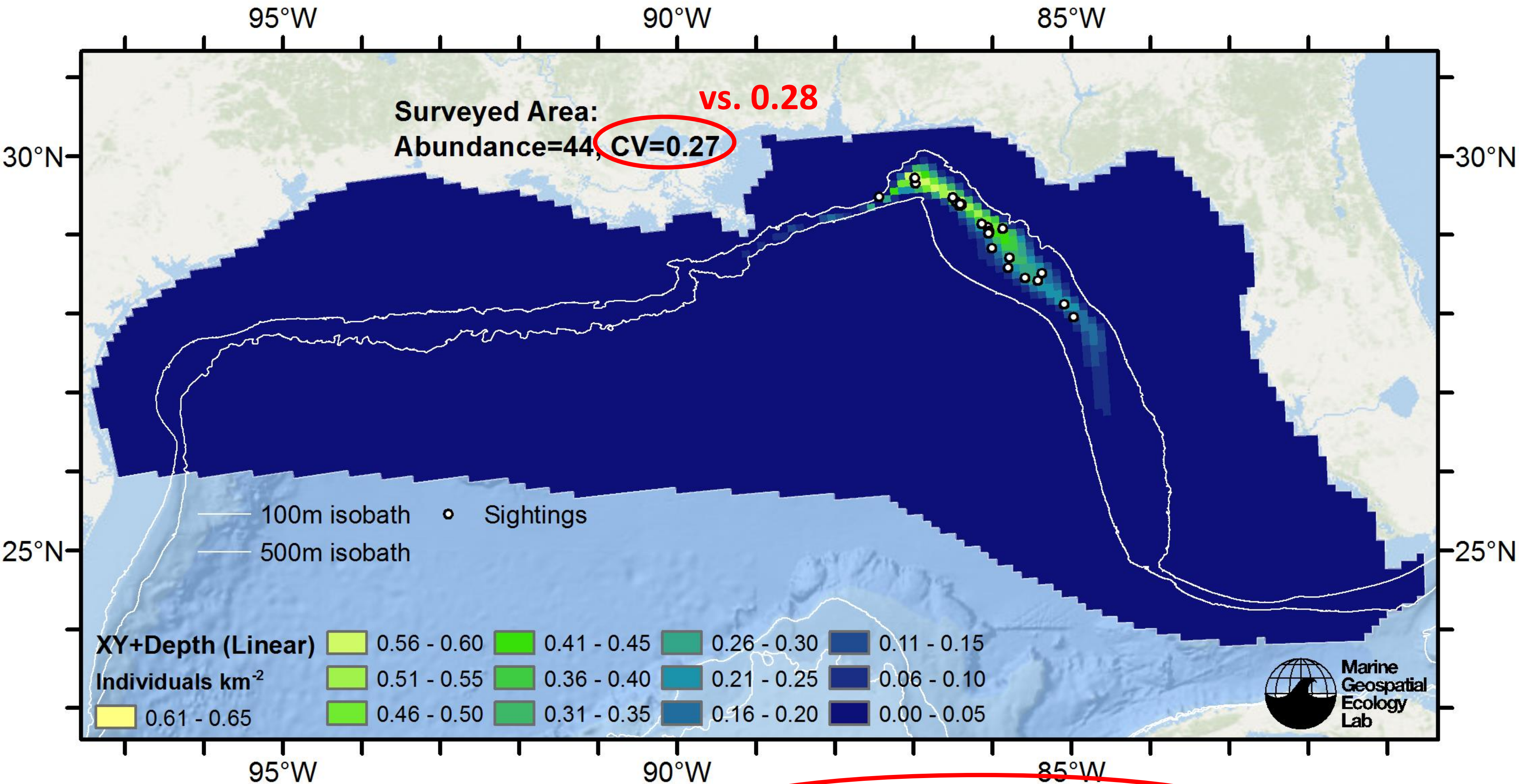
Bryde’s whales sighted on NOAA surveys in the Gulf of Mexico, 1994-2009




```
abundance ~ s(x,y, bs="ts", k=60) +
            offset(log(area_km2))

           edf Ref.df     F p-value
s(x,y) 10.09      59 0.368 0.00957 **

R-sq.(adj) =  0.012   Deviance explained =   49%
-REML = 165.54  Scale est. = 13.017    n = 13163
```




```
abundance ~ s(x,y, bs="ts", k=60) + s(log10(Depth), bs="ts", k=5) +
            offset(log(area_km2))
```

vs. 10.09

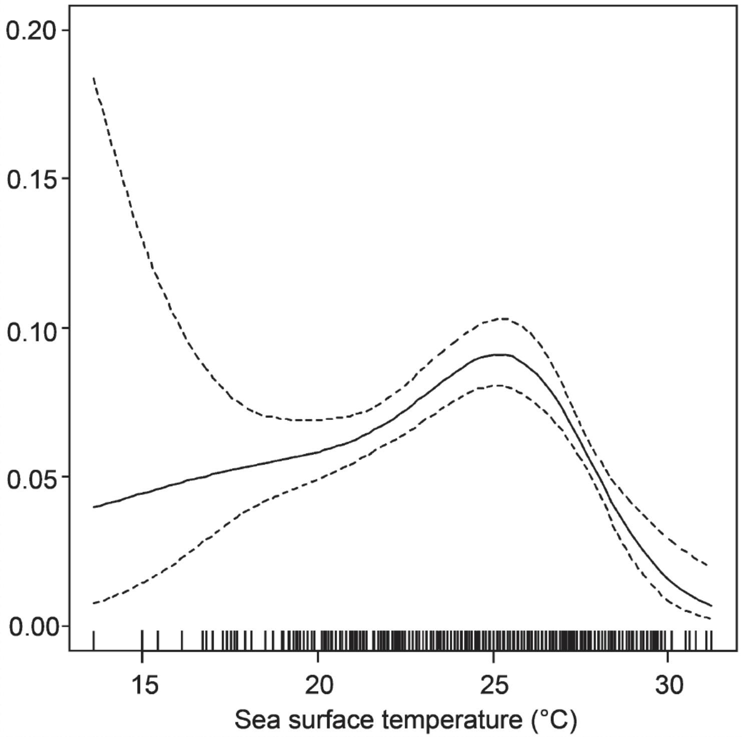
```
                  edf Ref.df     F p-value
s(x,y)          1.736     44 0.203 0.00615 **
s(log10(Depth)) 1.907      4 2.099 0.01087 *
```

vs. 165.54

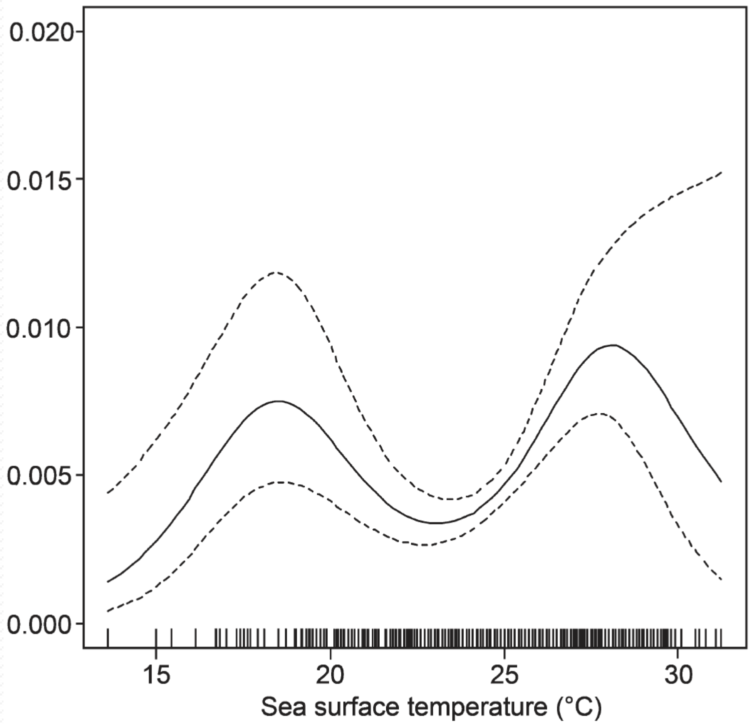
```
R-sq.(adj) =  0.0103   Deviance explained = 50.4%
-REML = 147.78  Scale est. = 12.394    n = 13163
```

vs. 49%

Can you interpret the term plots? Should you?



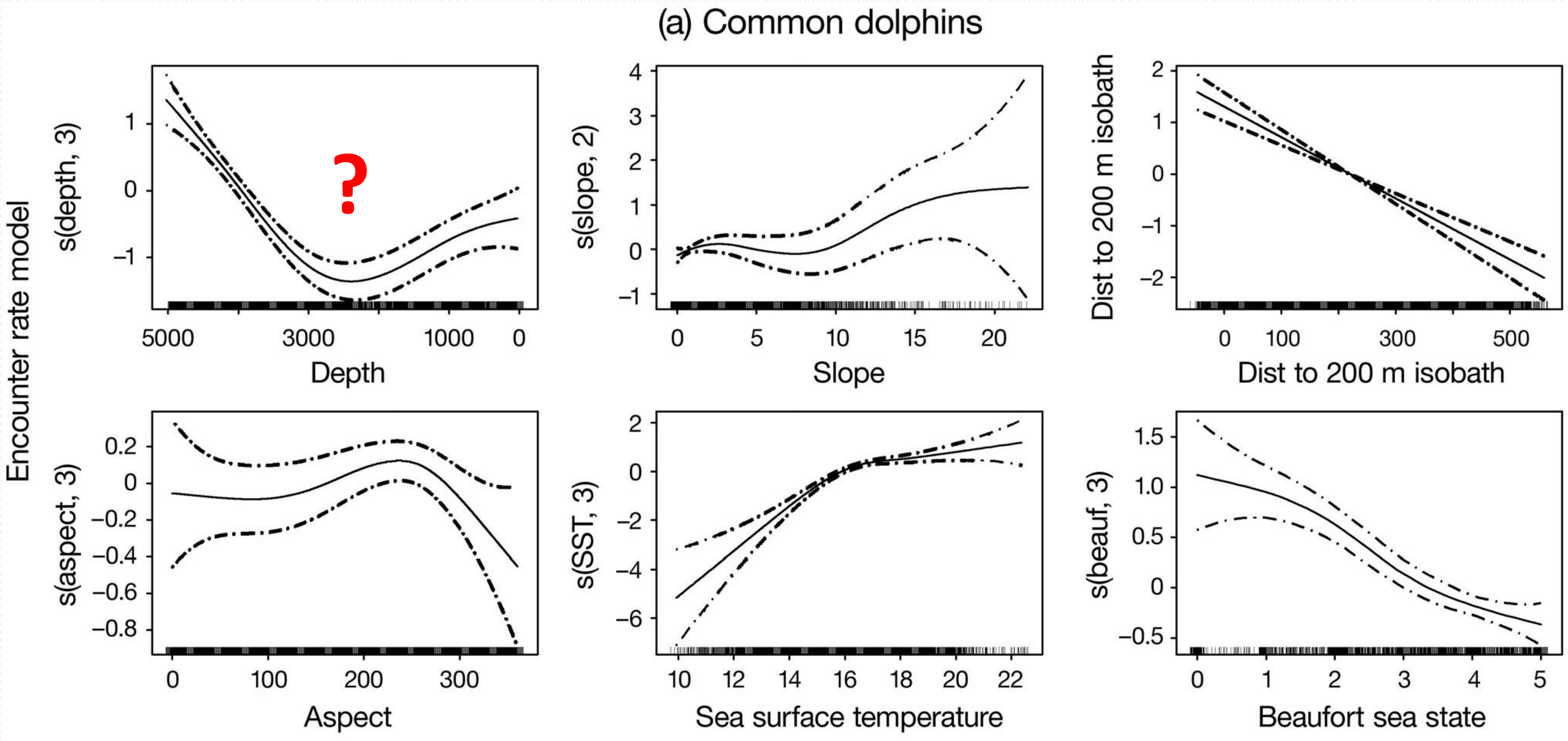
Classic unimodal response from niche theory



What does this mean?

Plots: Read et al. (2014)

How do you interpret the effects of each term in large additive models?



Plots: Becker et al. (2014)

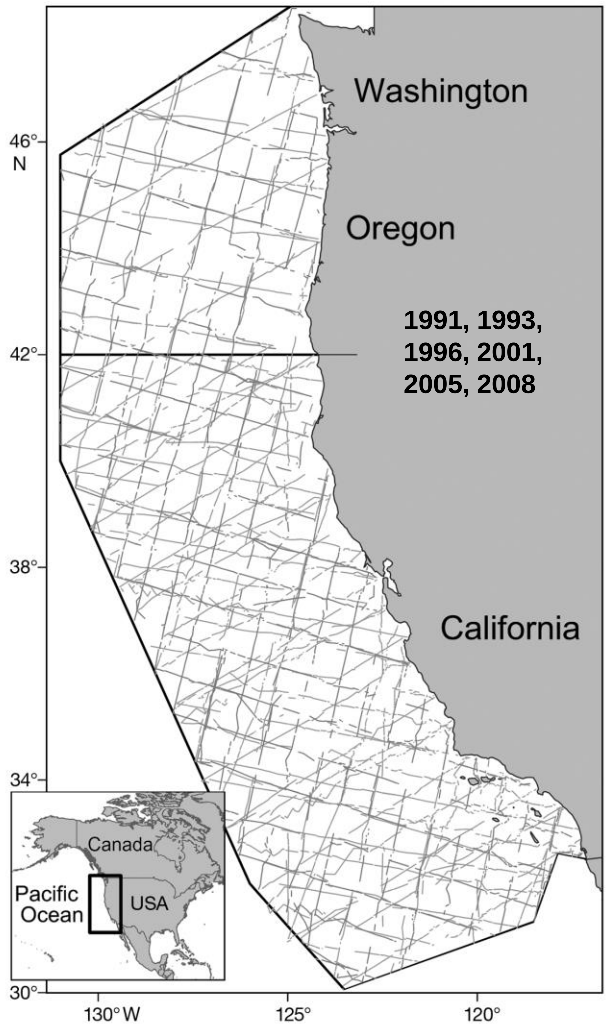
Why use covariates other than x and y?

Three common motivations:

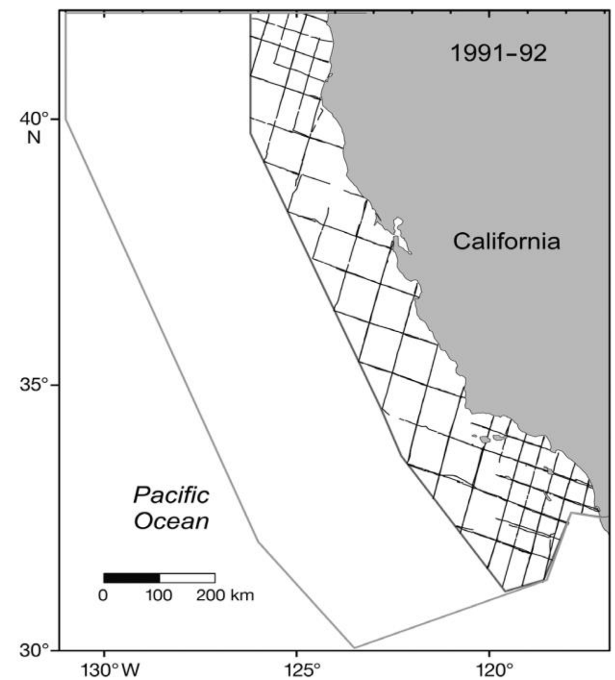
1. Desire for ecologically relevant covariates
   - Tie model to ecological theory (but correlation ≠ causation!)
   - Proximal variables → better correlations → better models
2. Desire to model temporal dynamics
   - E.g. migratory animals, especially in the ocean

Becker et al. (2014) Predicting seasonal density patterns of California cetaceans based on habitat models. *Endang Species Res* 23: 1-22.

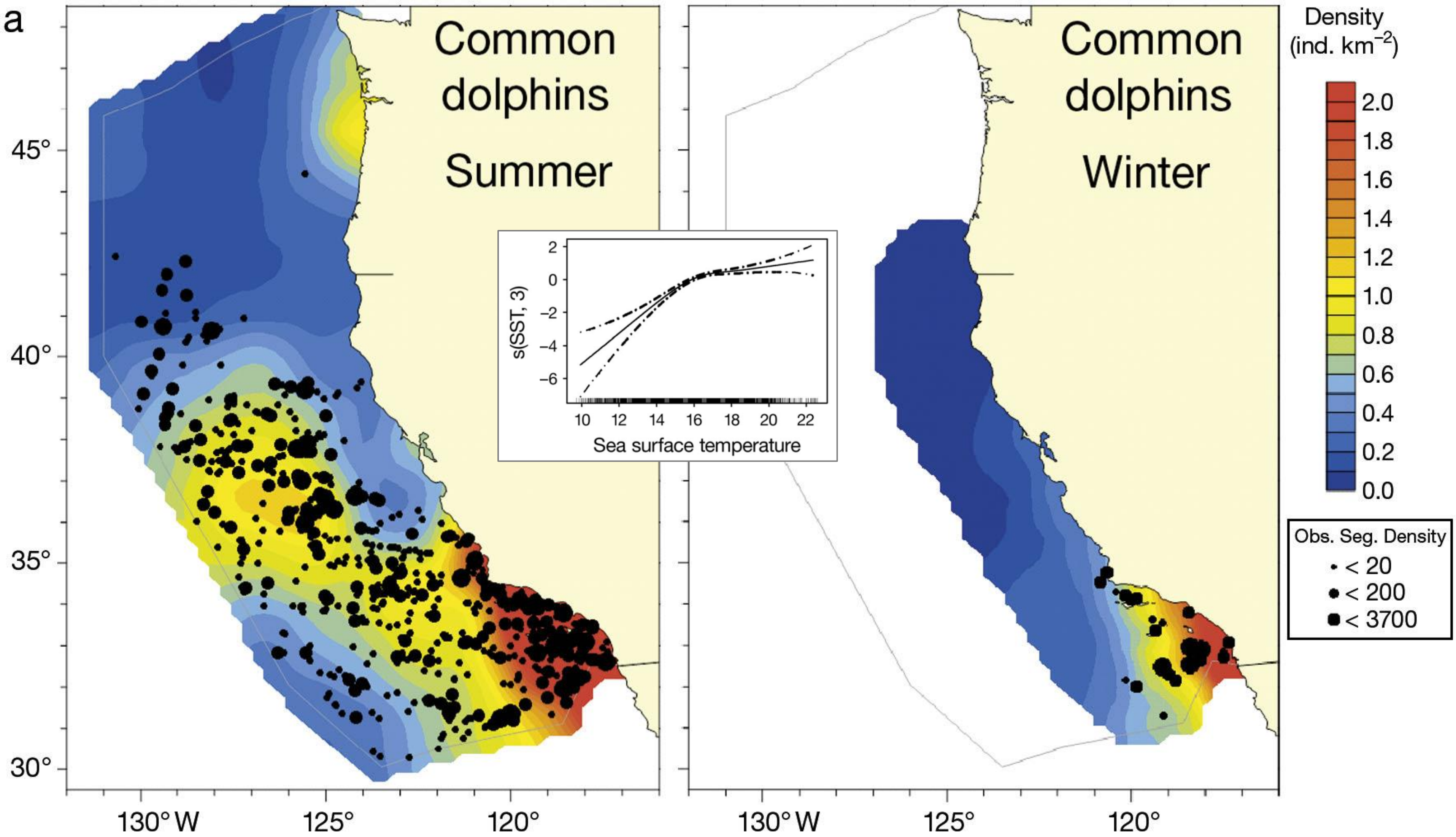
Summer shipboard surveys

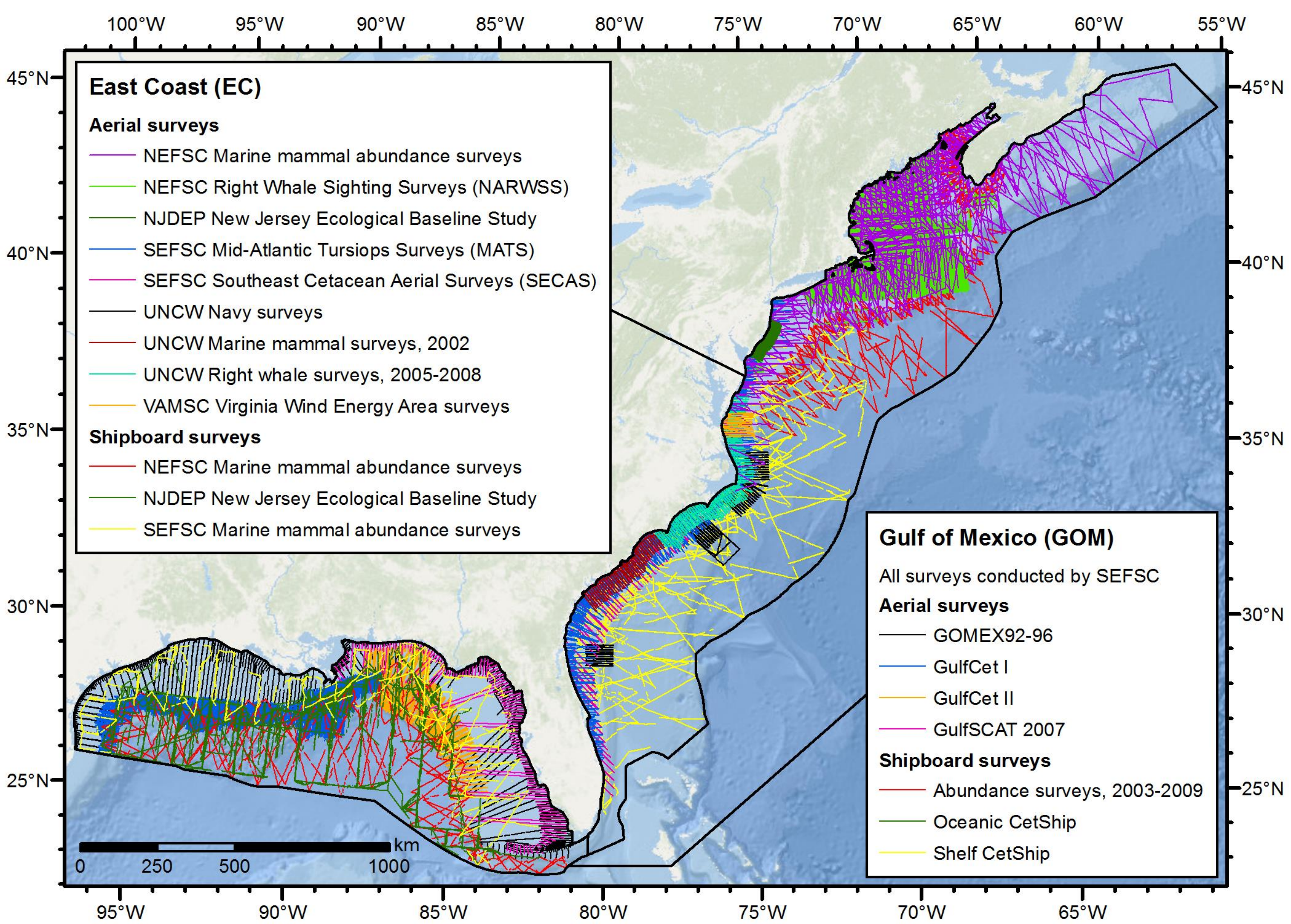


Winter aerial surveys



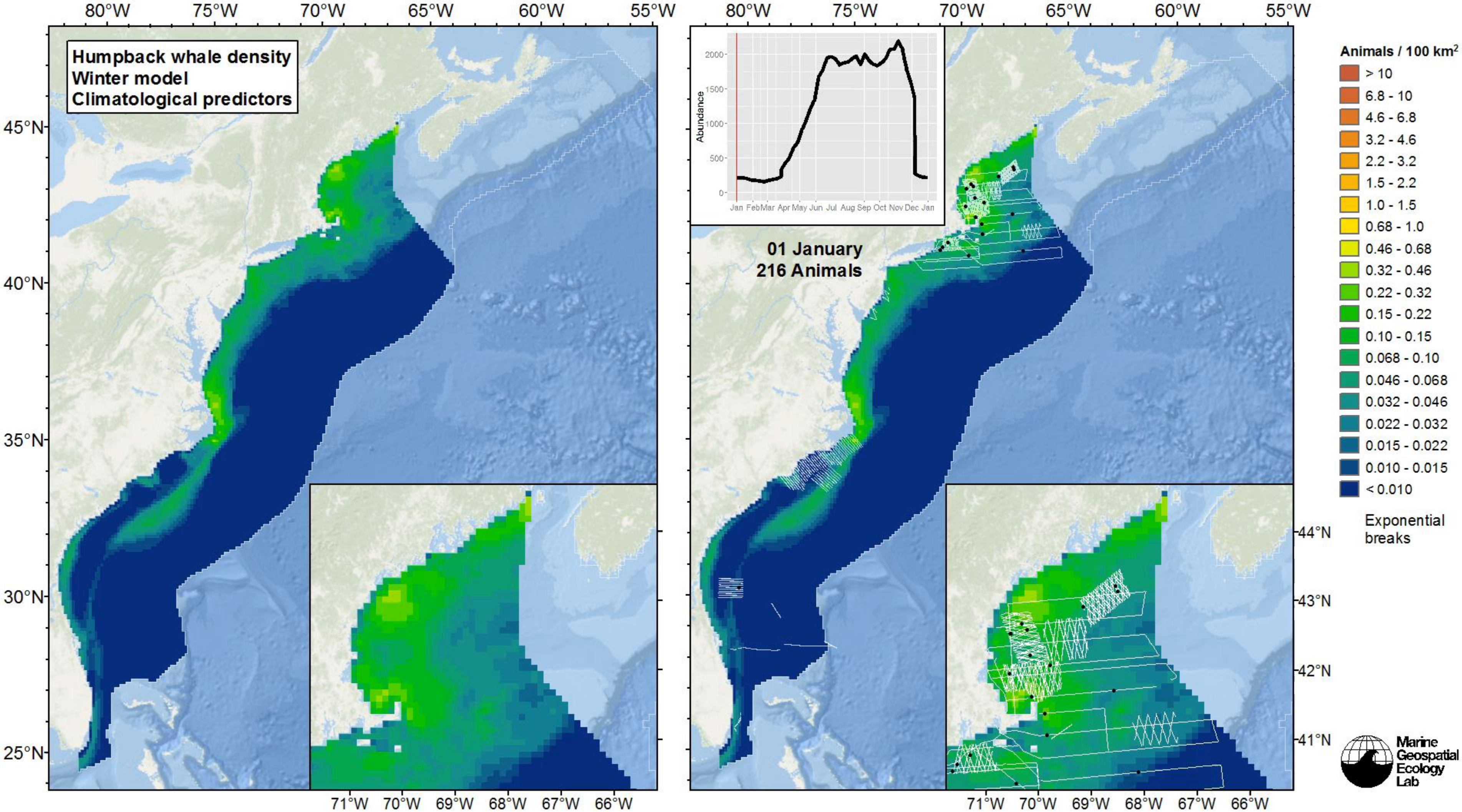
Becker et al. (2014) Predicting seasonal density patterns of California cetaceans based on habitat models. *Endang Species Res* 23: 1-22.





East Coast (EC)
Aerial surveys
NEFSC Marine mammal abundance surveys
NEFSC Right Whale Sighting Surveys (NARWSS)
NJDEP New Jersey Ecological Baseline Study
SEFSC Mid-Atlantic Tursiops Surveys (MATS)
SEFSC Southeast Cetacean Aerial Surveys (SECAS)
UNCW Navy surveys
UNCW Marine mammal surveys, 2002
UNCW Right whale surveys, 2005-2008
VAMSC Virginia Wind Energy Area surveys
Shipboard surveys
NEFSC Marine mammal abundance surveys
NJDEP New Jersey Ecological Baseline Study
SEFSC Marine mammal abundance surveys
Gulf of Mexico (GOM)
All surveys conducted by SEFSC
Aerial surveys
GOMEX92-96
GulfCet I
GulfCet II
GulfSCAT 2007
Shipboard surveys
Abundance surveys, 2003-2009
Oceanic CetShip
Shelf CetShip
km
0
250
500
1000
100°W
95°W
90°W
85°W
80°W
75°W
70°W
65°W
60°W
55°W
45°N
40°N
35°N
30°N
25°N

Roberts et al. (in prep) Habitat-based density models for the U.S. Atlantic and Gulf of Mexico.



Why use covariates other than x and y?

Three common motivations:

1. Desire for ecologically relevant covariates
   - Tie model to ecological theory (but correlation ≠ causation!)
   - Proximal variables → better correlations → better models
2. Desire to model temporal dynamics
   - E.g. migratory animals, especially in the ocean
3. Need to extrapolate beyond the surveyed area
   - Managers ask you to do this

Mannocci et al. (2014) Extrapolating cetacean densities beyond surveyed regions: habitat-based predictions in the circumtropical belt. *J. Biogeogr.* 42: 1267-1280.

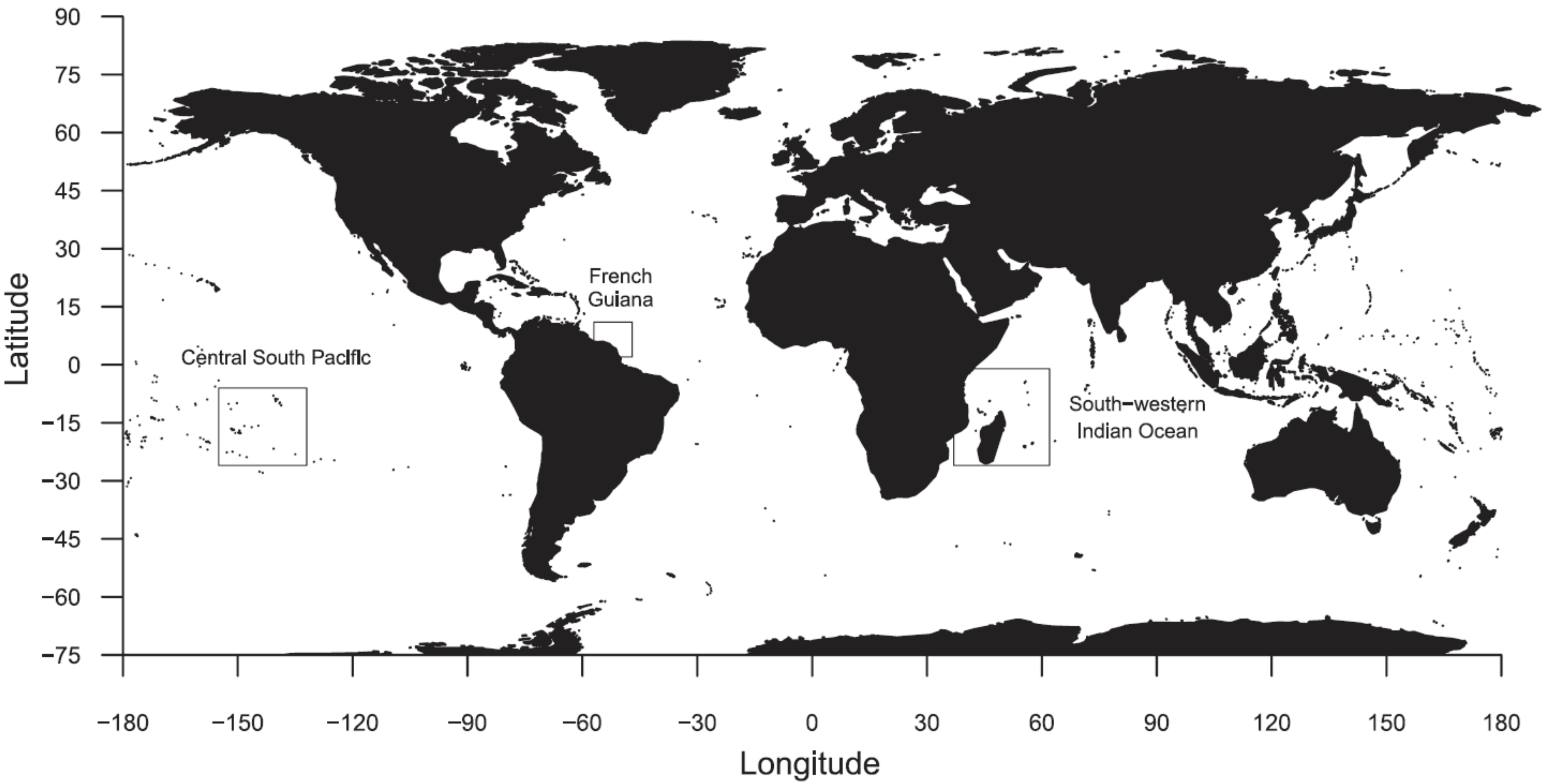
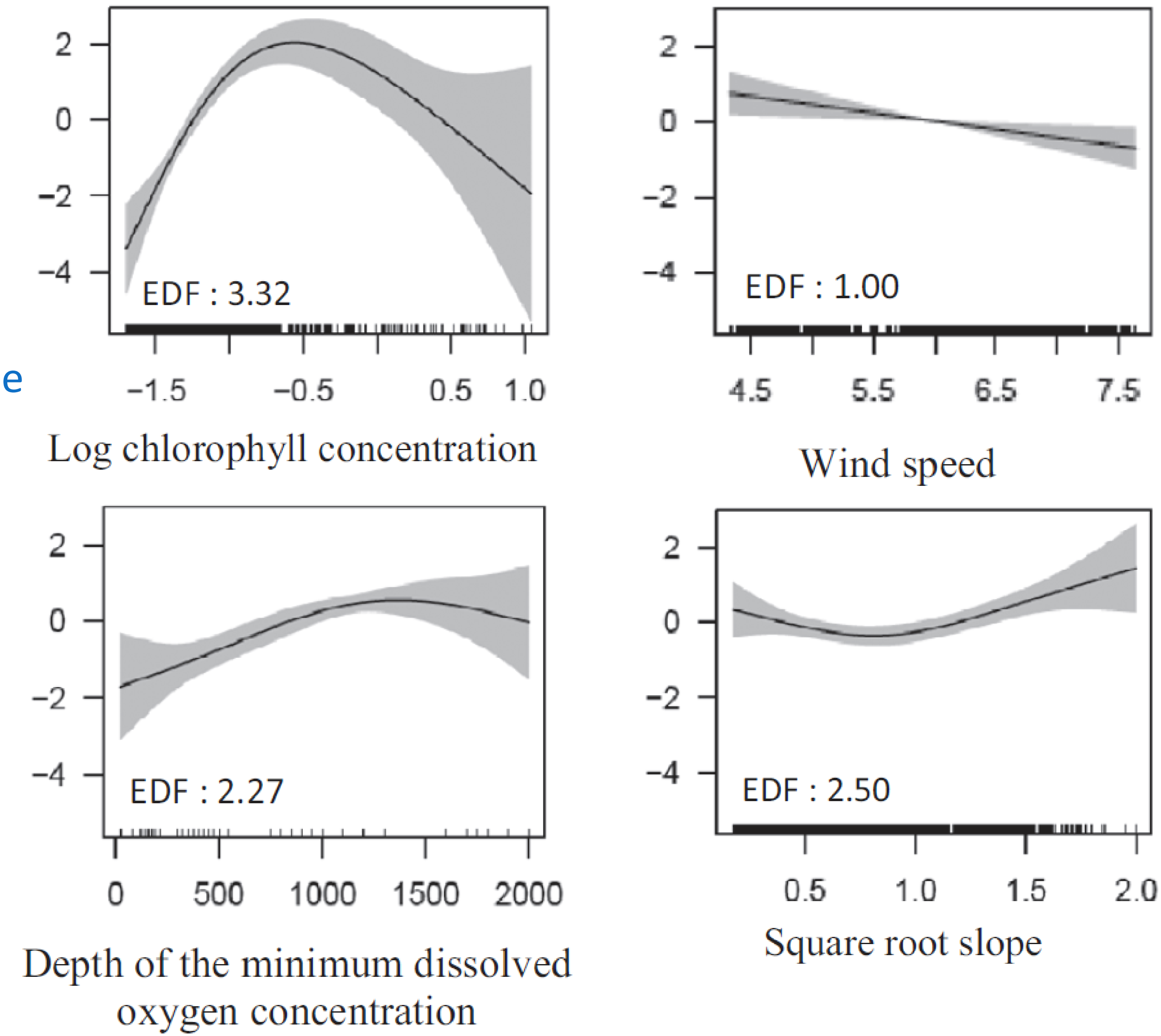


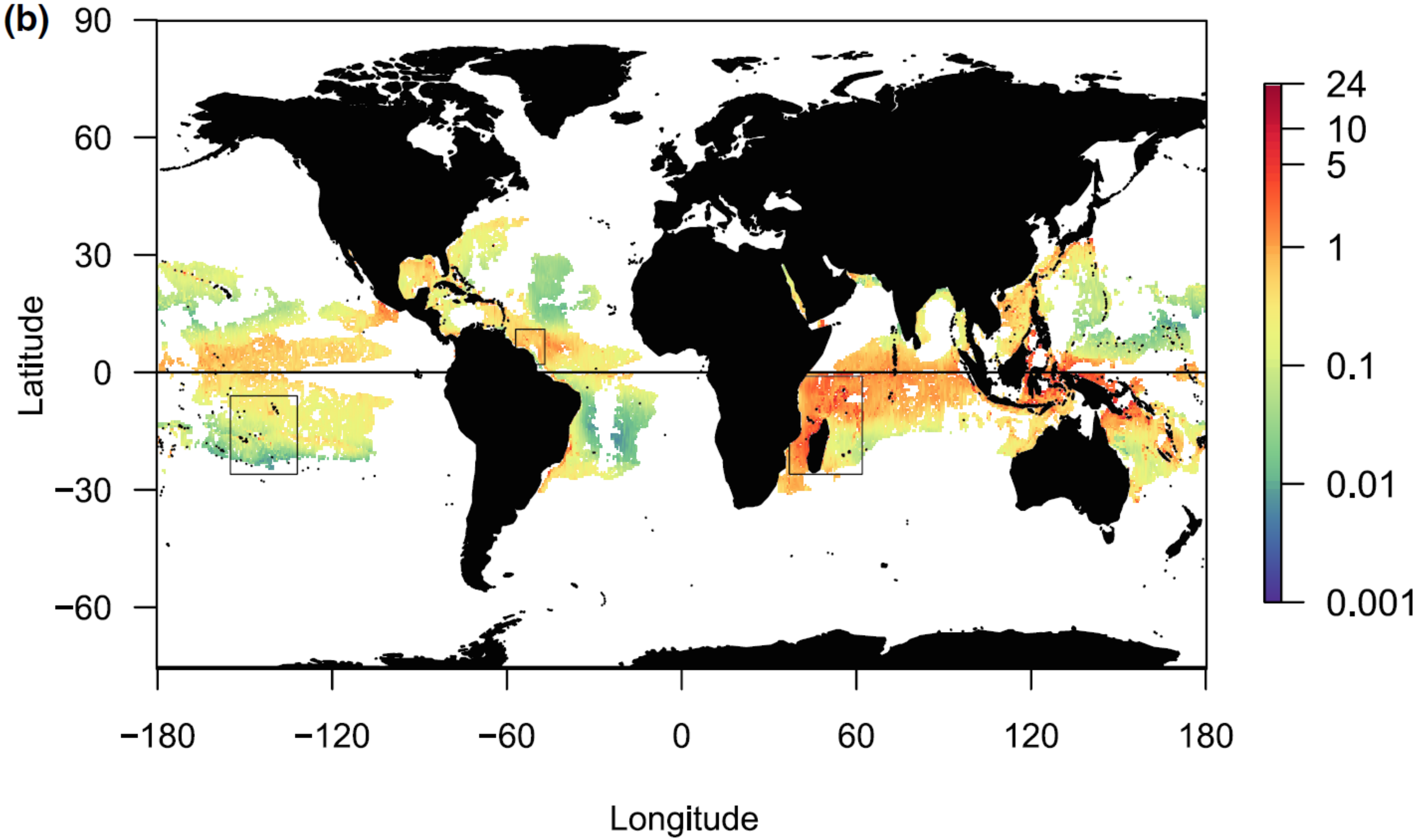
Figure 1 Locations of the three tropical surveyed regions: the central South Pacific, French Guiana and the south-western Indian Ocean. More details on the aerial surveys in each region can be found in Mannocci *et al.* (2013, 2014a,b).

Mannocci et al. (2014) Extrapolating cetacean densities beyond surveyed regions: habitat-based predictions in the circumtropical belt. *J. Biogeogr.* 42: 1267-1280.

Term plots for the Globicephalinae guild model



Mannocci et al. (2014) Extrapolating cetacean densities beyond surveyed regions: habitat-based predictions in the circumtropical belt. *J. Biogeogr.* 42: 1267-1280.



Globicephalinae density extrapolated to cells for which all covariates were within their sampled ranges.

What covariates can you use?

Commonly used:

- Time
- Temporally-varying covariates
- Spatially-varying covariates, a.k.a. static spatial covariates
- Spatiotemporally-varying covariates, a.k.a. dynamic spatial covariates

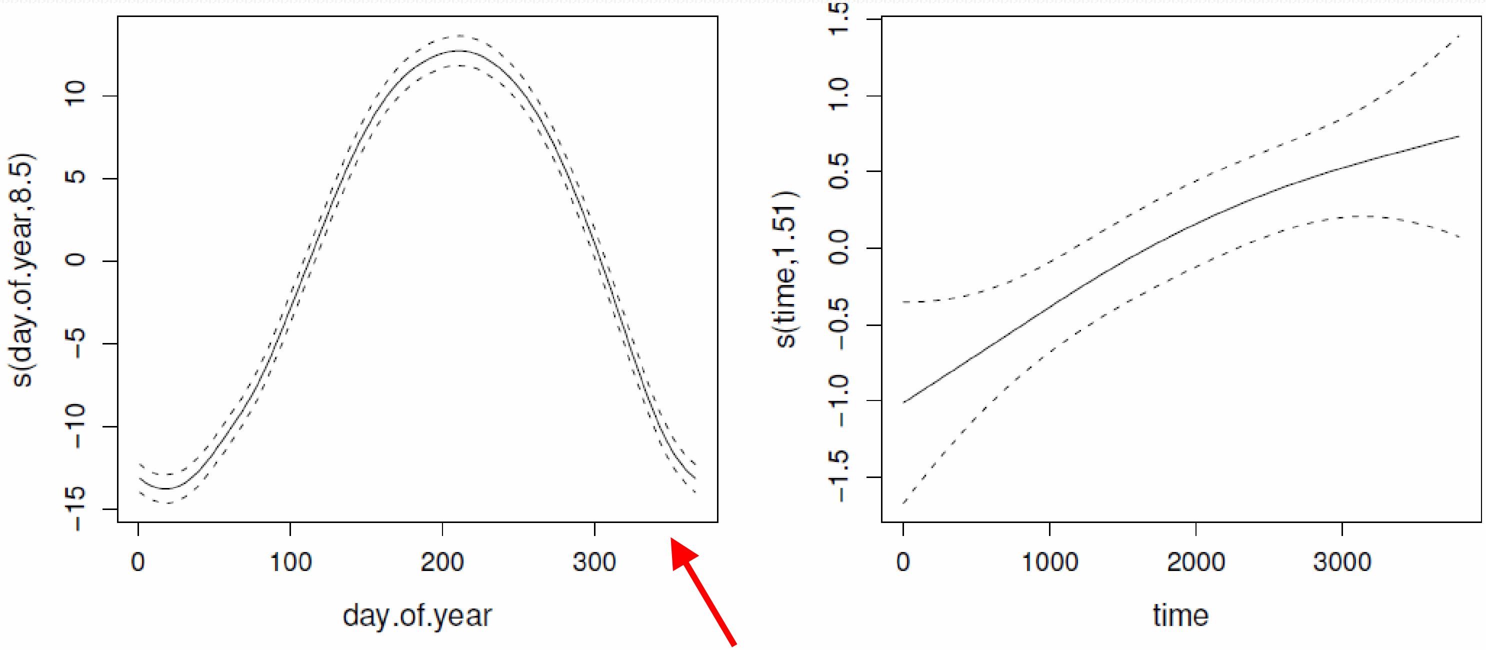
Not so common (discuss in later sessions, if interested):

- 2D smooths of environmental covariates (“interactions”)
- 3D smooth of x, y, time

Time, the usual ways

Inter-annual effect

Intra-annual effect



For year round data, consider a cubic cyclic spline

Figure 6.14 *GAMM terms for daily air temperature in Cairo, Egypt from January 1st 1995. The left panel is the estimated annual cycle: note that it has a fatter peak and thinner trough than a sinusoid. The right pattern is the estimated long term trend: there appears to have been a rise of around 1.5 F over the period of the data.*

Wood (2006)

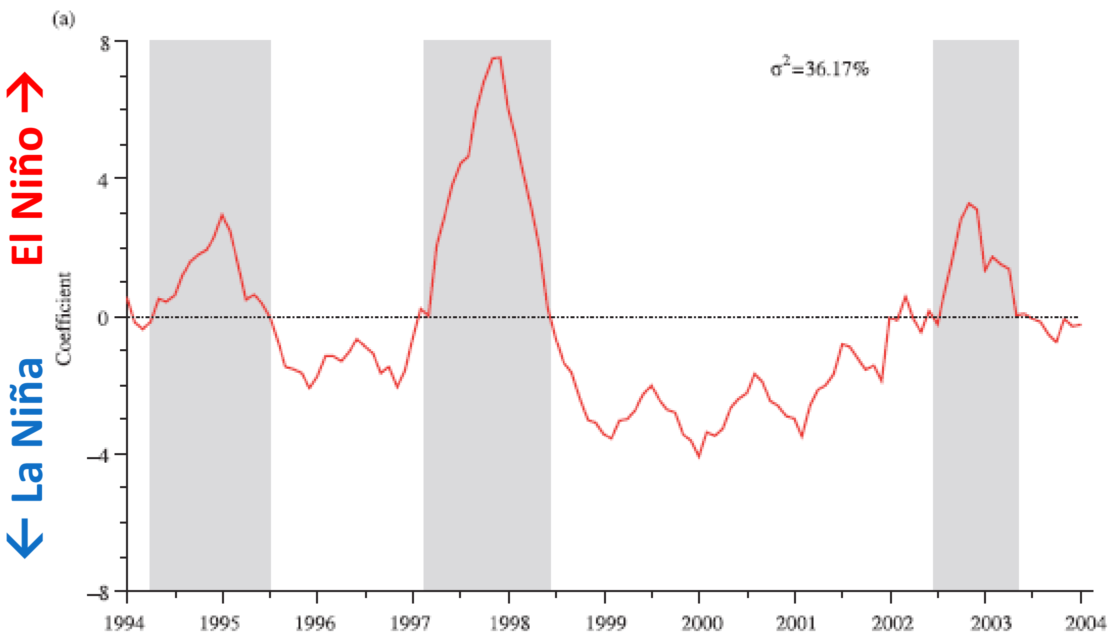
What about?

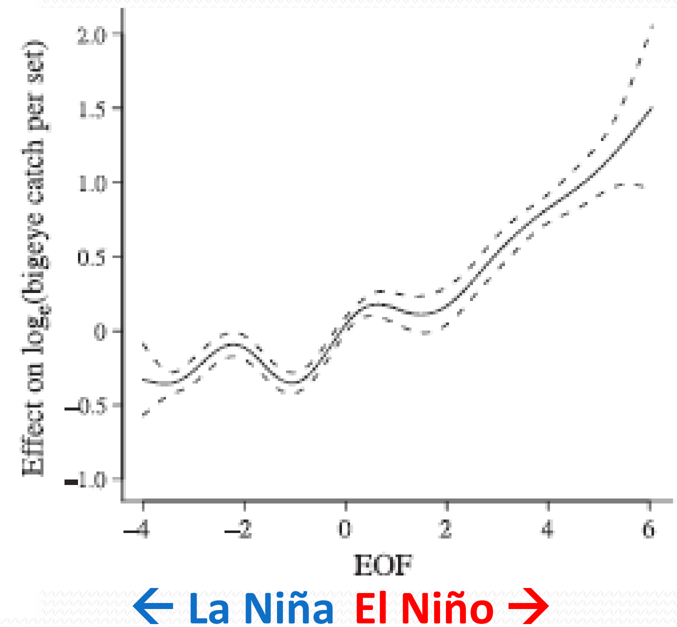
- What’s better: month (1 to 12) or day of year (1 to 365)?
  - Probably day of year. Why discard information?
- What’s better: year as an integer (e.g. 2002) or a higher resolution representation of time (e.g. previous slide)?
  - Probably the higher resolution representation
- Should I use time of day as a covariate?
  - Probably not in a density surface model. Generally we are trying to estimate abundance of a population, which we do not expect to vary diurnally.

Temporally-varying covariates

- Not common in marine models, in my experience

Figure 6. (a) The PC of the first mode of the EOF analysis for the equatorial Pacific, which explained 36.17% of the total variance of the system. The gray-shaded areas represent El Niño events. (b) The spatial pattern for the EOF analysis. The black square represents the boundaries of the region of this study (158–167°W, 4–9°N).

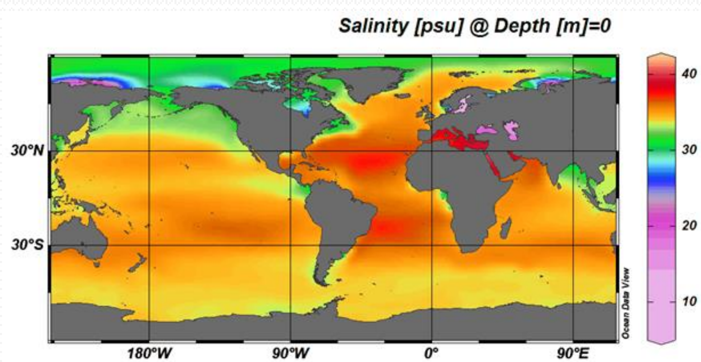


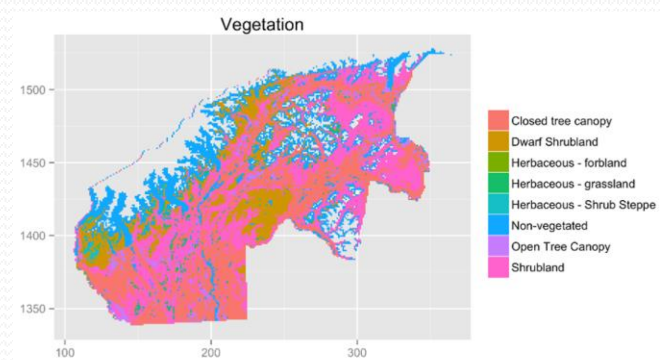


Howell EA, Kobayashi DR (2014) El Niño effects in the Palmyra Atoll region: oceanographic changes and bigeye tuna (*Thunnus obesus*) catch rate variability. *Fish. Oceanogr.* 15(6): 477-489.

Spatially-varying covariates

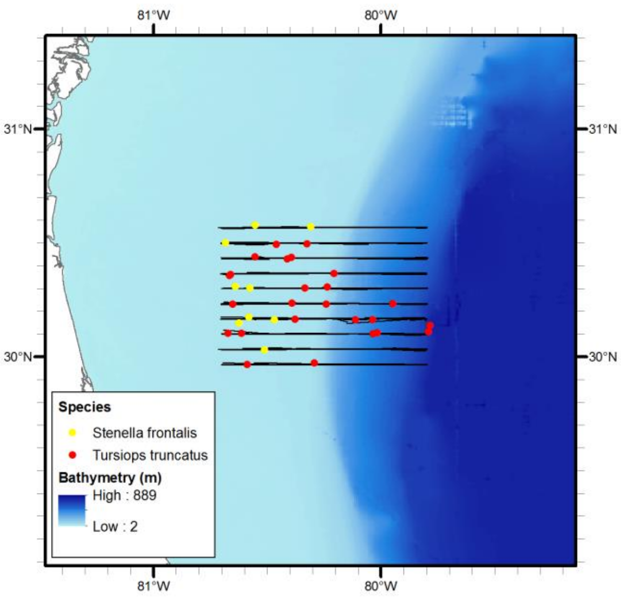
- Static maps of something, e.g.:
  - Elevation, bathymetry, and derivatives: slope, aspect, etc.
  - Cover type, soil type, seafloor type, and other classifications
  - Cumulative climatologies of dynamic processes, e.g. mean annual rainfall, mean primary production
- Generally easy to work with: exact values for your segments from a single image, fit your model, predict over that image





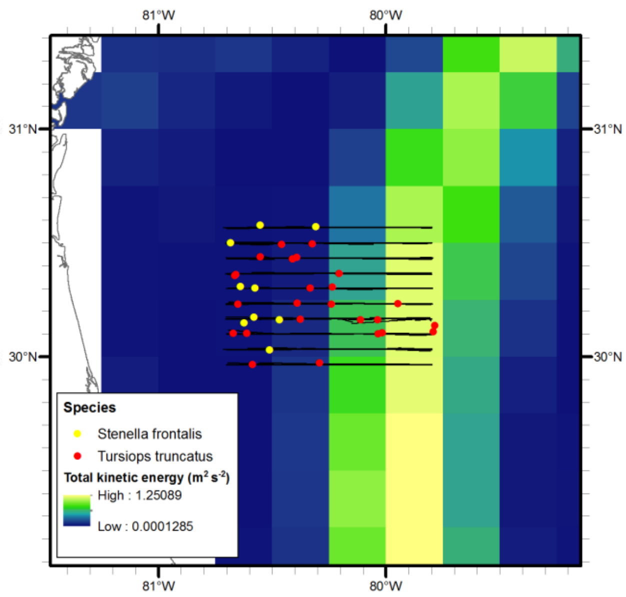
Spatial resolution can be a problem

Bathymetry (1/120°)



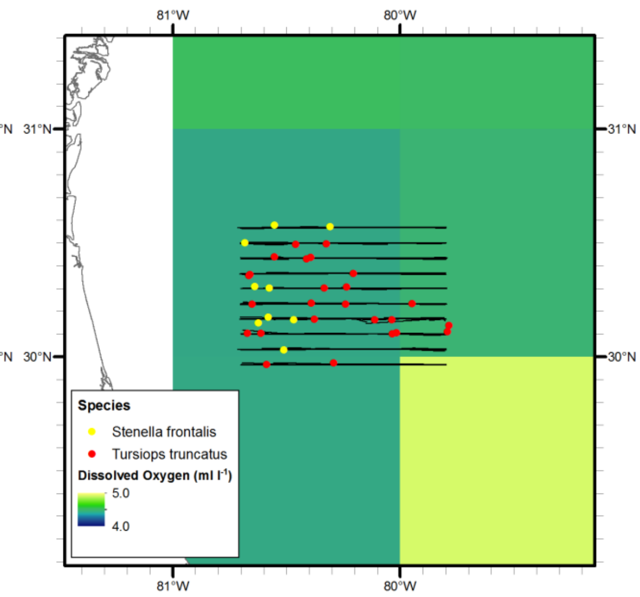
GOOD: survey extent spans many pixels.

Total kinetic energy (1/4°)



POOR: survey extent spans only four pixels.

Dissolved oxygen (1°)



BAD: survey extent spans one pixel. Can this covariate provide much useful information?

What if covariates have different spatial resolutions?

- A common problem in gridded marine data:
  - Regional bathymetry and derivatives (e.g. slope): 5-90 m
  - Global bathymetry and derivatives: 1-2 km
  - Popular remotely sensed sea surface temperature, ocean color, and primary productivity products: 4-9 km
  - Sea ice products: generally 6.25-25 km
  - Sea surface winds: 12.5-25 km
  - Sea surface height and derivatives (e.g. currents): 25 km
  - Salinity, chemistry, zooplankton, climate models: 1-5°

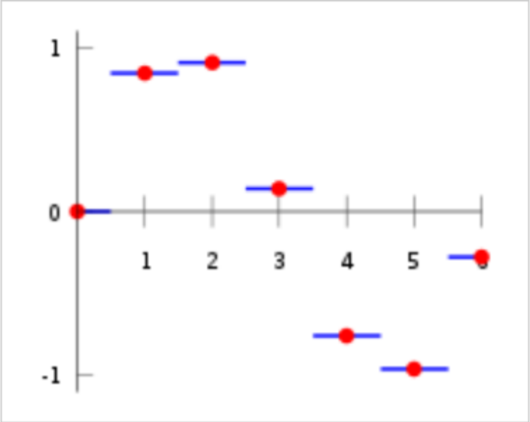
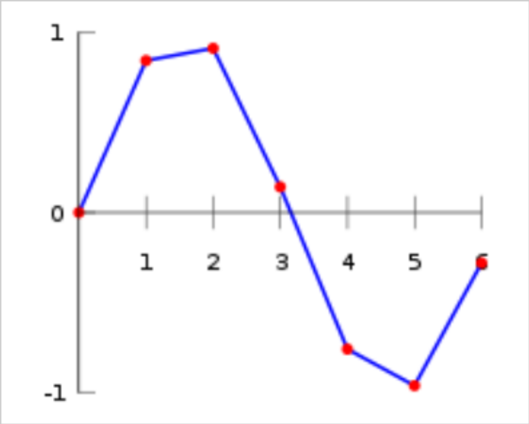
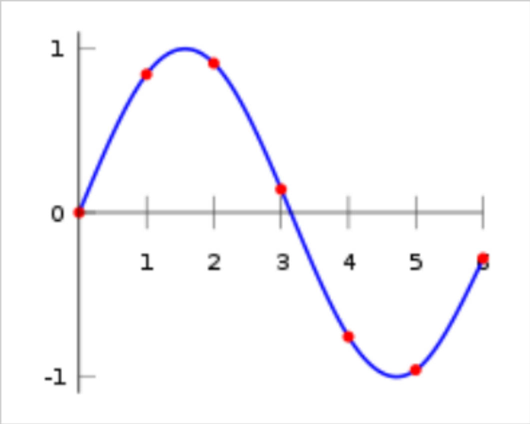
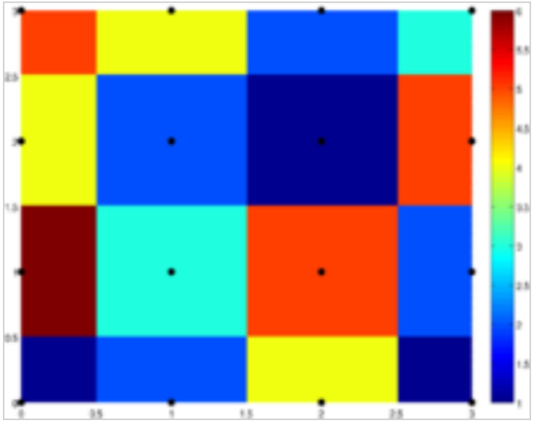
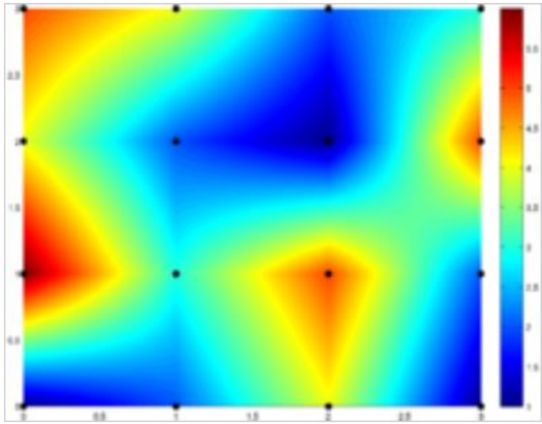
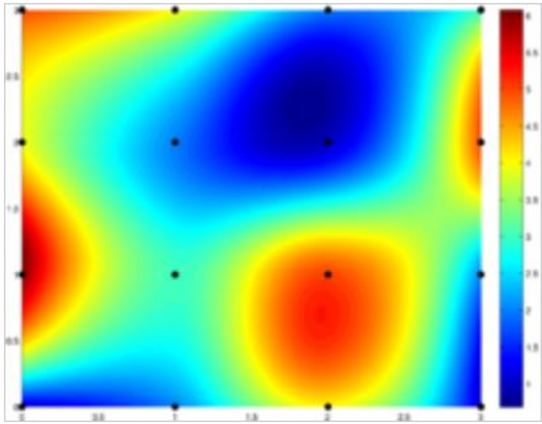
Resolution mismatch shows up twice

1. When you are sampling (a.k.a. interpolating) values of the covariates at your points
2. At prediction time, when it is necessary to obtain values of all covariates on grids that have the same extent, coordinate system, and cell size (and thus rows and columns)
   - This requires you to reproject the covariate images to your common “template” grid you’ll use for predictions
   - It may be desirable to have the cell size of this grid roughly match the effective area of your survey segments

Common approaches to this problem

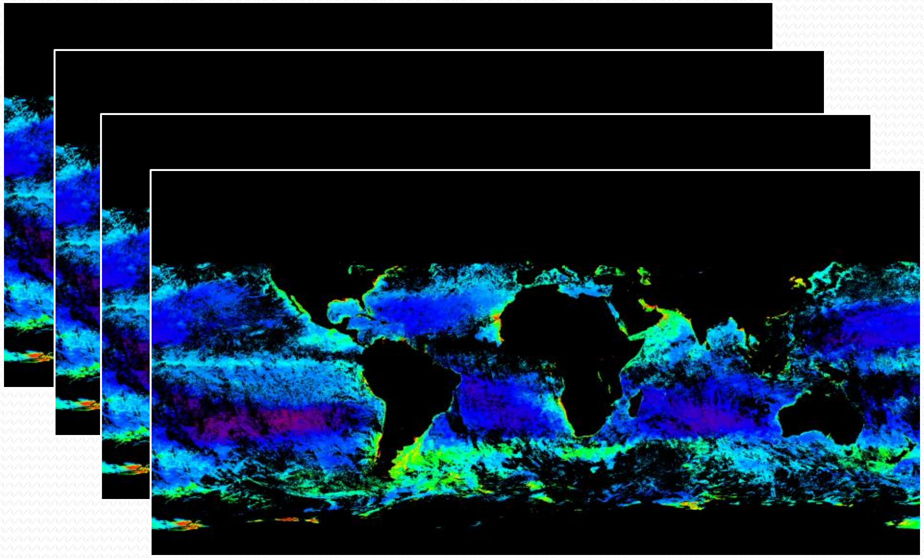
1. Rescale all of your covariates to the lowest resolution covariate (e.g. using a focal or block statistic in ArcGIS)
2. Leave them at original resolutions, and then:
   a. Sample / project them with the nearest neighbor interpolator
   b. Sample / project them with another interpolator, such as linear or cubic spline

The usual suspects

	Nearest Neighbor	Linear	Cubic Spline
1 Dimension			
2 Dimensions			

Spatiotemporally-varying covariates

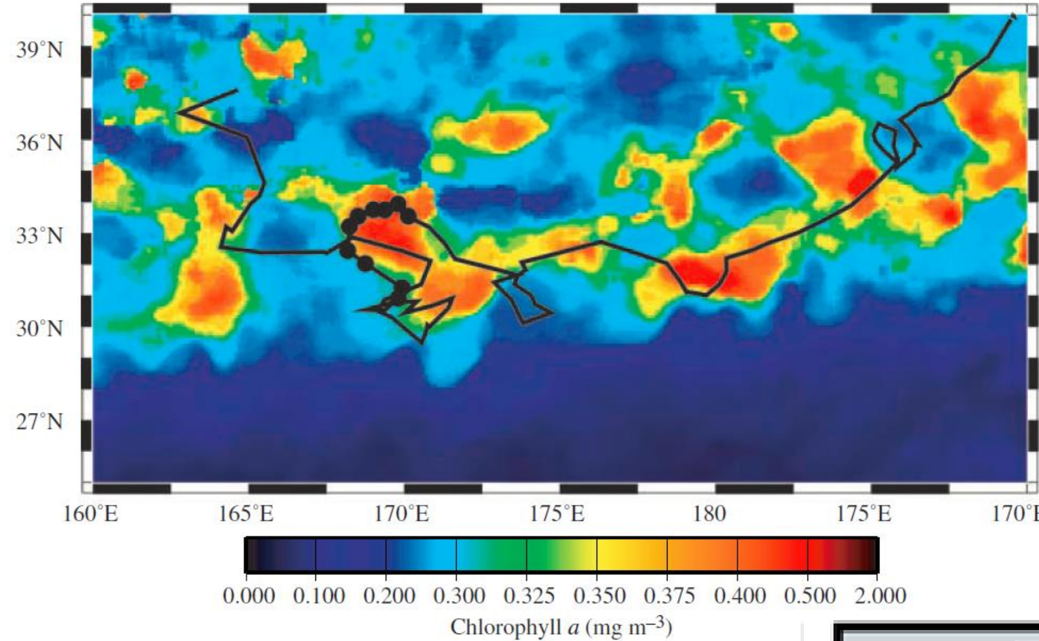
- Typically distributed as a time series of images
- Used very commonly in marine models
- Can be very complicated… let's look at some of the issues...



(Hint: the forecast
is mostly cloudy…)

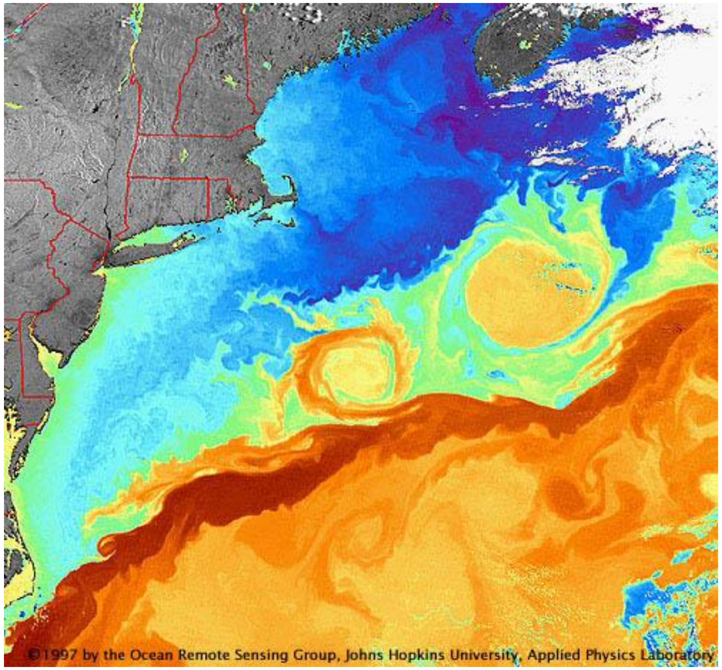
Dynamic spatial covariates: how hard can it be?

In the beginning, you saw this:



Polovina et al. 2004

and this:



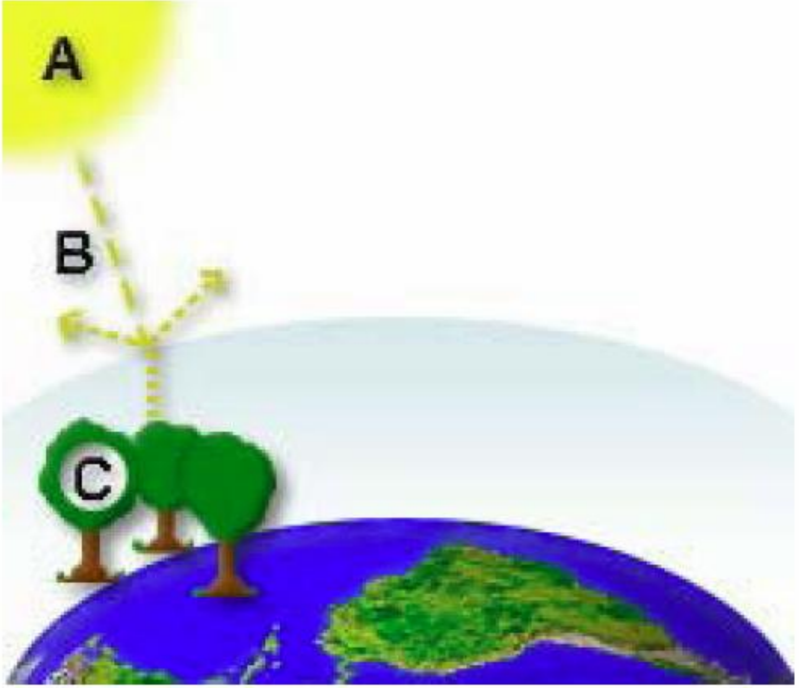
and thought: Wow!

Basic idea of remote sensing

A.Energy Source – An energy source generates electromagnetic radiation (EMR) that illuminates objects it encounters.

B.Radiation and the Atmosphere – As the EMR encounters the atmosphere, only a fraction of it passes through to the ground.

C.Radiation and the Surface – EMR is absorbed, transmitted, or reflected by objects on the Earth's surface.



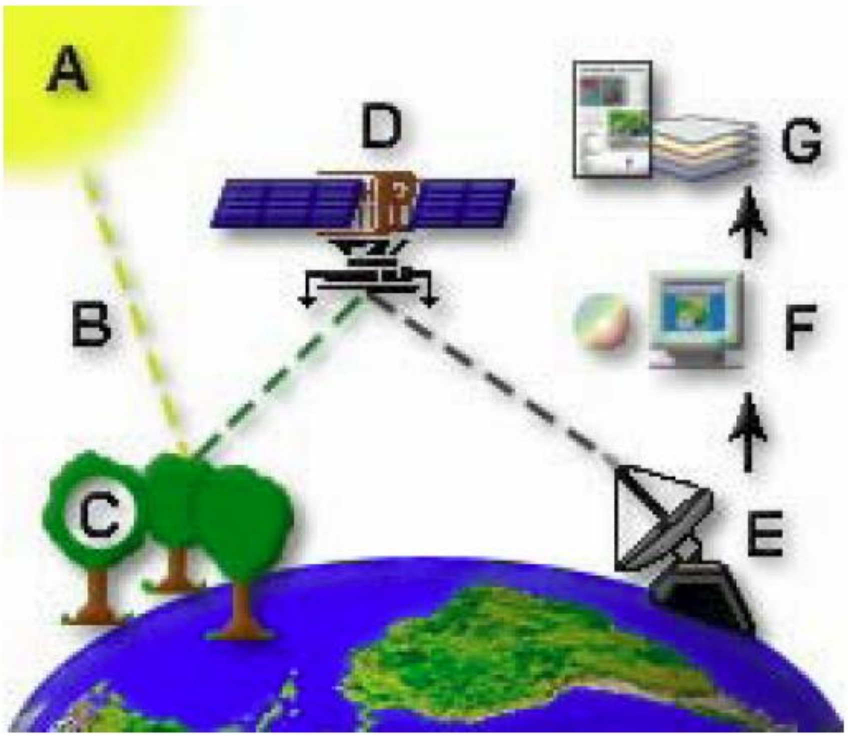
Basic idea of remote sensing

D. Sensor records Radiation – EMR that is reflected is then recorded by a sensor (via a satellite or other platform).

E. Transmitting Sensor Data – EMR data from the sensor is then transferred to a receiving center where it is transformed into an image.

F. Data Analysis – The data is analyzed and pertinent information is extracted.

G. Remote Sensing Application – The data is used to increase understanding about a particular locale or issue.



There are many sources of radiation

a. **Surface (skin) Emissions**

b. **Upwelling Radiation from within the Ocean**

c. **Reflected Solar Energy**

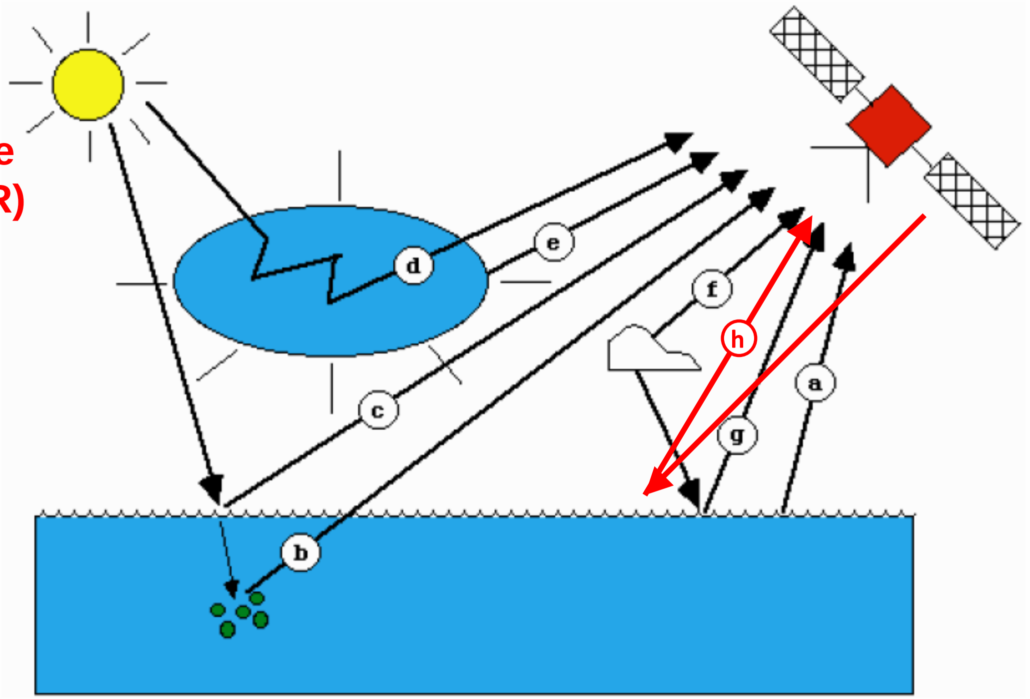
d. **Solar Emissions Scattered into the Satellite's FOV**

e. **Direct Atmospheric Emissions**

f. **Direct Cloud Emissions**

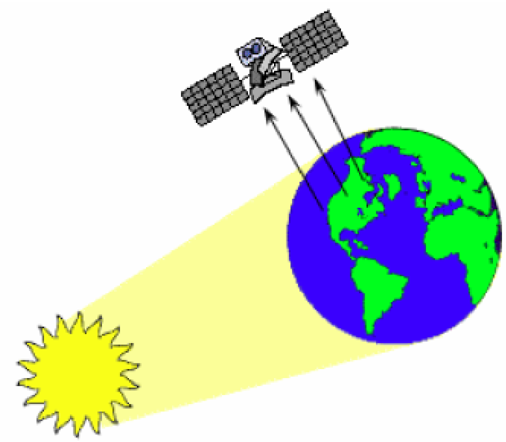
g. **Reflected Clouds Emissions**

h. **Reflected Emissions from the Satellite (e.g. LASER, RADAR)**

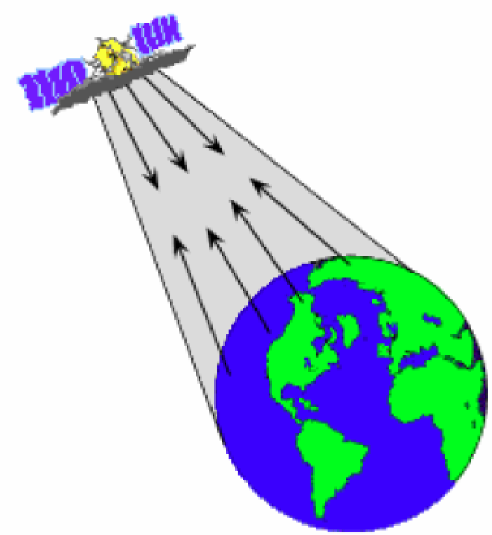


Passive and active sources

Remote Sensing uses electromagnetic energy from both natural and man-made sources.

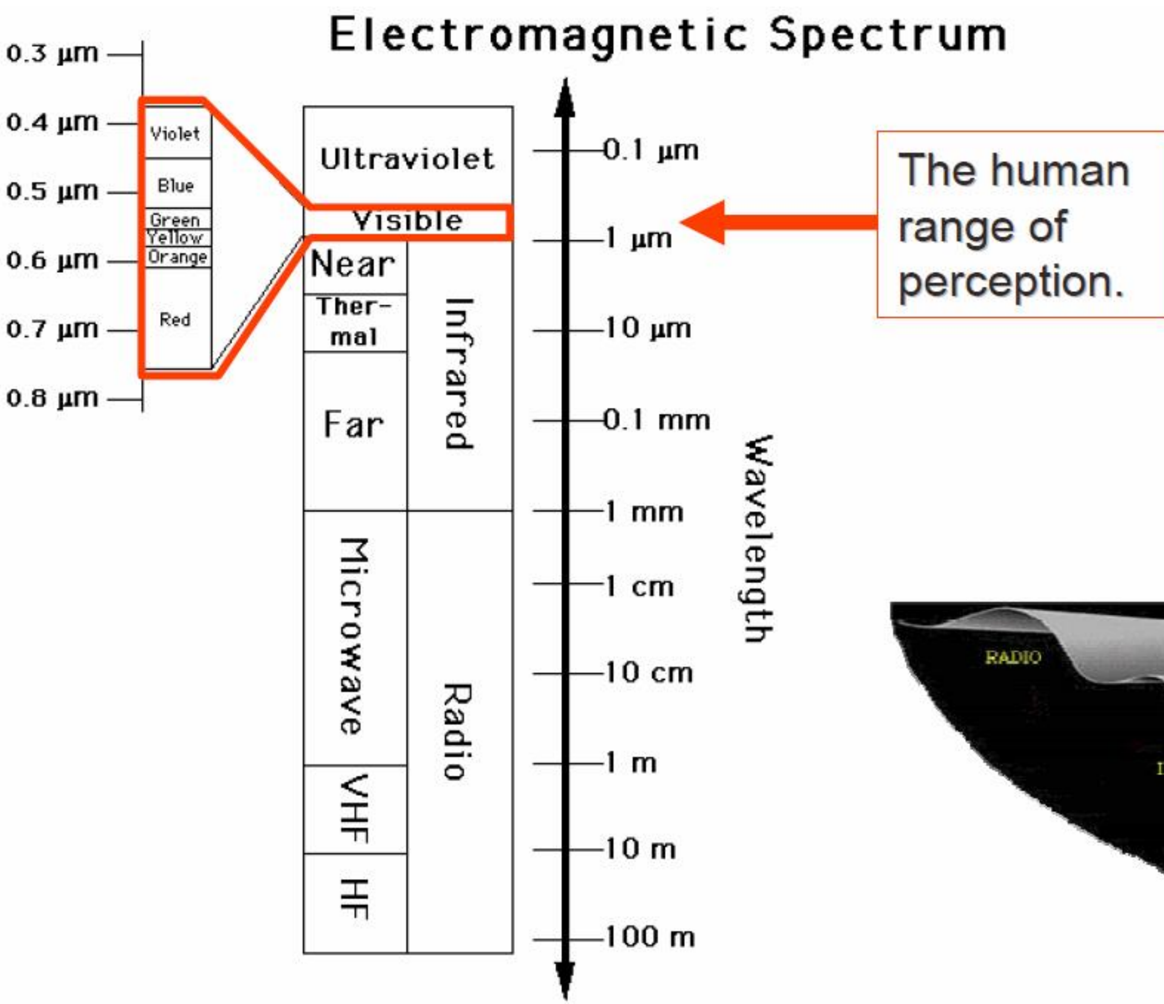


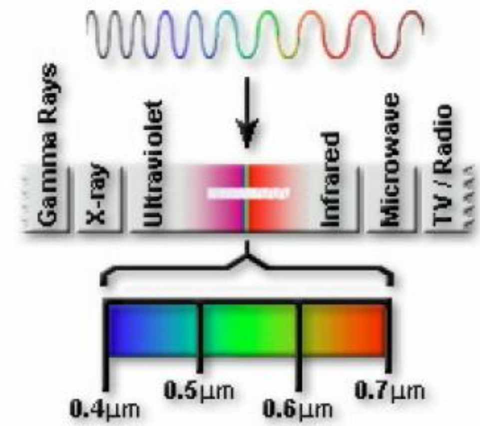
Those energy sources which occur naturally are often referred to as passive energy sources. Solar energy and radiant heat are examples of **passive energy sources.**

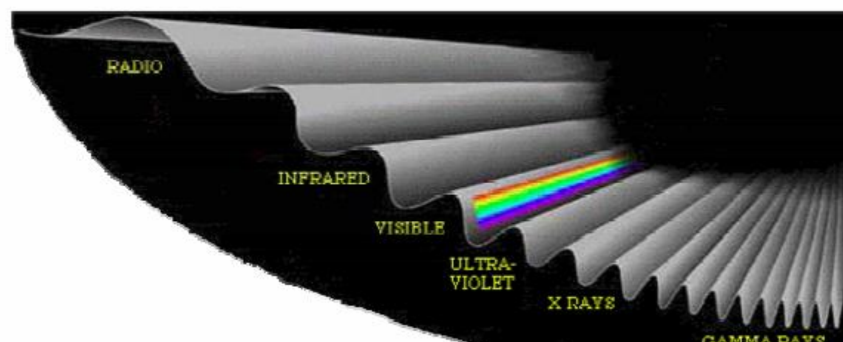


Remote sensing based on electromagnetic energy deriving from man-made sources is usually referred to as active. Radar and laser profilers are examples of **active energy sources**.

Radiation comes in many wavelengths





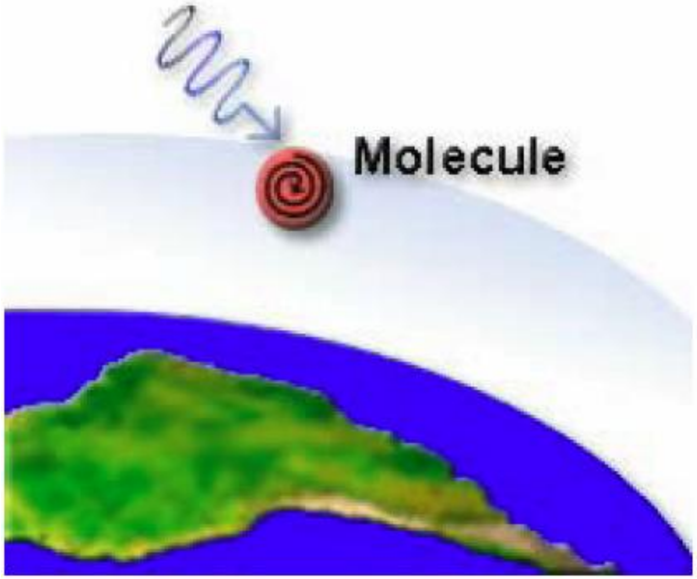


The atmosphere absorbs radiation

Absorption occurs when various wavelengths encounter atmospheric gases that absorb their energy.

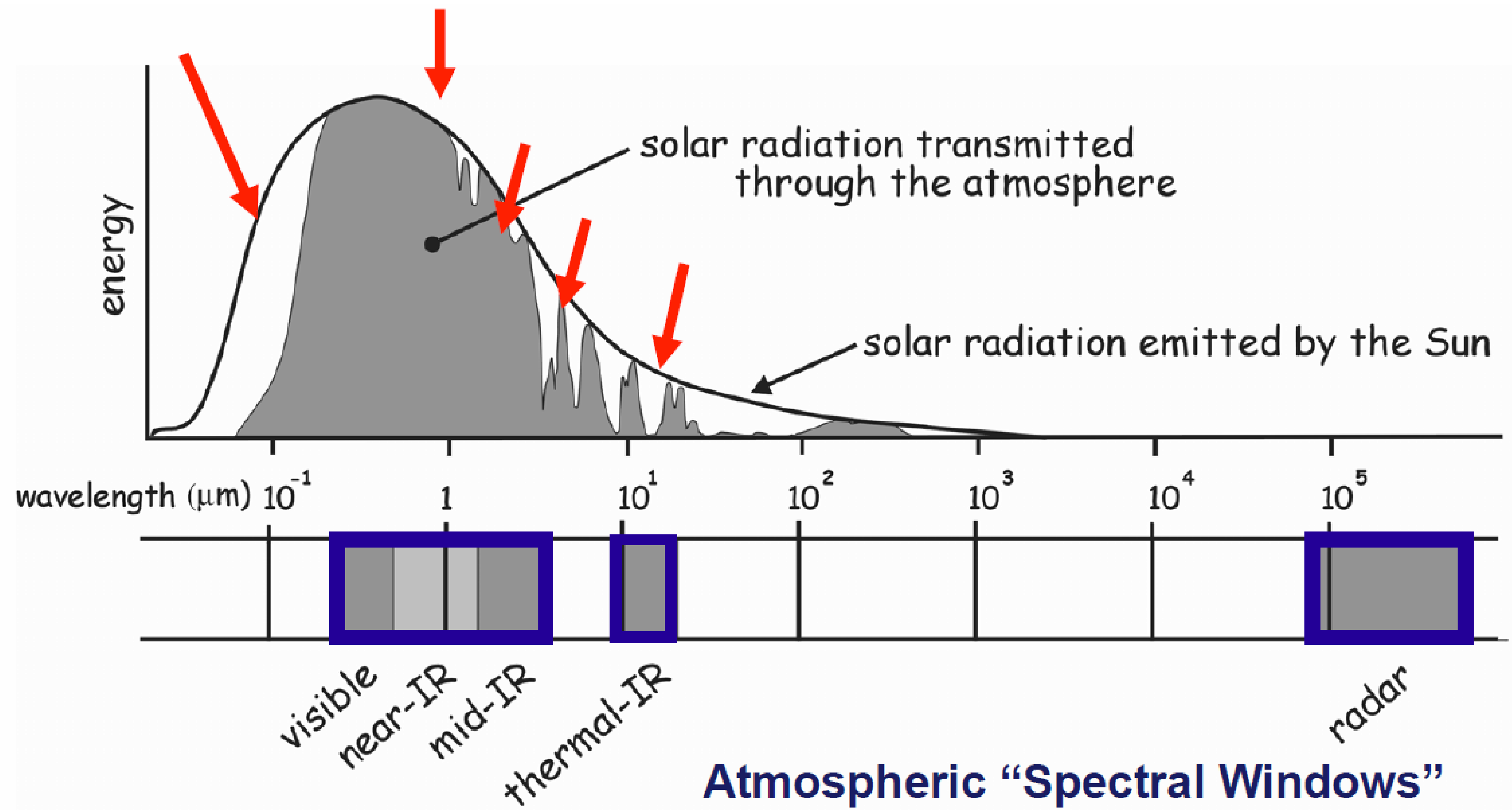
Absorption is mostly caused by:

- Ozone
- Carbon Dioxide
- Water Vapor

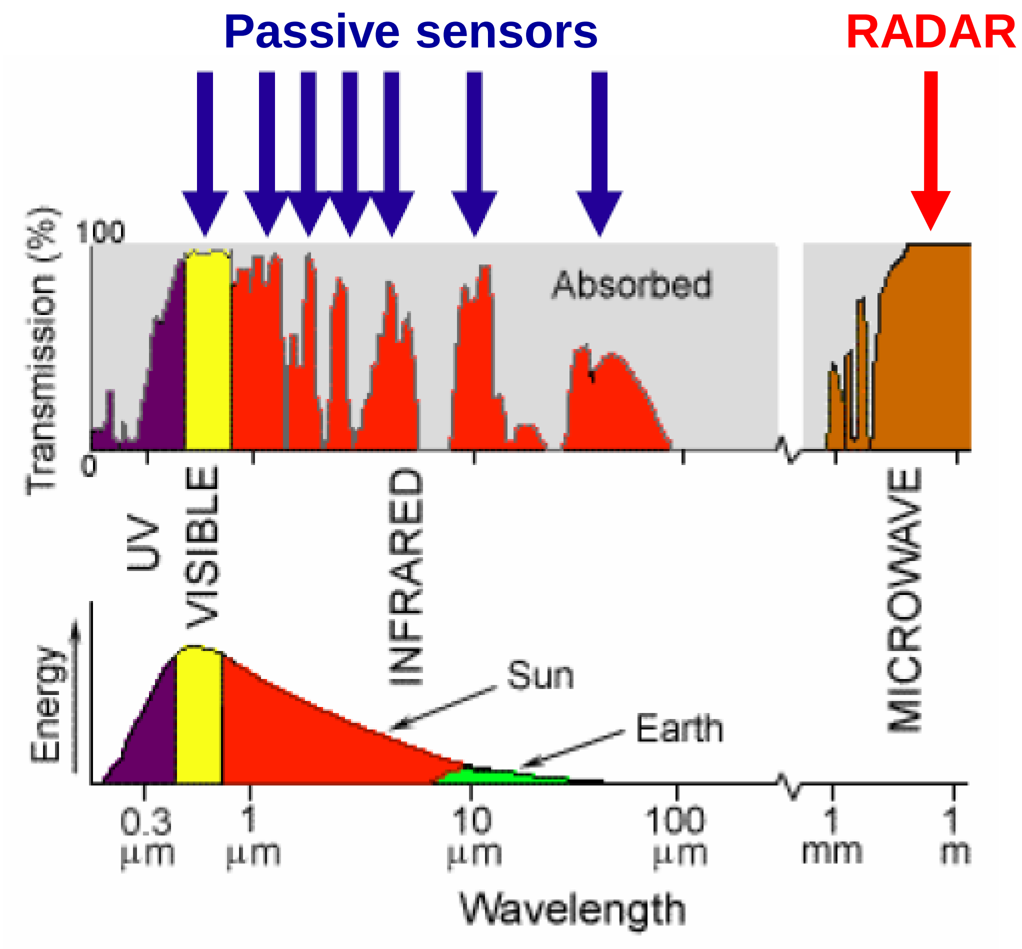


Clouds: a major problem for many sensors!

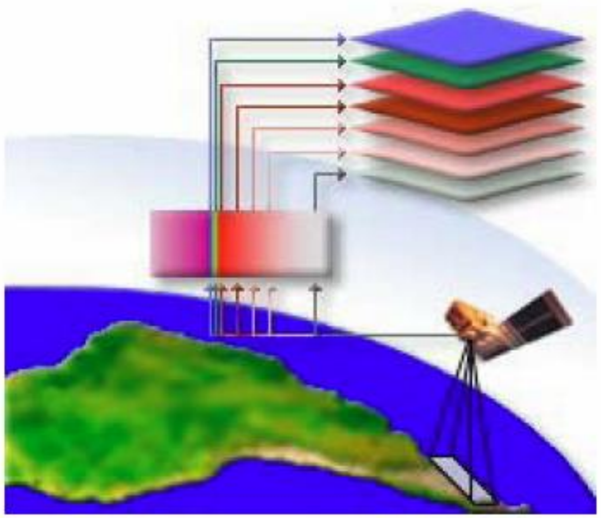
Level of absorption depends on wavelength

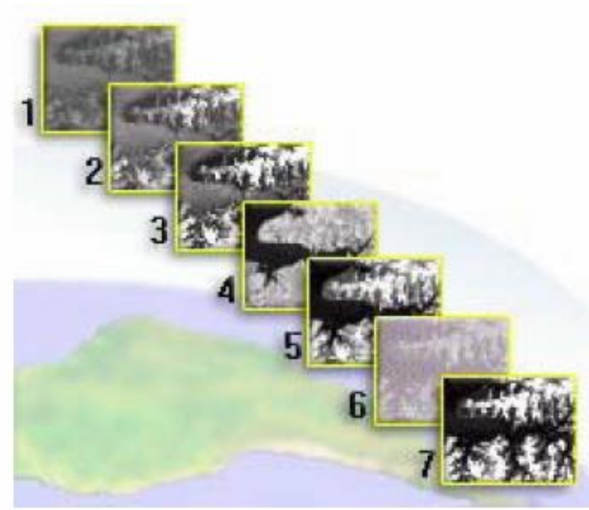


Sensors are designed to exploit this



Example: Landsat-TM: 7 wavelengths





Landsat-TM: 7 band multispectral

3 visible (1,2,3)

3 infrared (4,5,7)

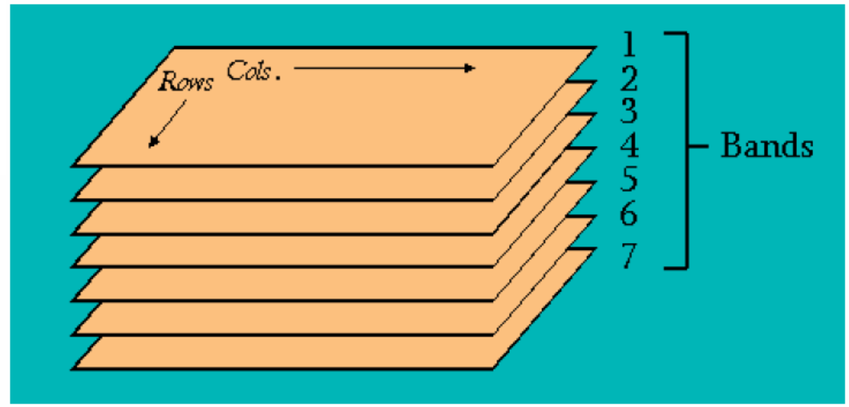
1 thermal (6)

These are called *bands*

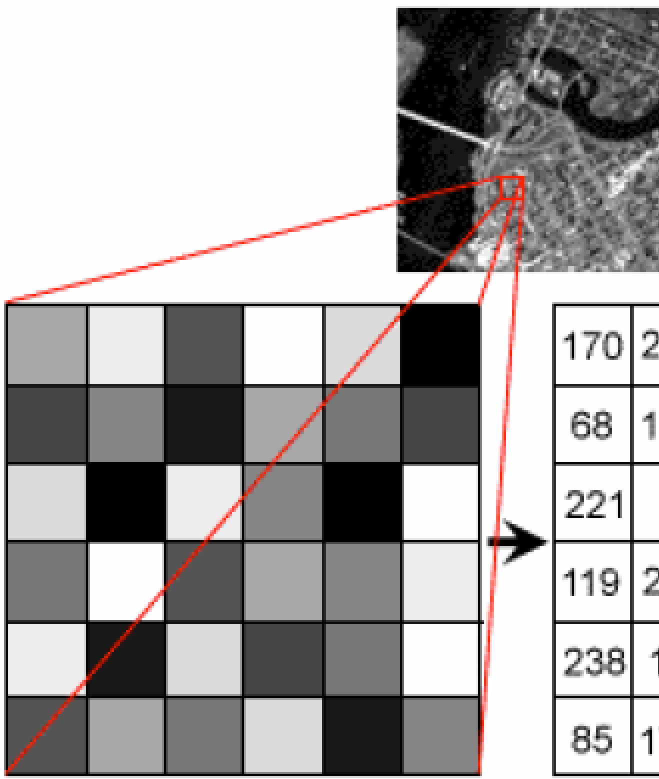
Spectral resolution can be described by the number and operating ranges of bands. For example, the 7 band LandSat-TM Thematic Mapper platform:

- ✓ ***band 1 0.45 - 0.52 microns VIS blue (Visible)***
- ✓ ***band 2 0.52 - 0.60 microns VIS green (Visible)***
- ✓ ***band 3 0.63 - 0.69 microns VIS red (Visible)***
- ✓ ***band 4 0.76 - 0.90 microns NIR (Near InfraRed)***
- ✓ ***band 5 1.55 - 1.75 microns SWIR (Mid InfraRed)***
- ✓ ***band 6 10.40 -12.50 microns TIR (Thermal InfraRed)***
- ✓ ***band 7 2.08 - 2.35 microns SWIR (Mid InfraRed)***

Digital image for each band



One band



170	238	85	255	221	0
68	136	17	170	119	68
221	0	238	136	0	255
119	255	85	170	136	238
238	17	221	68	119	255
85	170	119	221	17	136

Digital Numbers

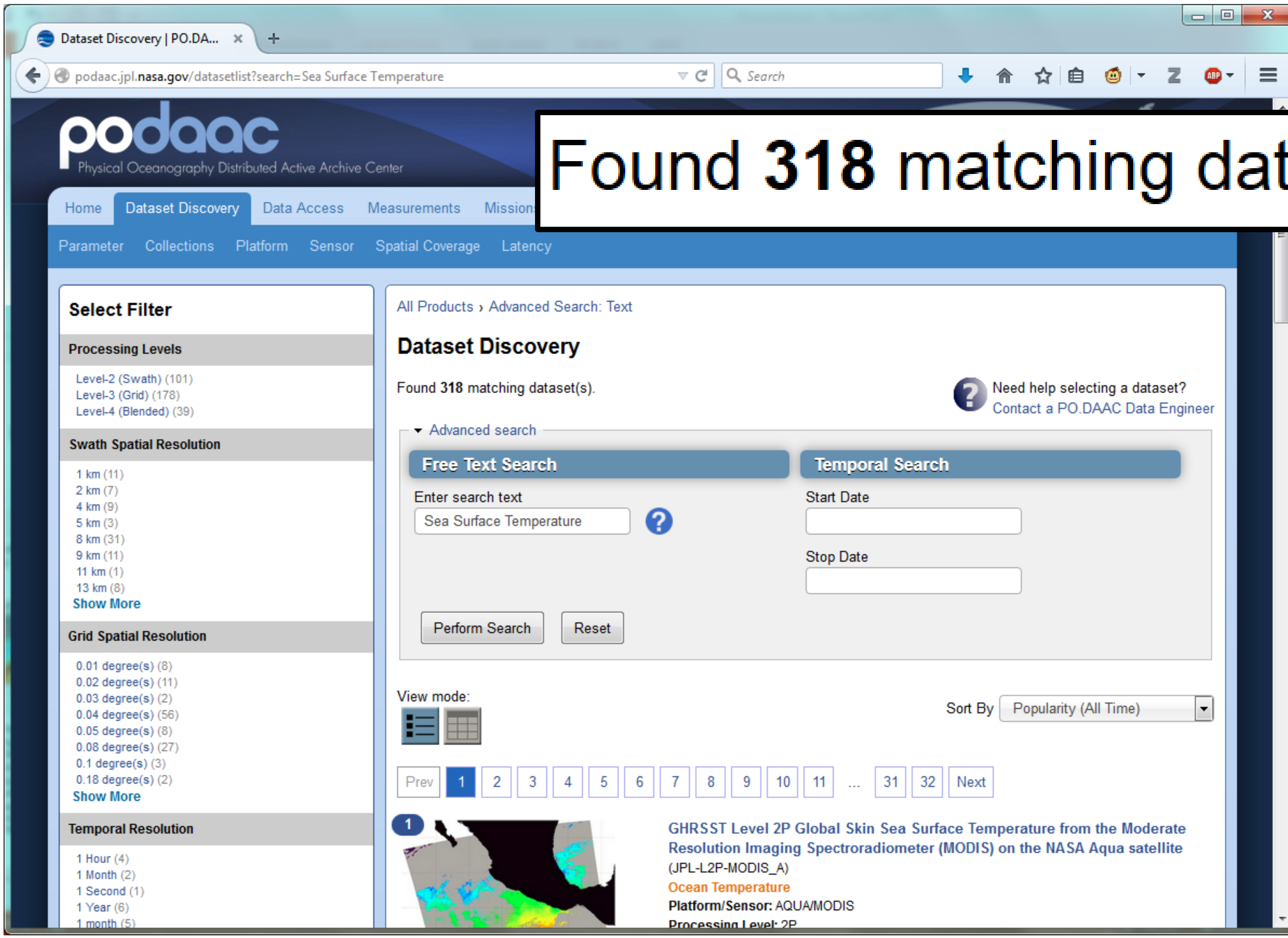
Environmental variables estimated by equations that combine bands

MODIS Band	Bandwidth (μm)	Radiance @ 300 K ($\mathrm{Wm^{-2}\mu m^{-1}sr^{-1}}$)	NEΔT (K)
20	3.660 - 3.840	0.45	0.05
22	3.929 - 3.989	0.67	0.07
23	4.020 - 4.080	0.79	0.07
31	10.780 - 11.280	9.55	0.05
32	11.770 - 12.270	8.94	0.05

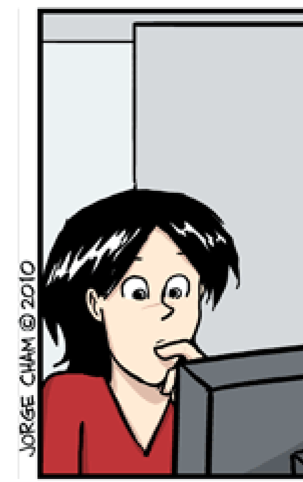
$$sst = k_0 + k_1 T_{31} + k_2 \langle T_{31} - T_{32} \rangle \, T_{env} + k_3 \langle T_{31} - T_{32} \rangle \, (1/\mu - 1)$$

SATELLITE	D/N	ALG.	COEFFICIENTS			
			k_0	k_1	k_2	k_3
Aqua	Day	*sst*	1.152	0.960	0.151	2.021
	Night	*sst*	2.133	0.926	0.125	1.198
	Night	*sst4*	0.987	1.031	0.349	1.766
Terra	Day	*sst*	1.052	0.984	0.130	1.860
	Night	*sst*	1.886	0.938	0.128	1.094
	Night	*sst4*	-0.065	1.034	0.723	0.972

You found NASA PO.DAAC on Google and clicked on Sea Surface Temperature…

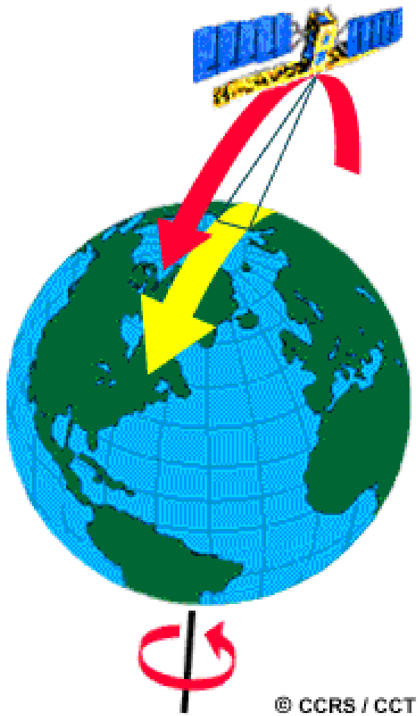


Found **318** matching dataset(s).



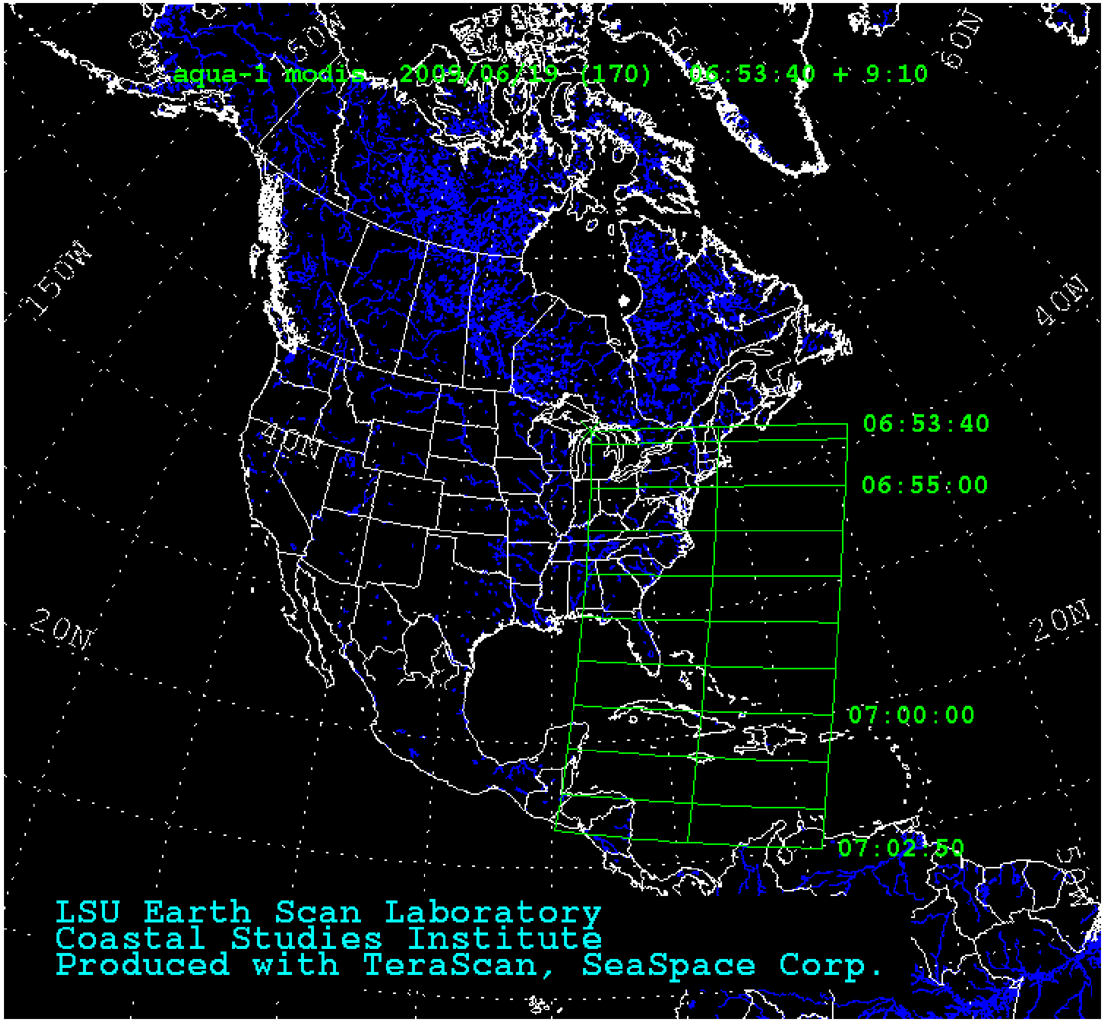
What “level” of data do you want?

Level	Description
0	~~Real time data feed from ground control station.~~
1	~~Files of calibrated, geolocated, at-aperture radiance values for swath segments, with quality flags and error estimates.~~
2	Files of geophysical values (e.g. SST) for swath segments, calculated from the Level 1 data by applying an algorithm to the radiance values. **Highest resolution – why not try it?**
3	Files of uniform grids of geophysical values, for various spatial and temporal scales, produced by accumulating and projecting Level 2 data.
4	Same as Level 3, but with missing data filled in via interpolation, modeling, integration of data from multiple sensors, or other means.

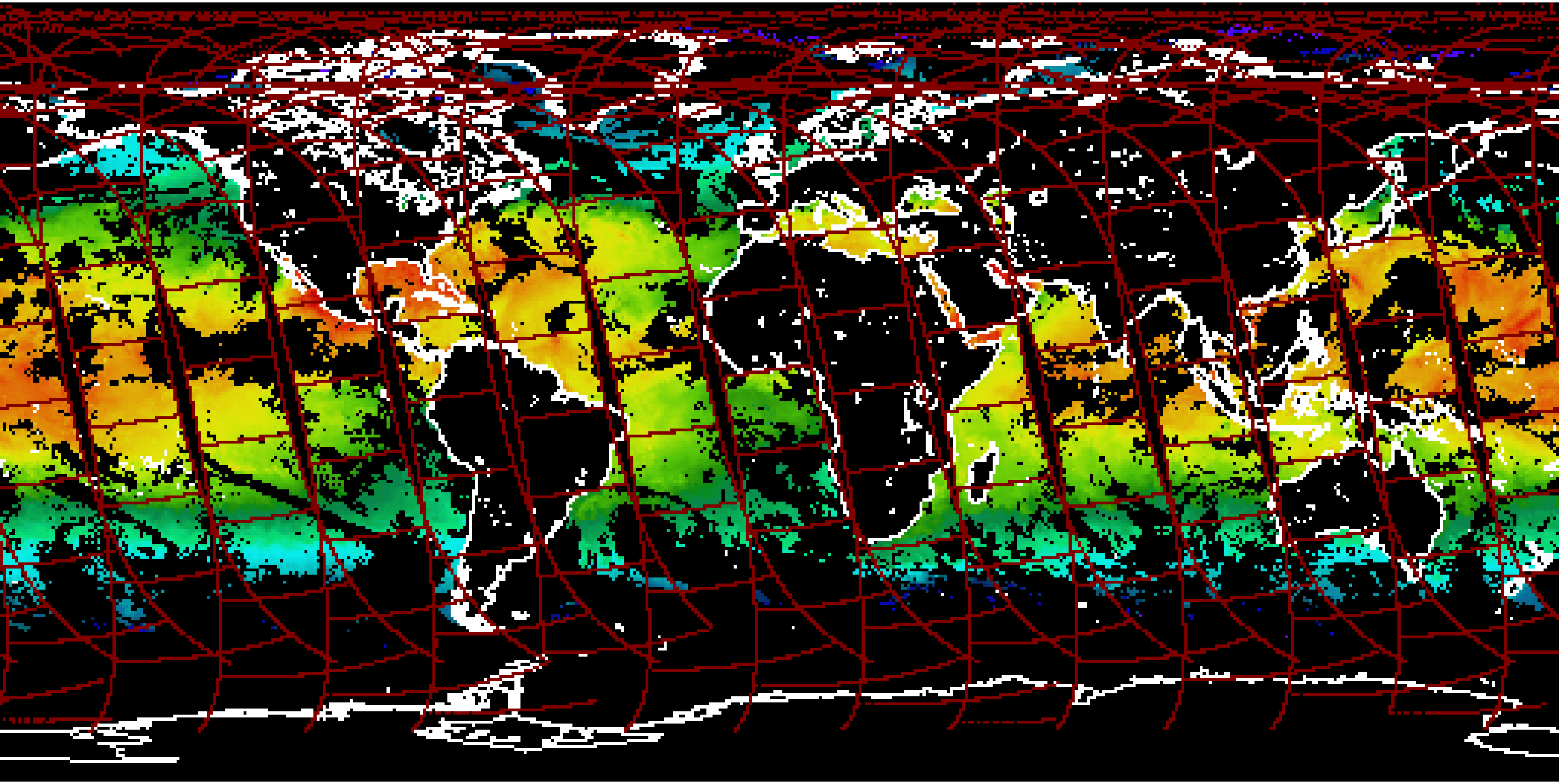


Like strips of photographs

A **swath** is the continuous strip of the planet's surface imaged by the polar orbiting sensor as it circles the plant



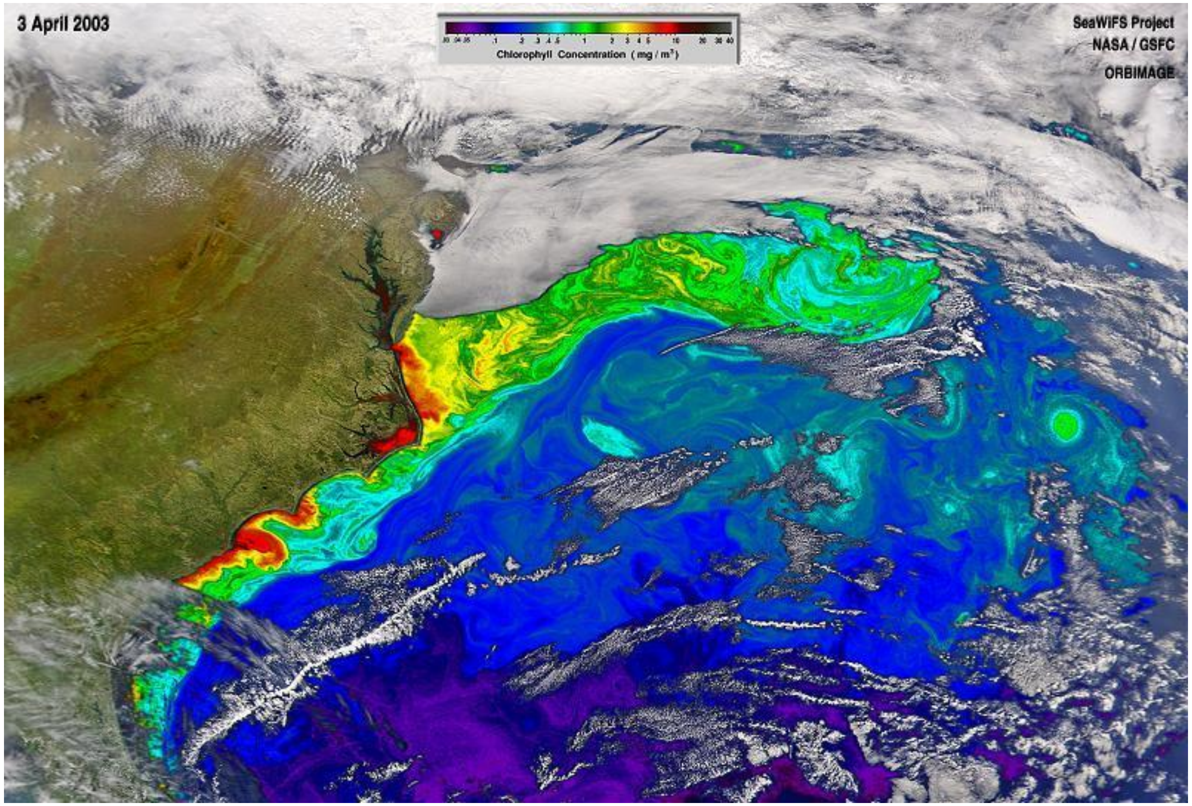
MODIS SST swath granules



Level 2: swath granules

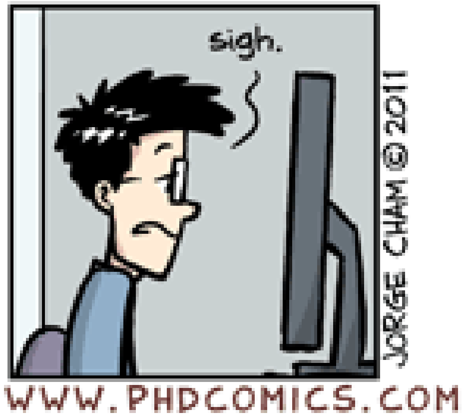
Advantages

- Highest resolution possible
- Multiple passes over locations at high latitudes



SeaWiFS Level 2 image of chlorophyll bloom along Gulf Stream

Level 2: swath granules



Ok, how about Level 3 daily images?

Disadvantages

- Irregular, non-rectangular grid cells
  - Can't represent as raster in ArcGIS; <u>not a projection issue</u>
  - Must treat data as points and interpolate your own grid
- Images overlap
  - A given point at a given date may have multiple images
  - How do you select which one to use?
- Must have a very large hard disk
  - Individual images may be hundreds of megabytes
- More difficult to download

What "level" of data do you want?

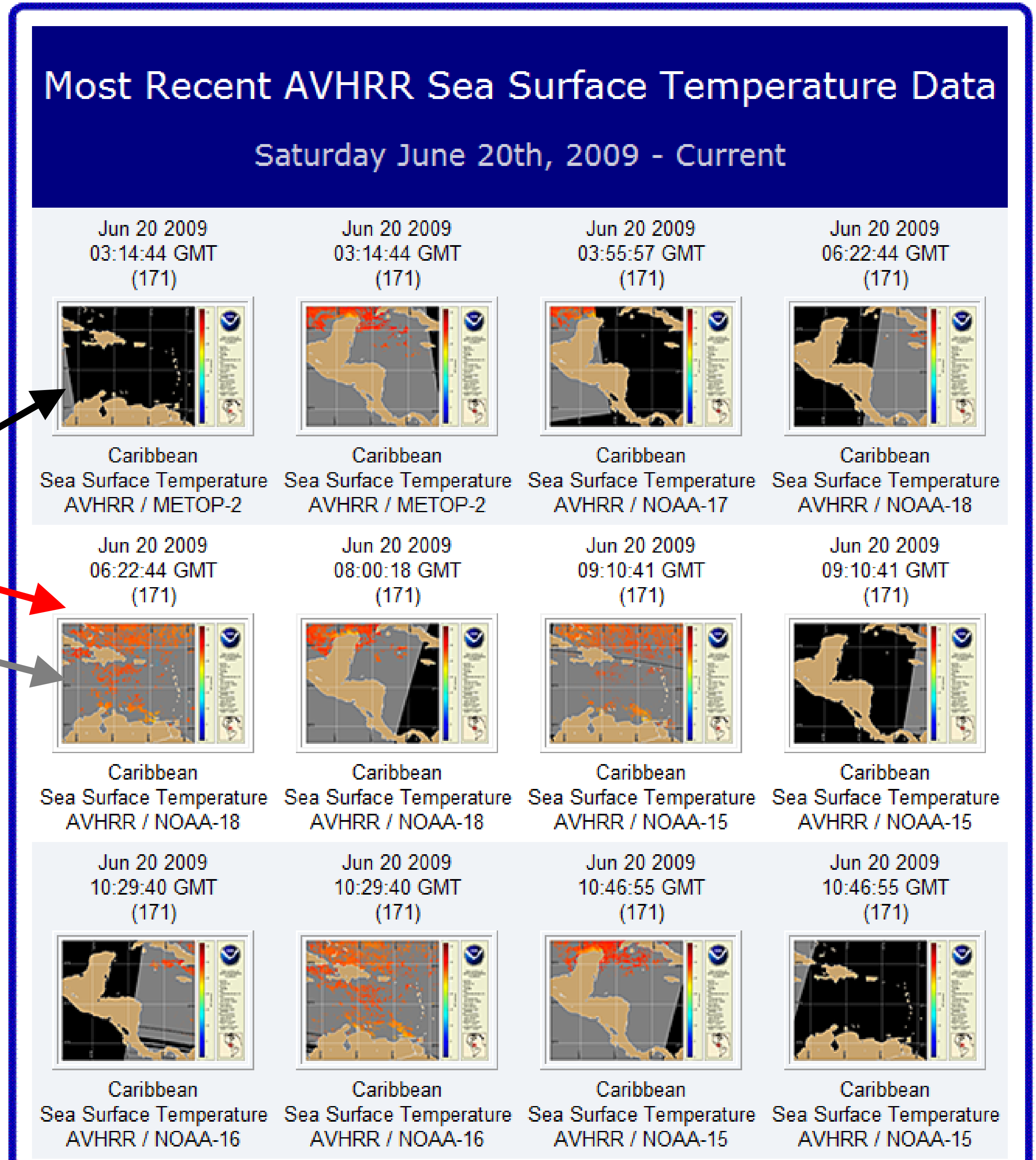
Level	Description
~~0~~	~~Real time data feed from ground control station.~~
~~1~~	~~Files of calibrated, geolocated, at-aperture radiance values for swath segments, with quality flags and error estimates.~~
~~2~~	~~Files of geophysical values (e.g. SST) for swath segments, calculated from the Level 1 data by applying an algorithm to the radiance values.~~
3	Files of uniform grids of geophysical values, for various spatial and temporal scales, produced by accumulating and projecting Level 2 data.
4	Same as Level 3, but with missing data filled in via interpolation, modeling, integration of data from multiple sensors, or other means.

Wow, look at all how much data is missing!

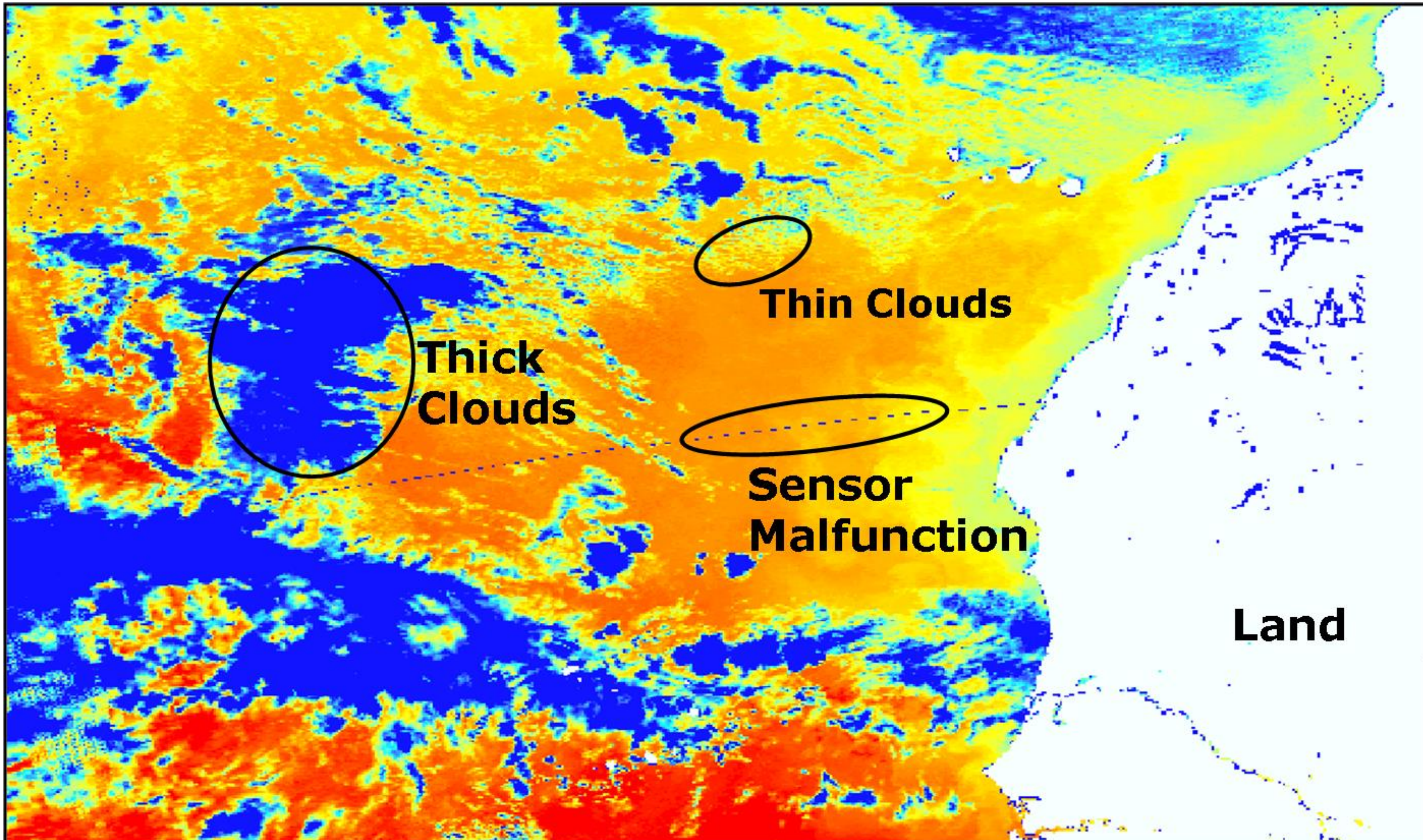
No coverage (black)

SST (orange)

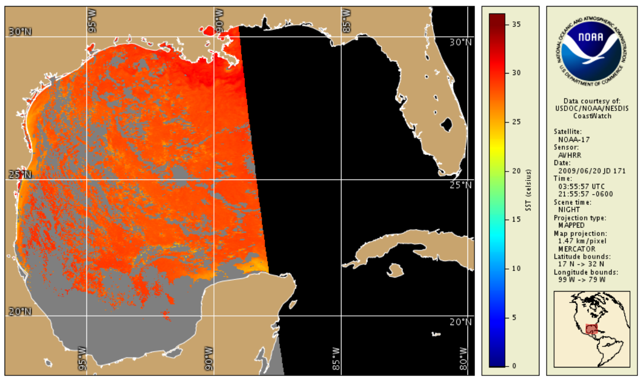
Clouds (gray)



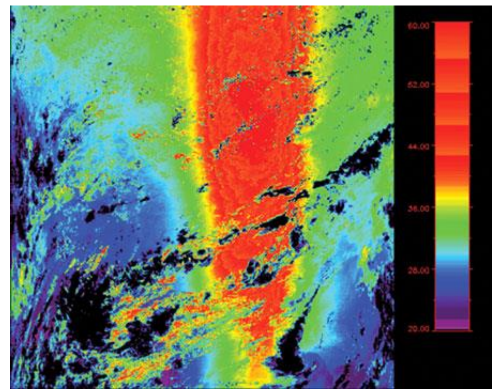
Reasons for missing data



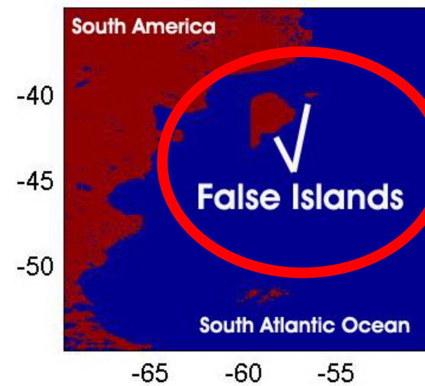
Reasons for missing data



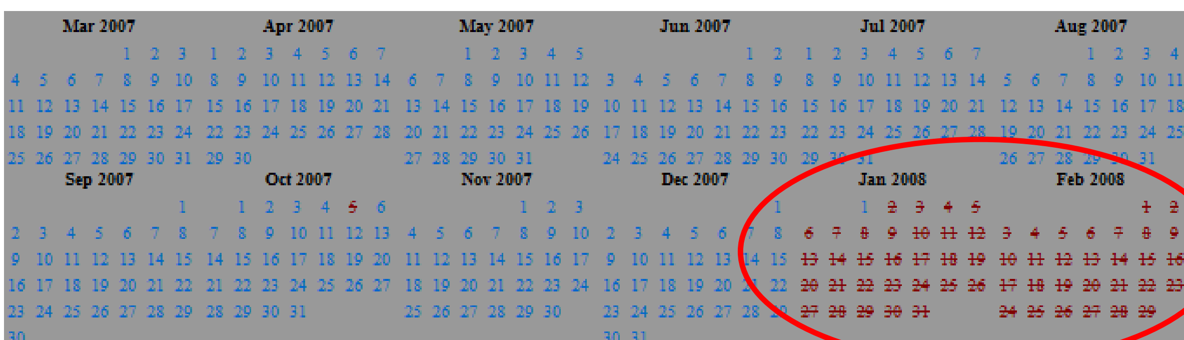
Partial coverage by satellite swath



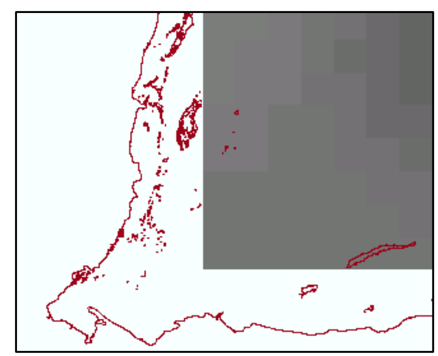
Sun glint



Errors in data provider's land mask

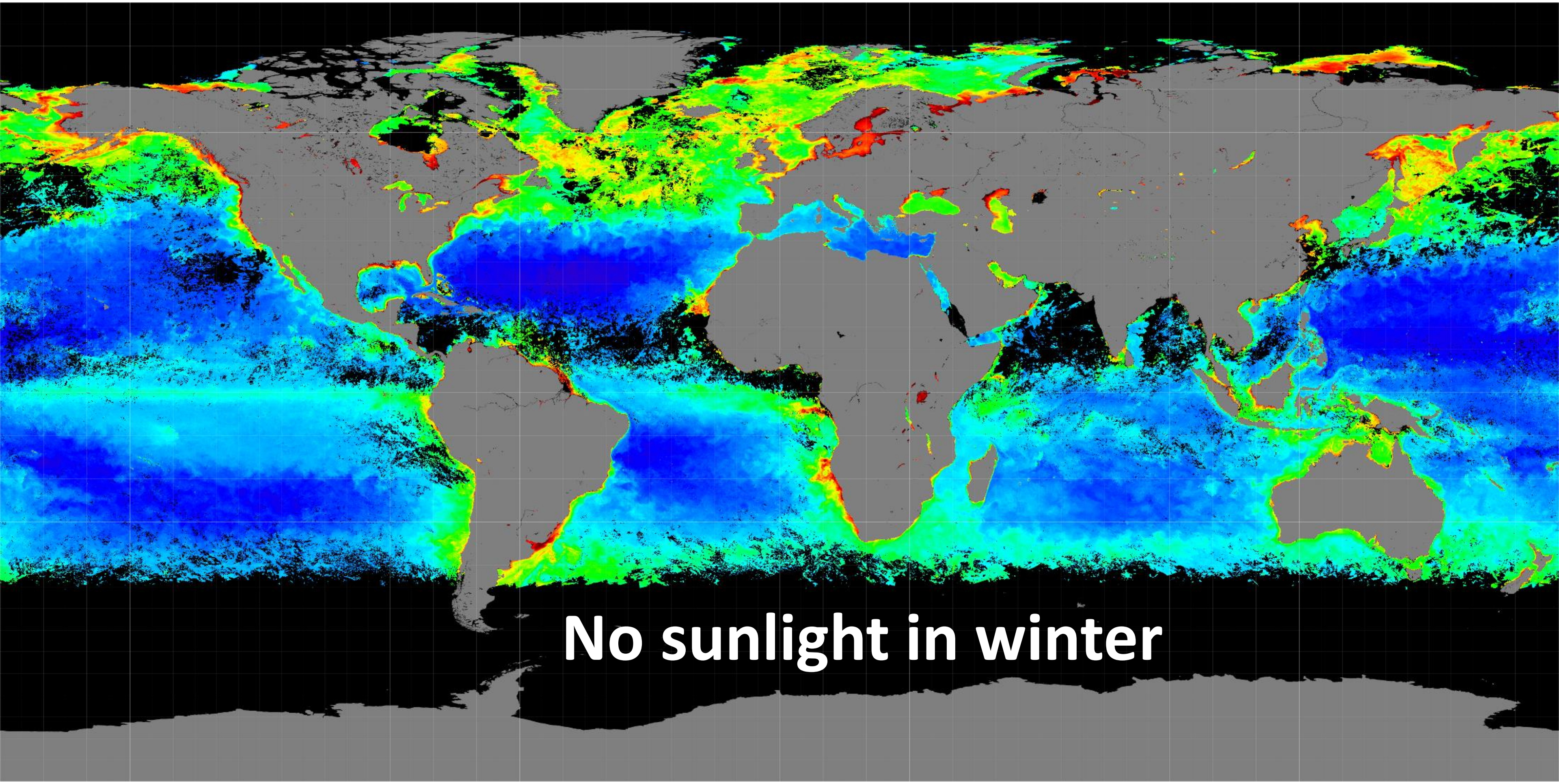


Long duration systematic satellite failure (e.g. SeaWiFS early 2008)



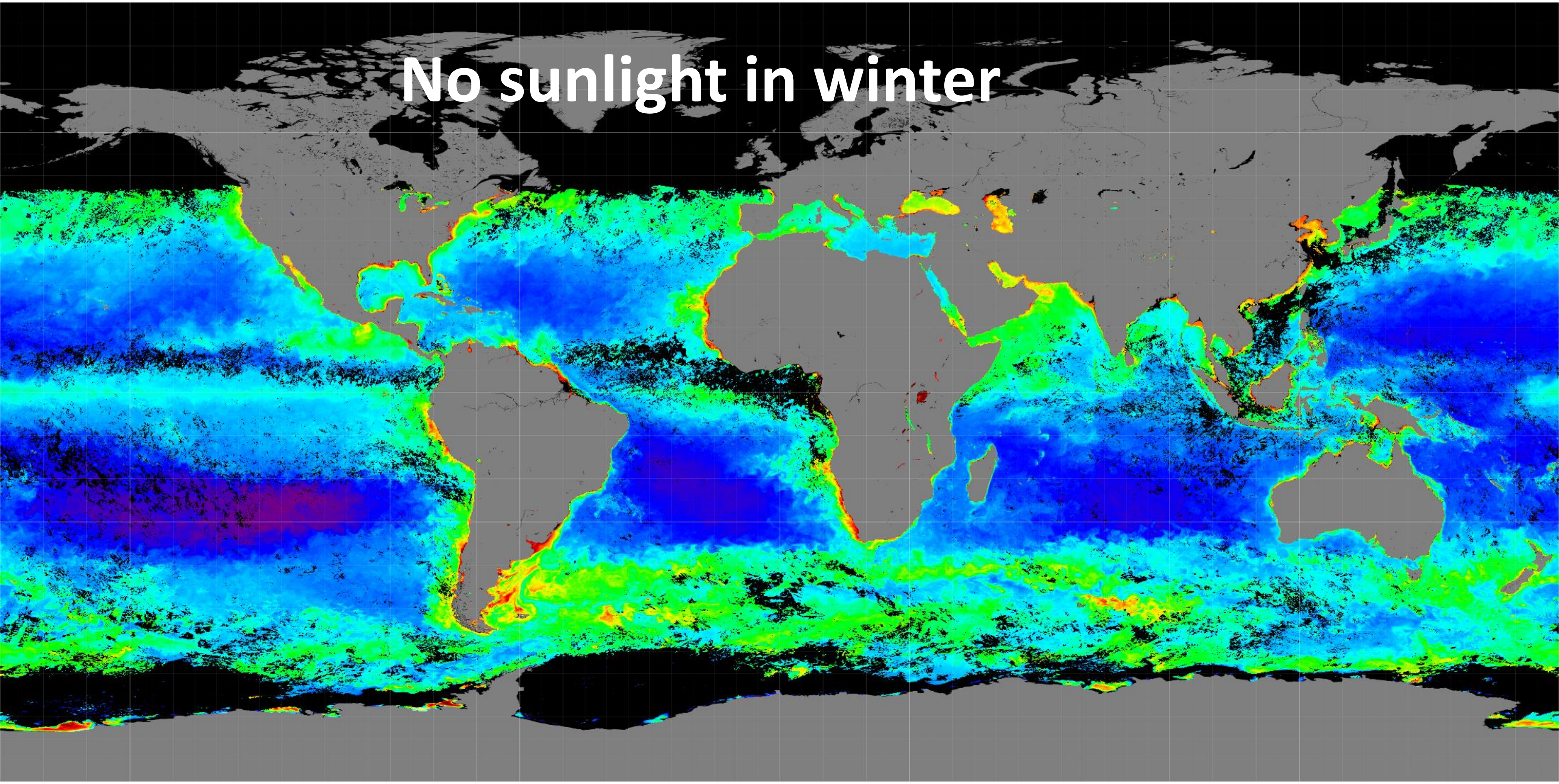
Proximity to land

Reasons for missing data



VIIRS chlorophyll concentration – June 2014

Reasons for missing data



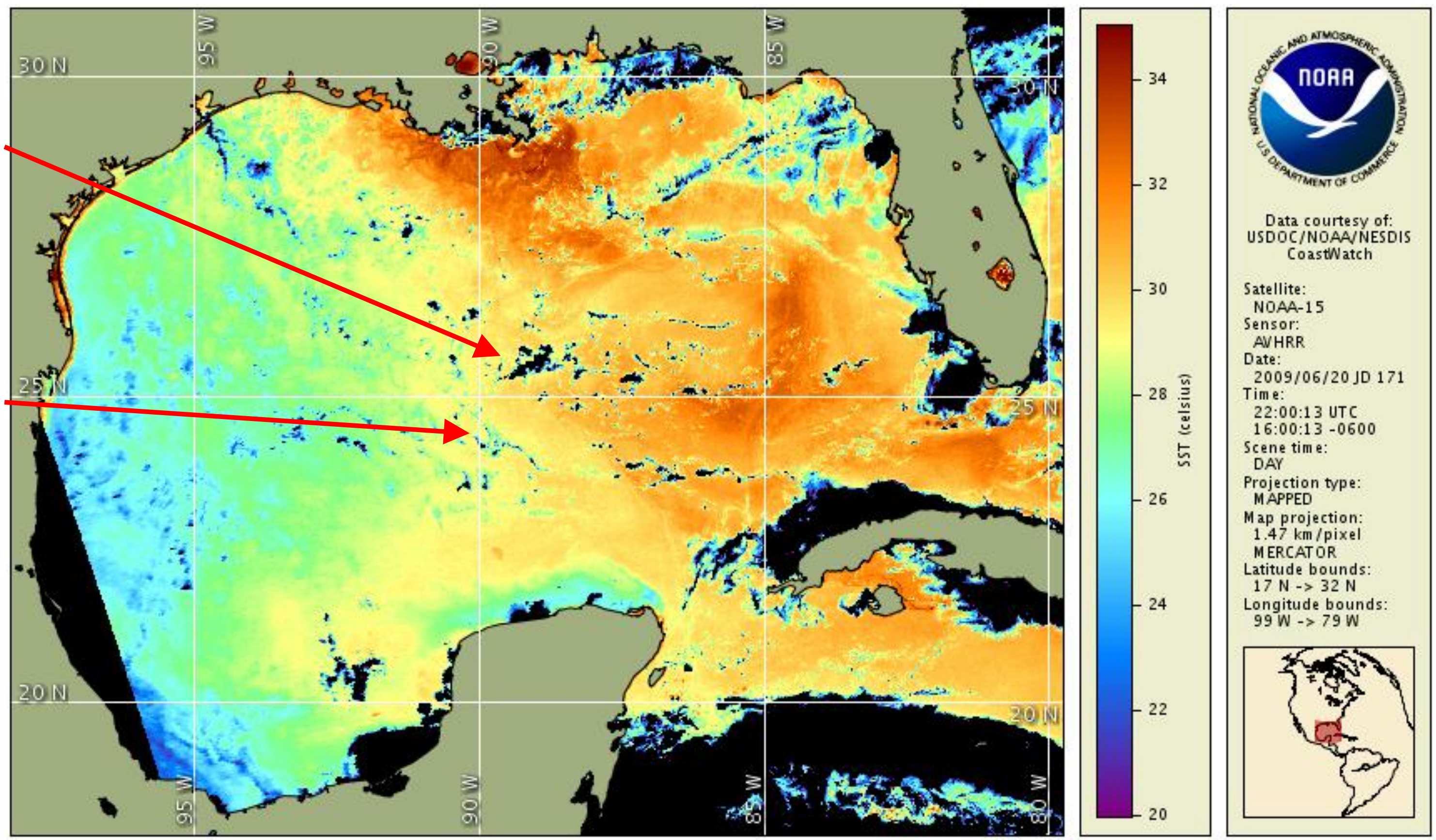
VIIRS chlorophyll concentration – December 2014

Then there can be problems with how the provider flags “bad” pixels

- Cloud detection algorithms fail to detect some clouds

Black areas identified as clouds

Some blue areas are probably clouds but were not identified as such

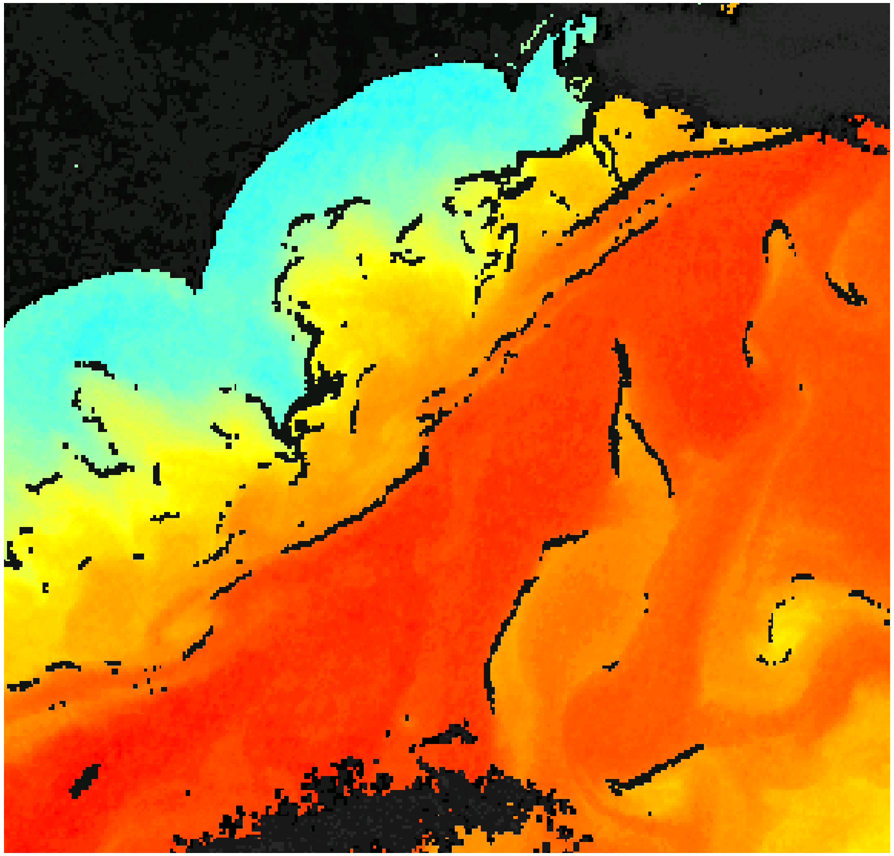


Problems with flagging “bad” pixels

- Cloud detection algorithms may classify fronts as clouds

SST fronts are mistaken as clouds because one of the provider’s tests for clouds is to look for strong gradients

Example from NOAA Pathfinder SST v5.2

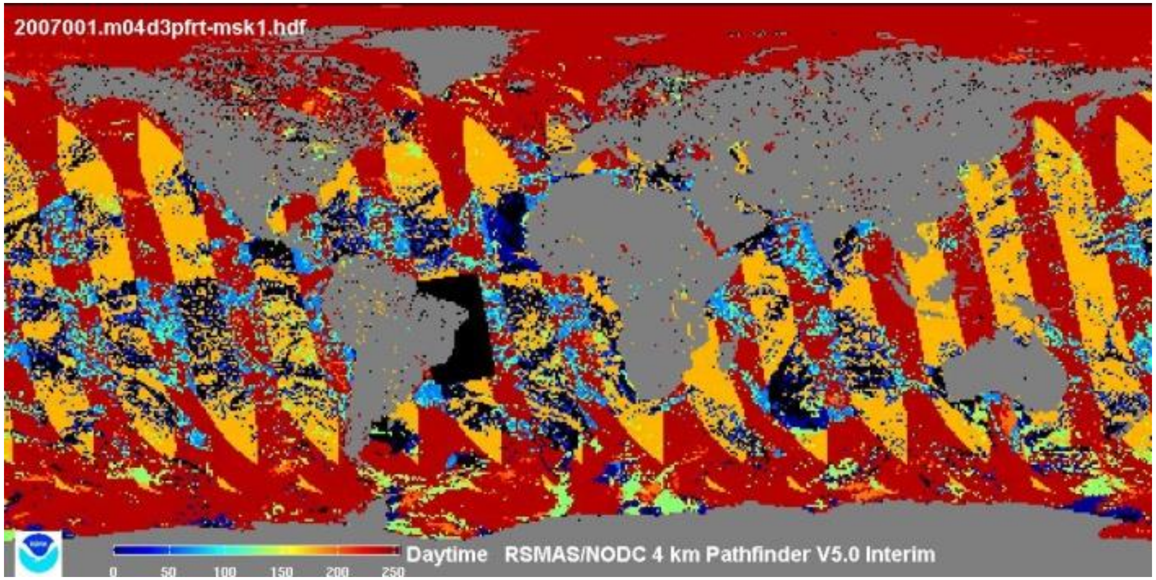


Problems with flagging "bad" pixels

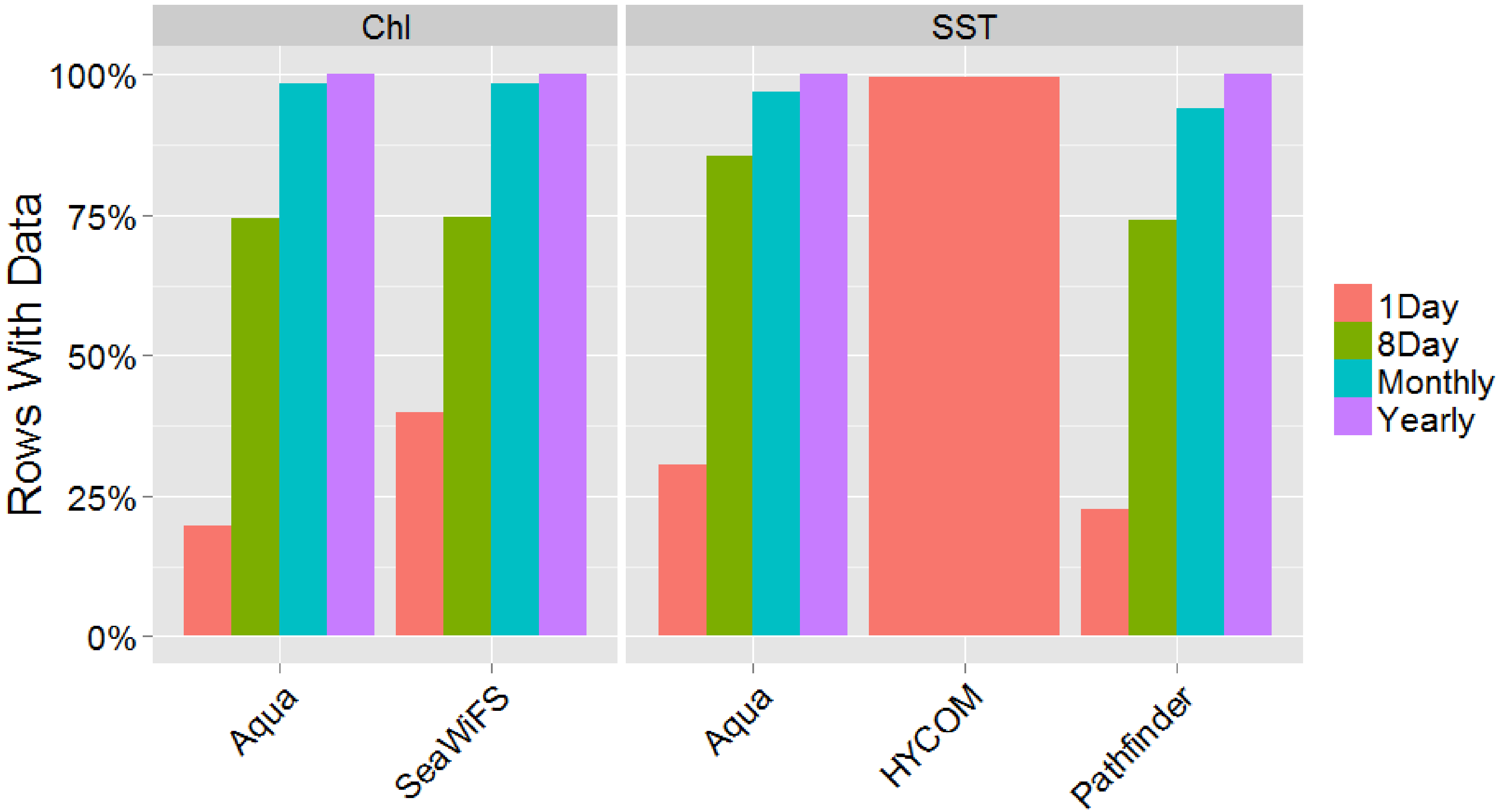
- The tests used by a provider may be too conservative
  - For example, fields that are close to the edge of the swath may be rejected because the signal passes through too much atmosphere, biasing the measurements, but you may be willing to tolerate this error

NOAA NODC 4km AVHRR Pathfinder SST "Mask 1"

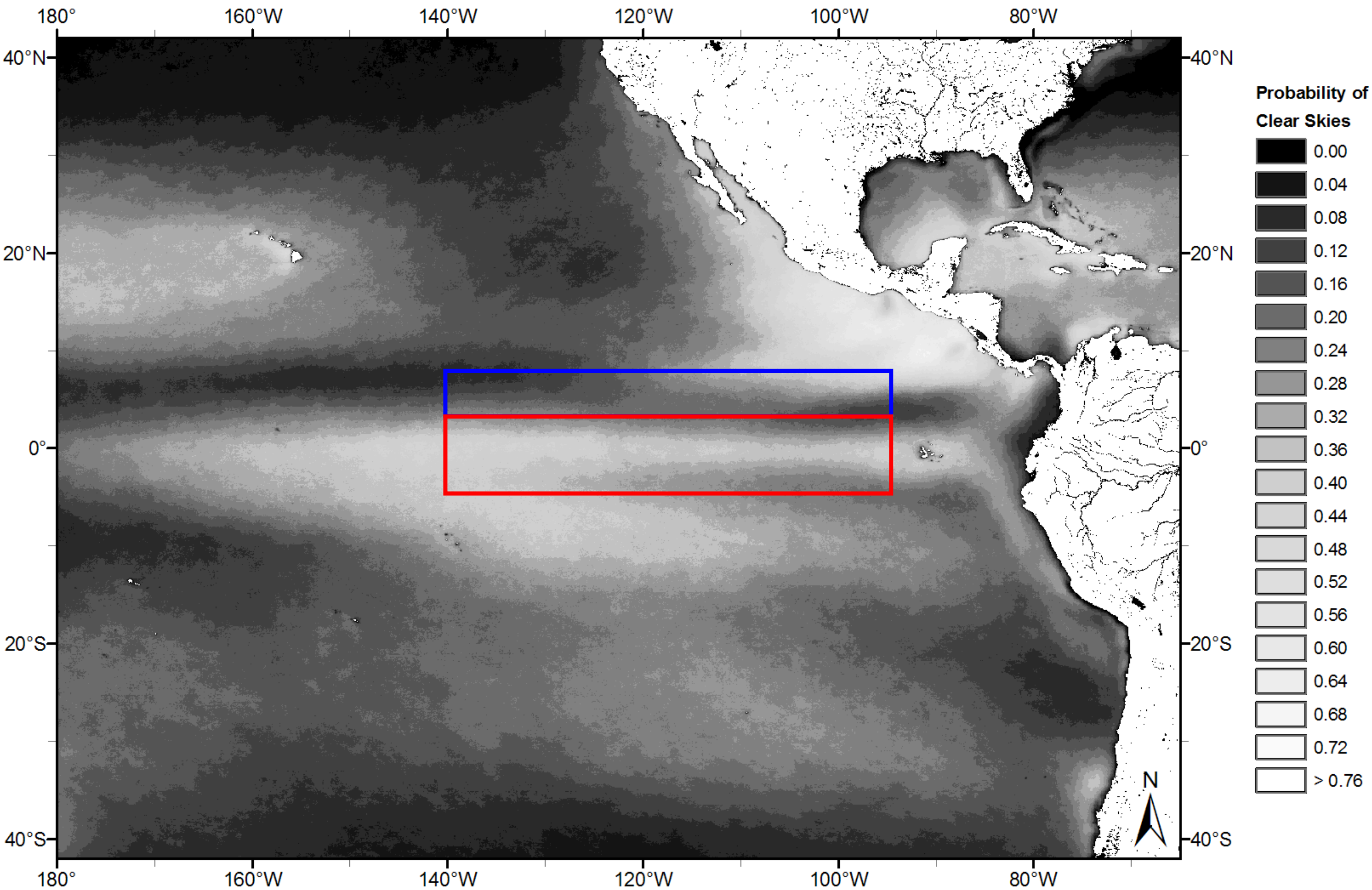
Yellow areas are observed by the sensor but are rejected because they appear at the edge of the field of view



<u>The bottom line:</u> you must discard 70-80% of your observations to use daily images



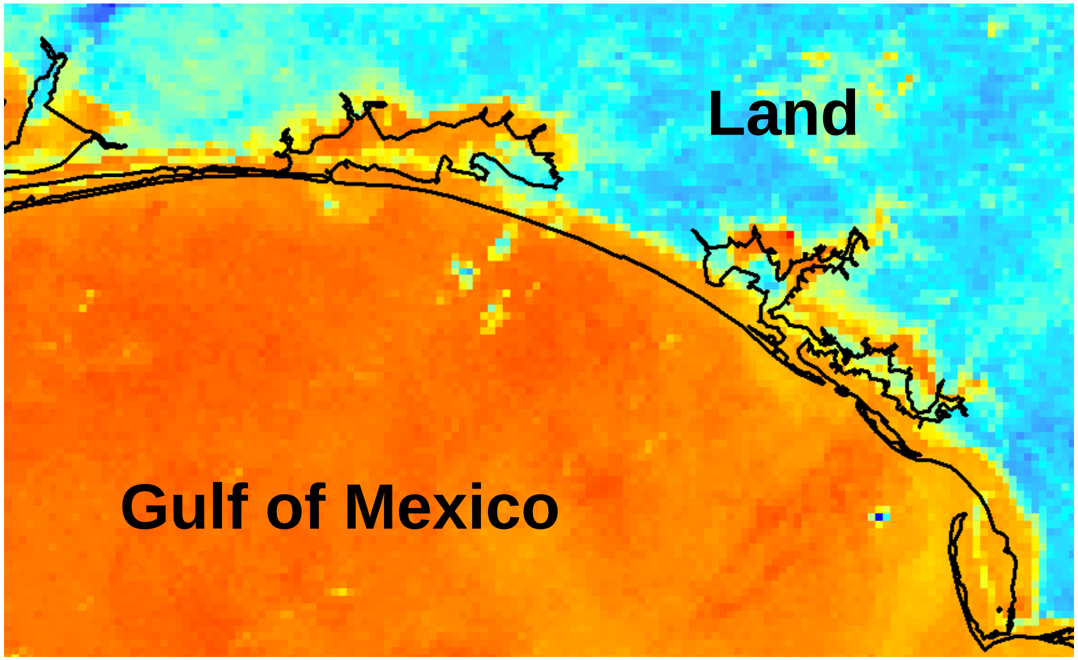
Clouds Climatology - February



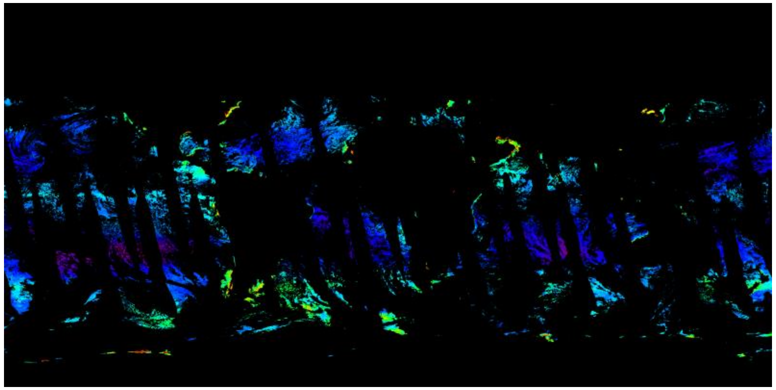
Probability that pixels were sufficiently cloud-free in AVHRR v5 images to identify SST fronts, 1985-2005

One of my favorites: geolocation error with NOAA CoastWatch SST

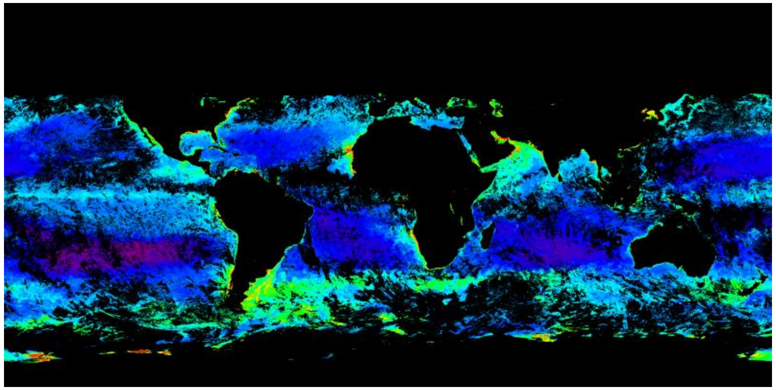
- You must check each image for “navigation” error!
- Automatic correction tools exist, but may not work well
- Manual review and editing is the only way to be sure



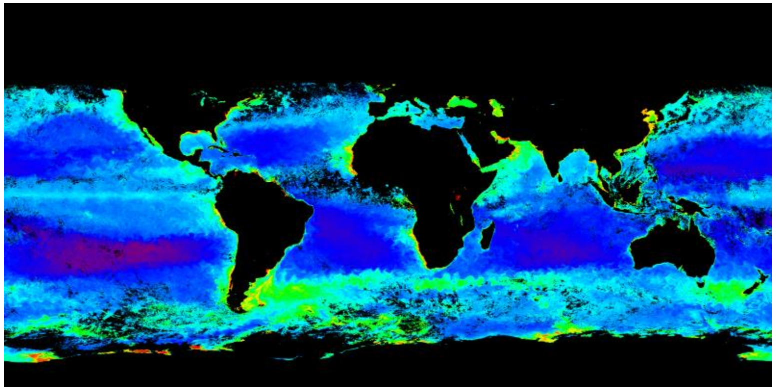
How about L3 composites instead?



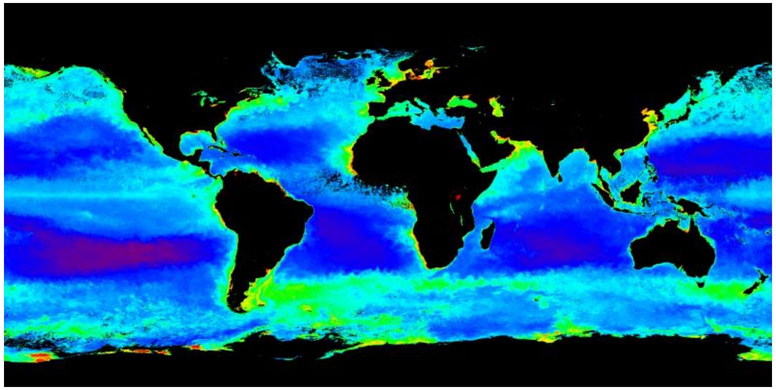
Daily MODIS Aqua Chlorophyll



8-Day

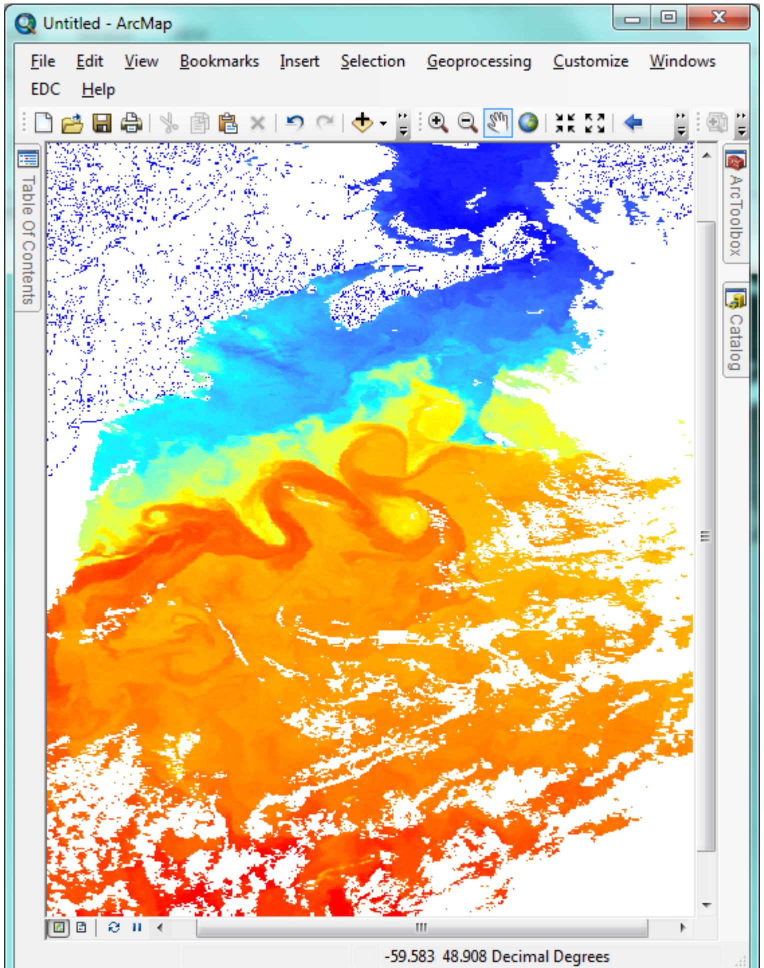


Monthly

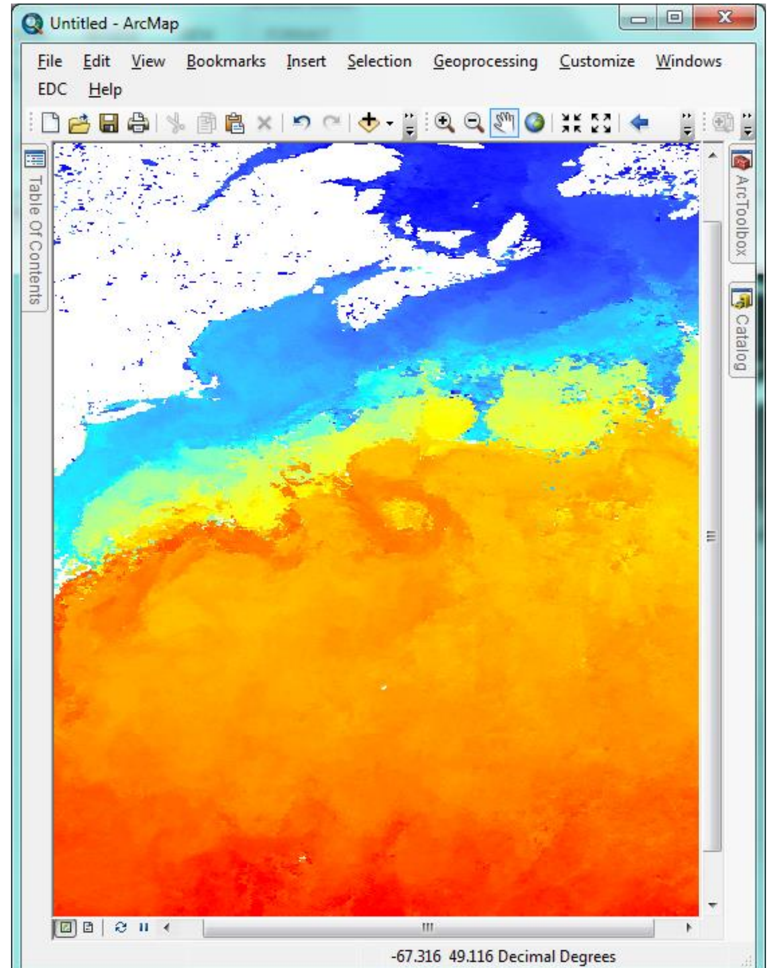


Seasonal

Problem: composites can smooth out ephemeral features…



19 April 2005 Daily



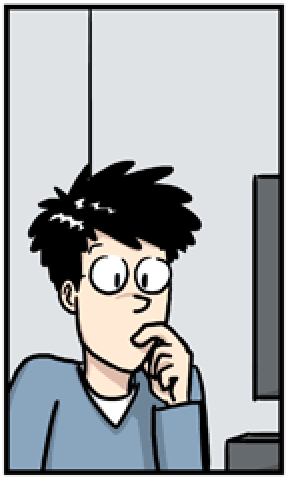
April 2005 Monthly

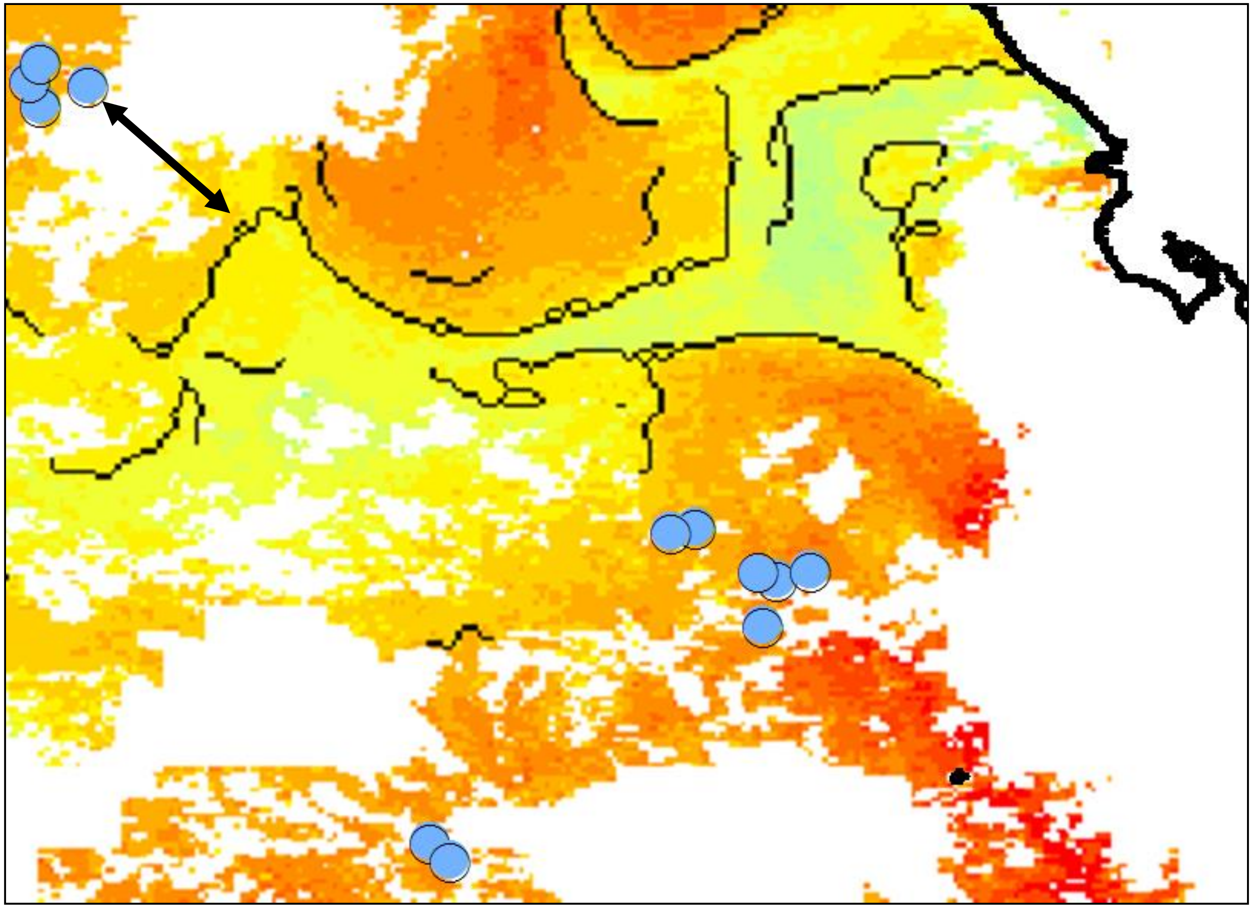
To detect fronts well, you must use daily. But clouds cause interesting problems…

There is a cloud between these points and the closest visible SST front.

There could be a front behind the cloud.

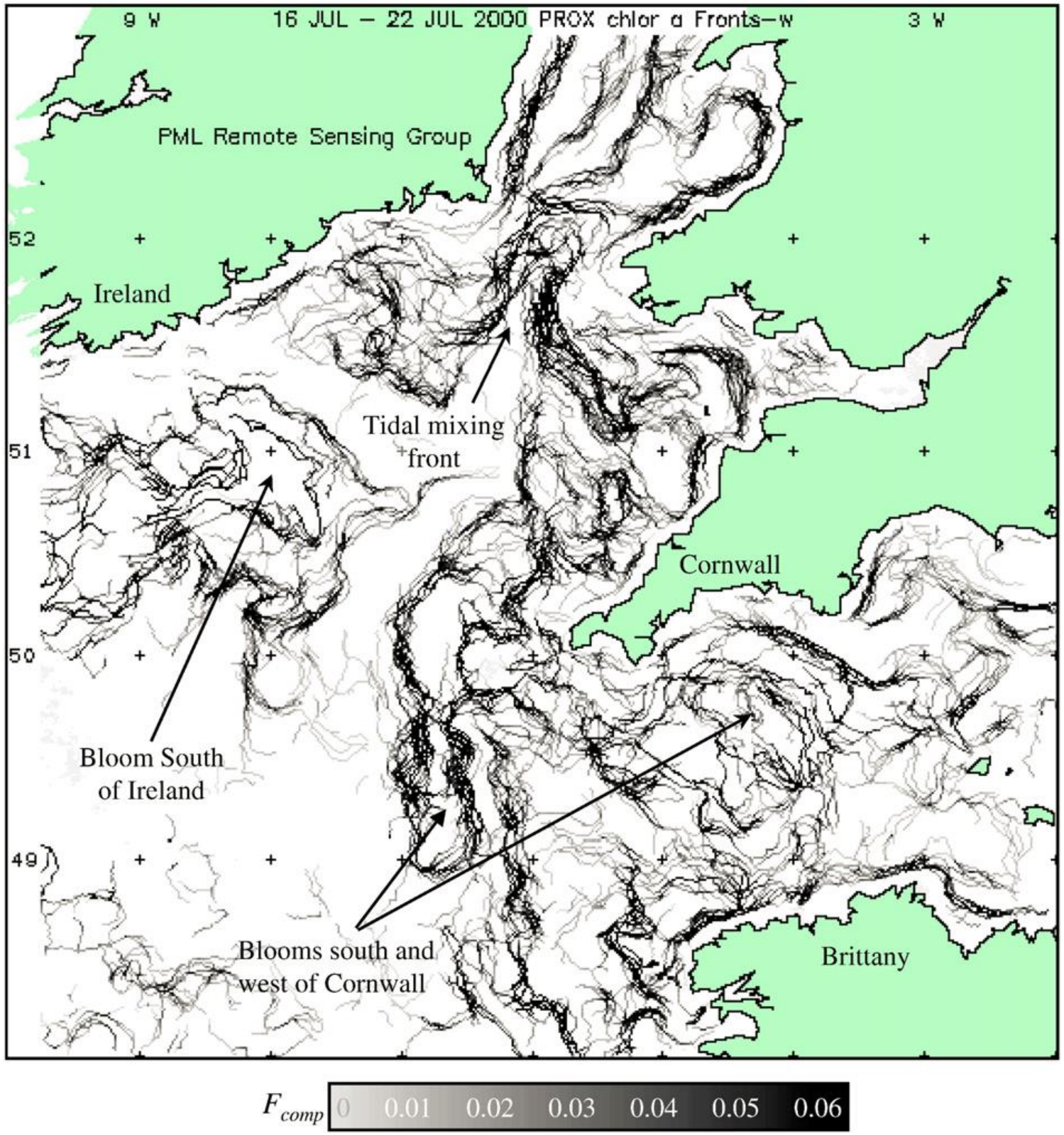
What do you do?





7-day, 1.47 km SST front composite

You probably need to first detect fronts in daily images, then create running composites…

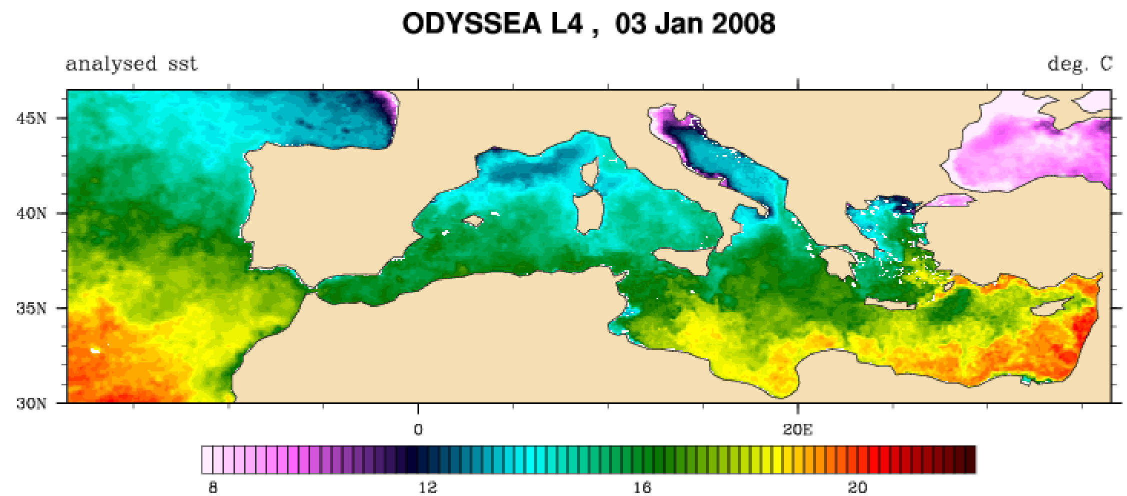


Miller (2009) J. Mar. Sys.

What “level” of data do you want?

Level	Description
0	~~Real time data feed from ground control station.~~
1	~~Files of calibrated, geolocated, at-aperture radiance values for swath segments, with quality flags and error estimates.~~
2	~~Files of geophysical values (e.g. SST) for swath segments, calculated from the Level 1 data by applying an algorithm to the radiance values.~~
3	Files of uniform grids of geophysical values, for various spatial and temporal scales, produced by accumulating and projecting Level 2 data. **Maybe use this**
4	Same as Level 3, but with missing data filled in via interpolation, modeling, integration of data from multiple sensors, or other means. **What about this?**

Level 4 products



Main advantage

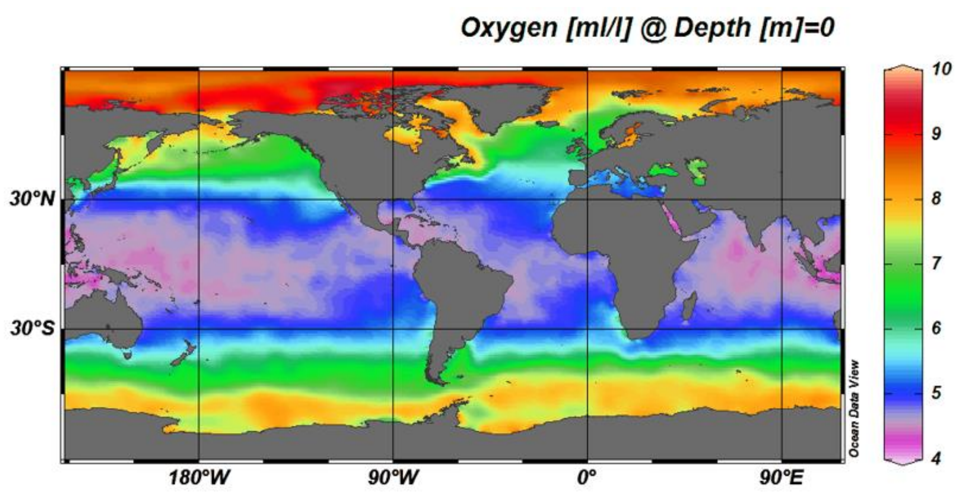
- No missing data!

Disadvantages

- Rely on recent satellites; do not go far back in time
- Often have reduced resolution
- Some products have very high resolution but may show fine spatial structure that is based on <u>interpolation or modeling, not actual observations</u>

Then you discovered CLIMATOLOGIES!

- Long term averages assembled from a long series of repeated observations
- Often no missing data!
- Often available at seasonal or monthly resolution
- Often provide sophisticated covariates
  - These can be low resolution: 0.5^o to 5^o



- Should you use them?

Practical advice

How do you choose dynamic covariates?

- Unless you have lots of survey data, programming skill, time, and storage space, avoid L2 images
  - The difficulties are just not worth it
  - Use them to make interesting figures
  - Do not mislead readers into believing that you are using them for analysis (unless you actually are)

Practical advice

- This leaves four general choices:
  - Daily level 3 images
  - Daily level 4 images and ocean model predictions
  - Composite level 3 images (8-day, monthly, …)
  - Climatological images
- This choice is mainly about deciding what temporal resolution is appropriate
- With marine data, you usually can and should decide that before addressing spatial resolution

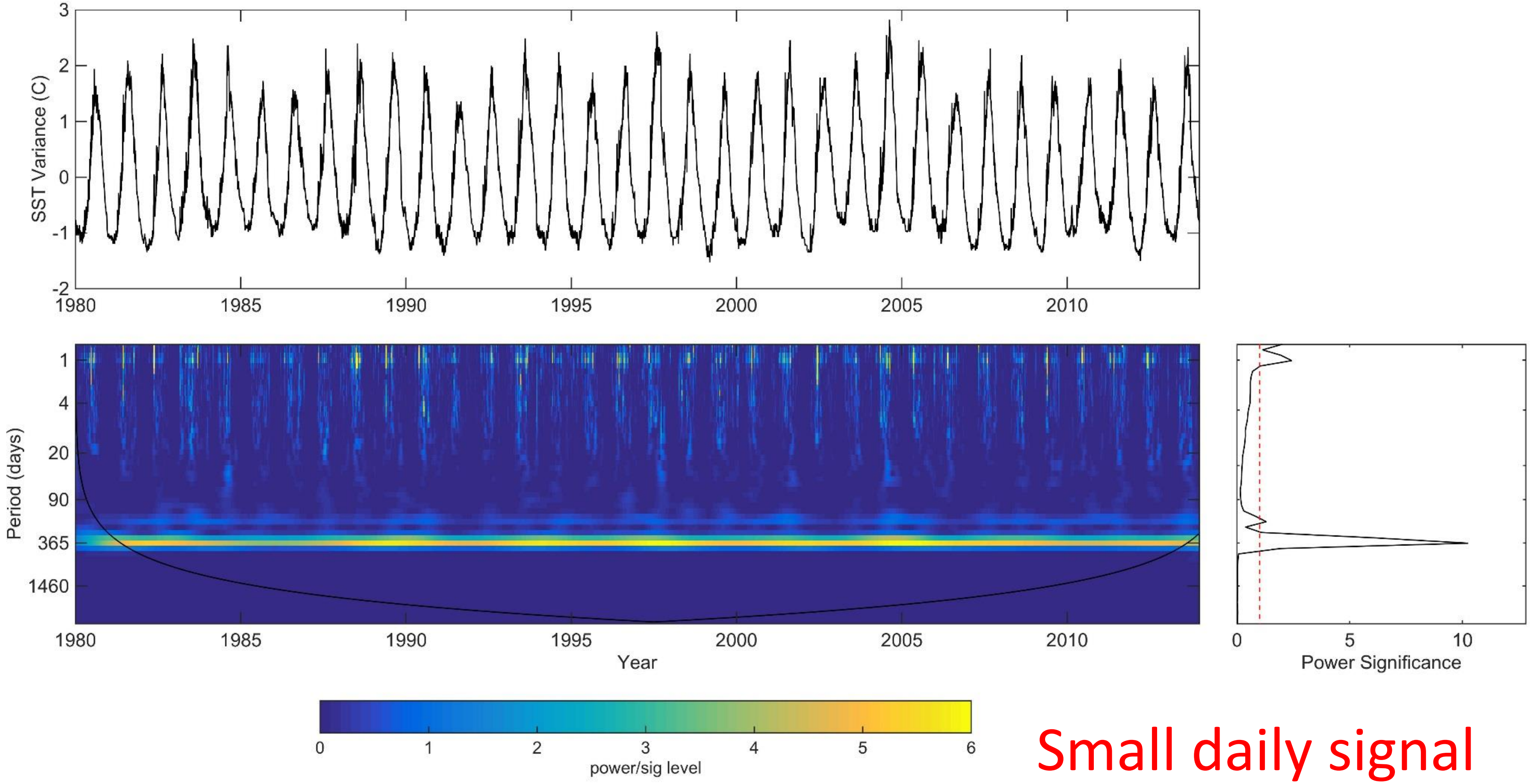
Practical advice

- First review the known ecology of your organism
  - Is there *really* a plausible argument that it is responding to ephemeral or mesoscale features?
  - If so, is that process happening at spatial and temporal scales you can detect remotely?
  - And does your survey provide enough temporal coverage to detect it?
- Unless you answered a solid “yes” to all three, you probably don’t need daily resolution

Practical advice

- Also review the known dynamics of the dominant oceanographic processes in your ecosystem
  - What are these processes?
  - At what temporal (and spatial) scales do they operate?

Long term temporal analysis of buoy SSTs: Gulf of Alaska

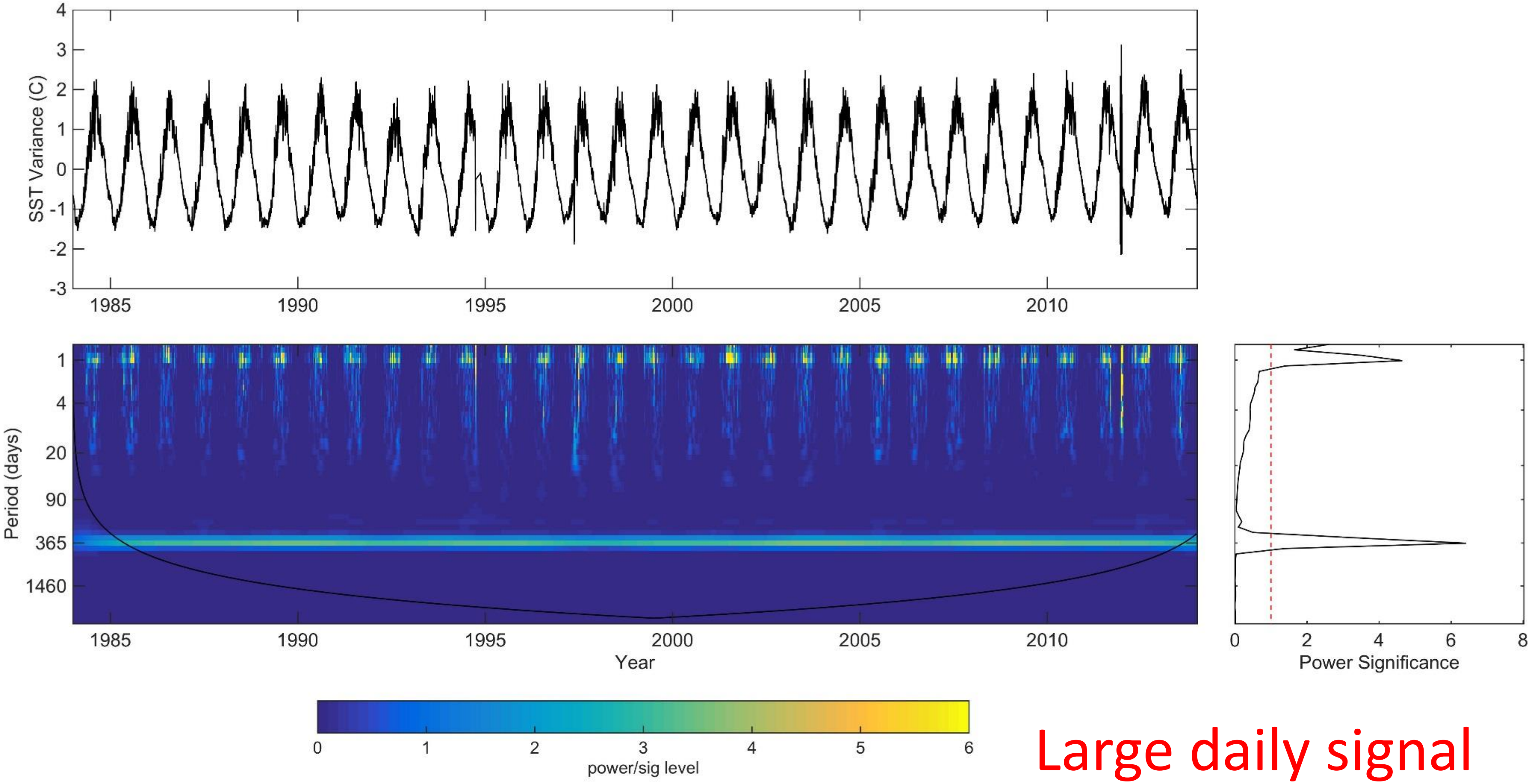


Small daily signal

Large annual signal

Figures: Andre Boustany

Long term temporal analysis of buoy SSTs: Gulf of Maine near Boston

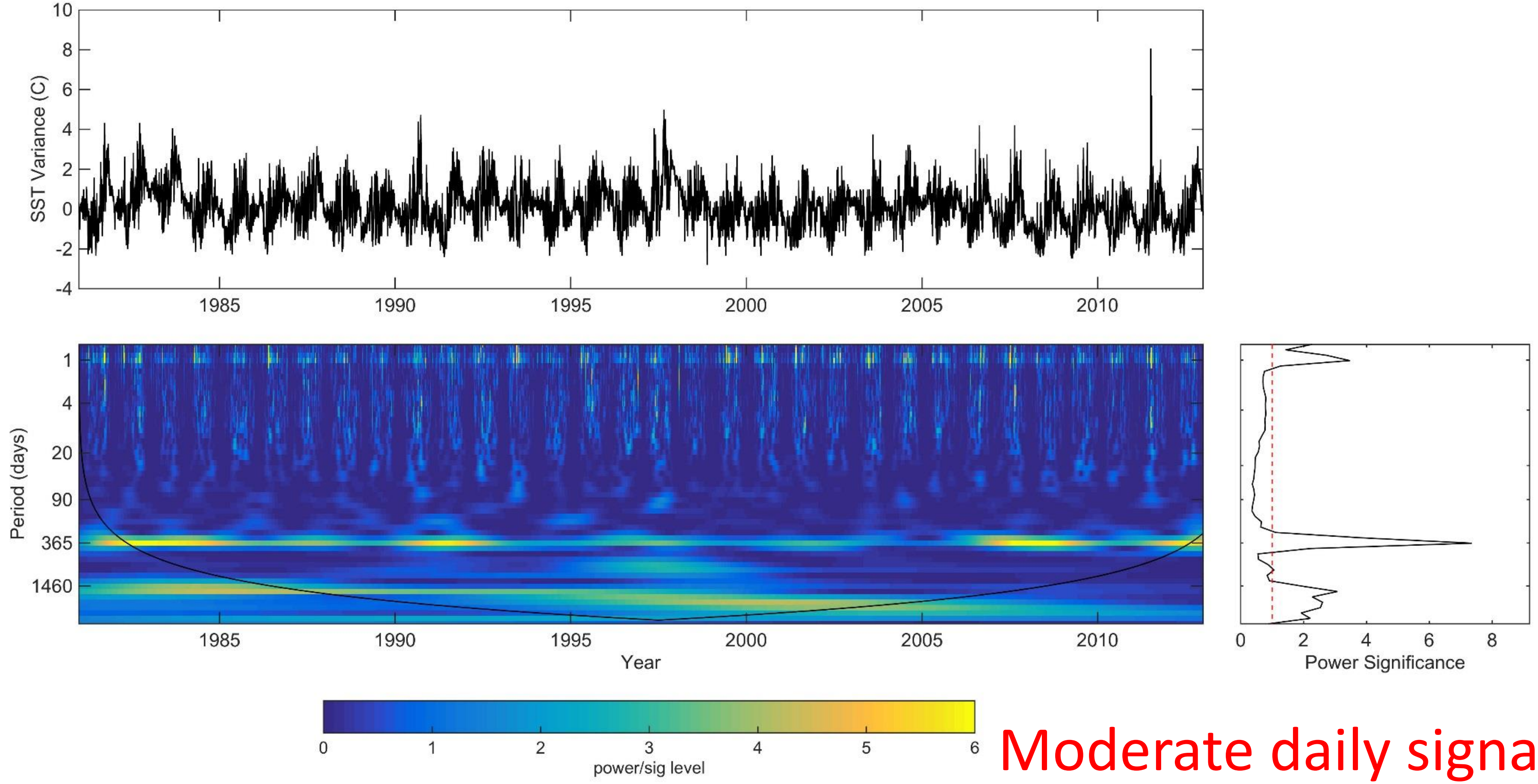


Large daily signal

Large annual signal

Figures: Andre Boustany

Long term temporal analysis of buoy SSTs: Southern California



Moderate daily signal
Large annual signal
Moderate ENSO signal

Figures: Andre Boustany

Contemporaneous vs. climatological estimates of covariates

- Inter-annual variation in California
  - Suggests that contemporaneous estimates are needed to model this ecosystem
- Only intra-annual variation in Alaska and Boston
  - Suggests that climatological estimates may be sufficient to capture seasonal dynamics
- Daily variation results from tides
  - Probably not relevant to most density models

Practical advice

- If you cannot build a case for contemporaneous estimates, use climatological estimates
  - Can use old surveys without worrying whether the satellite was launched yet
  - No loss of data due to clouds
  - Access to sophisticated covariates
  - Only a few images to sample and predict with
- Monthly and seasonal climatologies readily available

Practical advice

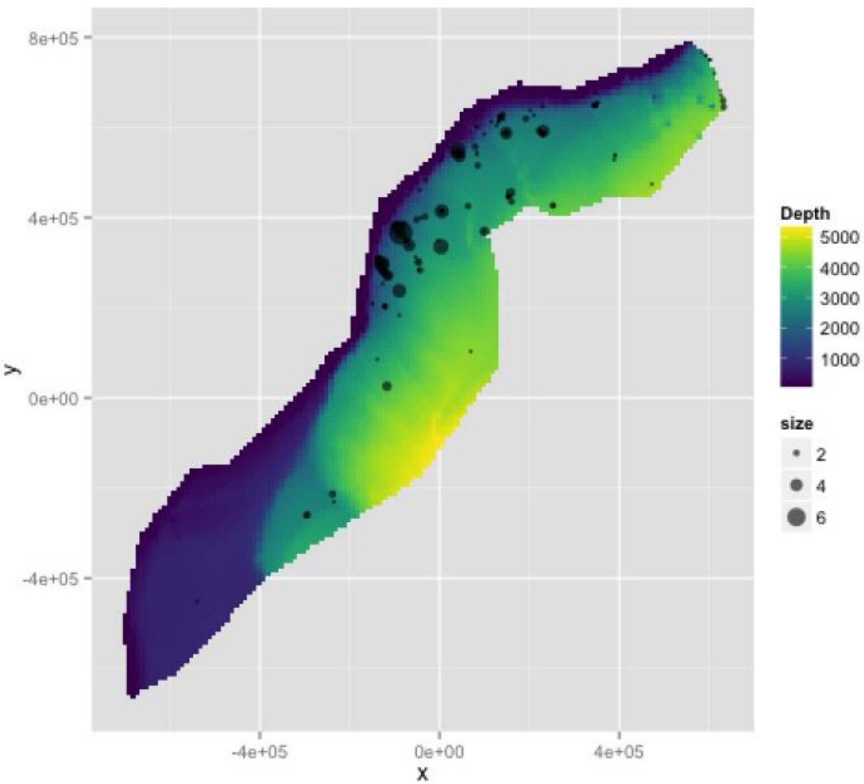
- For contemporaneous estimates, we are down to:
  - Daily level 3 images
  - Daily level 4 images and ocean model predictions
  - Composite level 3 images (8-day, monthly, …)
- Only use daily level 3 images if you must use daily and there are no good Level 4 or ocean models
- Otherwise weigh tradeoff between:
  - Level 4 and ocean models are totally cloud free
  - Composites are fewer, so easier to sample and predict

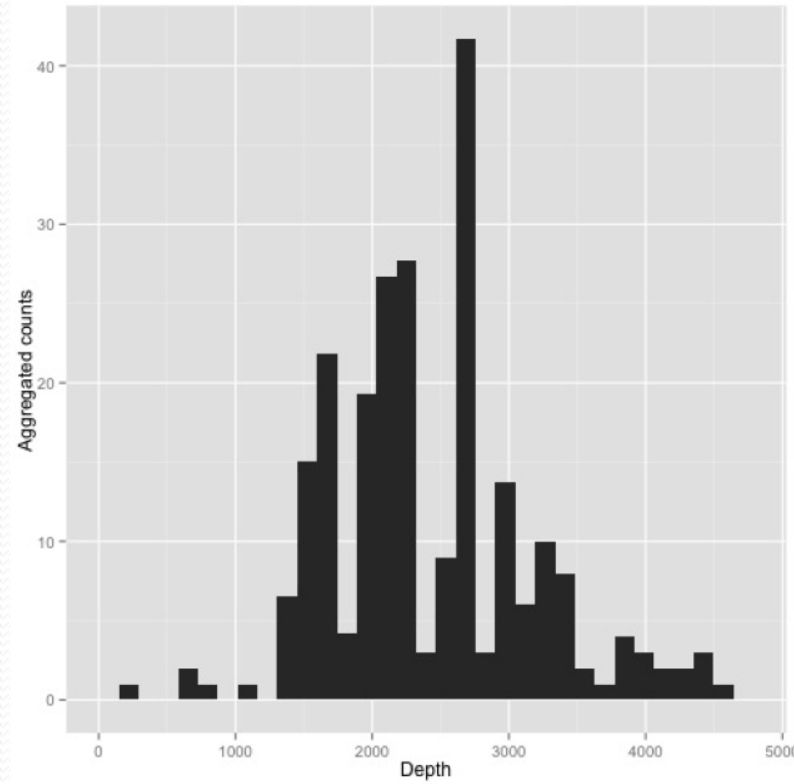
Practical advice

- Once you have settled the temporal resolution question, then consider spatial resolution

Covariates for our sperm whale model

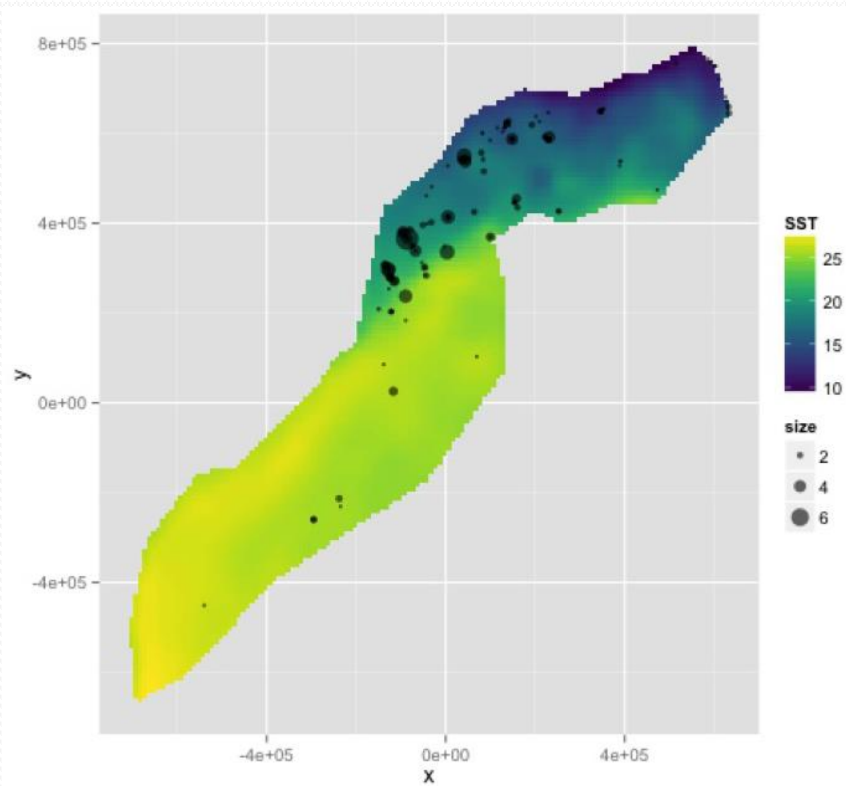
- Sperm whales are characterized as deep diving squid-eaters
- Exploratory analysis suggested depth may be useful covariate

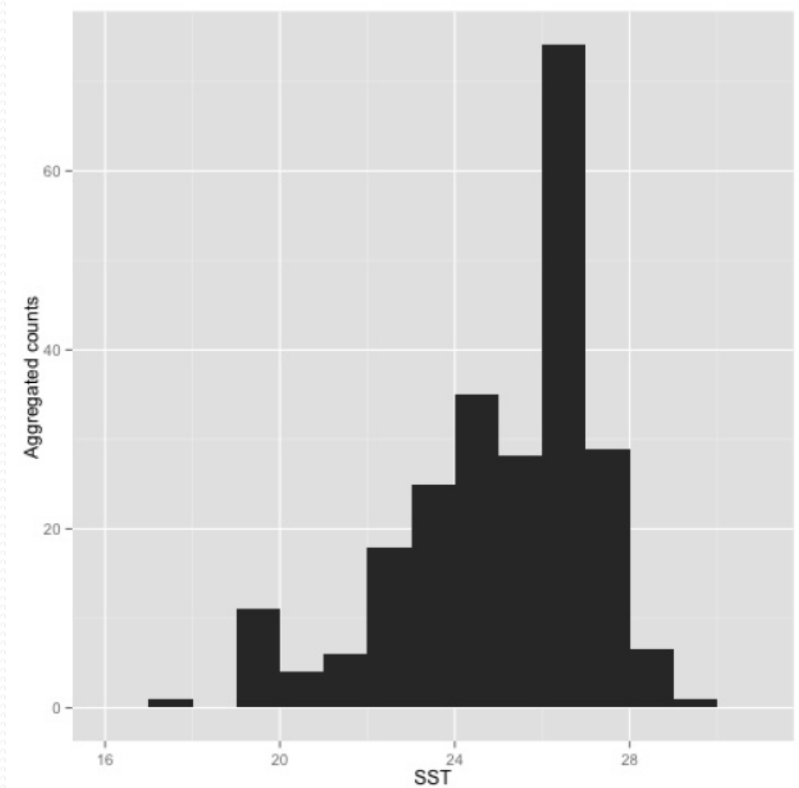




Covariates for our sperm whale model

- Sea surface temperature also looked promising





Canyons may be important

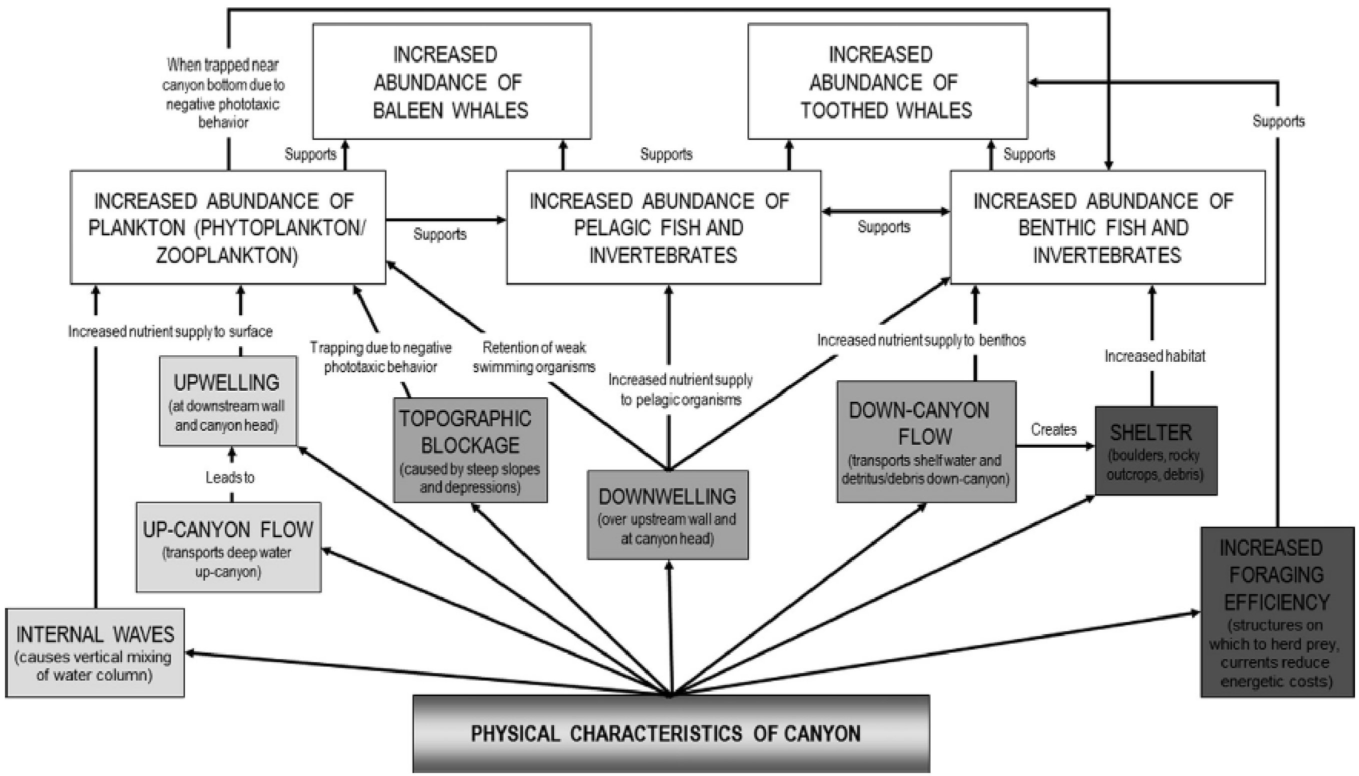
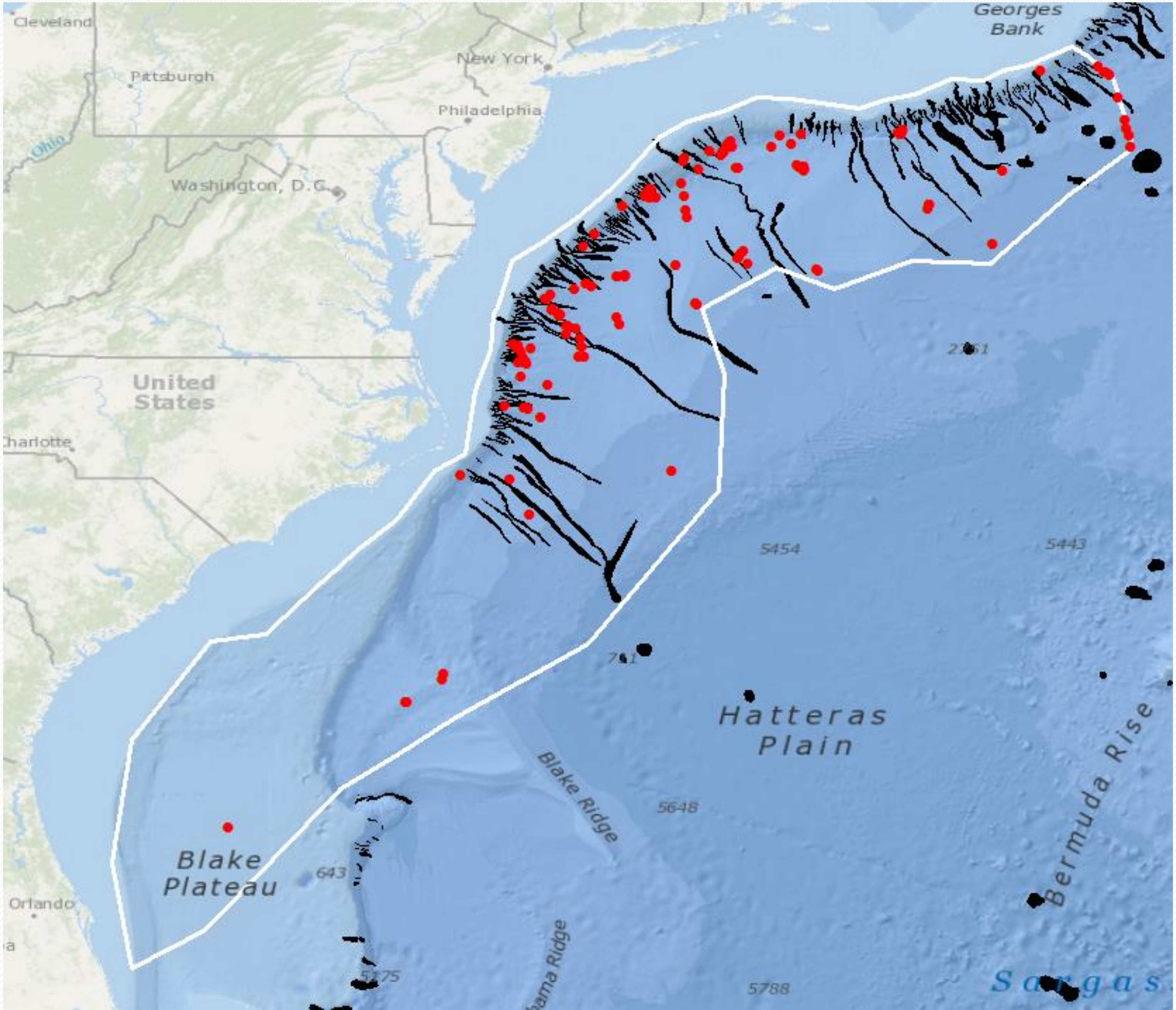


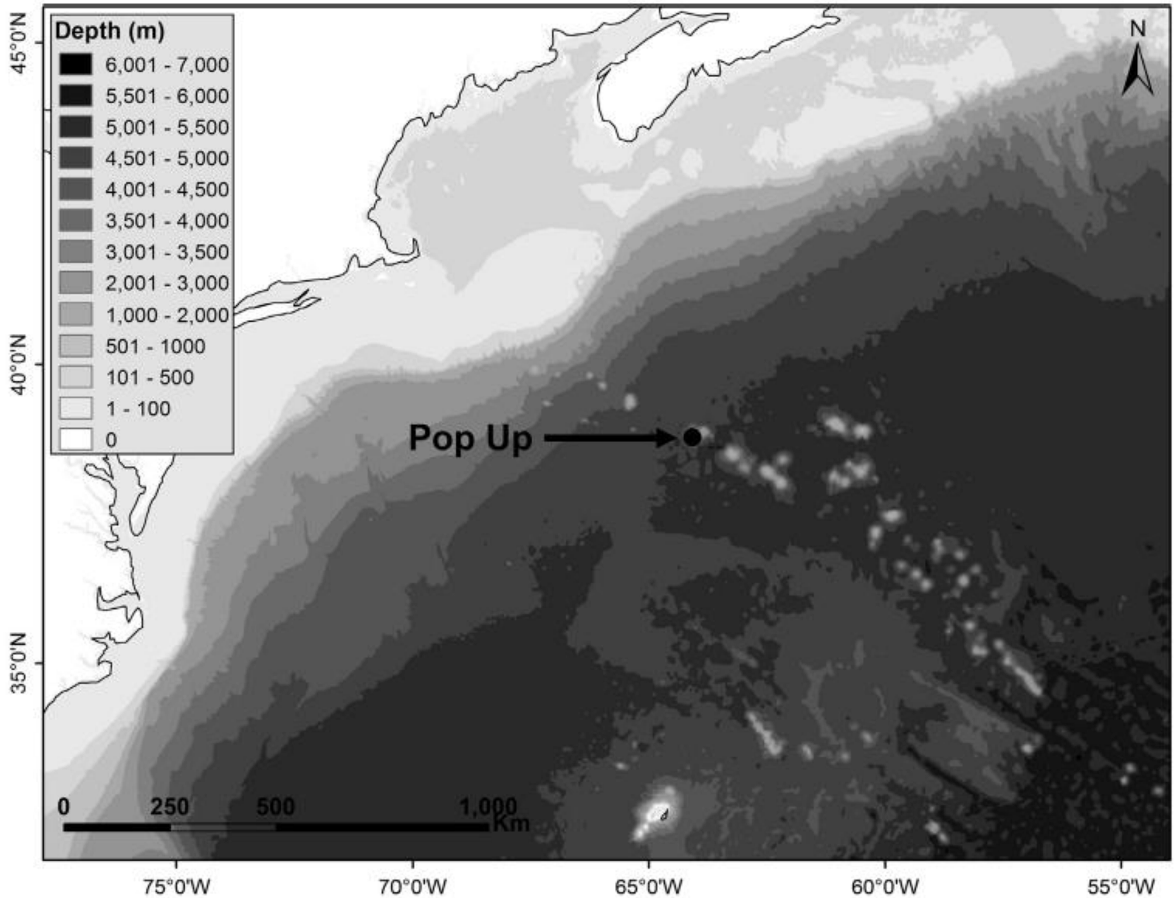
Fig. 1. Summary of mechanisms through which submarine canyons may attract cetaceans. Light gray boxes indicate enrichment processes, medium gray boxes indicate concentrating processes and dark gray boxes indicate aggregating processes.

Moors-Murphy HB (2014). Submarine canyons as important habitat for cetaceans, with special reference to the Gully: A review. *Deep Sea Research II* 104: 6–19.

Sightings with canyons and seamounts



Seamounts



Wong SNP, Whitehead H (2014) Seasonal occurrence of sperm whales (*Physeter macrocephalus*) around Kelvin Seamount in the Sargasso Sea in relation to oceanographic processes. *Deep Sea Research I* 91: 10-16.

Productivity and mesoscale eddies

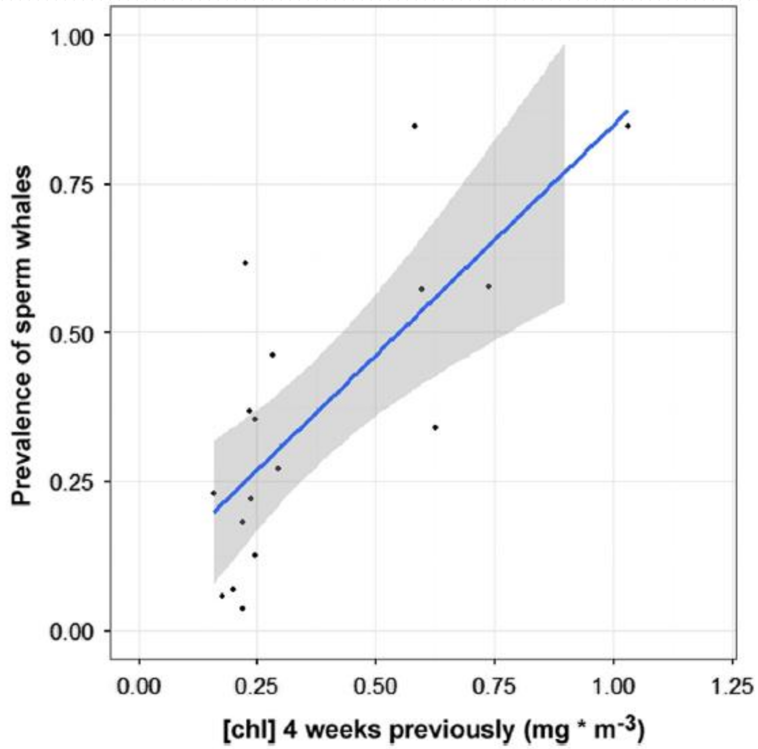


Fig. 3. Relationship between mean prevalence of sperm whales and mean chlorophyll-*a* concentration (mg m^{-3}) in an area 66 km from pop-up (deployed on Kelvin Seamount). Prevalence of sperm whales is the number of two-minute recordings per week in which sperm whales were detected/number of two-minute recordings per week.

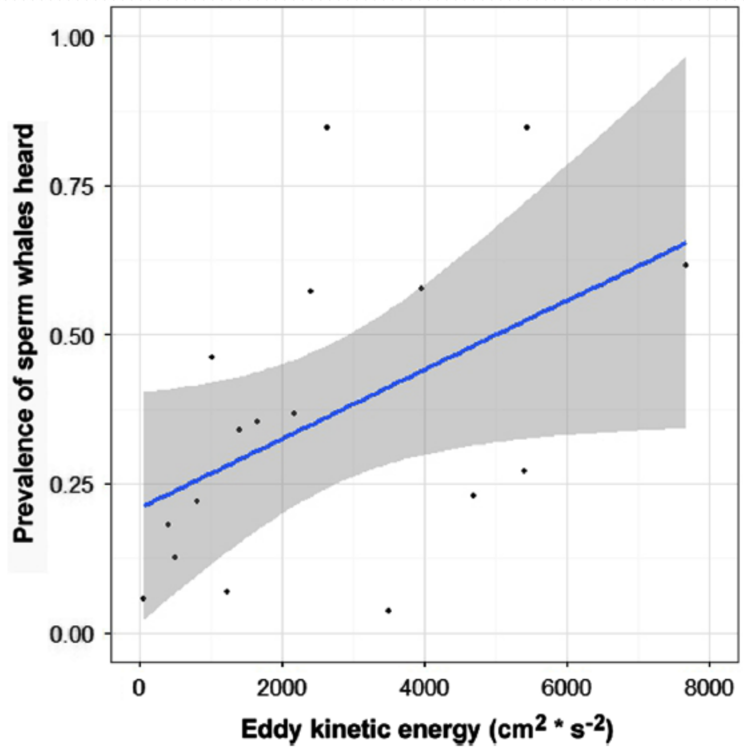
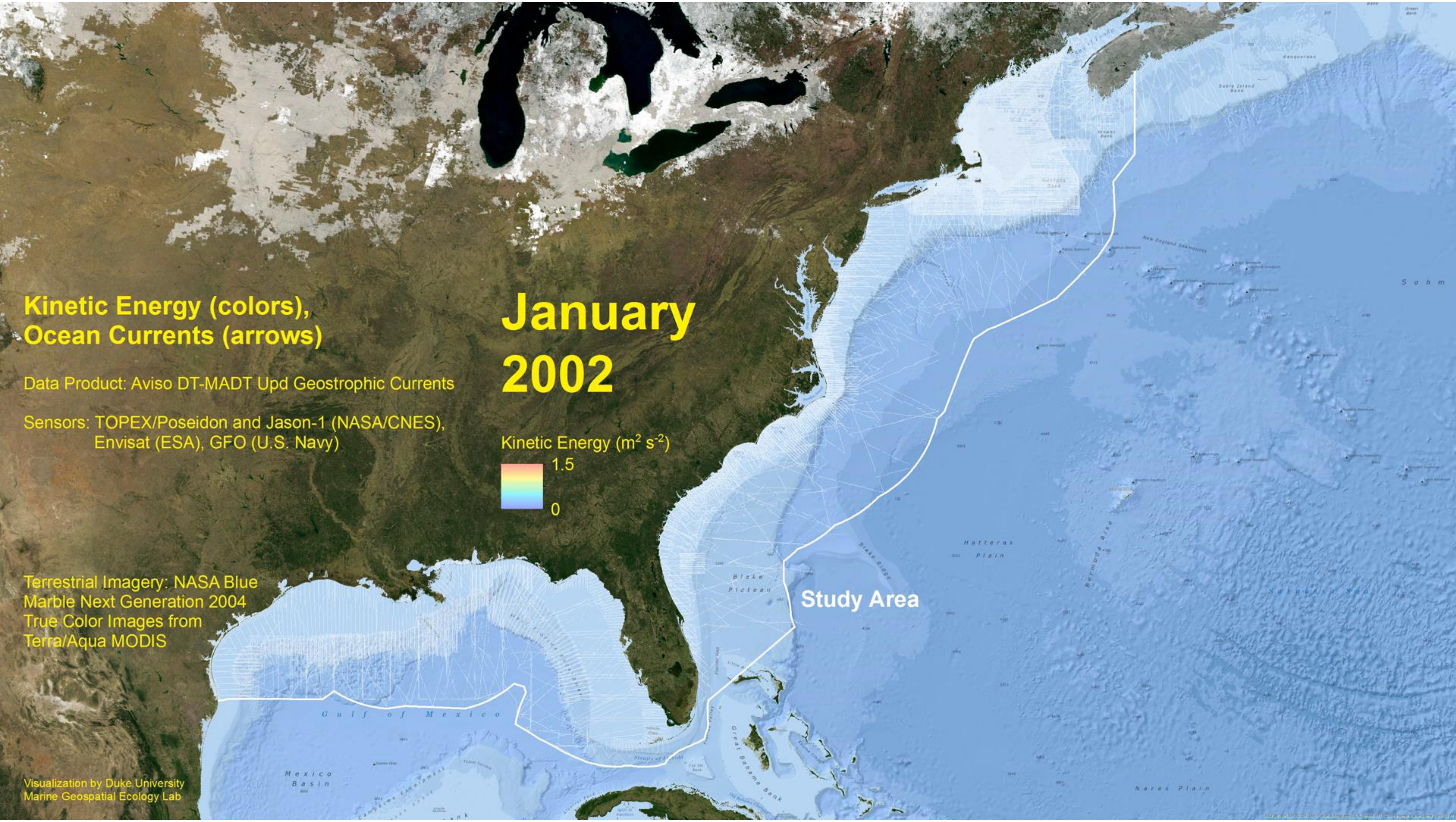


Fig. 4. Relationship between mean prevalence of sperm whales and eddy kinetic energy (cm^2 s^{-2}) around the pop-up (deployed on Kelvin seamount). Prevalence of sperm whales is the number of two-minute recordings per week in which sperm whales were detected/number of two-minute recordings per week.

Wong SNP, Whitehead H (2014) Seasonal occurrence of sperm whales (*Physeter macrocephalus*) around Kelvin Seamount in the Sargasso Sea in relation to oceanographic processes. *Deep Sea Research I* 91: 10-16.


Kinetic Energy (colors),
Ocean Currents (arrows)
Data Product: Aviso DT-MADT Upd Geostrophic Currents
Sensors: TOPEX/Poseidon and Jason-1 (NASA/CNES),
Envisat (ESA), GFO (U.S. Navy)
January
2002
Kinetic Energy (m² s⁻²)
1.5
0
Terrestrial Imagery: NASA Blue
Marble Next Generation 2004
True Color Images from
Terra/Aqua MODIS
Study Area
Blake
Plateau
Hatteras
Plain
Gulf of Mexico
Mexico
Basin
Nares Plain
Visualization by Duke University
Marine Geospatial Ecology Lab

Covariates for our sperm whale model

- Depth
  - SRTM-30 PLUS global 30 arc-second bathymetry
- Distance to closest canyon or seamount
  - Derived from the Harris et al. (2014) geomorphology, 30 arc-sec
- Sea surface temperature
  - GHRSST CMC 2.0 L4 daily 0.2^o SST
- Eddy kinetic energy
  - Derived from AVISO DT-MSLA daily 0.25^o geostrophic current anomalies
- Primary productivity
  - Oregon State 8-day 9 km VGPM from MODIS Aqua