Ocean Satellite Observations and Challenges

Robert N. Miller

College of Oceanic and Atmospheric Sciences
Oregon State University
Corvallis, OR 97330

Outline

- Introduction
- Satellite observations:
  - What are they,really?
  - What information do they really contain?
- Interdisciplinary oceanography
- Representation Error
- Challenges

A Modeler's View of Satellite Data

The good news:

- Extensive spatial coverage
- A variety of data types:
  - SST
  - SSH
  - Wind
  - Color
  - Salinity (coming soon!)
- Instruments keep getting better!

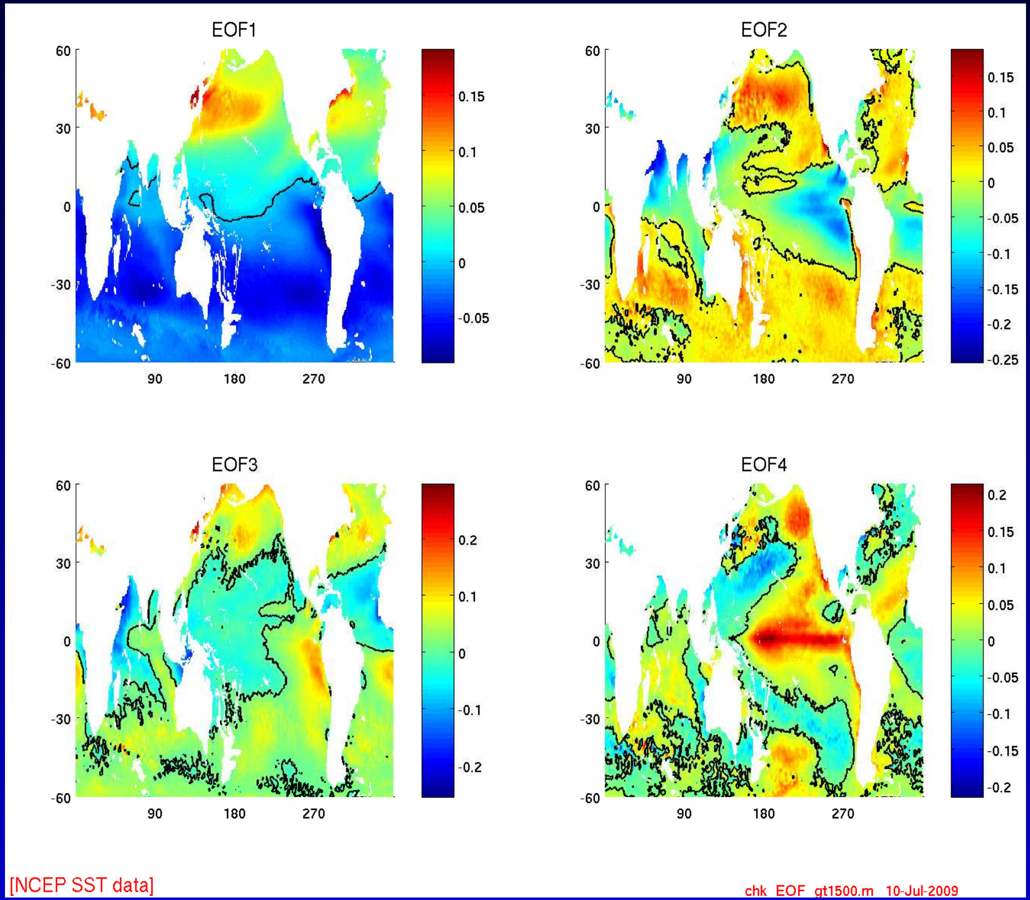
A Modeler's View of Satellite Data

The not-so-good news:

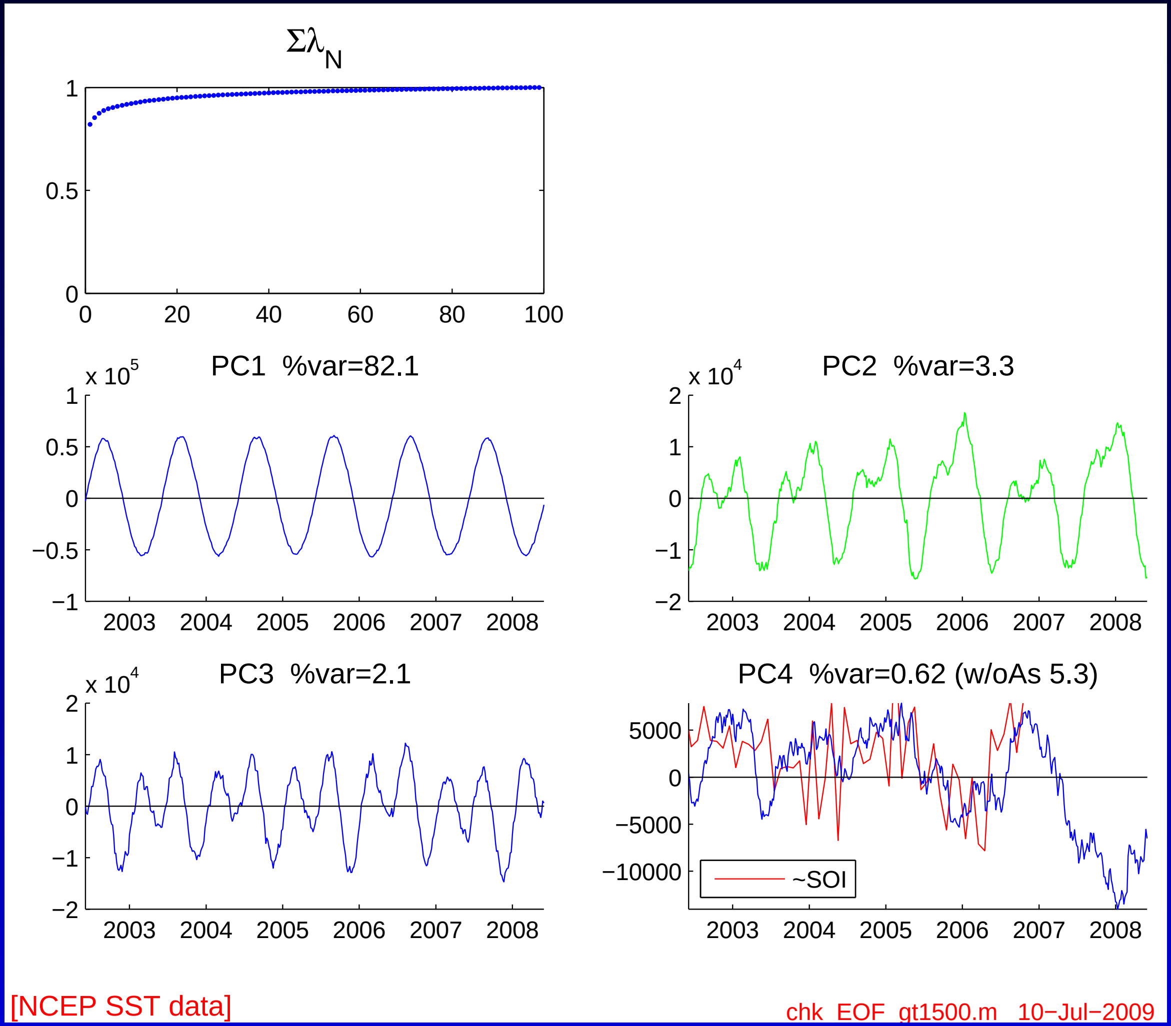
- Surface only
- Short and often gappy time series
- Changes in instruments make statistics difficult
- Some measured quantities are hard to relate to model state variables
- One exception is altimetry, but altimeter data are not synoptic

What Will Satellite Data Tell Us?

begin with SST:

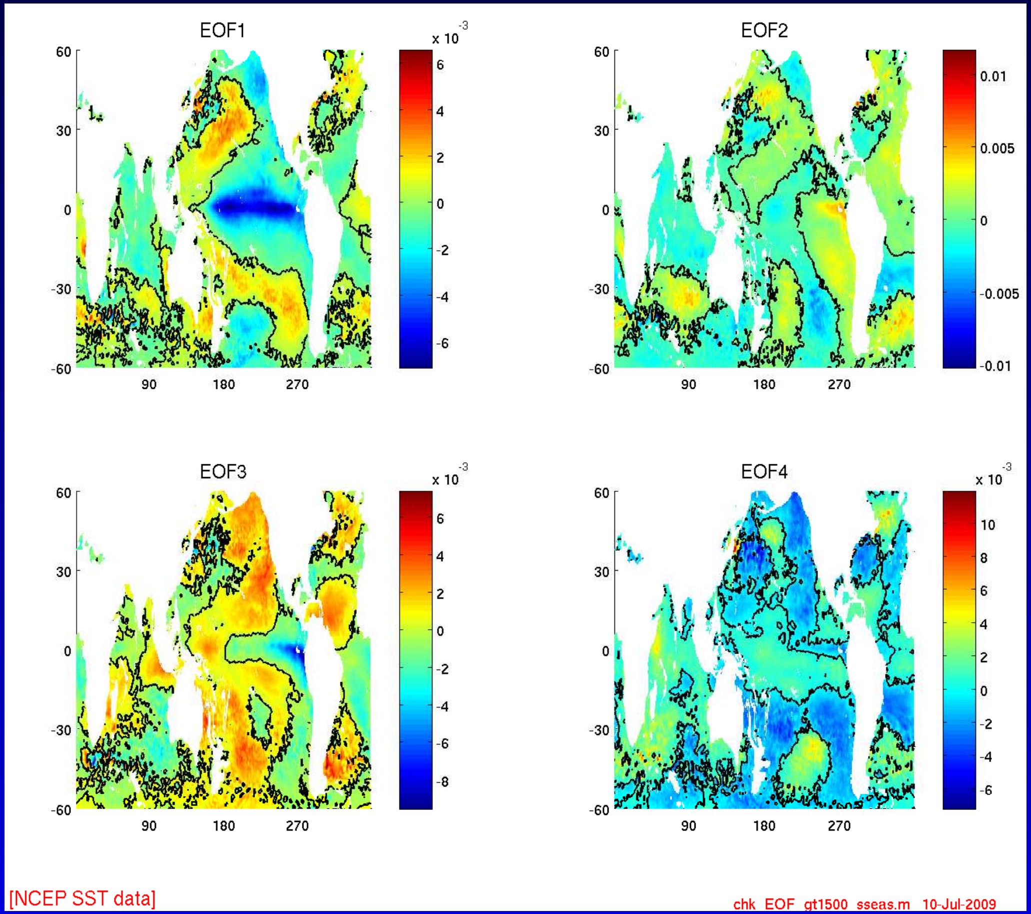


EOF Amplitudes

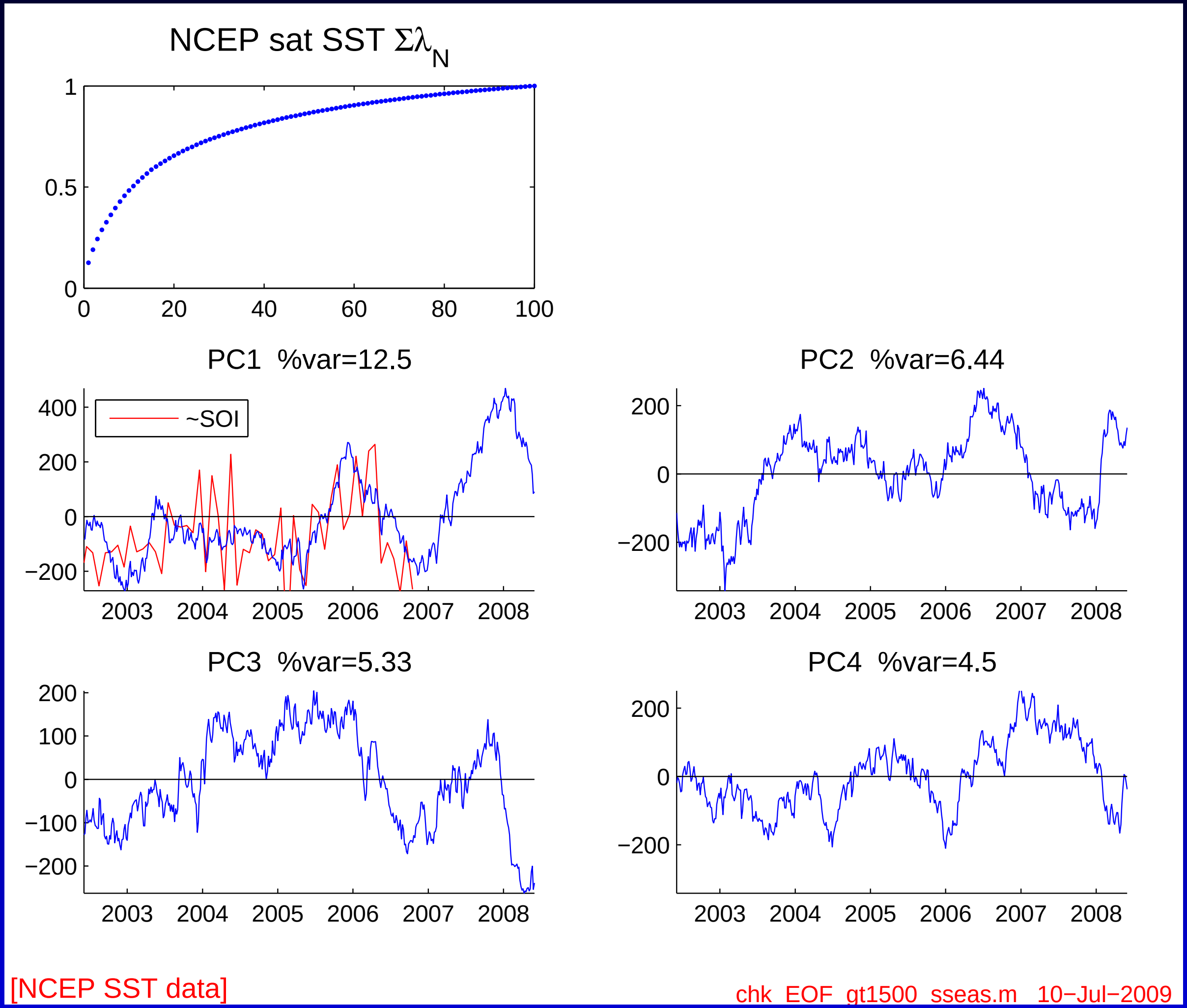


What Will Satellite Data Tell Us?

Now take out the seasonal cycle:



EOF Amplitudes



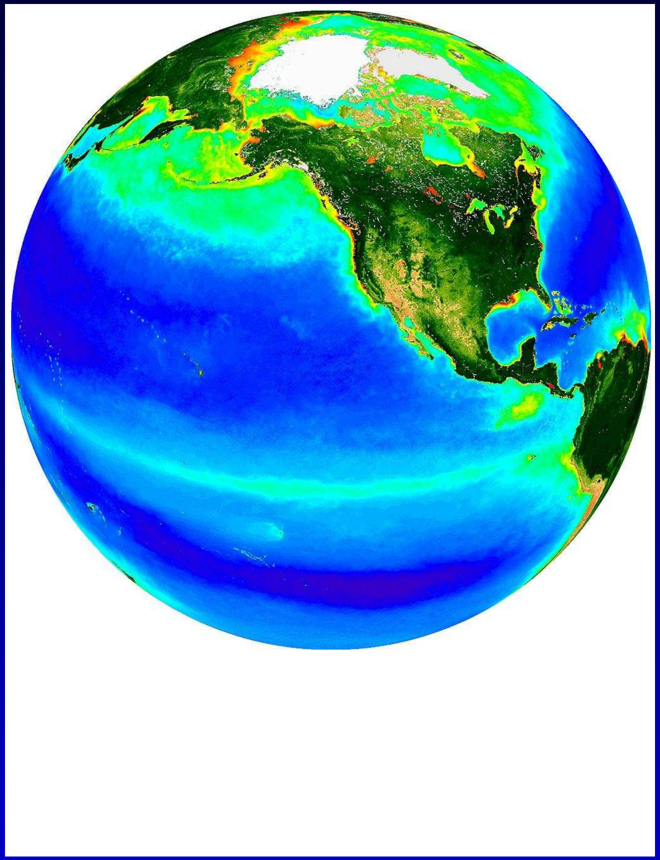
Satellite Observations: What are They?

Usual DA setup:

- Model state vector $\mathbf{x}$
- Modeled value of observed quantity is $H\mathbf{x}$
- What do we use for H?

Ocean Color

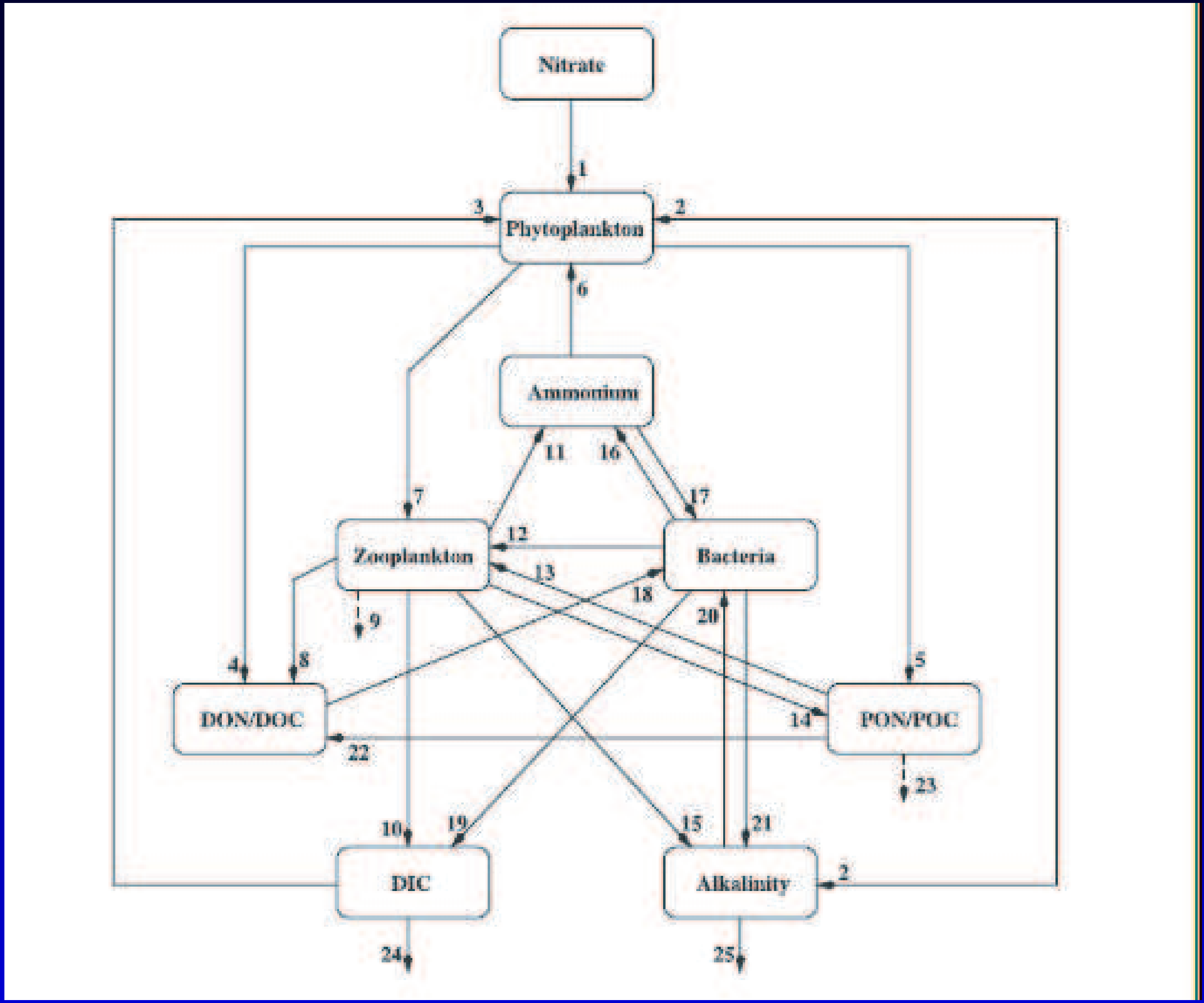
1247271251 (JPEG Image, 1024x1024 pixels) - Scal... http://oceancolor.gsfc.nasa.gov/cgi/biosphere_glob...



1 of 1 07/10/2009 06:19 PM

... how can we use it?

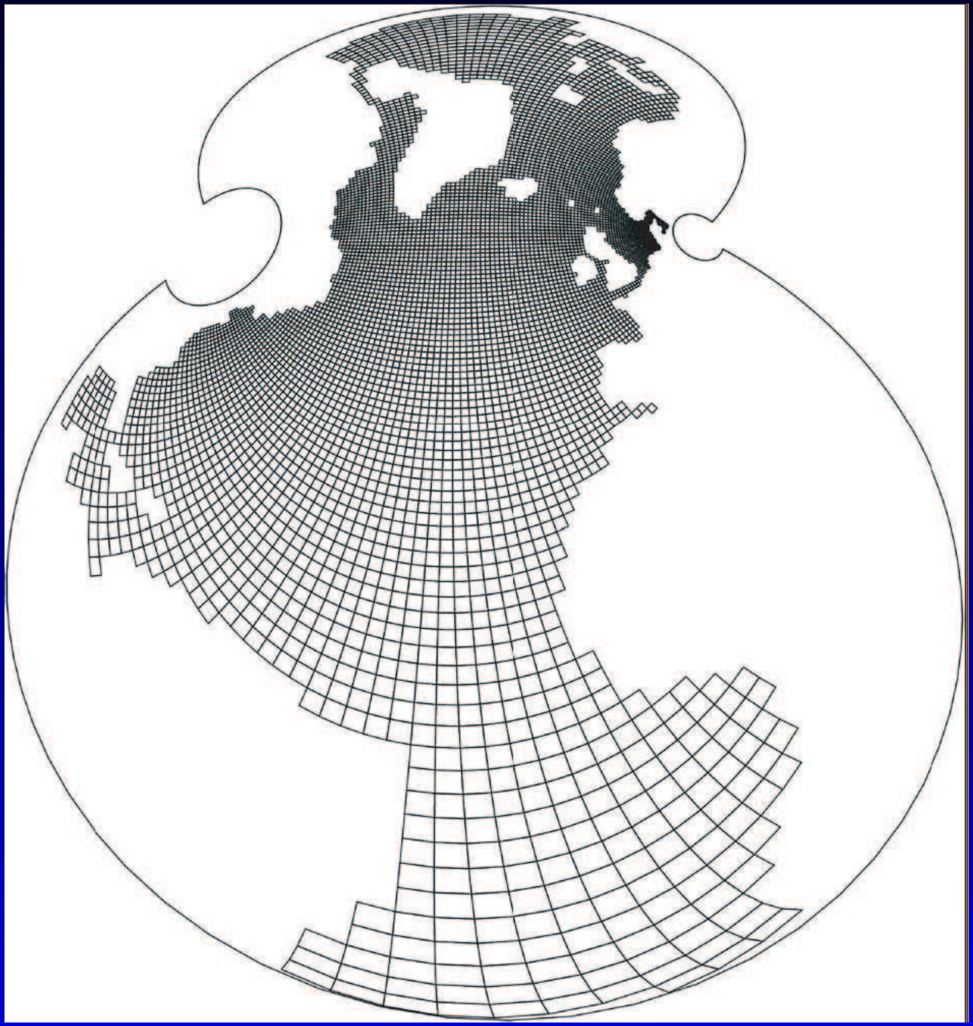
Ocean Ecology Models



The Physical Model

- A version of the Micom isopycnic model of the north Atlantic
- 17 layers
- Uppermost layer divided into two biochemical layers
- Curvilinear grid
- Biological model does not affect physical model

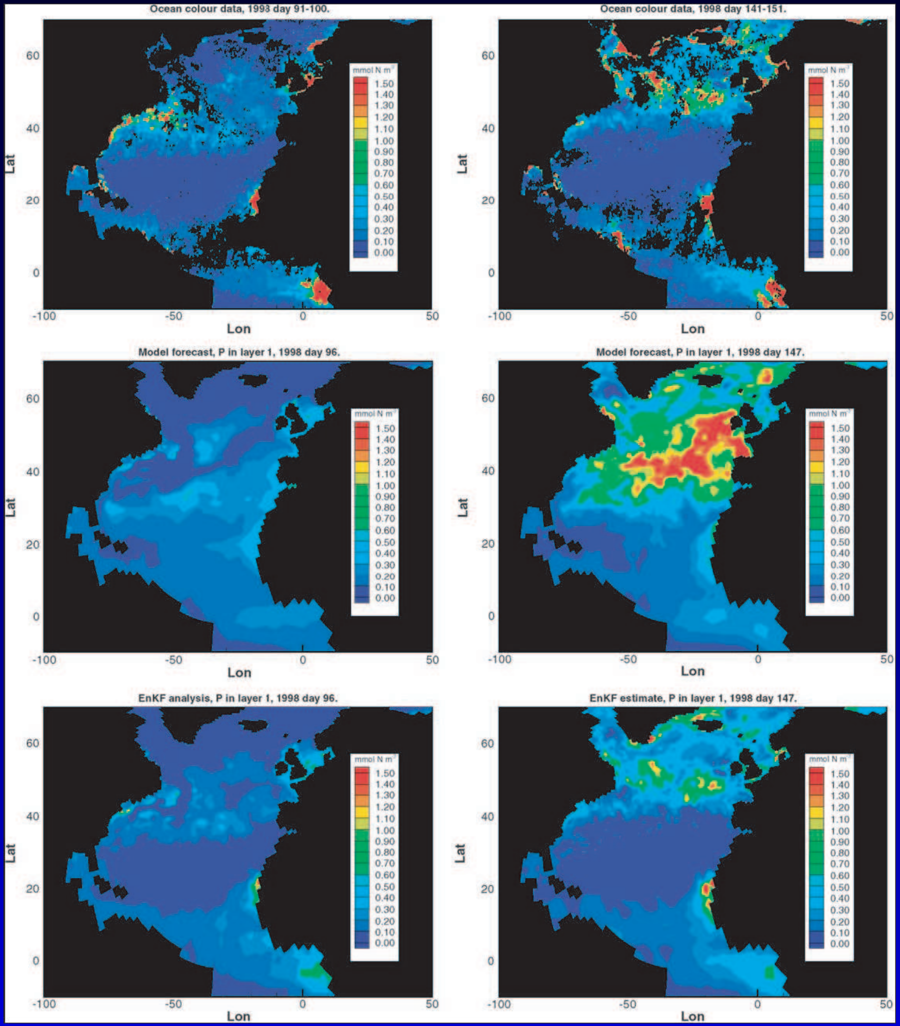
Grid for Coupled Model



The Ensemble Kalman Filter

- Initial conditions and forcing perturbed in physical model
- No perturbations added to biological state variables in order to avoid the possibility of negative concentrations
- Chlorophyll data converted to phytoplankton biomass by: $C = \rho_{max}(Chl_a/(Chl_a + K_{1/2}))Chl_a$. C relates to P by a fixed ratio.
- 100 ensemble members in the main experiment; similar results obtained with much smaller ensembles

EnKF Results from Coupled Model



Particle Filters

- Estimate the (possibly non-Gaussian) distribution with an ensemble
- Update the ensemble when data become available
- One way to do this is to re-sample, with replacement:
  1. For each member, calculate the probability that the member could have given rise to the actual observation
  2. Normalize the probabilities so that their sum is 1.
  3. Assign each member a sub-interval of the unit interval
  4. Generate a uniformly distributed random number between 0 and 1.

The SIR Filter

- Resampling, continued:
  1. Find the member's sub-interval into which the random number falls, and choose that ensemble member for the next interval
  2. Continue this process to fill the ensemble
- This is *Sequential Importance Resampling*

Particle Filters

- The SIR filter has been applied with some success in ocean modeling
- Like most particle filters it requires very large ensembles to avoid sample impoverishment

Representation Error

The big question:
What is truth?

Data Assimilation: Assumptions

Given

- A model: $\mathbf{u}_t - L\mathbf{u} = \mathbf{f}$
- Chosen to mimic the "*true*" state $\mathbf{u}^{(t)}$ which evolve according to:

$$\mathbf{u}_t^{(t)} - L\mathbf{u}^{(t)} = \mathbf{f} + \mathbf{b};\ \mathbf{b}\ \mathrm{r}andom$$

- Observations $\mathbf{z} = H\mathbf{u}^{(t)} + \mathbf{e}_{obs}$; $\mathbf{H}$ defines the relation between the state vector and the observed quantities
- Question for Today: What, *precisely*, is $\mathbf{u}^{(t)}$?

In Search of the *True* State

- The ocean measured by instruments doesn't know about physical approximations, coarse resolution or their consequences
- It is not subject to the limitations in computing power that restrict models to coarse resolution
- Measurements are not subject to the same requirements for approximate physical parameterizations

So ask: What quantity in nature is the "true" value of the model state?
No specific answers today; Rather a suggestion for what to do while we are waiting.

Representation Error

- Data assimilation makes use of data misfits, aka *innovations*: $\mathbf{z} - H\mathbf{u}^{(f)}$
- $\mathbf{u}^{(f)}$ is the forecast state
- Let $\tilde{\mathbf{u}}^t$ be the “true” ocean, as the instruments measure it.

Representation Error

Write the innovation:

$$
\begin{aligned}
\mathbf{z} - H\mathbf{u}^{(f)} &= \mathbf{z} - \mathbf{z}^{(t)} + \mathbf{z}^{(t)} - H\mathbf{u}^{(f)} \\
&= \epsilon^0 + \mathbf{H}(\tilde{\mathbf{u}}^{(t)} - \mathbf{u}^{(t)}) + \mathbf{H}(\mathbf{u}^{(t)} - \mathbf{u}^{(f)})
\end{aligned}
$$

- $\epsilon^0 = \mathbf{z} - \mathbf{z}^{(t)}$,the *instrument error*
- $\mathbf{H}(\tilde{\mathbf{u}}^{(t)} - \mathbf{u}^{(t)})$ is *representation error*
- Estimates of its statistics must appear in the terms reserved for instrument error
- $\mathbf{u}^{(t)} - \mathbf{u}^{(f)}$ is the *forecast error*

Estimating Representation Error

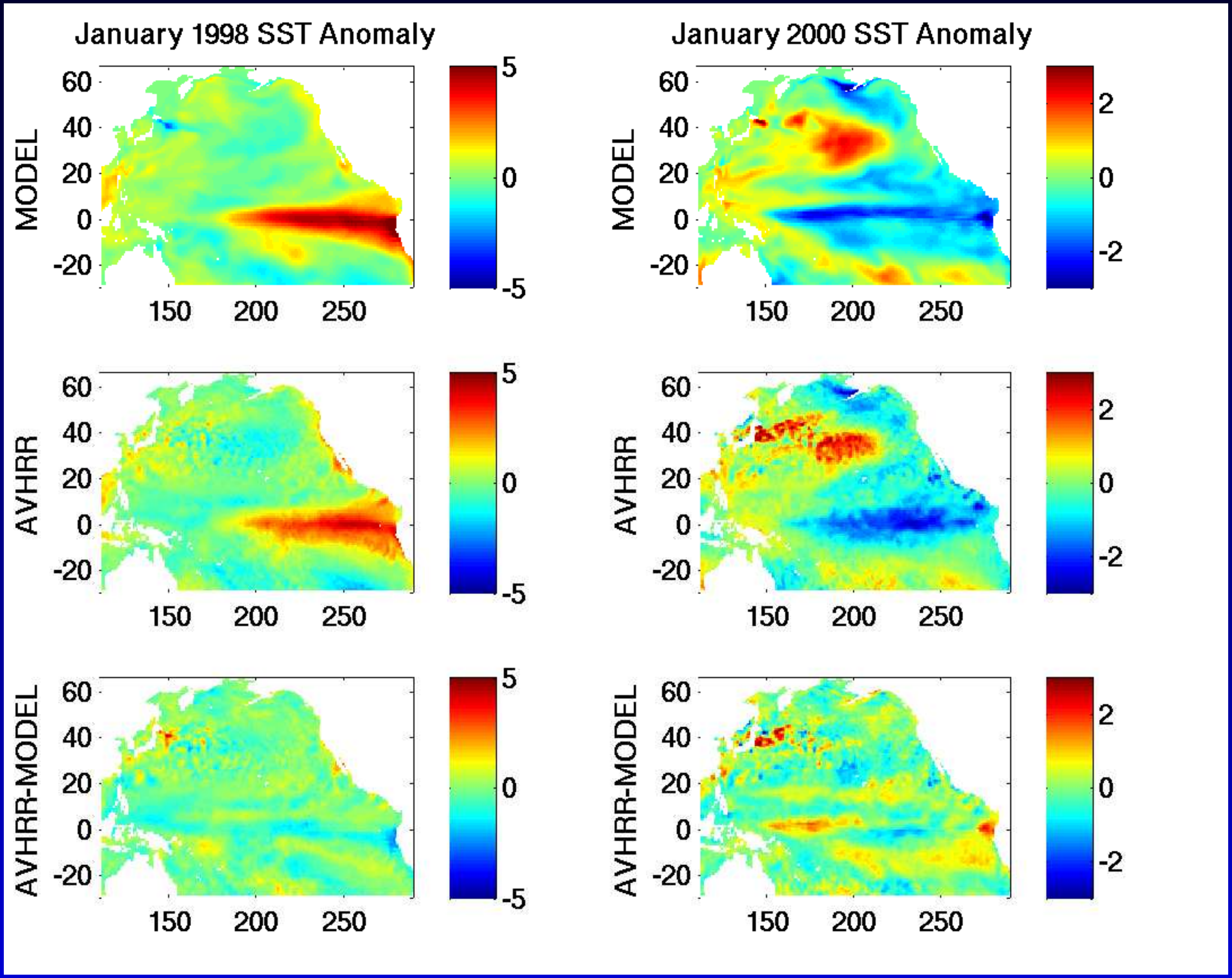
Our method for estimating the representation error for SST:

1. Generate a long model run
2. Calculate EOFs of the model run, considered as a matrix whose (i, j) element is the value of state element j at time i
3. Determine the number of meaningful degrees of freedom
4. Project the innovations on the meaningful singular vectors
5. Subtract the result from the innovations.
6. The difference is an estimate of the representation error

Pacific Circulation Model

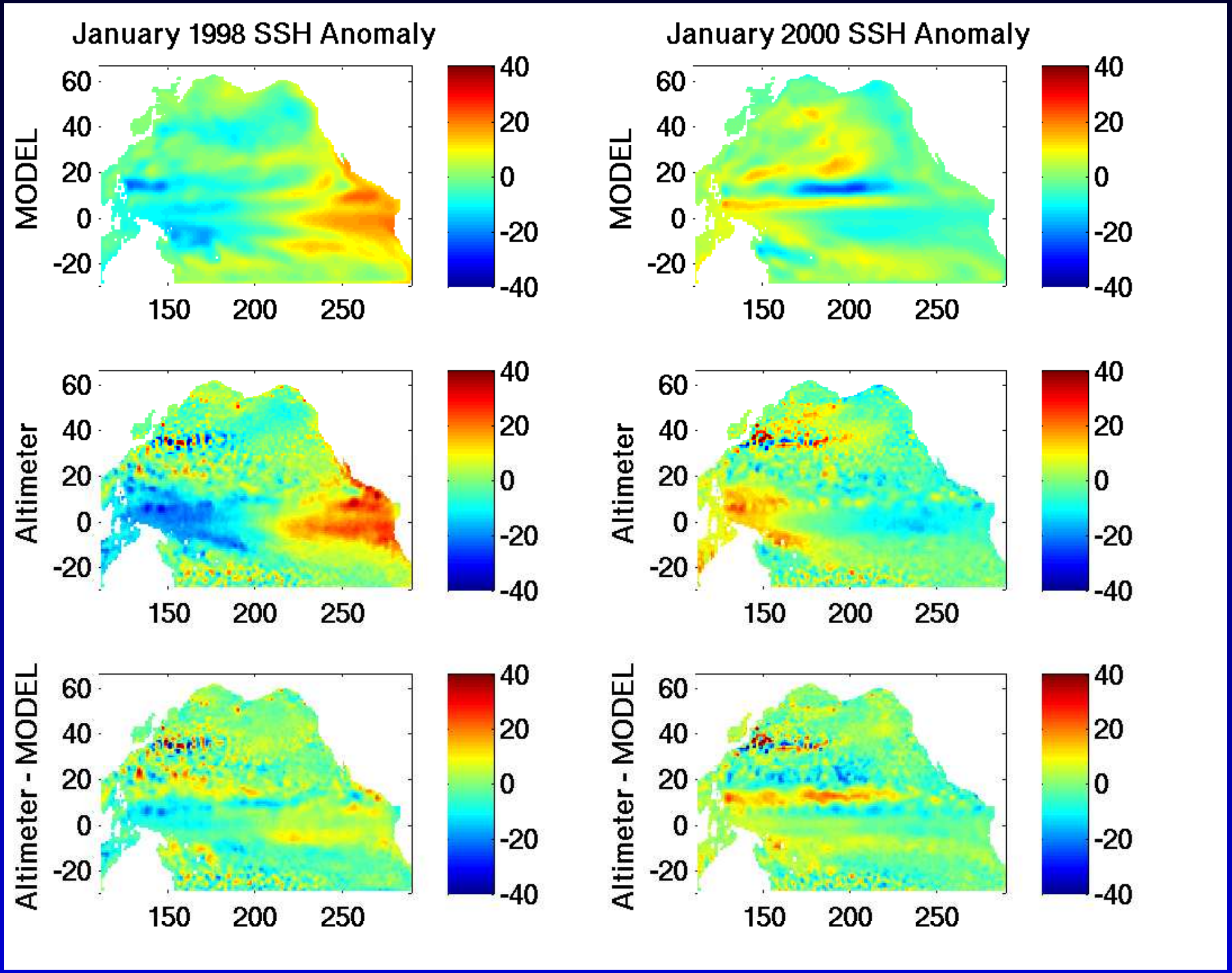
- Parallel Ocean Program (POP)
- Domain:
  - $105^o E$ to $85^o W$
  - $30^o S$ to $64^o N$
- Resolution
  - 1^o at the Equator, Mercator projection
  - 0.5^o average resolution
  - 50 vertical levels, 25 in top $500m$
- 25 years (1978-2002), forced by NCEP/NCAR reanalysis
- Initialized from Levitus, 30 year spinup

Model and AVHRR Anomalies



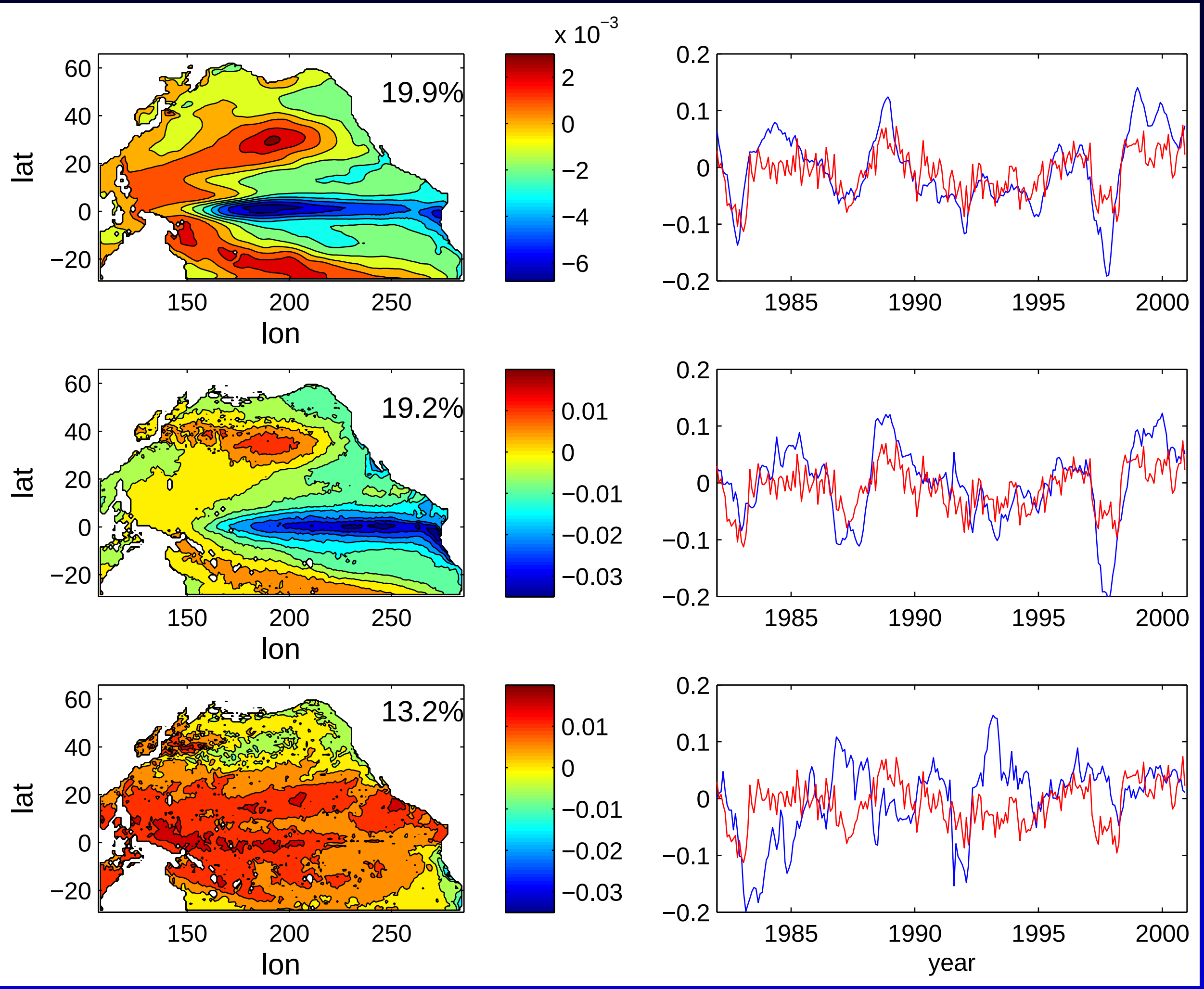
Model SST anomalies (top), AVHRR SST (middle) and data-model misfit (bottom) for 2 different years

Model and SSH Anomalies



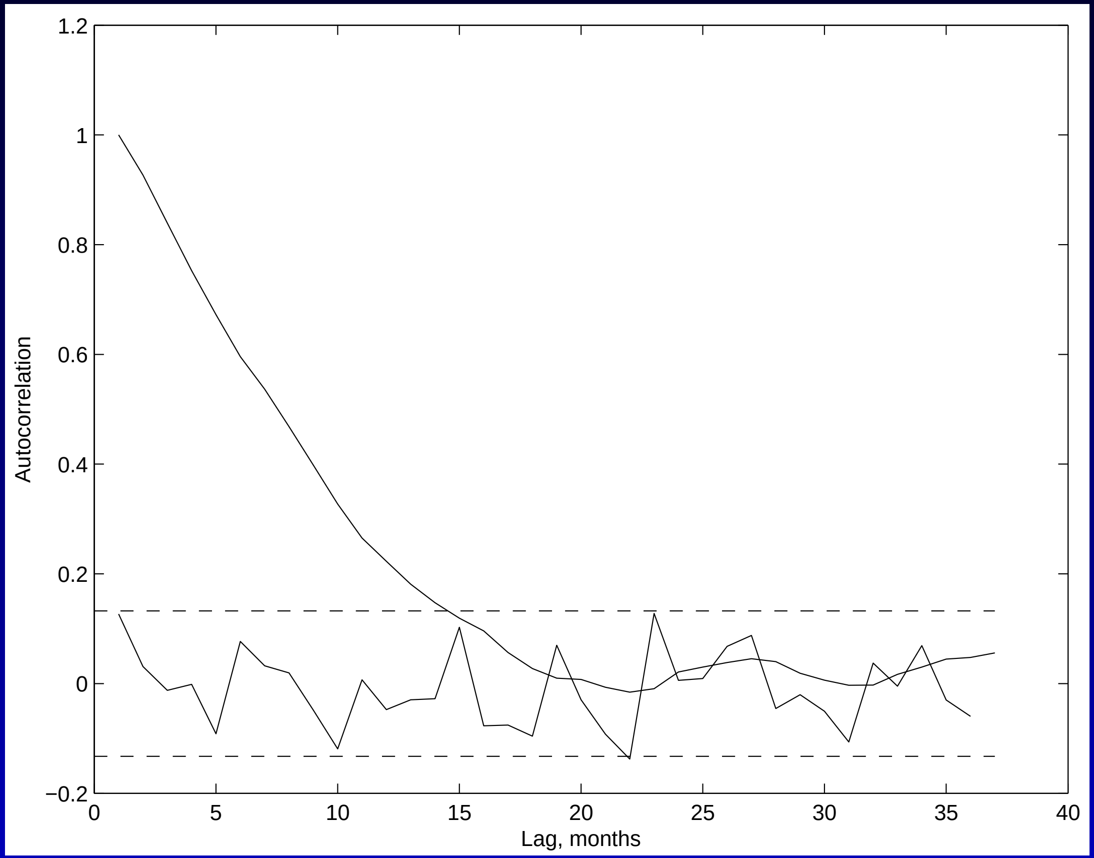
Model SSH anomaly (top) AVISO SSH (middle) and data-model misfit (bottom) for 2 different years

Leading EOFs



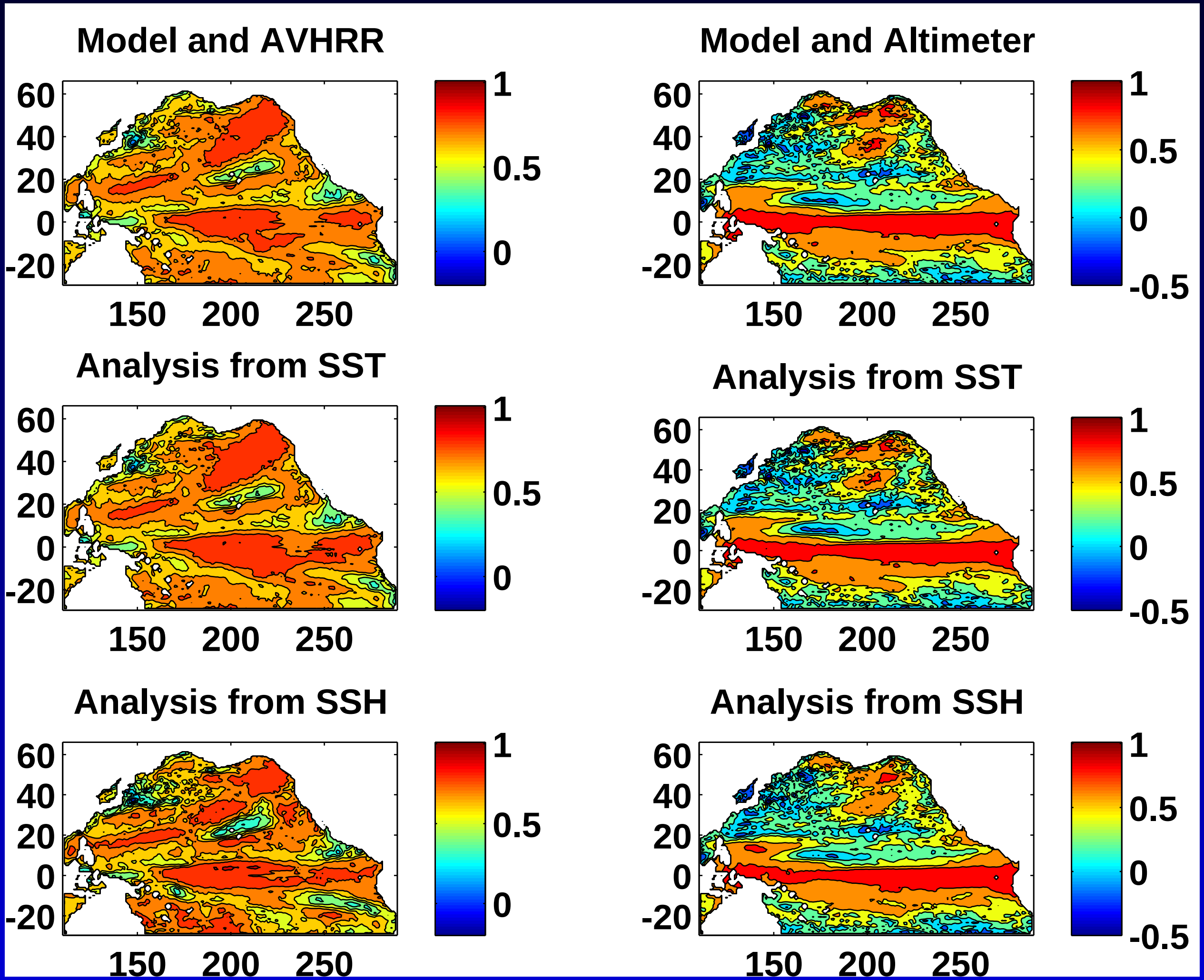
Leading EOFs and PCs for (top to bottom) Multivariate model SST; AVHRR SST; Misfit. Red = SOI.

Misfit EOF Time Series



Autocorrelations of time series of SST misfit PCs. Upper curve: lead PC; Lower curve: residuals from a fitted AR(1) process

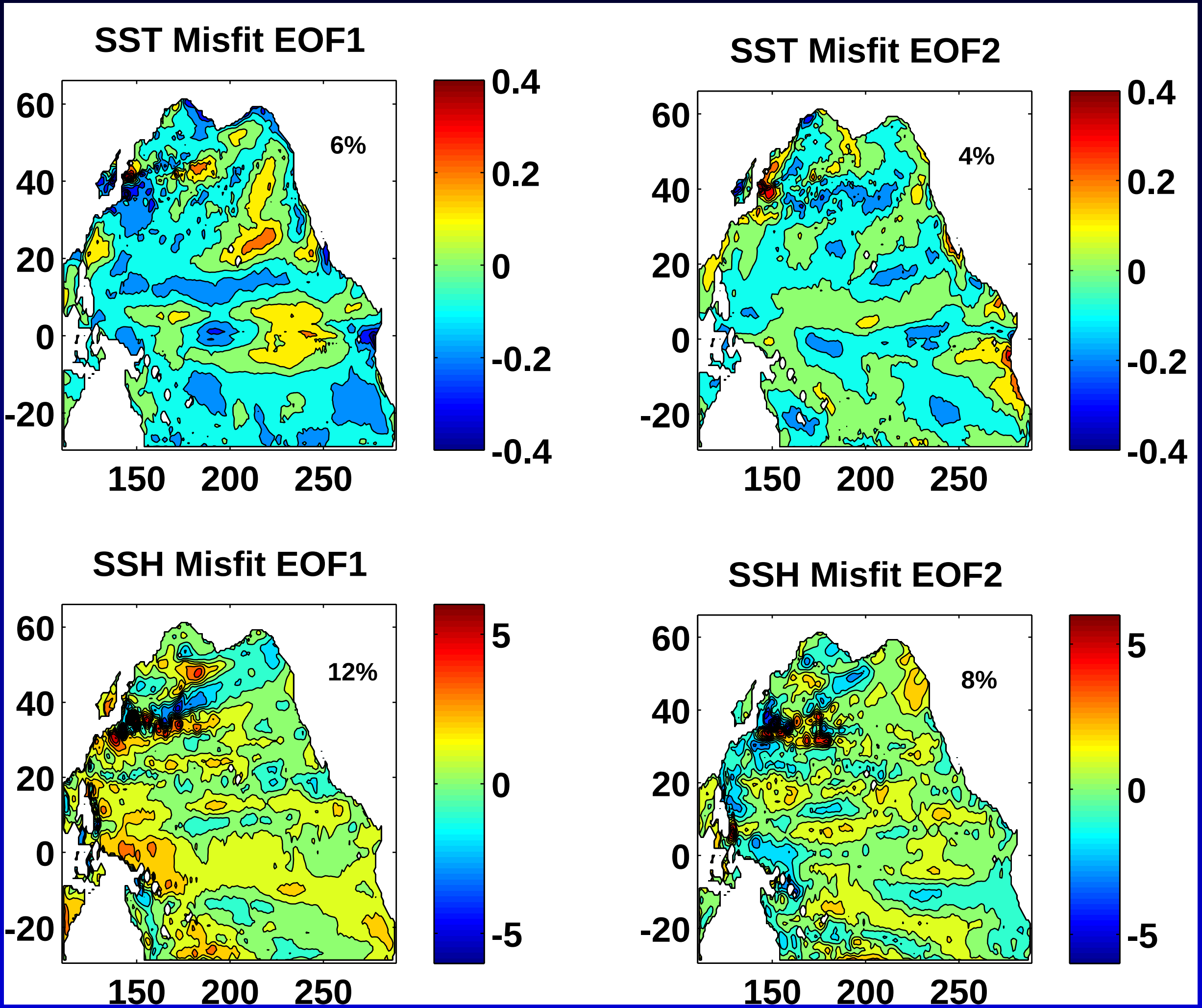
Data Assimilation Results



Correlation with observed SST (left) and SSH (right)

Top to bottom: No assim; assim of SST; SSH

Representation Error EOFs



EOFs of the SST and SSH representation error. Percentages of total orthogonal space variance are shown.

Summary of 1^o Pacific Model Results

- Lead EOFs of model and observed SST dominated by SOI
- Lead EOF of model-data misfit is well-modeled by an AR(1) process
- Estimates of representation error are as we expect
- Our estimates lead to ensembles with statistical properties of SST misfits

Summary

- New remotely sensed data sets will present new challenges
- Much satellite data comes in forms that are not easily represented in terms of dynamical variables
- Least-squares methods may be insufficient; we may have to work with explicitly nonlinear and/or Bayesian methods such as particle filters
- We will need to deal explicitly with representation error
- Satellite data sets provide new and great opportunities in interdisciplinary oceanography