Representation Error in Ocean Data Assimilation

Robert N. Miller
James G. Richman
Yvette H. Spitz
David Behringer

Outline

- Data assimilation works by using model-data misfits to correct the model state of the system
- The causes of some of the observed variability are not reflected in the model, and that portion of the observed variability cannot be usefully assimilated
- We propose a method for constructing statistical error estimates that account for representation error explicitly
- We describe progress toward implementing our methods within the framework of the operational climate forecast system.

Representation Error

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

Representation Error

Write the innovation:

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

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

Estimating Representation Error

1. Compute multivariate EOFs of a long model run
2. Estimate number of significant degrees of freedom (DOF) by the Preisendorfer test. The span of the significant DOF is the "model space"
3. Project a series of innovations into the model space. Assume: innovations - their projections on the model space = instrument error + representation error.
4. The significant EOFs of the the innovations - their projections into model space are assumed to span the space of representation errors.

Summary of Previous Results

- EOFs of model, observations and SST innovations were calculated for a 23 year run of POP, implemented at 1^o for the north Pacific.
- The model space had $O(35)$ dimensions, by the Preisendorfer test.
- Representation error accounted for much of the variability in the innovation sequence.
- An OI scheme was devised, based on projections of innovations into the model space.
- Statistics of estimated model SST variability + an ensemble of simulated representation errors were identical to statistics of the SST data.

Error for the NCEP CFS

Analyze output of 16 year run of ocean component of NCEP CFS:

- MOM4, 0.5^o driven by NCEP reanalysis
- 5 day averages of 16 year run, 1993-2008 incl.

Getting Started

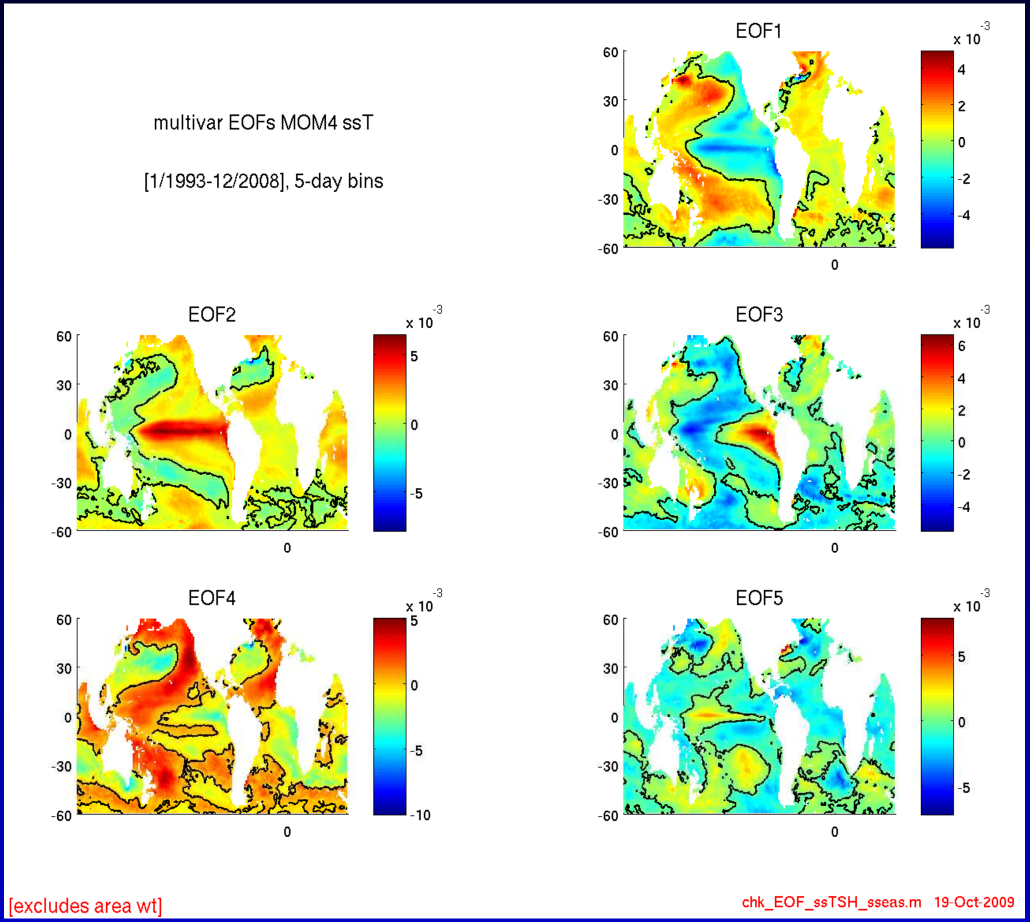
Restrict attention to:

- $60^o S$ - $60^o N$
- Consider only upper $1500m$ and points where depth $\geq 1500m$
- Remove the seasonal cycle

...still an enormous calculation:

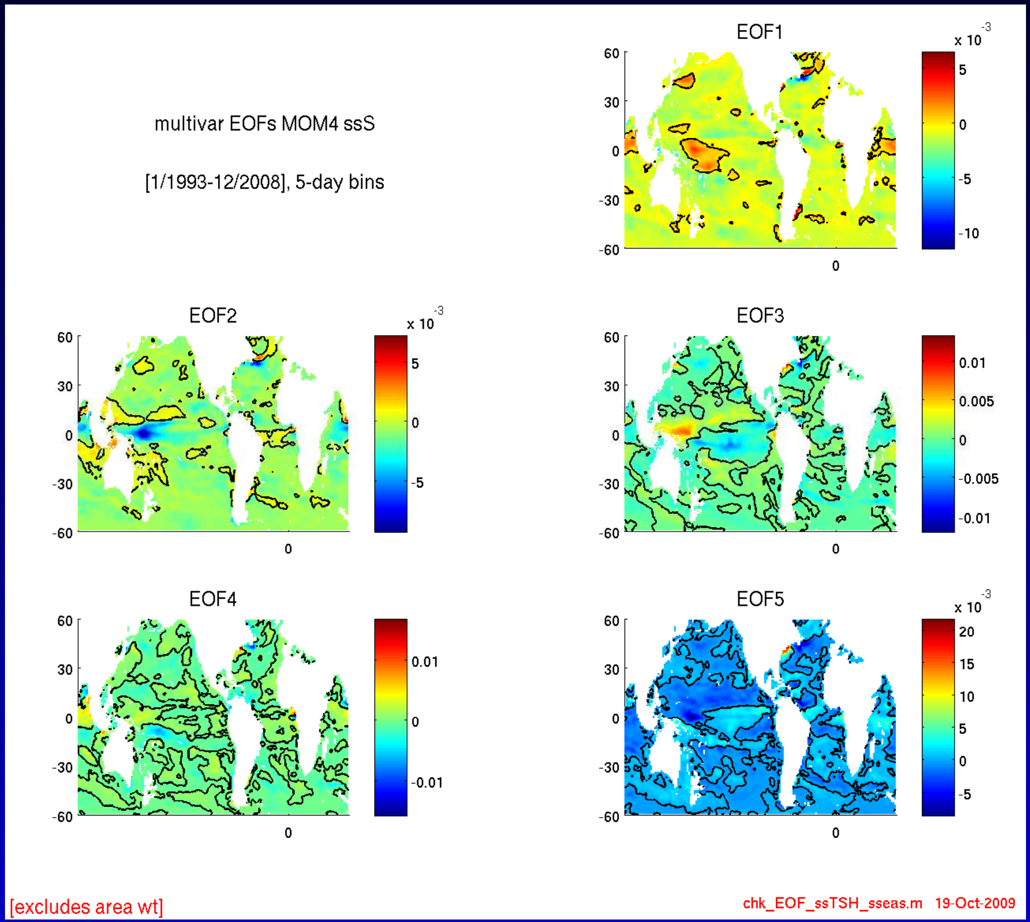
- $\approx 150K$ points horizontal, 33 vertical levels
- 3D T,S,U,V and 2D H

EOFs: Surface Variablity



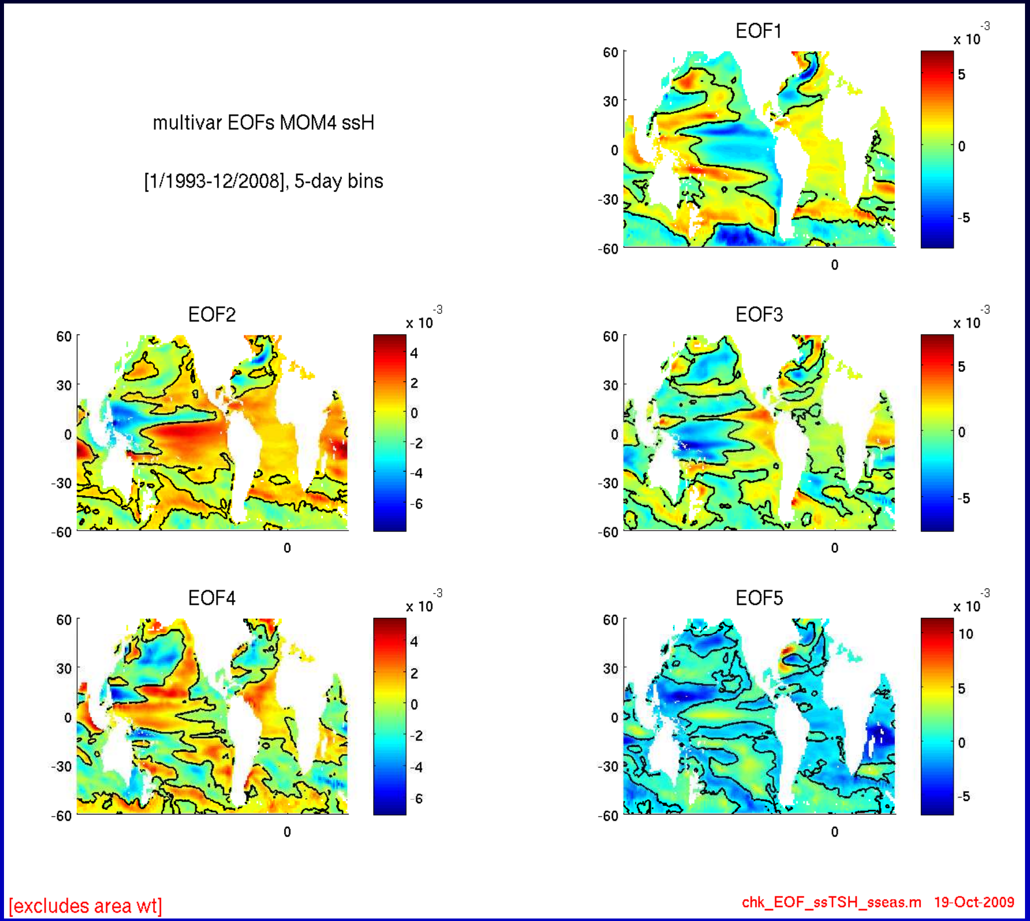
Temperature maps from multivariate EOF

EOFs: Surface Variablity



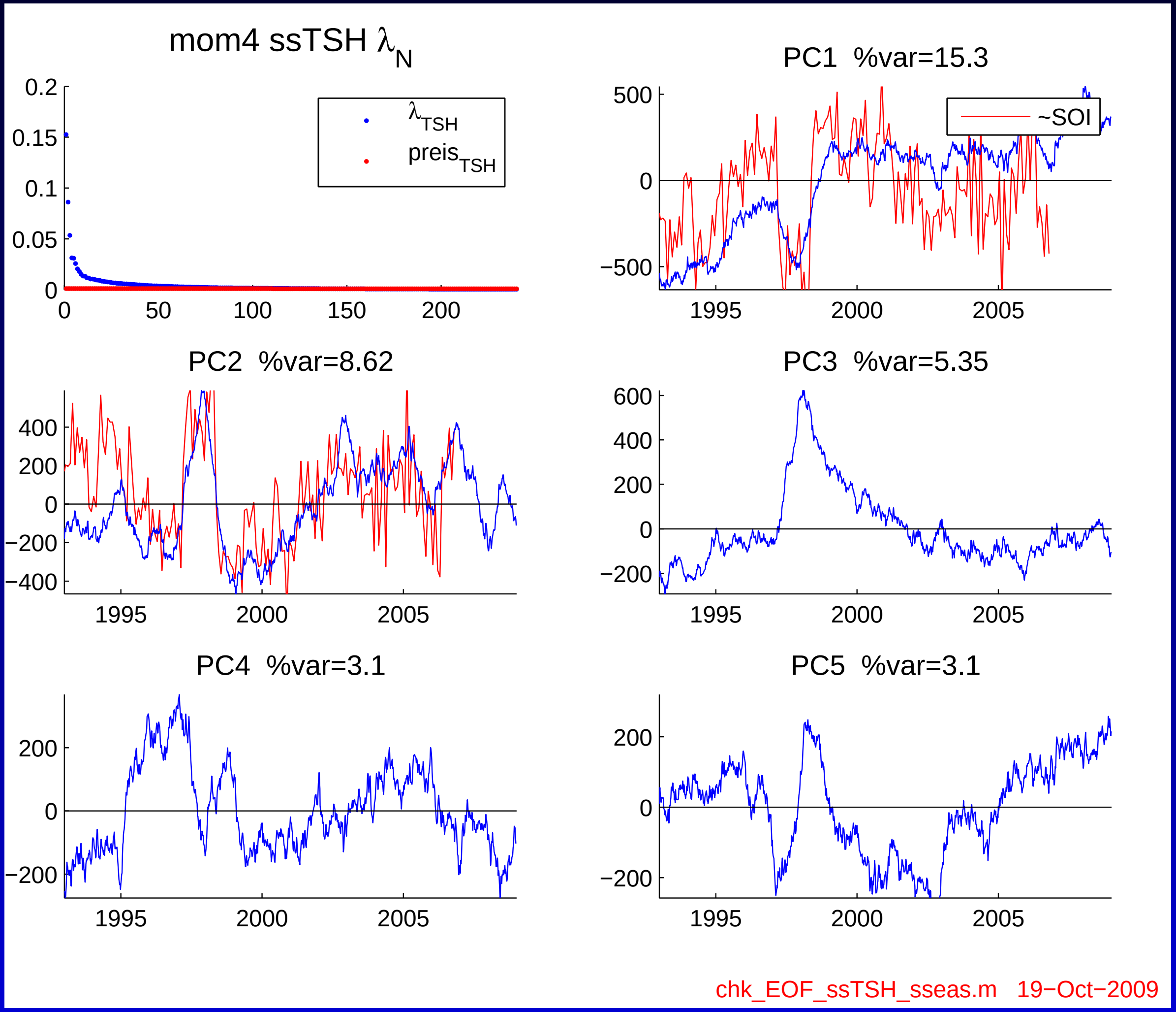
Salinity maps from multivariate EOF

EOFs: Surface Variablity



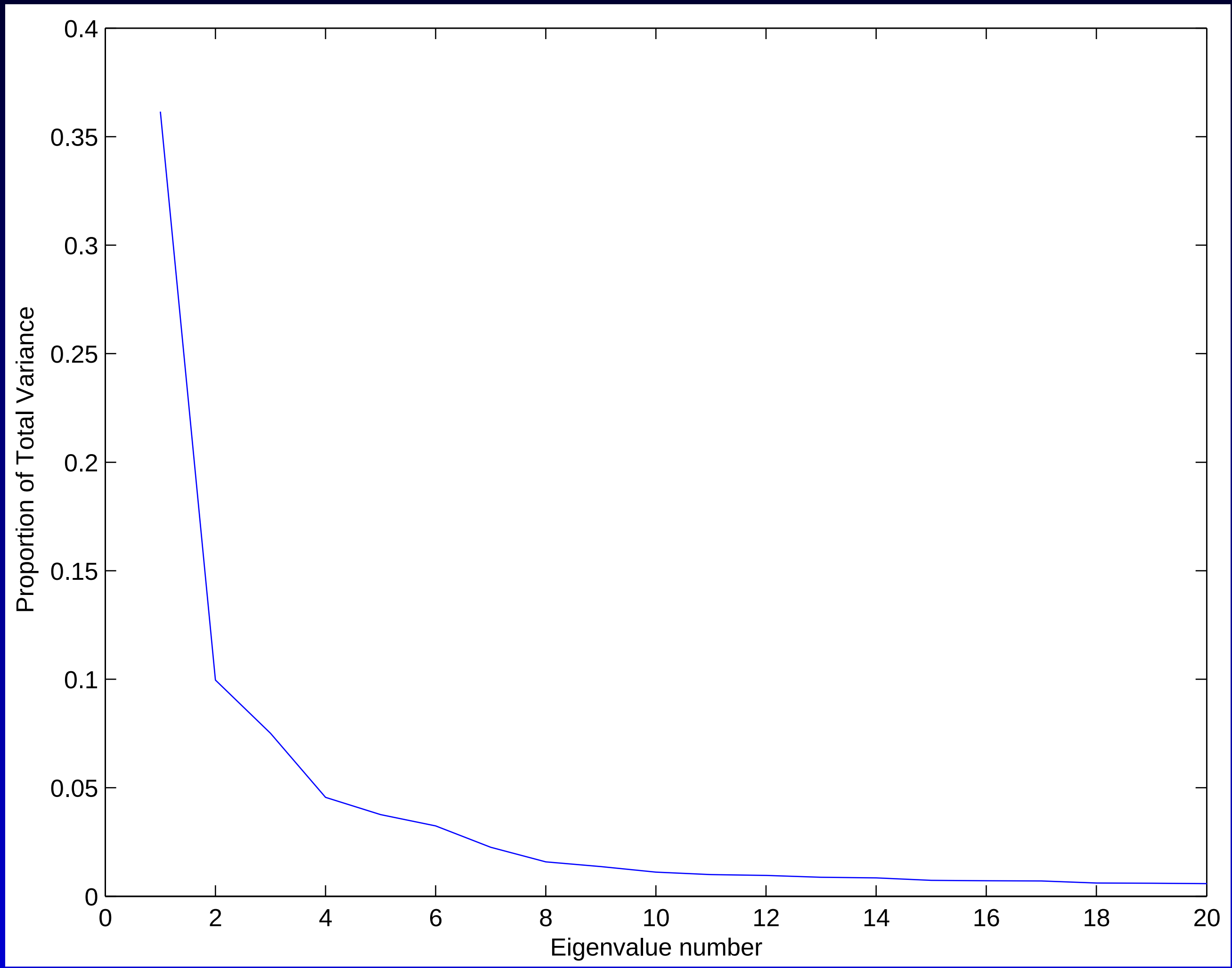
Height anomaly maps from multivariate EOF

PCs: Surface Variablity



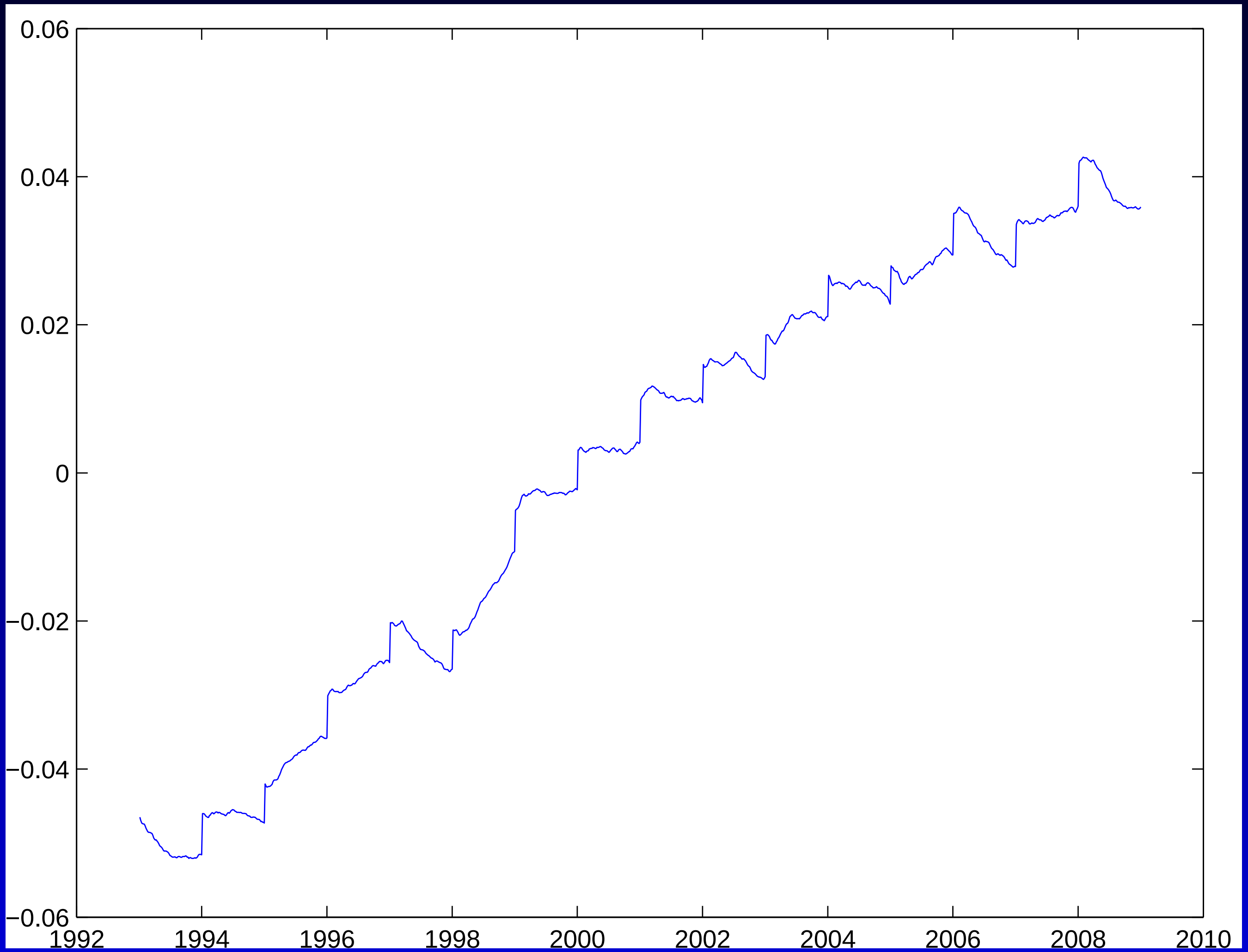
Blue curves: PCs. Red curves: SOI

Characteristic Values: 3D Temperature Variability



Eigenvalues of 3D T covariance, normalized

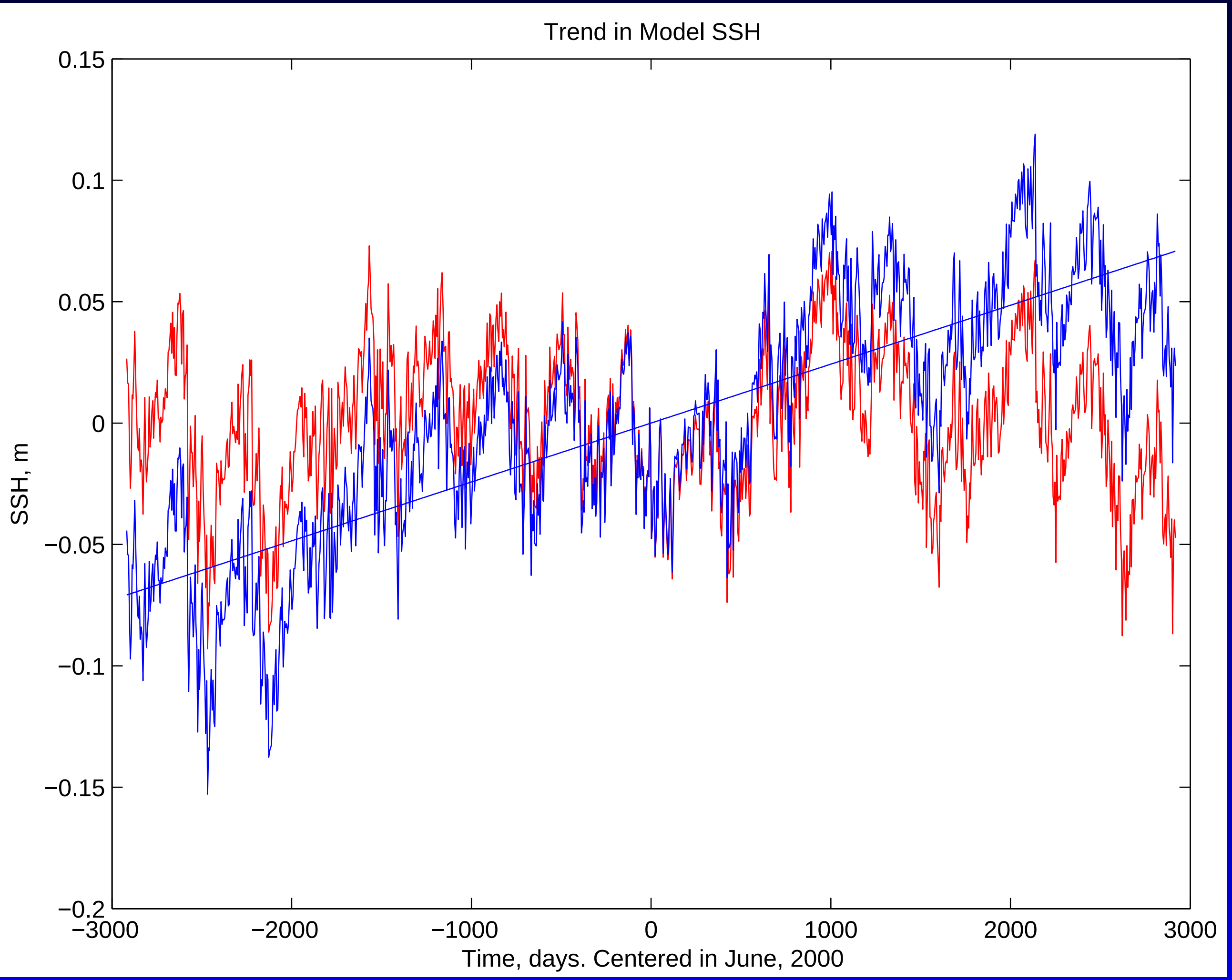
PCs: 3D Temperature Variability



Lead PC, 3D temperature covariance

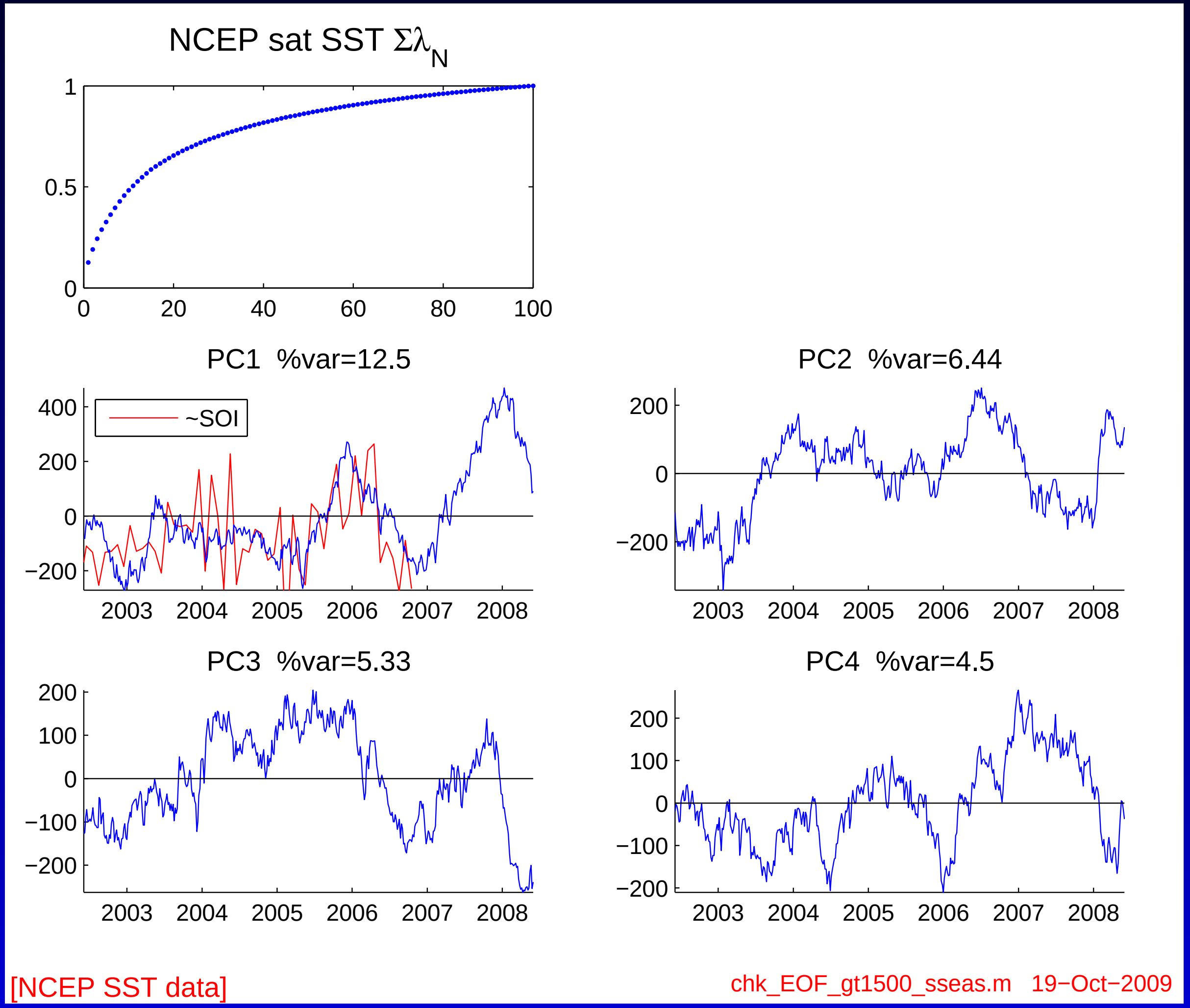
Model Sea Surface Height

...somewhere in the north Atlantic



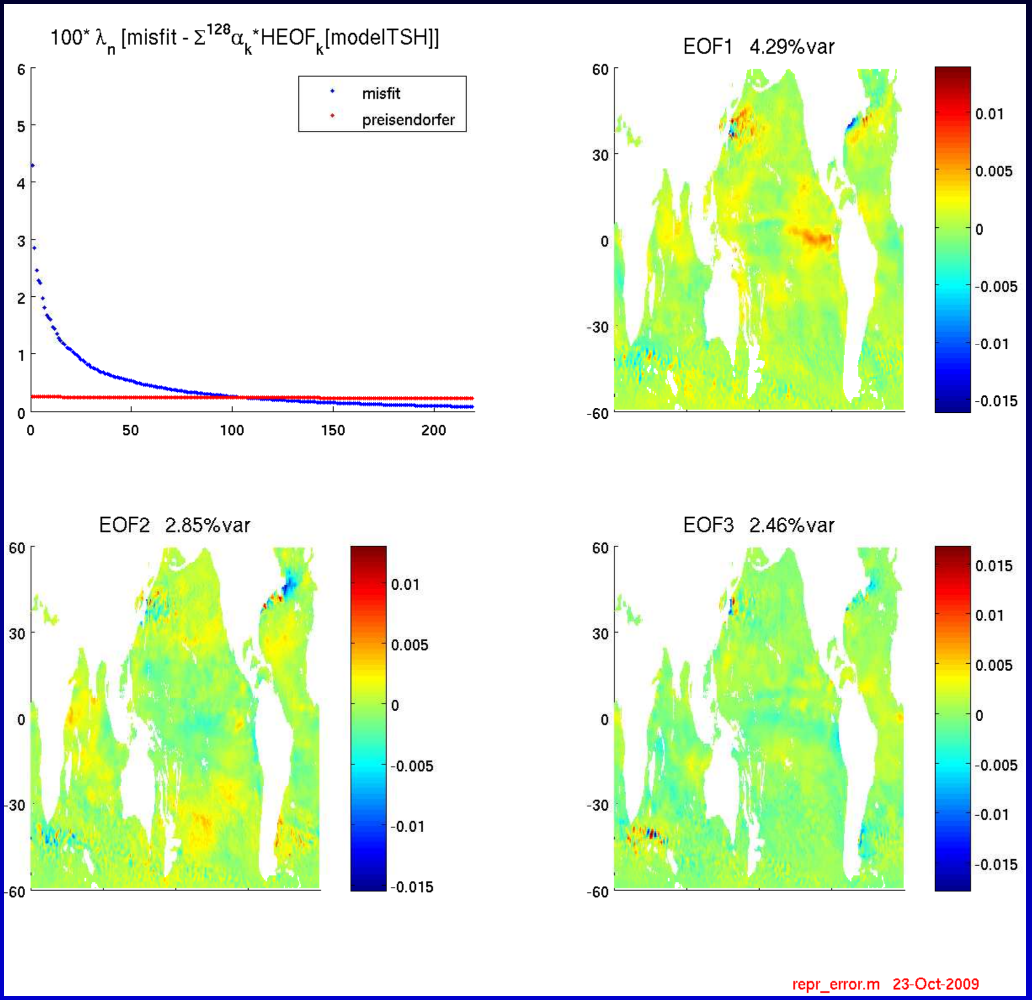
blue curve: SSH; blue line: trend, $\approx 1cm/yr$; red curve: detrended SSH

Was there a trend in the data?



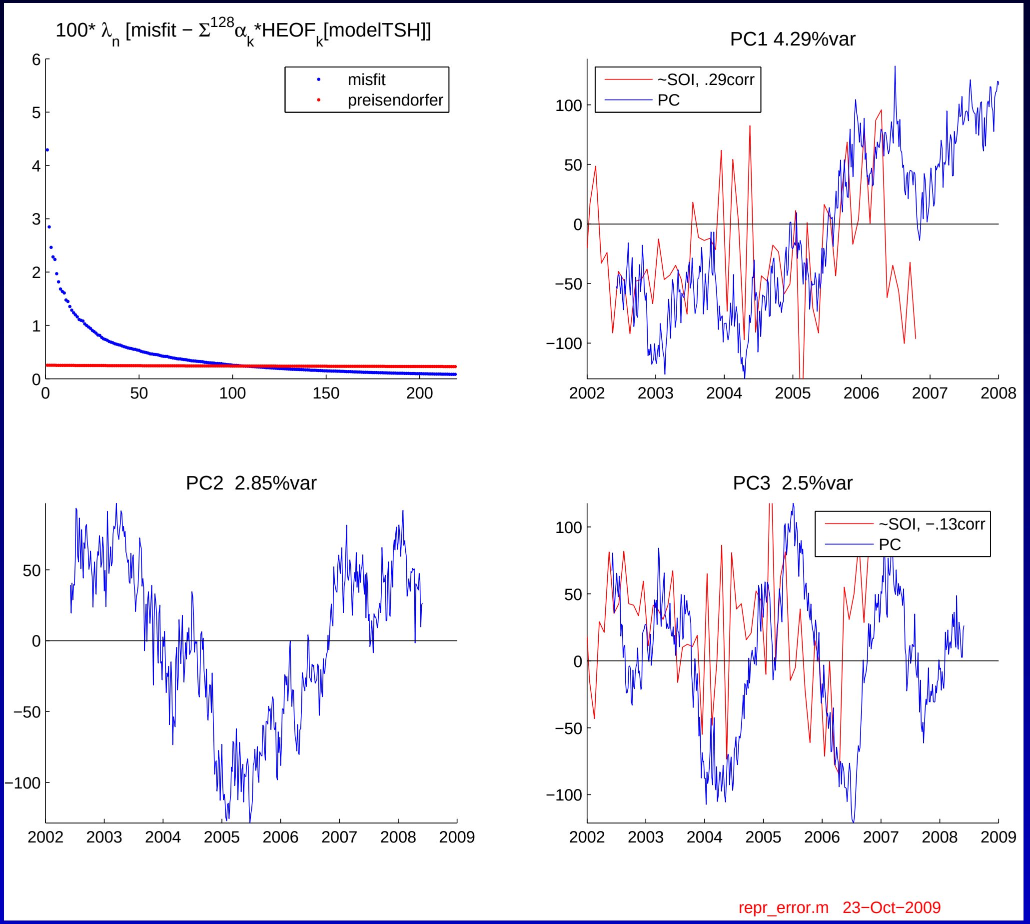
...nope. No trend in 23 year SST time series either.

Representation Error EOF



Upper left: Preisendorfer test; upper right and bottom row: EOF loadings.

SST Representation Error PC



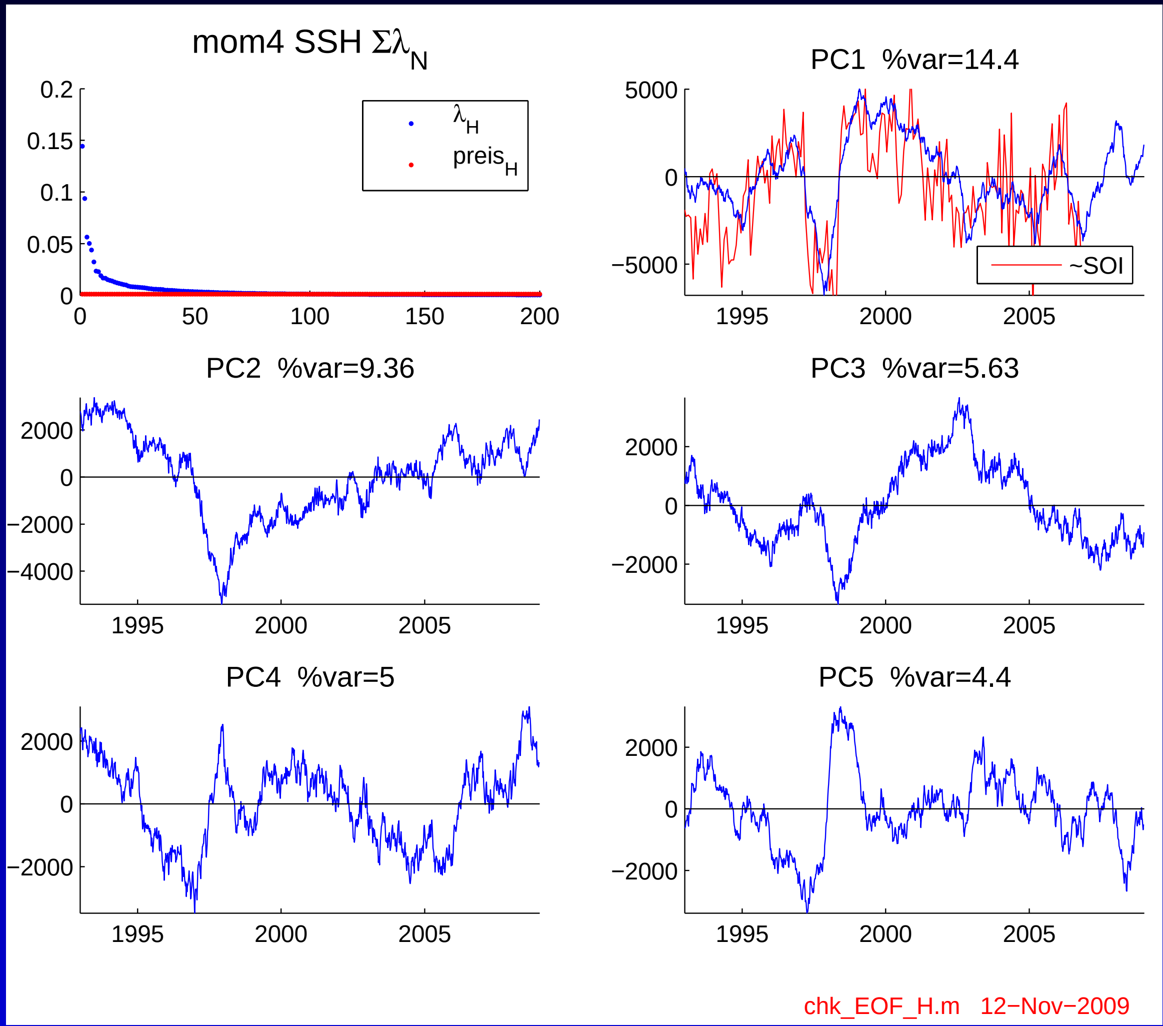
Upper left: Preisendorfer test. Red curves on PC1 & PC3 plots: scaled SOI

So where did a trend come from?

...probably an internal adjustment process.

- Initial condition is the result of CFS data assimilation; initial state unbalanced
- Model is adjusting according to its own internal physics
- Is it “right?” No way to know, given available data
- How long will adjustment take?
- What do we do about it?
  - For now, remove a least-squares trend

Model Sea Surface Height PC



Univariate PC, mean, trend and seasonal cycle removed. trend $\approx 1cm/yr$

The Plan

- Assimilate data by 3DVAR:

$$
\begin{aligned}
J &= \frac{1}{2}(\mathbf{x} - \mathbf{x}^{(b)})^T E^{-1}(\mathbf{x} - \mathbf{x}^{(b)}) + \\
&\quad \frac{1}{2}(z - H\mathbf{x})^T F^{-1}(z - H\mathbf{x}) \\
\delta J &= \delta\mathbf{x}^T \left(E^{-1}(\mathbf{x} - \mathbf{x}^{(b)}) - H^T F^{-1}(z - H\mathbf{x}) \right) \\
\nabla J &= E^{-1}(\mathbf{x} - \mathbf{x}^{(b)}) - H^T F^{-1}(z - H\mathbf{x})
\end{aligned}
$$

- Obs error covariance F appears only in the term $H^T F^{-1}(z - H\mathbf{x})$
- Write $F = D + uu^T$ where D is diagonal and columns of u span representation error space.

Effect of Augmentation of F

- Calculation of F^{-1} will be inexpensive.
- Let $F = \sigma^2 I + uu^T$, σ^2 the obs error variance and u a vector with $u^T u = r^2 > \sigma^2$. The Sherman-Morrison formula:

$$(\sigma^2 I + uu^T)^{-1} \; = \; \frac{1}{\sigma^2}\left(I - \frac{uu^T}{\sigma^2 + u^T u}\right)$$

- Multiplying the innovation by F^{-1} will thus have the effect of damping the component parallel to u by a factor of $\sigma^2/(r^2 + \sigma^2) << 1$

The Way Forward

1. Implement CFS data assimilation system with augmented F
2. Does this help? Greatest effect should be less shock to the system by assimilation, and smaller initialization transients
3. Construct model error covariance from projection of innovations on full 3D model space.
4. Use the result to construct an OI scheme, an ensemble scheme or both.

The Way Forward ...

1. Augment the ocean state with an ensemble of simulated representation error time series and study the effect on the model atmosphere
2. Construct an ensemble of coupled model outputs and study its properties

...but for now we'll be happy to see the results from the augmented obs error covariance in the next few weeks.