Model Representation Error Estimation for Ocean Data Assimilation

James G. Richman
Ocean Dynamics and Prediction Branch (Code 7323)
Naval Research Laboratory
Stennis Space Center, MS 39529

and

Robert N. Miller
College of Oceanic and Atmospheric Sciences
Oregon State University
Corvallis, OR 97331

Presented at
National Center for Environmental Prediction
Camp Springs, MD
November 4, 2009

Representation Error

- Data assimilation maps the difference between a model simulation and observations into the model state space
- We estimate the present state of the system using the observations with a filter

$$x^a = x^f + P^f H^T (HP^f H^T + R)^{-1} (y^o - Hx^f) \quad (1)$$

P^f is forecast error covariance

x^f is the model forecast

$\varepsilon^f = x^t - x^f$ is the forecast error

x^a is the model analysis

y^o is the observation

x^t is the “true” model

R is the observation error covariance

H is the mapping operator which maps the model state to the observation space

Observation Error

The model-data misfit, innovation. can be decomposed into instrument error, forecast error and representation error in the following way:

$$y^o - Hx^f = (y^o - y^t) + (y^t - Hx^t) + (Hx^t - Hx^f) \qquad (2)$$

$$= \quad \varepsilon^o + \varepsilon^R + H\varepsilon^f \qquad (3)$$

y^t is the "true" value of the observed quantity. The three terms on the right hand side of (3) are the instrument error, the representation error and the forecast error mapped into observation space

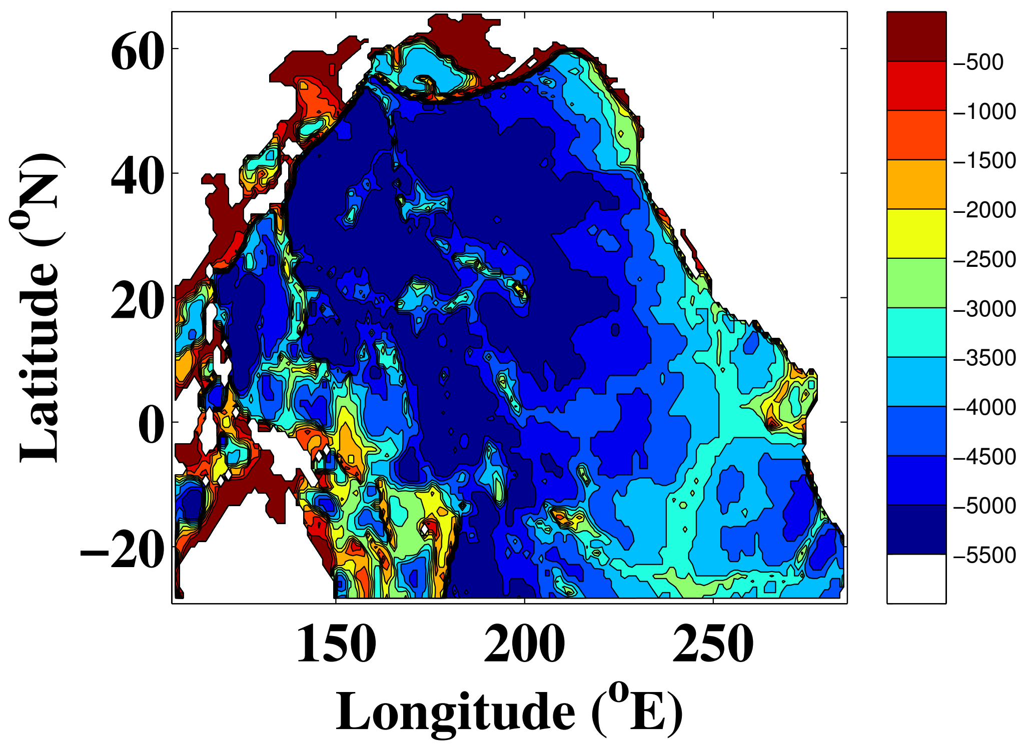
Representation Error Outline

- Results from North Pacific Climate Model
- information content and Reduced state space filter design
- Definition of Observation Error Subspace and estimation of Observation Error covariance
- Posterior statistical analysis of representation error
- Preliminary results for the ocean component of the Climate Forecast System

North Pacific Circulation Model

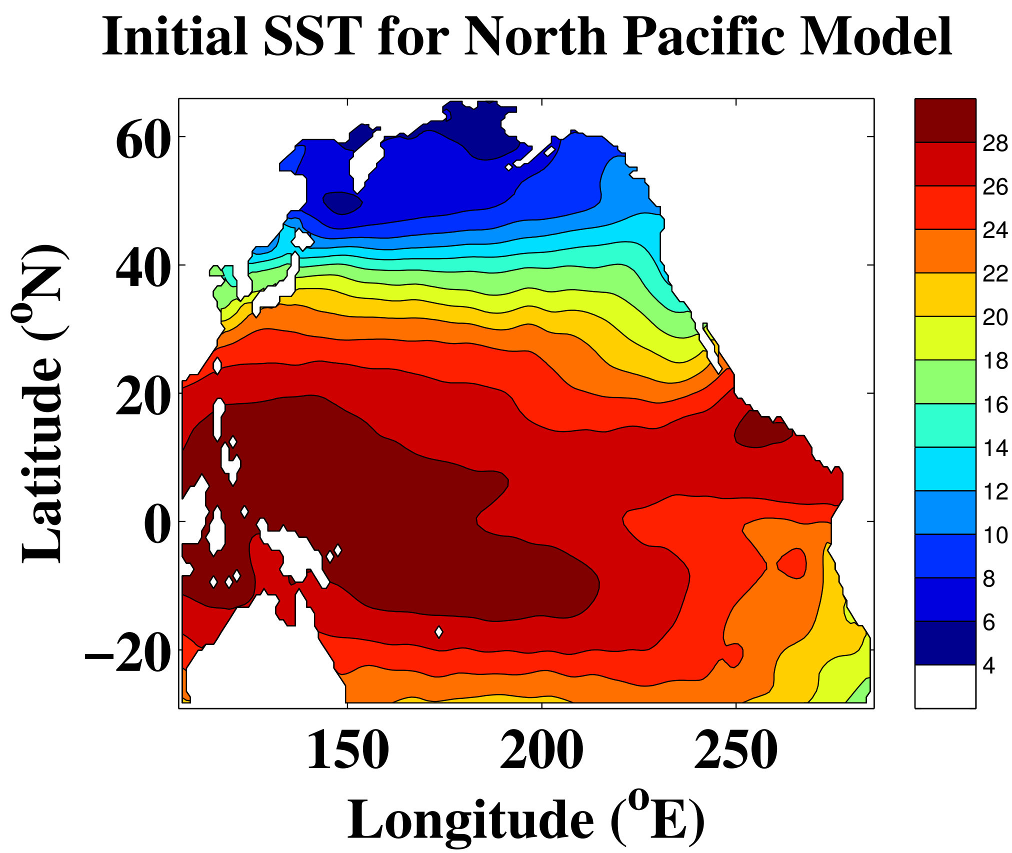
- Model:
  Parallel Ocean Program (POP) model
- Domain:
  105° E to 85° W
  30° S to 64° N
- Resolution:
  - **1° at Equator on mercator projection**
  - **0.5° average resolution**
  - **50 vertical levels with 25 in top 500 m**

Bathymetry of North Pacific Model

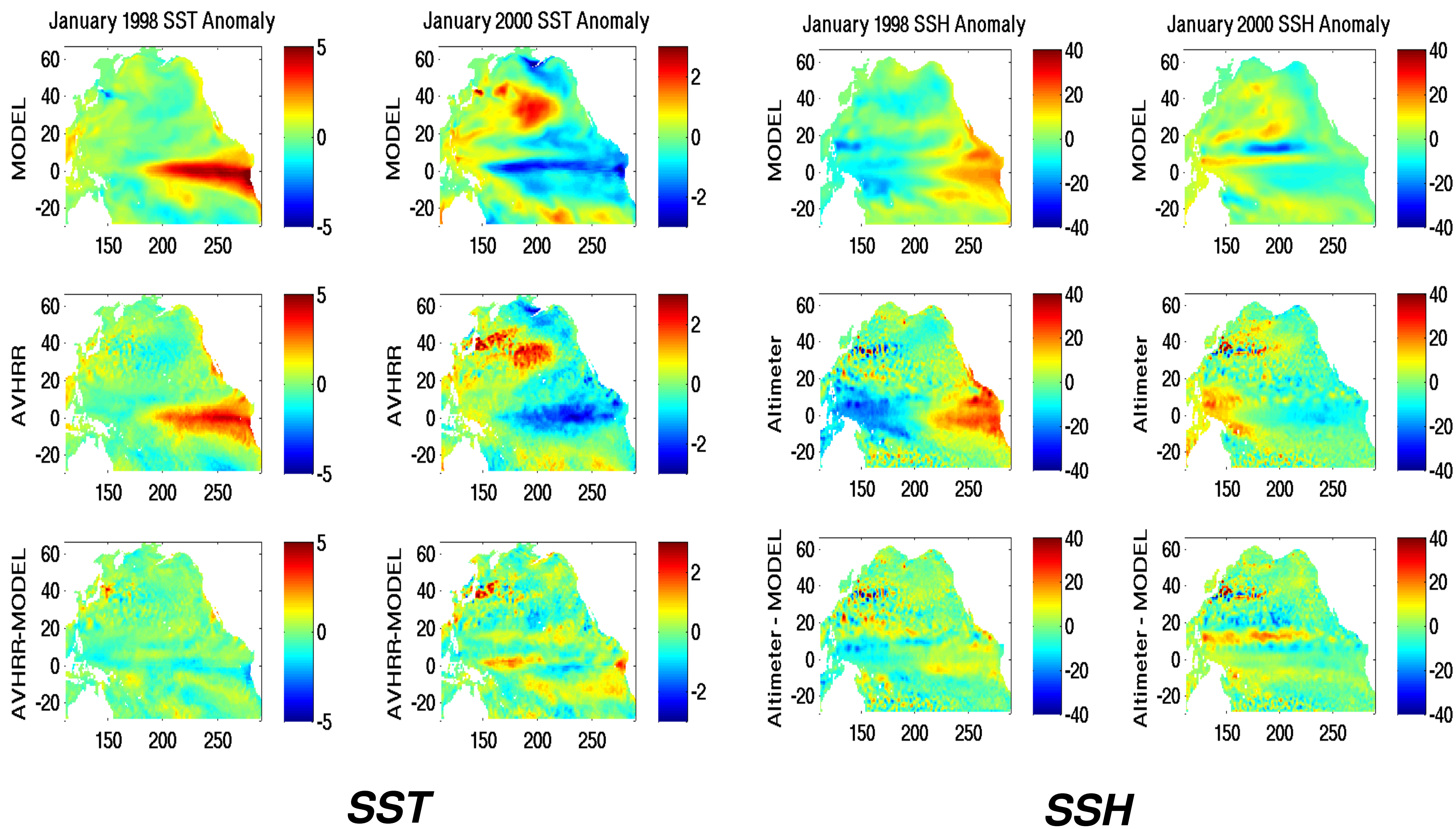


North Pacific Upper Ocean Model

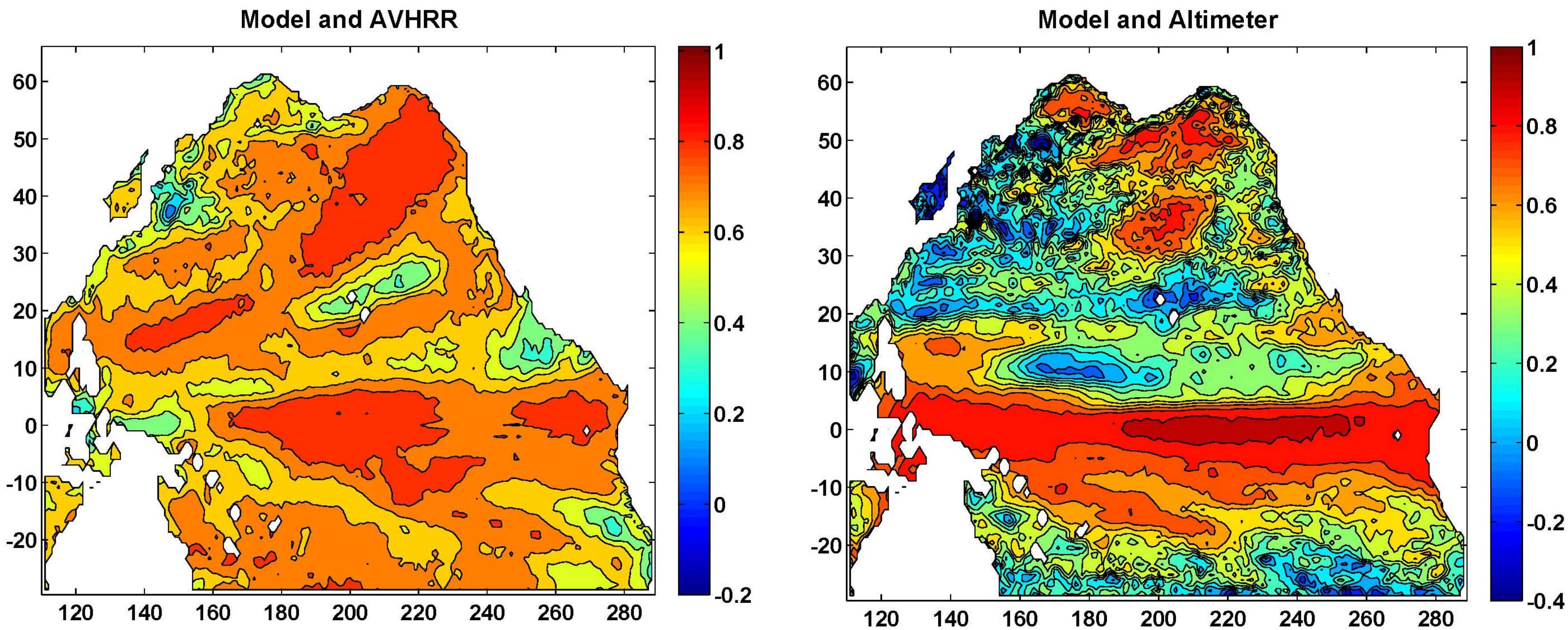
- Model initialized from Levitus WOA98 temperature and salinity
- 26 years (1979 thru 2004) of NCEP/DOE Reanalysis Fields are used to force the model
- Model is restored to the WOA98 surface salinity with 30 day restoration time
- Mixing is the upper ocean with the KPP mixed layer model of Large et al (1994)



Model and Data Comparison for a non-El Nino year (Jan 1996) and an El Nino year (Jan 1998)



Correlation between model forecast and the remotely sensed SST and Sea Level Observations



Large number of state variables prohibits solving the full system

→ ***Reduced State Space Kalman Filter***

1) **Compute the multivariate empirical orthogonal functions (EOF's) of our 26 year time series of deviations from the seasonal cycle,**

2) A statistical test is performed in order to estimate the number of significant degrees of freedom. (Preisendorfer, 1988) (35 modes accounting for 59% of the total variance)

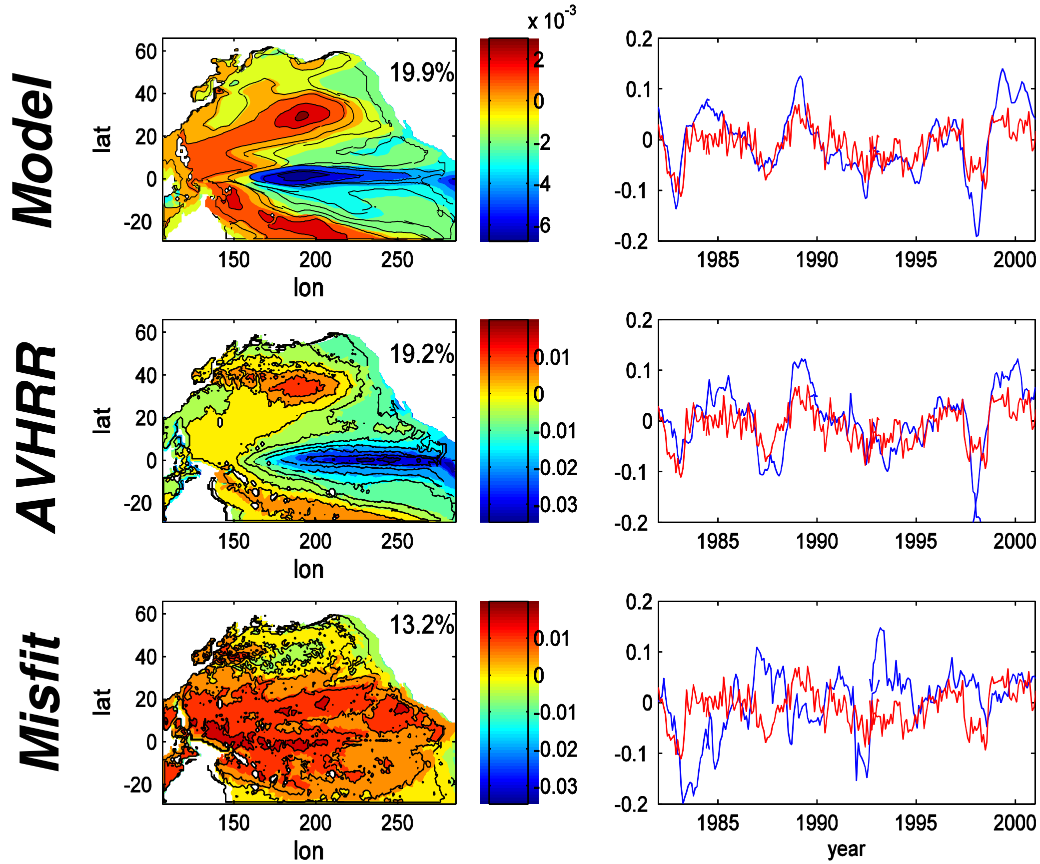
3) Recast the Kalman filter problem in terms of a Reduced State Space of approximately 35 EOFs instead of 10^5 discrete points

4) We estimate the multivariate model error covariance P^f by performing linear regressions to fit the EOF's of the SST model data misfits with the temperature component of the model multivariate EOF's.

5) Using the estimated model covariance, we calculate the Kalman gain and the update the model to combine with the observations.

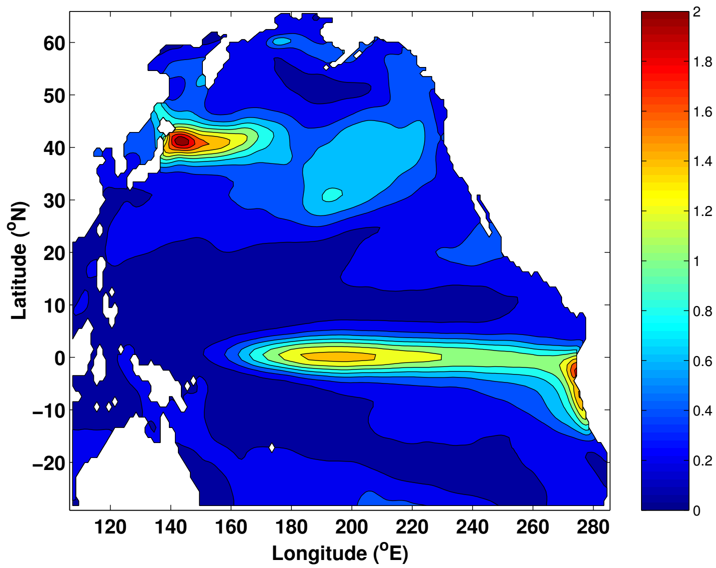
Information Content of North Pacific Ocean Model

- Using a 26 year simulation we calculate the EOFs of the model, observations and innovations (data-model misfits)

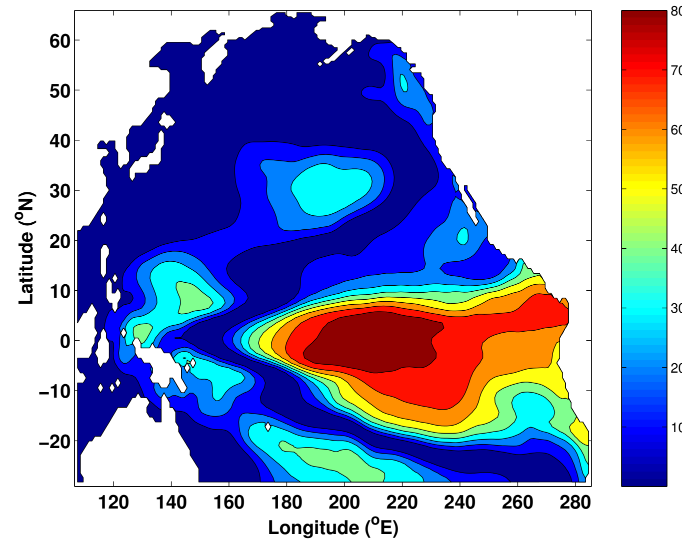


Variance described by Model SST EOFs

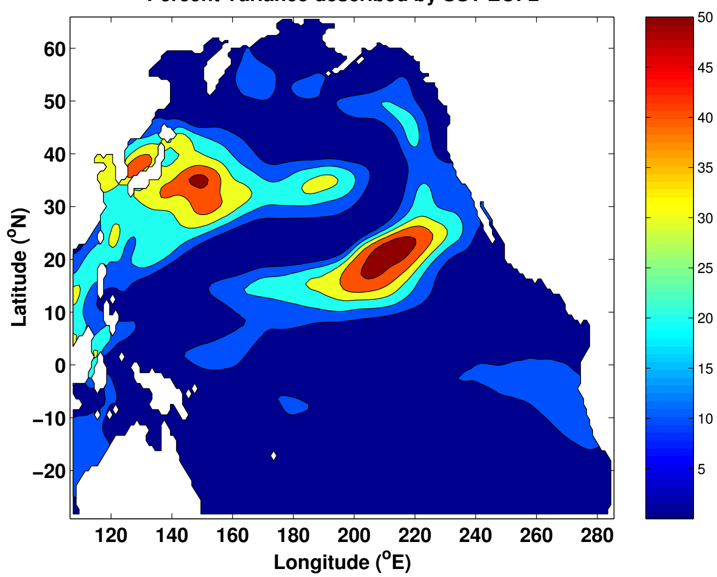
SST Variance ($^oC^2$)



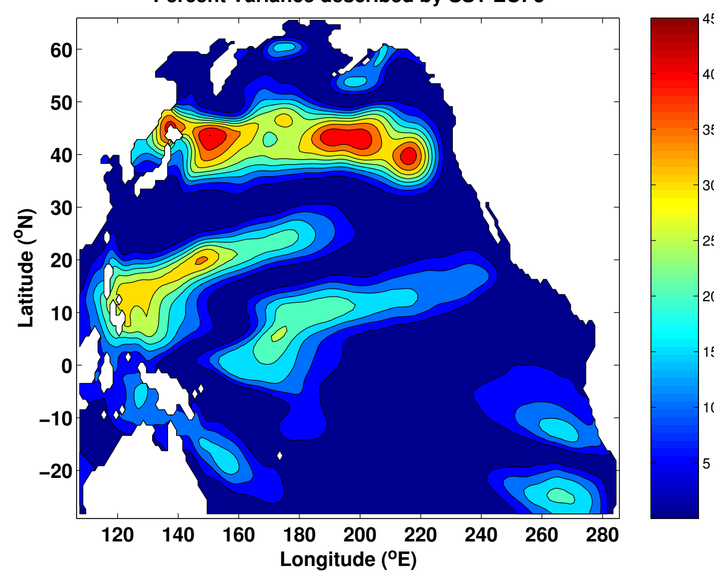
Percent Variance described by SST EOF1



Percent Variance described by SST EOF2

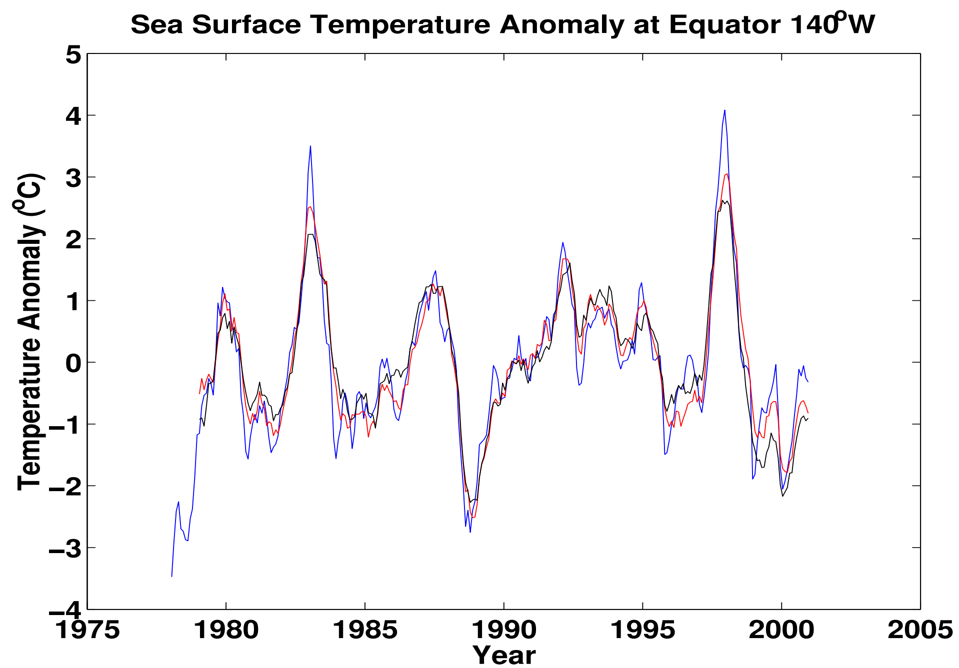


Percent Variance described by SST EOF3



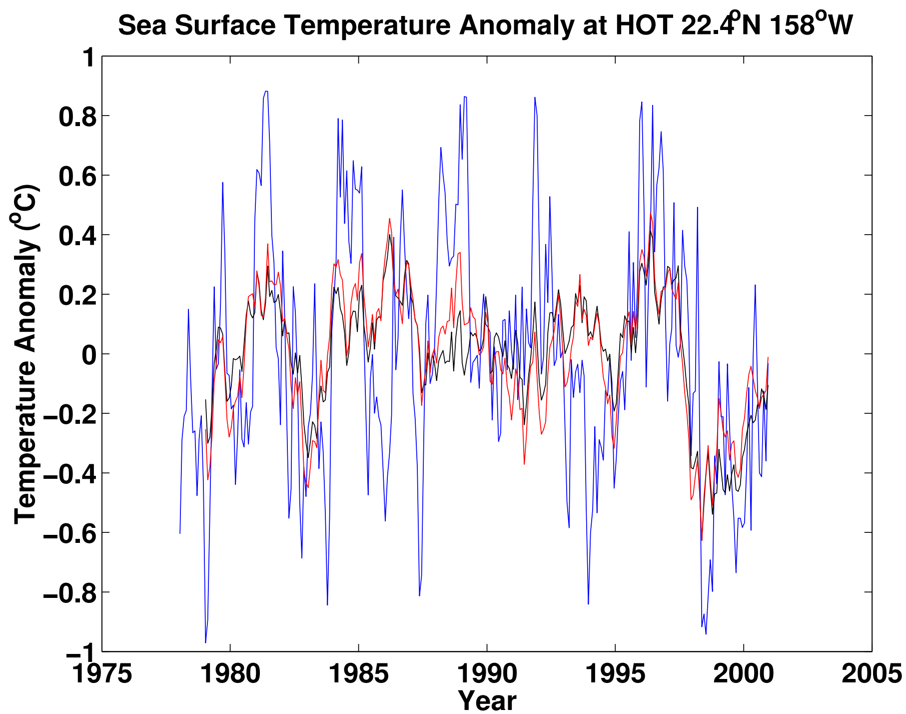
EOF Analysis of Sea Surface Temperature Anomalies

- The first EOF which describes 7% of the total variance is dominated by equatorial variability of the El Nino cycles. In the equatorial region, this mode describes 60-80% of the SST variance. The SST anomaly at 140W (blue) can be described by the first EOF (red) with the next two EOFs (black) making an insignificant contribution to the temperature.



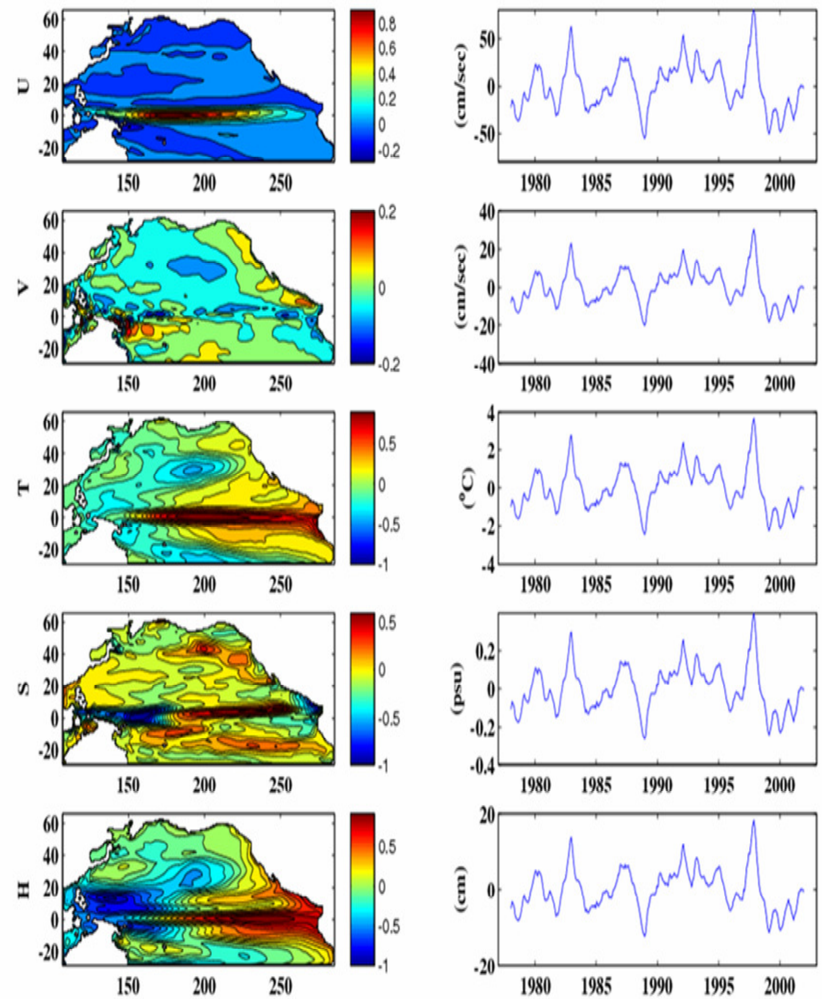
EOF Analysis of Sea Surface Temperature Anomalies

- The second EOF of the SST with 4% of the total variance described is dominated by variability in the strength of the subtropical gyre. In the subtropical gyre, this mode describes 30-50% of the SST variance. The SST anomaly at HOT (blue) is dominated by the second mode (red) with little contribution by the other two modes (black)



Model Multivariate EOF

- The first EOF of the surface velocity, temperature, salinity and sea level
- The first EOF is dominated by ENSO



Large number of state variables prohibits solving the full system

$\longrightarrow$ ***Reduced State Space Kalman Filter***

1) Compute the multivariate empirical orthogonal functions (EOF's) of our 26 year time series of deviations from the seasonal cycle,

2) **A statistical test is performed in order to estimate the number of significant degrees of freedom. (Preisendorfer, 1988) (35 modes accounting for 59% of the total variance)**

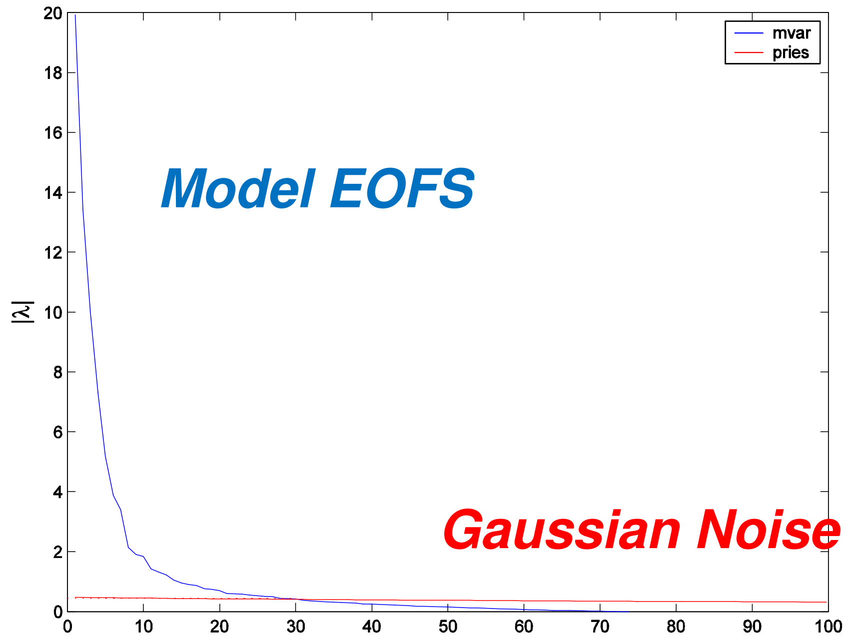
3) Recast the Kalman filter problem in terms of a Reduced State Space of approximately 35 EOFs instead of 10^5 discrete points

4) We estimate the multivariate model error covariance P^f by performing linear regressions to fit the EOF's of the SST model data misfits with the temperature components of the model multivariate EOF's.

5) Using the estimated model covariance, we calculate the Kalman gain and the update the model to combine with the observations.

Information Content

- The spectrum of the model EOFs is compared to the spectrum of gaussian noise with the same variance as the model
- Data assimilation only uses the projection of the innovations onto the model state space



$$x^a = x^f + P^f H^T (HP^f H^T + R)^{-1} (y^o - Hx^f)$$

Large number of state variables prohibits solving the full system

$\longrightarrow$ ***Reduced State Space Kalman Filter***

1) Compute the multivariate empirical orthogonal functions (EOF's) of our 26 year time series of deviations from the seasonal cycle,

2) A statistical test is performed in order to estimate the number of significant degrees of freedom. (Preisendorfer (1988)) (35 modes accounting for 59% of the total variance)

3) **Recast the Kalman filter problem in terms of a Reduced State Space of approximately 35 EOFs instead of 10^5 discrete points**

4) **We estimate the multivariate model error covariance P^f by performing linear regressions to fit the EOF's of the SST model data misfits with the temperature components of the model multivariate EOF's.**

5) **Using the estimated model covariance, we calculate the Kalman gain and the update the model to combine with the observations.**

Estimation of the forecast error covariance

- We estimate the multivariate model error covariance matrix $\mathbf{P}^f = \mathbf{VDV}^T$, where **V** is a matrix whose columns are linear combinations of the multivariate EOFs of the model and **D** is a diagonal matrix whose (i,i)th entry is the variance associated with the ith EOF of the model-data misfits.
- The coefficients α_{ij} in the linear combination
- $$\mathbf{V}_i = \Sigma_j \, \alpha_{ij} \, \mathbf{X}_j, \qquad (3)$$
- where the $\mathbf{X}_j$ is the jth multivariate EOF of the model, are chosen to minimize
- $$(\mathbf{U}_i - \Sigma_j \, \alpha_{ij} \, \mathbf{HX}_j)^T \; (\mathbf{U}_i - \Sigma_j \, \alpha_{ij} \, \mathbf{HX}_j) \qquad (4)$$
- where $\mathbf{U}_i$ is the ith EOF of the model-data misfits, $\mathbf{V}_i$ is the ith column of **V** and **H** is the matrix that maps the state vector into the SST or SSH field
- The first eight (30) $\mathbf{U}_i$ contain about 15 (37)% of the total variability of the SST model-data misfit variance and 13 (26)% of the SSH mode-data misfit variances.

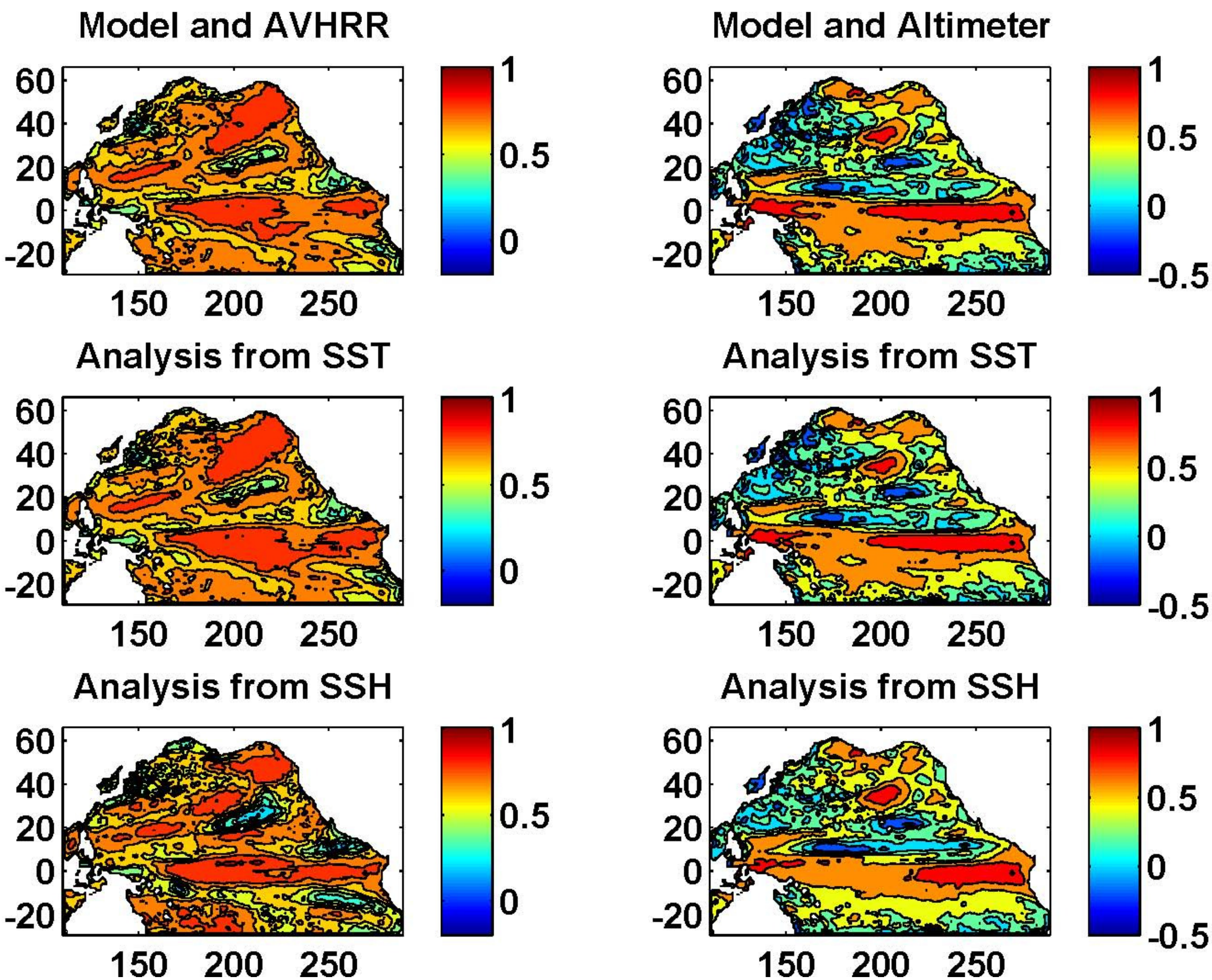
Estimation of the forecast error covariance

- The estimate of $\mathbf{P}^f$ based on (3) is used along with the approximation:
- $$(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{y}^o - \mathbf{Hx}^f) \approx \mathbf{U\check{D}}^{-1}\mathbf{U}^T (\mathbf{y}^o - \mathbf{Hx}^f) \qquad (5)$$
- to implement a data assimilation scheme based on the formula for optimal interpolation given in (1). We consider the matrix $\mathbf{P}^f \approx \mathbf{V\,DV}^T$ to be fixed, and do not run the model between assimilation steps. In this experiment, we only update surface values of the velocity components, the temperature and the salinity, as well as the sea surface height anomaly. We do not update the model state below the surface.
- With these assumptions, the gain matrix becomes:
- $$\mathbf{K} = \mathbf{VD}(\mathbf{HV})^T\mathbf{U\check{D}}^{-1}\mathbf{U}^T \qquad (6)$$

Estimation of the forecast error covariance

- We may examine the updating process by writing the analysis increment as
- $\mathbf{VD} * (\mathbf{HV})^T * \check{\mathbf{D}}^{-1}\mathbf{U}^T (\mathbf{y}^o - \mathbf{Hx}^f) \qquad (7)$
- The last term is the projection of the innovation vector on the leading EOFs of the model-data misfits, with the result weighted by the inverses of variances contributed by each EOF. The second term is the projection of the lead EOFs of the misfits onto the $\mathbf{HV_i}$, themselves linear combinations of the multivariate EOFs of the model output, so the second and third terms amount to a projection of the innovation vector into the space defined by the $\mathbf{HX_i}$. Forming the product of these projections with the first term **V** maps the projections back into the multivariate model state space.
- The only assimilated variability is that which projects into the model state space.

Model and Data Correlations before and after Reduced State Space OI



Representation error

The Kalman filter blending of the model and the observations made a modest improvement of the model ouputs

Why was not there a bigger impact?

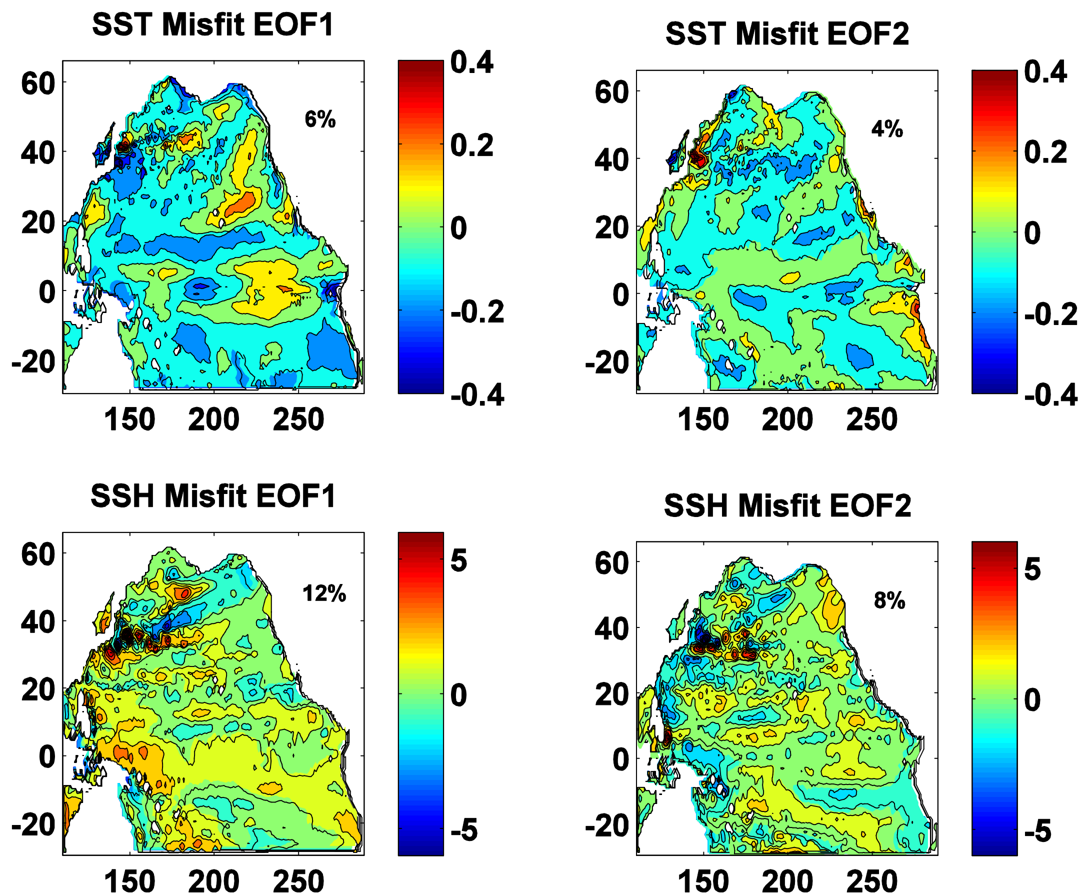
The model cannot represent all of the variability observed in the data.

Using the Reduced State Space, we can estimate this error of representation

The difference between the model data misfit and the EOF representation of this misfit (error of representation) gives us information on where improvement is needed.

Representation Error

- The innovations are projected onto the model state space.
- The remainder of the innovations can be decomposed into EOFs to show the spatial variability of the representation error



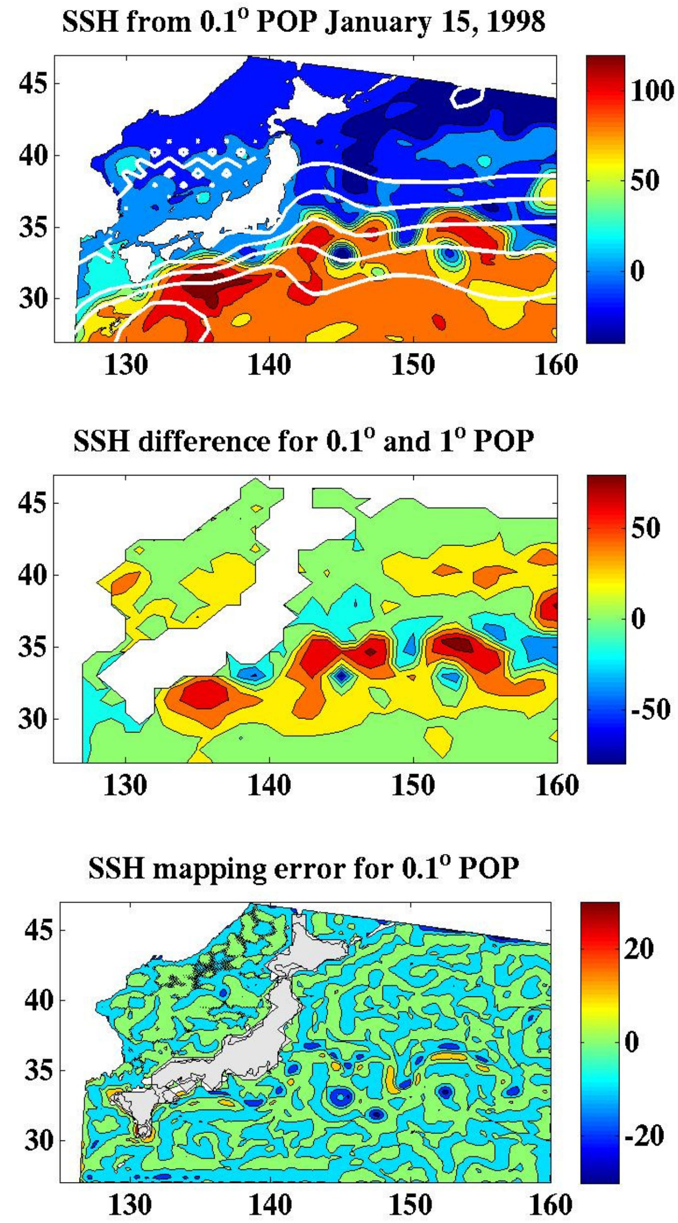
$$R = U\,D^2\,U^T$$

Representation Error is not the same as interpolation error

- Representation error often is defined as mapping or interpolation error for unresolved scales
- Interpolation error can be found by examining the difference between resolved observations mapped onto the coarse model grid and then remapped onto the finer observation grid
- Interpolation error does not account for missing model physics

Interpolation Error

- Results from 1/10° POP model are compared to 1° POP
- The 1° POP doesn't generate meanders or eddies
- The 1/10° POP features have scales which are mapped reasonably well on the 1° POP grid

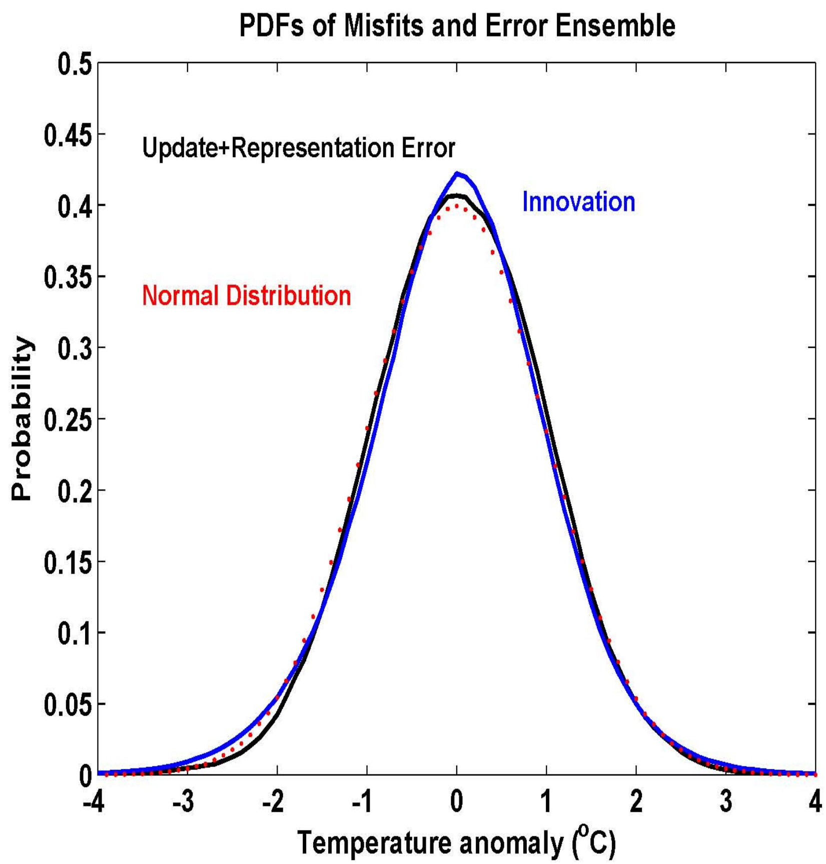


POSTERIOR STATISTICAL EVALUATION

- Estimate of the covariance of the innovation:
- $< (\mathbf{y}^o - \mathbf{Hx}^f) (\mathbf{y}^o - \mathbf{Hx}^f)^T > \approx (\sigma^o)^2\mathbf{I} + \mathbf{WW}^T + \mathbf{HV}^T\mathbf{VH}^T$ (10)
- **W** is the matrix whose columns are the eight leading EOFs of the representation error, weighted by their corresponding singular values, $< \varepsilon^o \varepsilon^{oT} >$ is assumed to be a multiple of the identity matrix **I**.
- Assume no cross correlation for the errors
- $< \mathbf{H}\varepsilon^f \varepsilon^{oT} > = < \mathbf{H}\varepsilon^f \varepsilon^{RT} > = 0$ (11)
- Assume that ε^o is determined by the properties of the instrument, rather than those of the physical system.
- ε^f and ε^R arere constructed to be orthogonal, but they may not be uncorrelated since the small scale variability may be linked to larger scale phenomena, so, e.g., the rate at which eddies are generated may be related to large scale factors.
- We test the hypotheses (10) and (11) by an ensemble experiment.

ESTIMATION OF REPRESENTATION ERROR STATISTICS

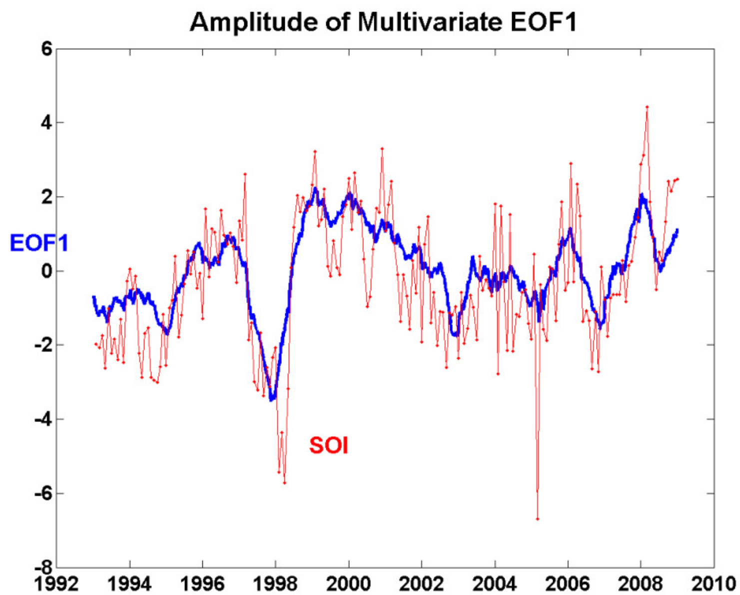
- We can generate a Monte Carlo estimate of the representation error from the EOFS of the representation error
- The resulting pdf of the Monte Carlo estimate of the SST misfit is indistinguishable from the actual SST misfit pdf

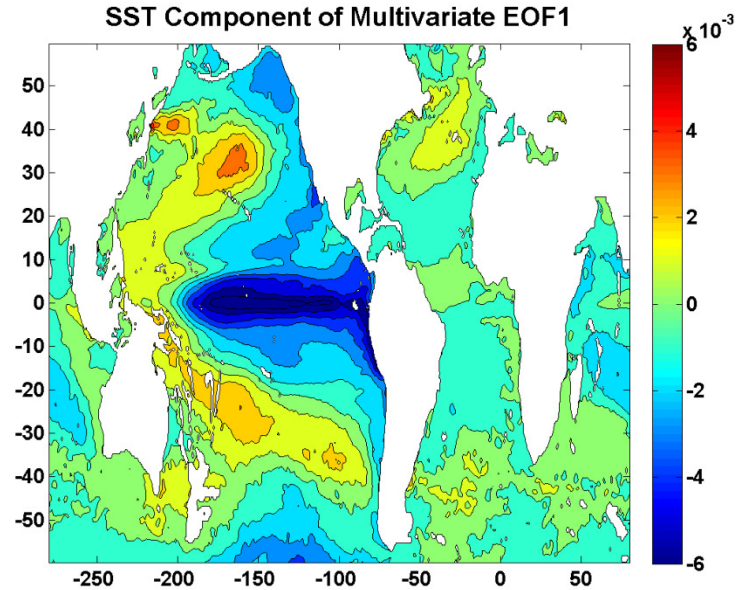


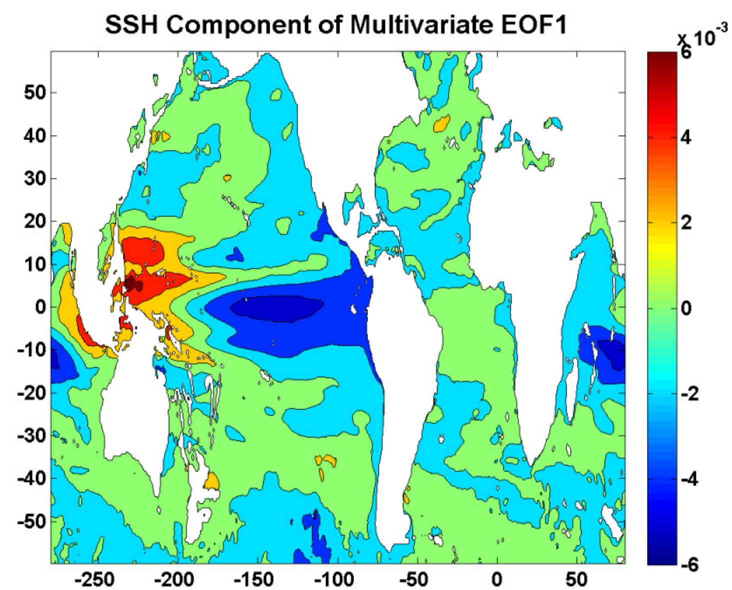
Error Estimation for 16 year 0.5° CFS model run

- Using a 16 year free run of the 0.5° CFS, we calculate the mutivariate eofs
- The lead eof accounts for 12.2% of the anomaly variance and is correlated with the SOI at 0.76
- The SST innovations are projected onto the SST component of the multivariate eofs
- The residuals are the orthogonal error space

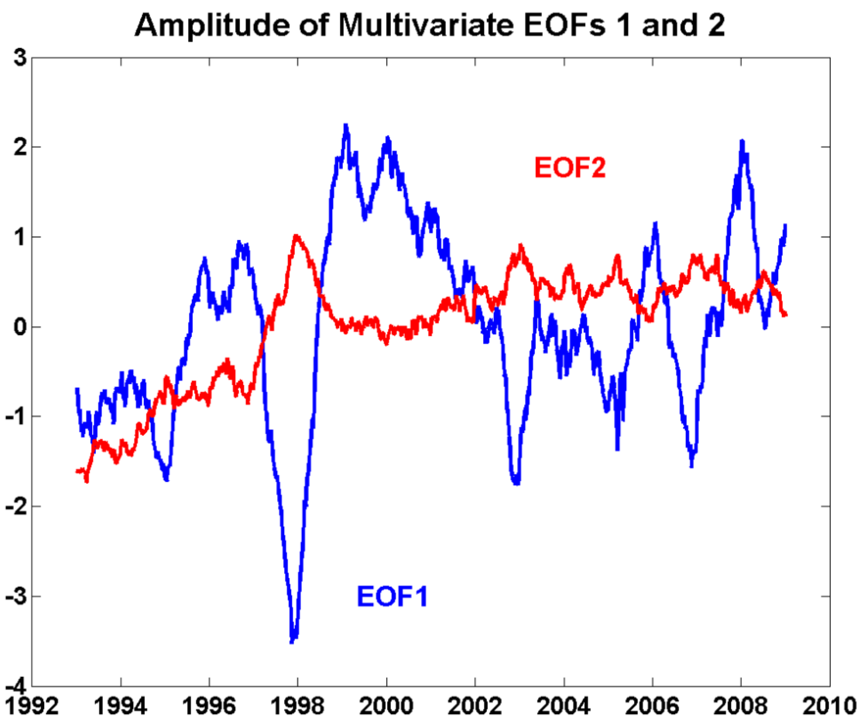
Multivariate EOFs for 0.5° CFS

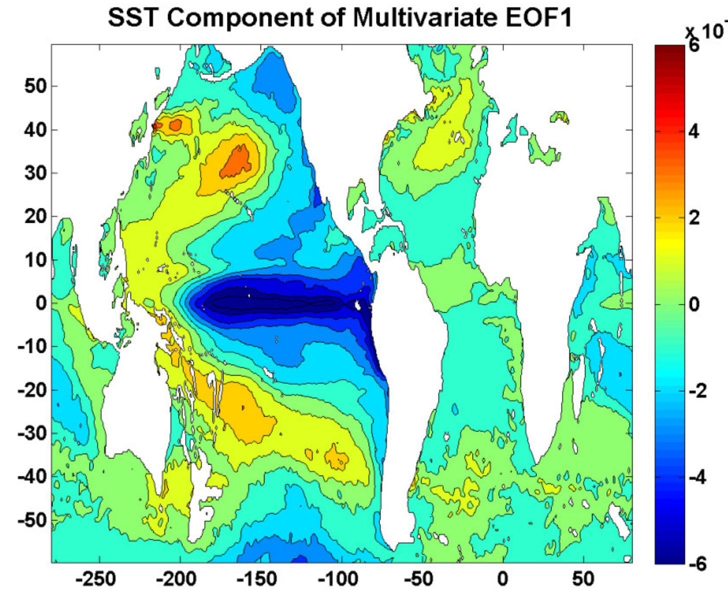


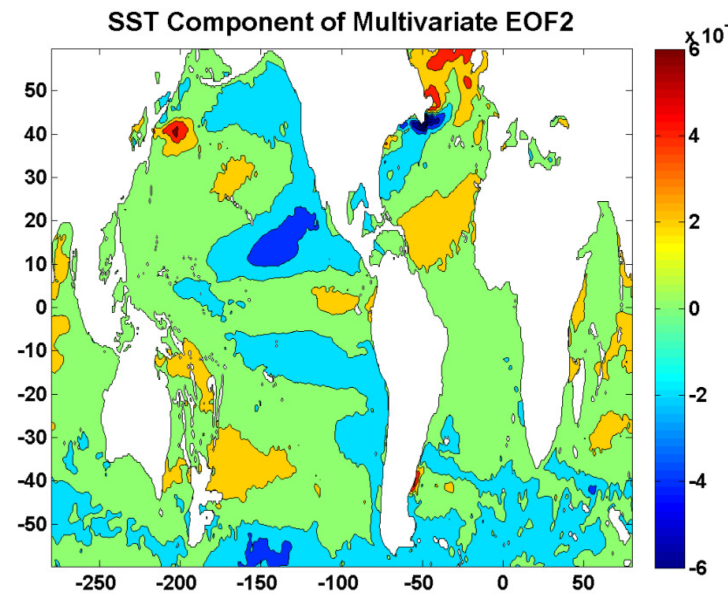




Multivariate EOFs for 0.5° CFS

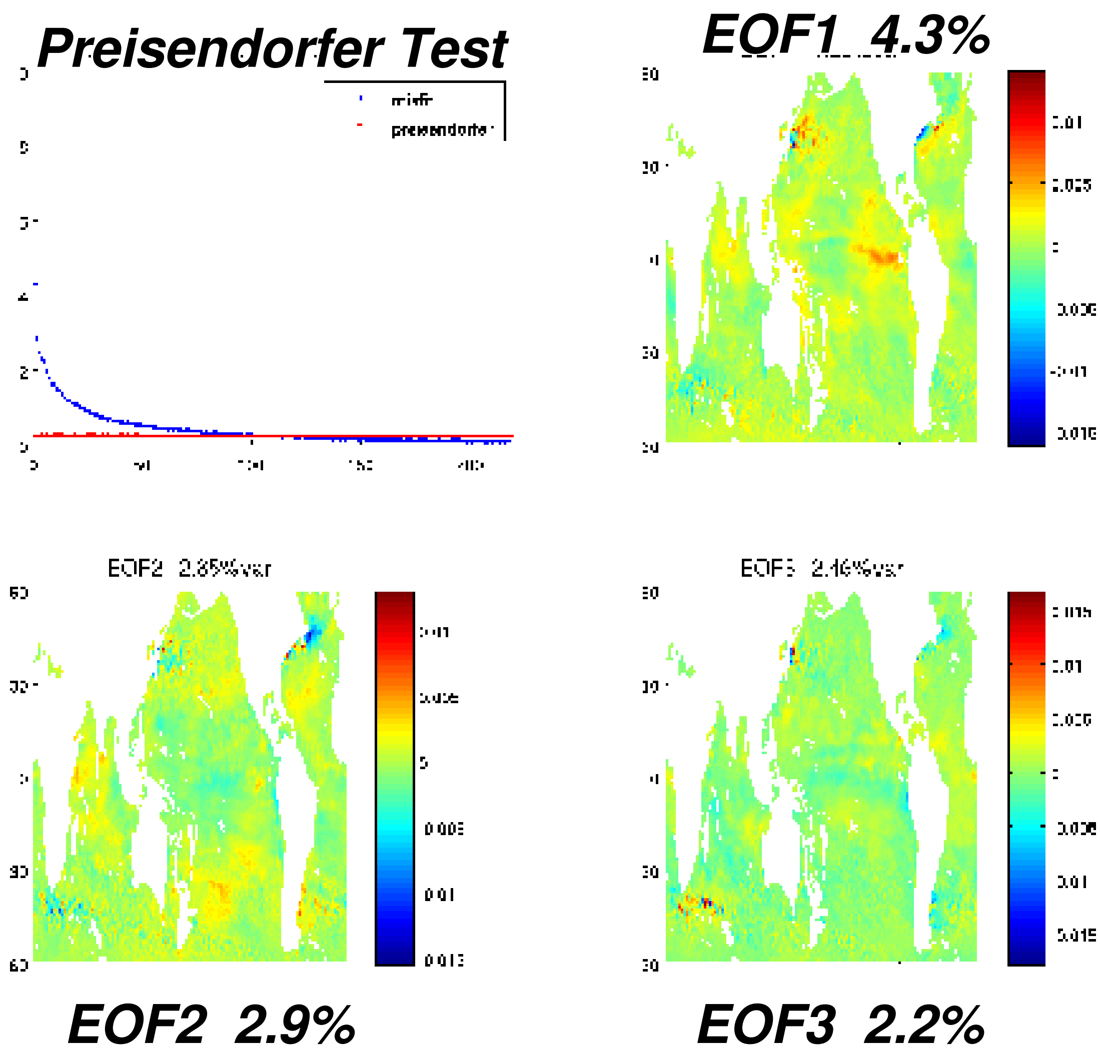






Representation Error for 0.5° CFS

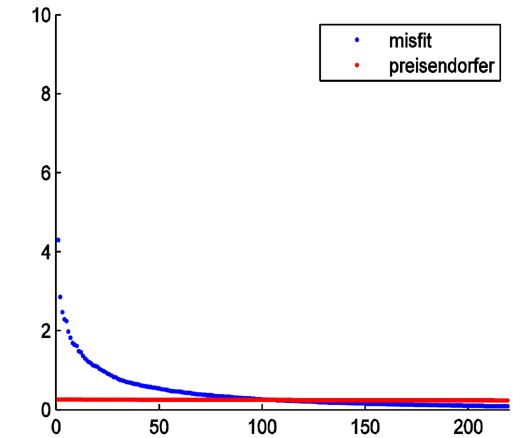
- The eofs of the SST misfits orthogonal space
- Approximately 52 modes pass the Preisendorfer test



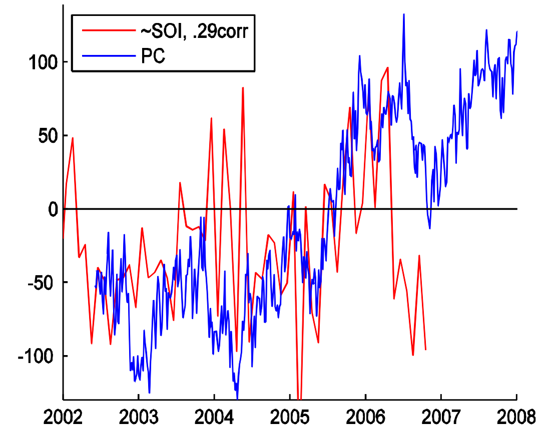
Representation Error for 0.5° CFS

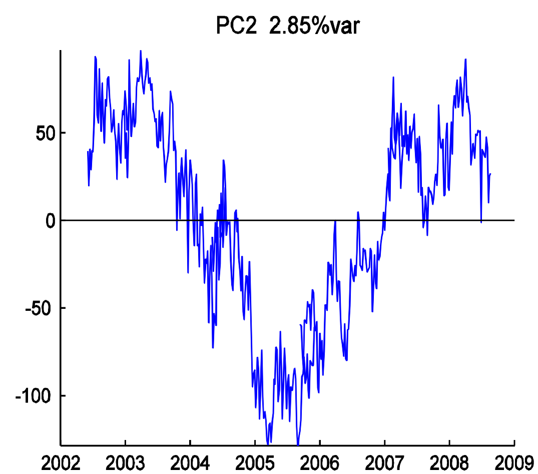
- The amplitudes of the eofs of the SST misfit orthogonal space

Preisendorfer Test

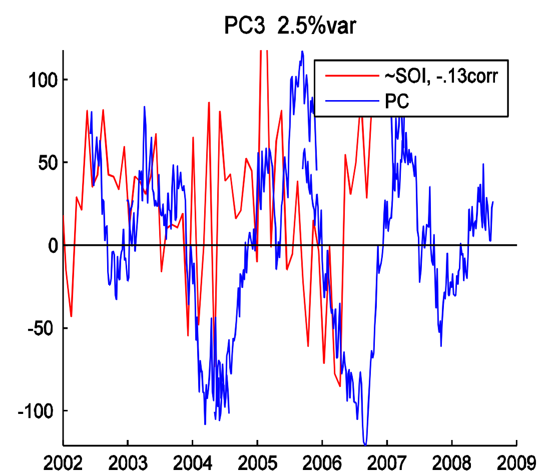


PC1 4.3%





PC2 2.9%



PC3 2.2%

Information Content and Representation Error

- We have developed a technique to determine the information that a model can represent (information content) and the information which the model cannot describe due to lack of resolution or inadequate model physics (representation error)
- The technique tested with a coarse resolution climate model, but can be generalized to any model

Conclusions

- Using an ensemble approach, we can define the information content of a model using the Priesendorfer test to separate significant eofs from noise
- Analyses from ensemble techniques are linear combinations of the ensemble members
- The portion of the model-data misfits (innovations) that does not lie in the ensemble eof space is observation error
- The significant eofs of the observation or representation error can be used to determine the spatial correlations
- The representation error is model and resolution dependent, but differs from the interpolation error.
- A posterior test of the representation error shows that an ensemble constructed from our error eofs is indistinguishable from the actual innovations.