High resolution and uncertainties: Combining scales in signal and error

Alexey Kaplan

Lamont-Doherty Observatory of Columbia University (LDEO)

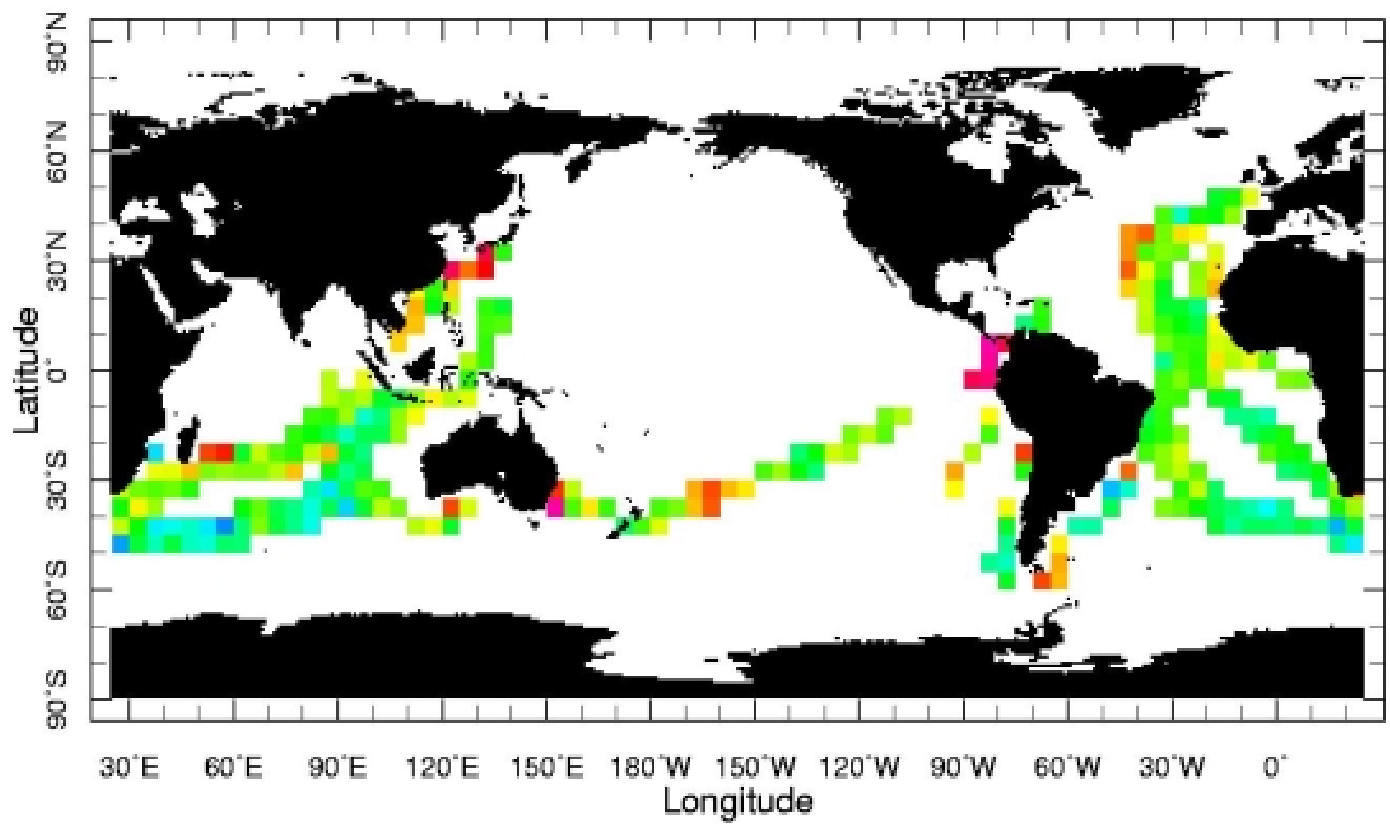
In collaboration with:

Craig Johns (University of Colorado @ Denver)

Nick Rayner (Hadley Centre)

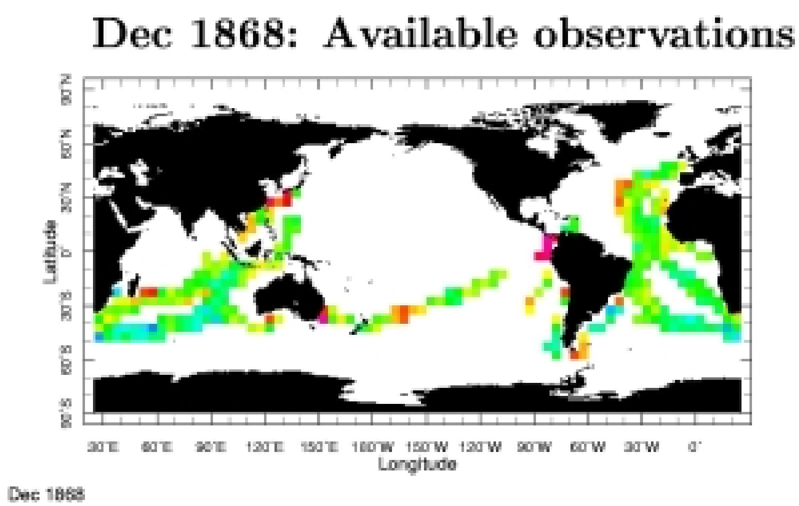
Mark Cane, Yochanan Kushnir (LDEO)

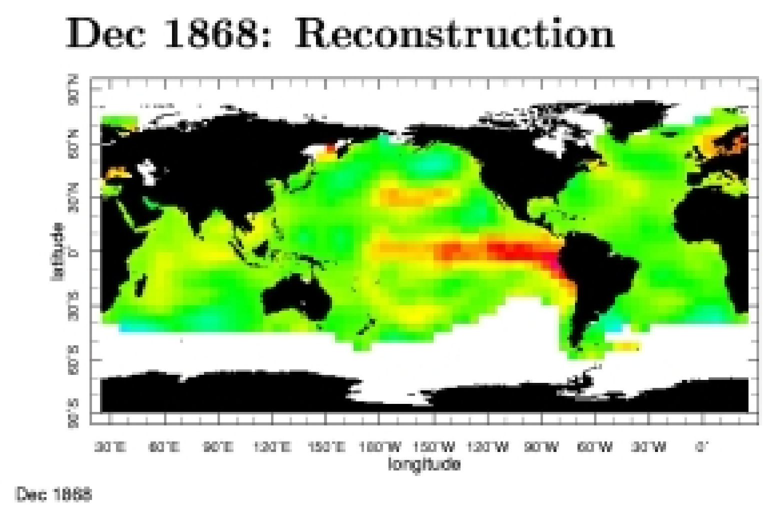
Dec 1868: Available observations



Dec 1868

People prefer right to left





Example of Optimal Interpolation

$$T = T_B + e_B$$

$$HT = T_o + e_o$$

$$<e_B> = < e_o > = < e_B e_o^T > = 0$$

$$< e_B\, e_B^T > = C$$

$$< e_o\, e_o^T > = R$$

Solution minimizes the cost function

$$S[T] = (HT - T_o)^T R^{-1} (HT - T_o) + (T - T_B)^T C^{-1} (T - T_B)$$

$$\boxed{T = (H^T R^{-1} H + C^{-1})^{-1} (H^T R^{-1} T_o + C^{-1} T_B)}$$

Projection of OI solution on eigenvectors

$C=EDE^T$

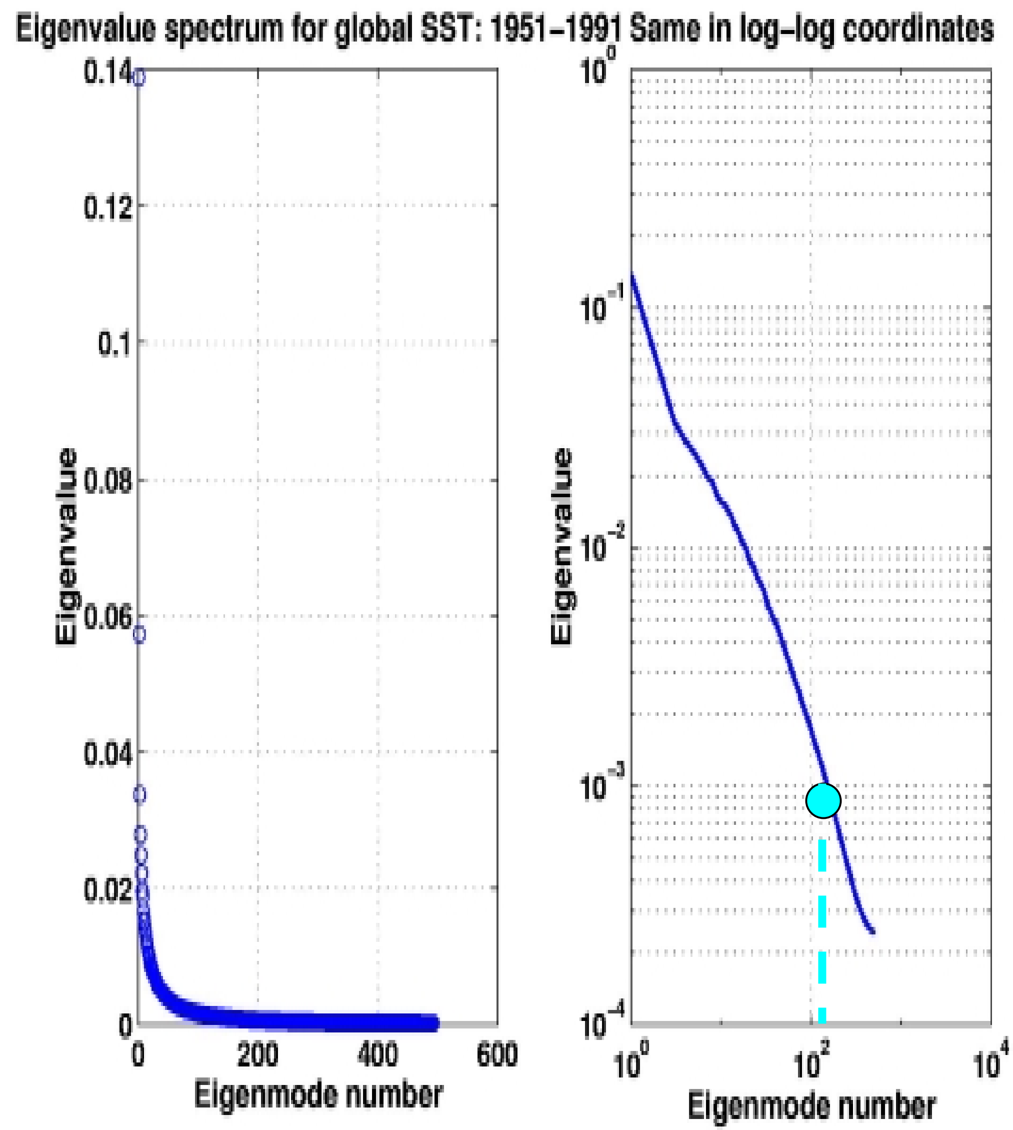
$T=Ea$

For simplicity: $H=I$, $R=rI$, $T:=T-T_B$

Then $a=\underline{D(D+R)^{-1}}E^T T_o$

$D(D+R)^{-1}=\text{diag}[\, d_i/(d_i+r)\,]$

In many applications (for spectrally red signals) diagonal elements of this matrix decrease from ~1 to ~0

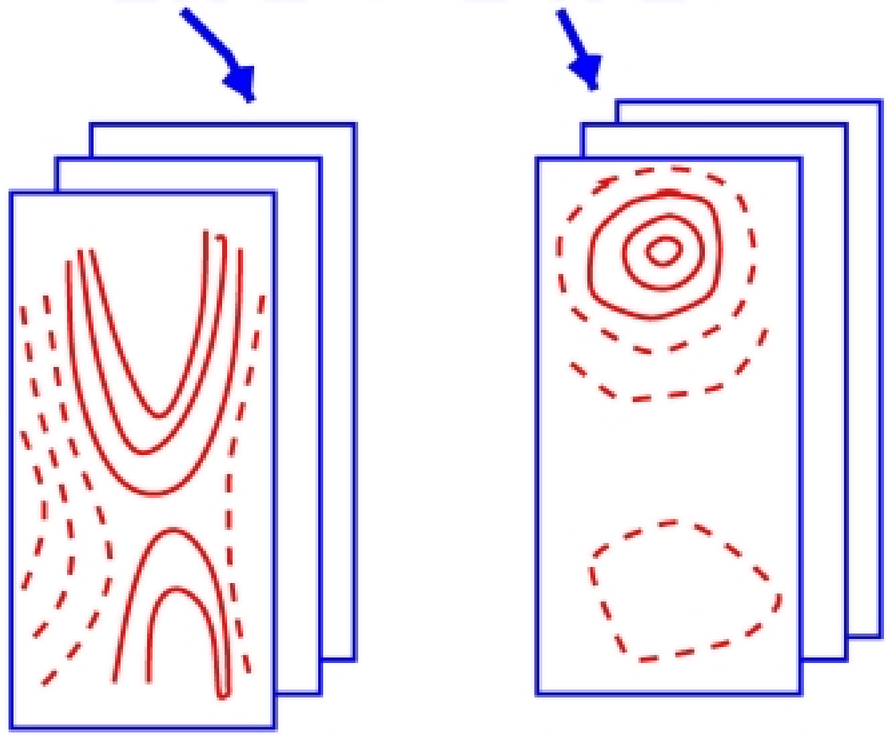

Eigenvalue spectrum for global SST: 1951–1991
Same in log–log coordinates
Eigenvalue
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
200
400
600
Eigenmode number
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^0
10^2
10^4

3 corollaries:

- The first is good: the tail (strongly dampened) modes can be filtered from the solution, i.e. the solution can be effectively approximated by a linear combination of a few leading (only slightly dampened) modes

APPROXIMATING COVARIANCE

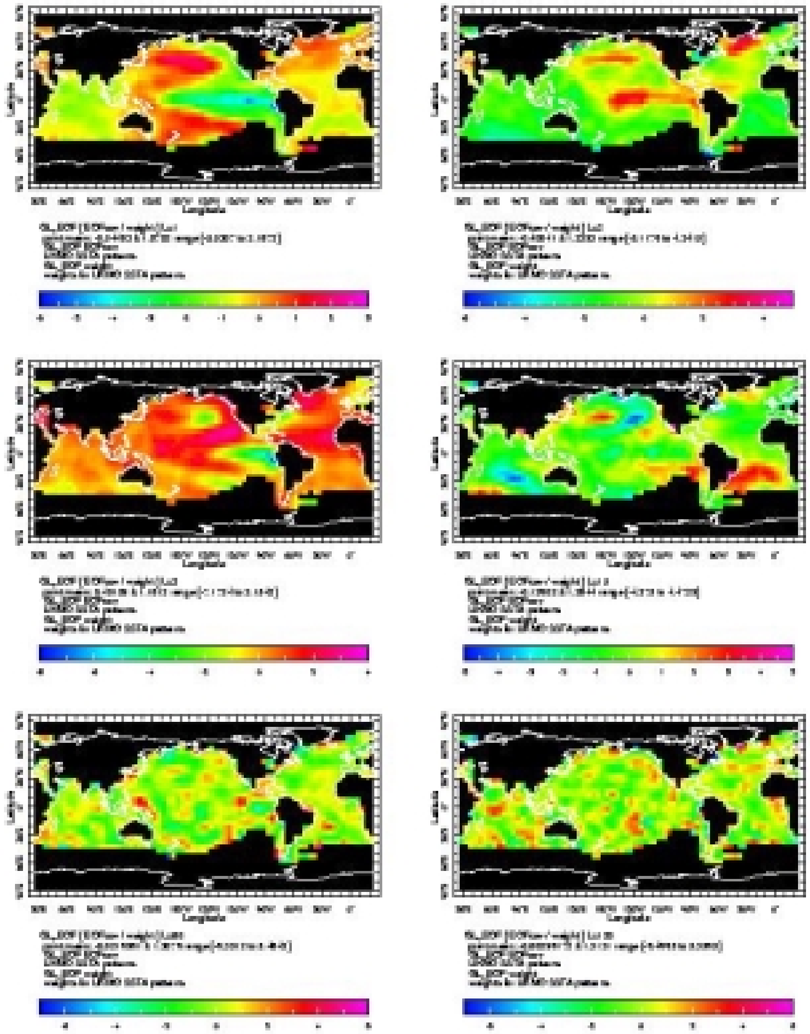
$$C = E\Lambda E^T + E'\Lambda' E'^T$$



Reduced space optimal analysis

Successive corrections; Kriging

EOFs of SST (#1,2,3,15,80,120)



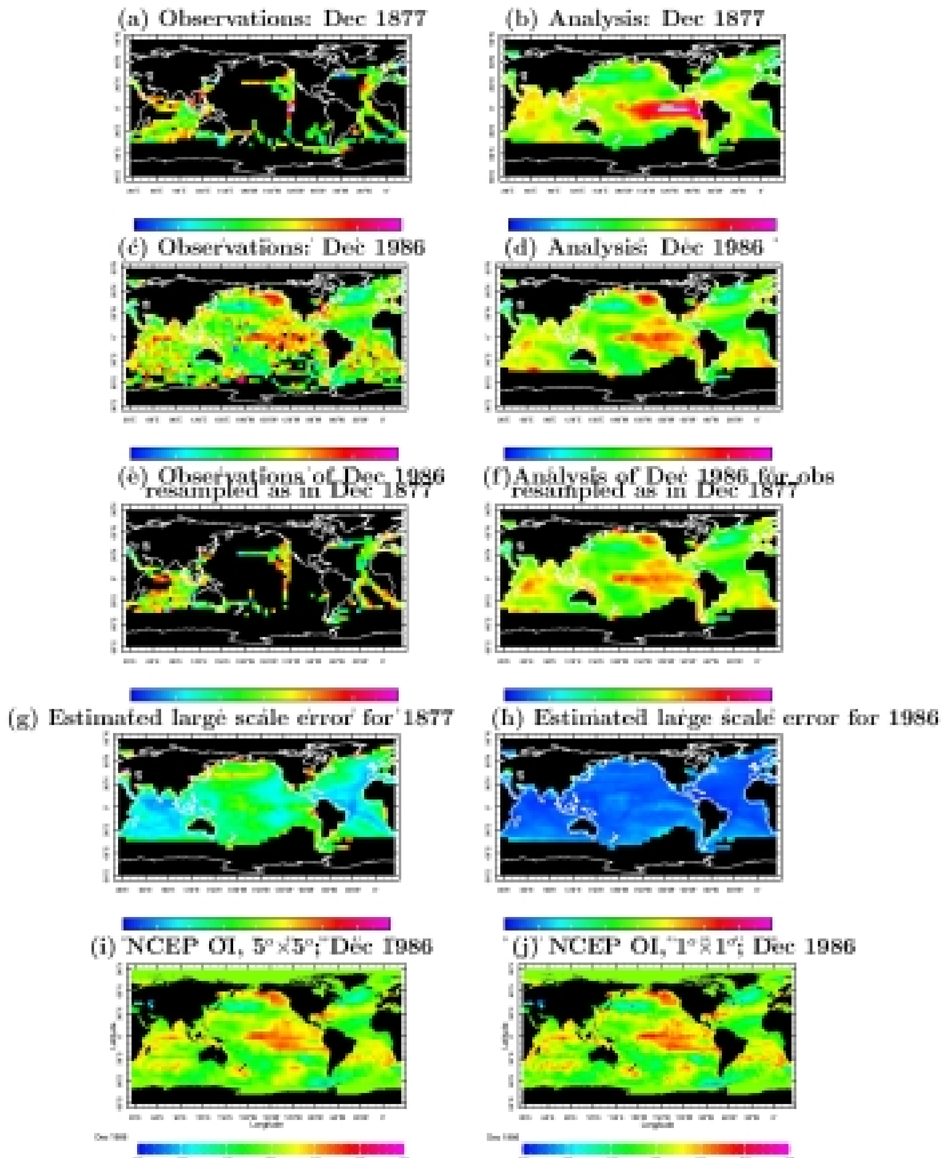


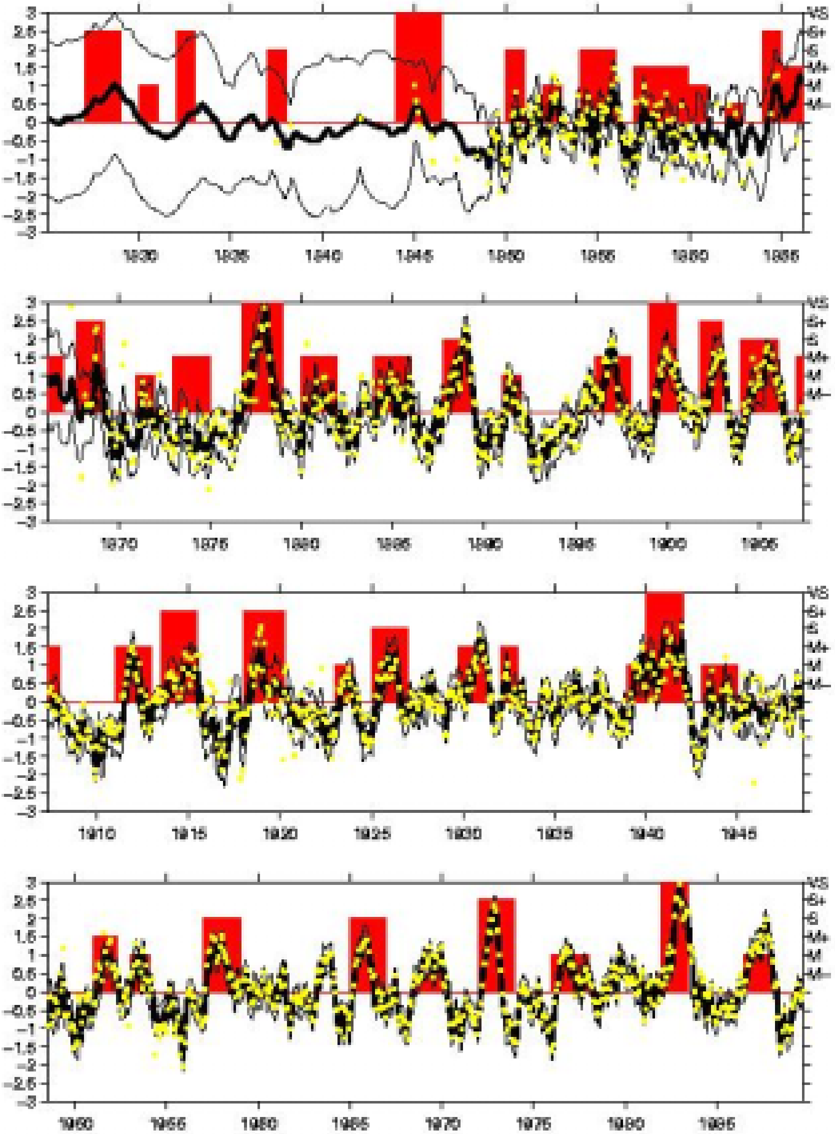
Figure 2: Available SST observations and their RS OS analysis for December 1877 (panels (a) and (b)) with verification through the experiment with 1986 data: simulated OS analysis for December 1986 using the data distribution of 1877 (panels (e) and (f)) versus the standard OS analysis for December 1986 with all available data (panels (c) and (d)). Also shown are large scale errors in the two reconstructions (panels (e) and (f)) and the NCEP OI December 1986 field presented in (i) $5^o\times5^o$ and (j) $1^o\times1^o$ resolution. Units are °C.

3 corollaries:

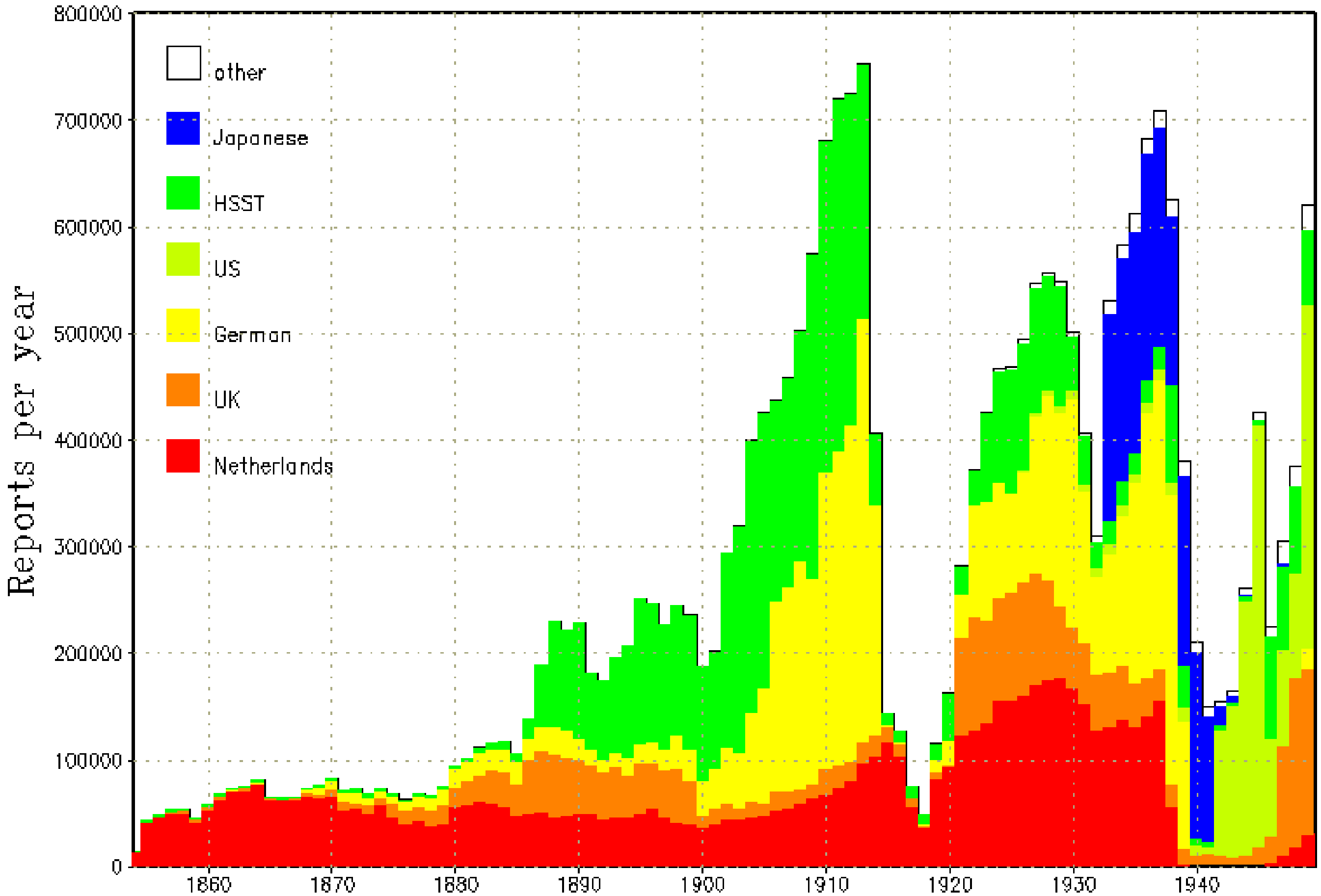
- The first is good: the solution can be effectively approximated by a linear combination of a few leading modes.
- The second is bad: the solution always has less variance than the true field.

In fact , $C=\langle TT^T \rangle + P$

SST El Nino indices vs Quinn's historical rankings



Number of observations in COADS



3 corollaries:

- The first is good: the solution can be approximated by a few leading modes.
- The second is bad: the solution always has less variance than the true field.
- The third is ugly: the solution is always redder than the truth (because of predominant dampening of tail modes).

Again, it helps to remember that

$$C=\langle TT^T \rangle+P$$

Scale separation in a field estimate.

OI problem: estimating a field $\mathcal{T}$ from a first-guess (background) solution $\mathcal{T}^b$ and an incomplete set of observations $\mathcal{T}^o$ is given by:

$$\mathcal{T}^b = \mathcal{T} + \varepsilon^b, \quad \langle \varepsilon^b \rangle = 0, \quad \langle \varepsilon^b \varepsilon^{b\,T} \rangle = C, \tag{1}$$

$$\mathcal{T}^o = H\mathcal{T} + \varepsilon^o, \quad \langle \varepsilon^o \rangle = 0, \quad \langle \varepsilon^o \varepsilon^{o\,T} \rangle = R. \tag{2}$$

The solution to this OI problem is a minimizer $\hat{\mathcal{T}}$ of the cost function

$$\mathbf{S}[\mathcal{T}] = (H\mathcal{T} - \mathcal{T}^o)^T R^{-1} (H\mathcal{T} - \mathcal{T}^o) + (\mathcal{T} - \mathcal{T}^b)^T C^{-1} (\mathcal{T} - \mathcal{T}^b).$$

$$\hat{\mathcal{T}} = CH^T (R + HCH^T)^{-1} \mathcal{T}^o.$$

$$C = E\Lambda E + E'\Lambda' E' = E\Lambda E + C'$$

$$\hat{\mathcal{T}} = E\hat{\alpha} + \Delta\hat{\mathcal{T}}.$$

$$\mathcal{T}^o = HE\alpha + \tilde{\varepsilon}^o, \qquad \langle \alpha\alpha^T \rangle = C', \qquad \langle \tilde{\varepsilon}^o \tilde{\varepsilon}^{o\,T} \rangle = HC'H^T + R.$$

$$\hat{\alpha} = \Lambda E^T H^T (HE\Lambda E^T H^T + HC'H^T + R)^{-1} \mathcal{T}^o.$$

Observational residual: $\Delta\mathcal{T}^o = \mathcal{T}^o - HE\hat{\alpha}$

$$\Delta\mathcal{T}^o = H\Delta\mathcal{T} + \varepsilon^o, \qquad \langle \Delta\mathcal{T}\Delta\mathcal{T}^T \rangle = C', \qquad \langle \varepsilon^o \varepsilon^{o\,T} \rangle = R.$$

$$\Delta\mathcal{T} = C'H(HC'H^T + R)^{-1} \Delta\mathcal{T}^o$$

Synthetic example

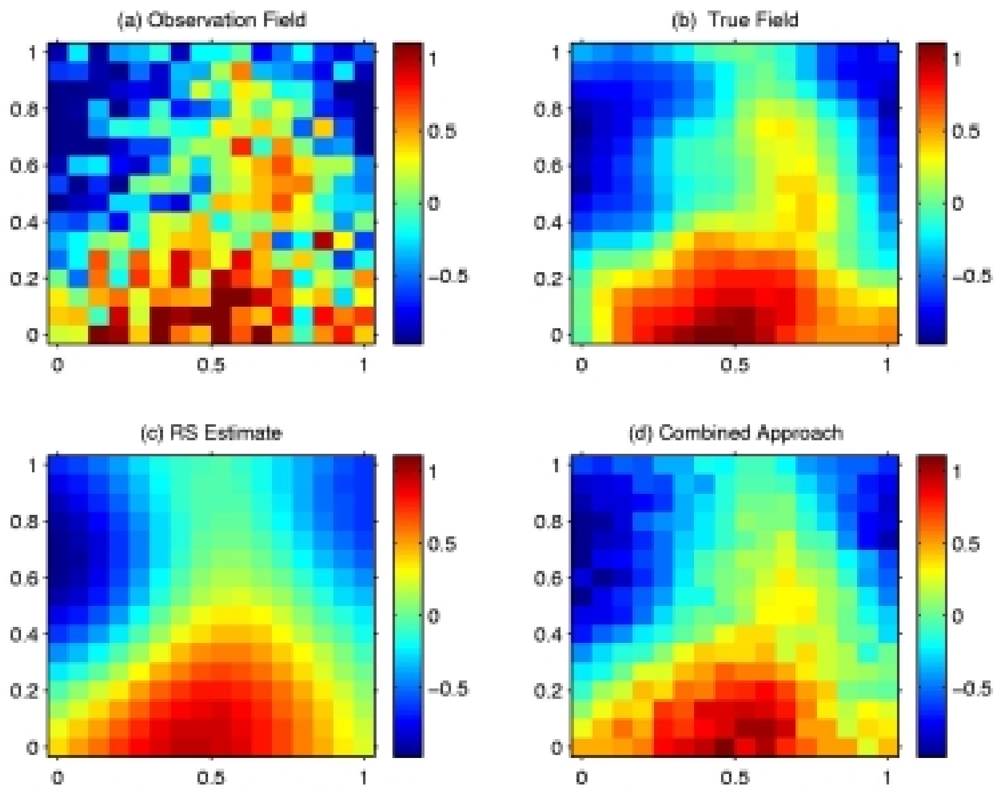
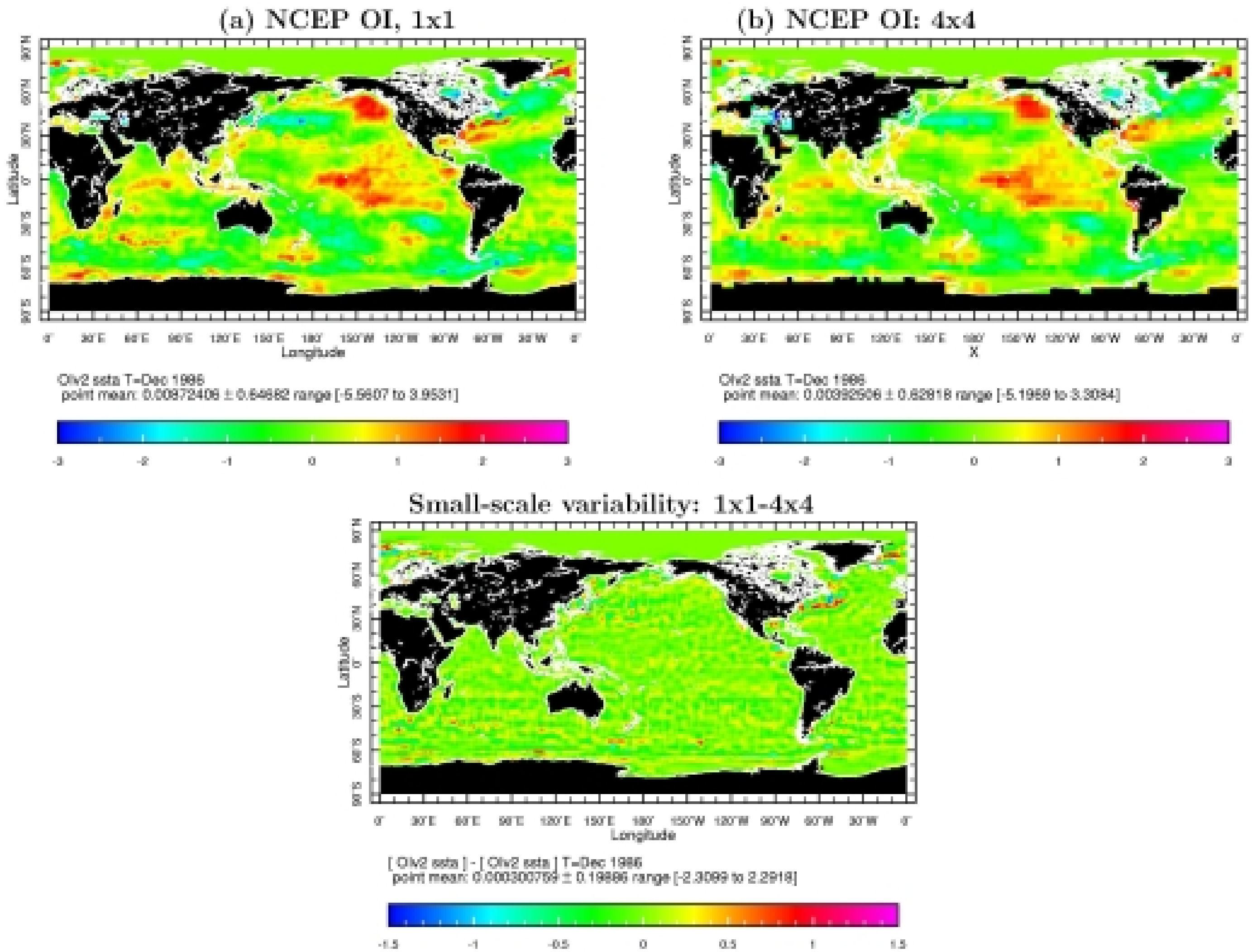
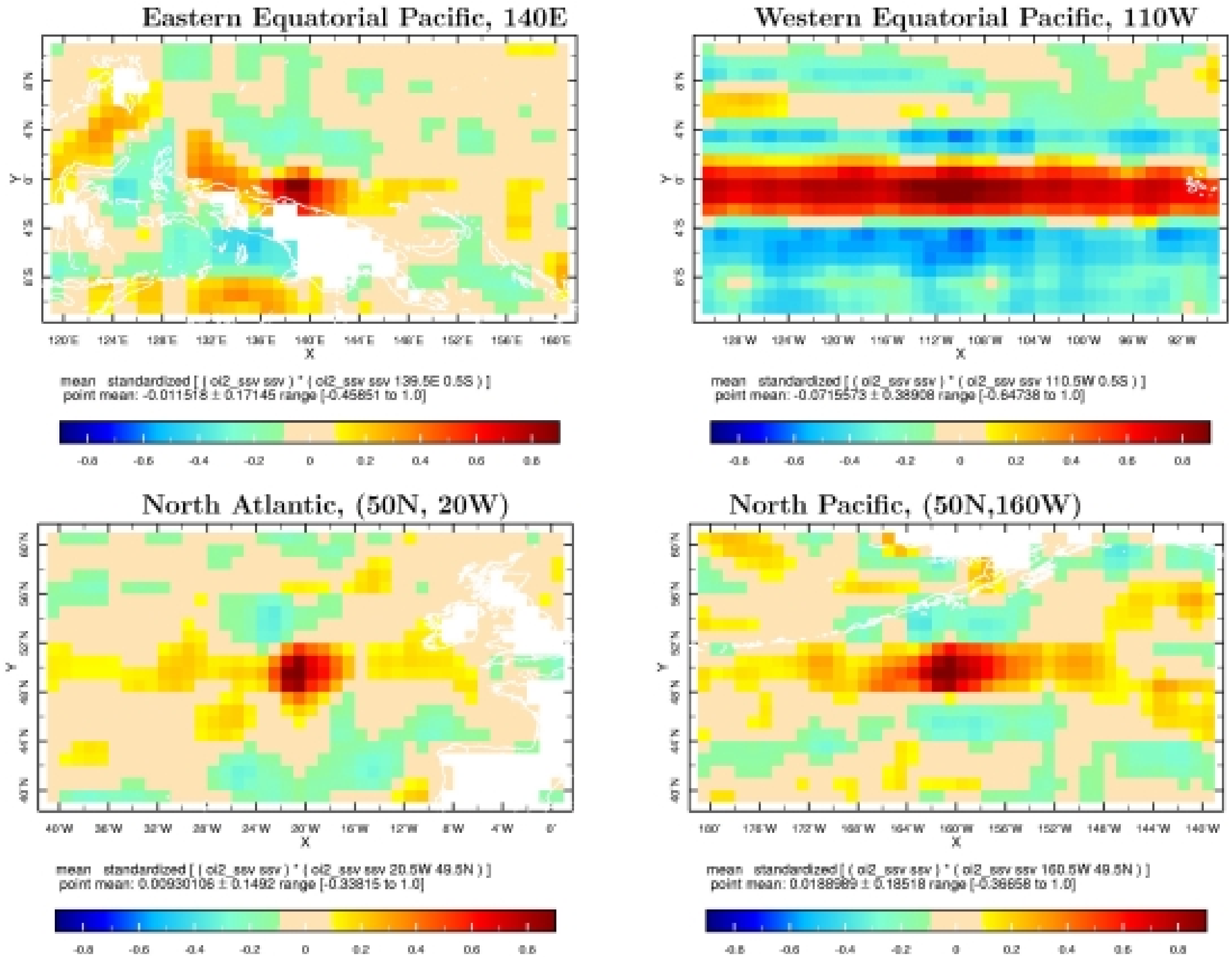


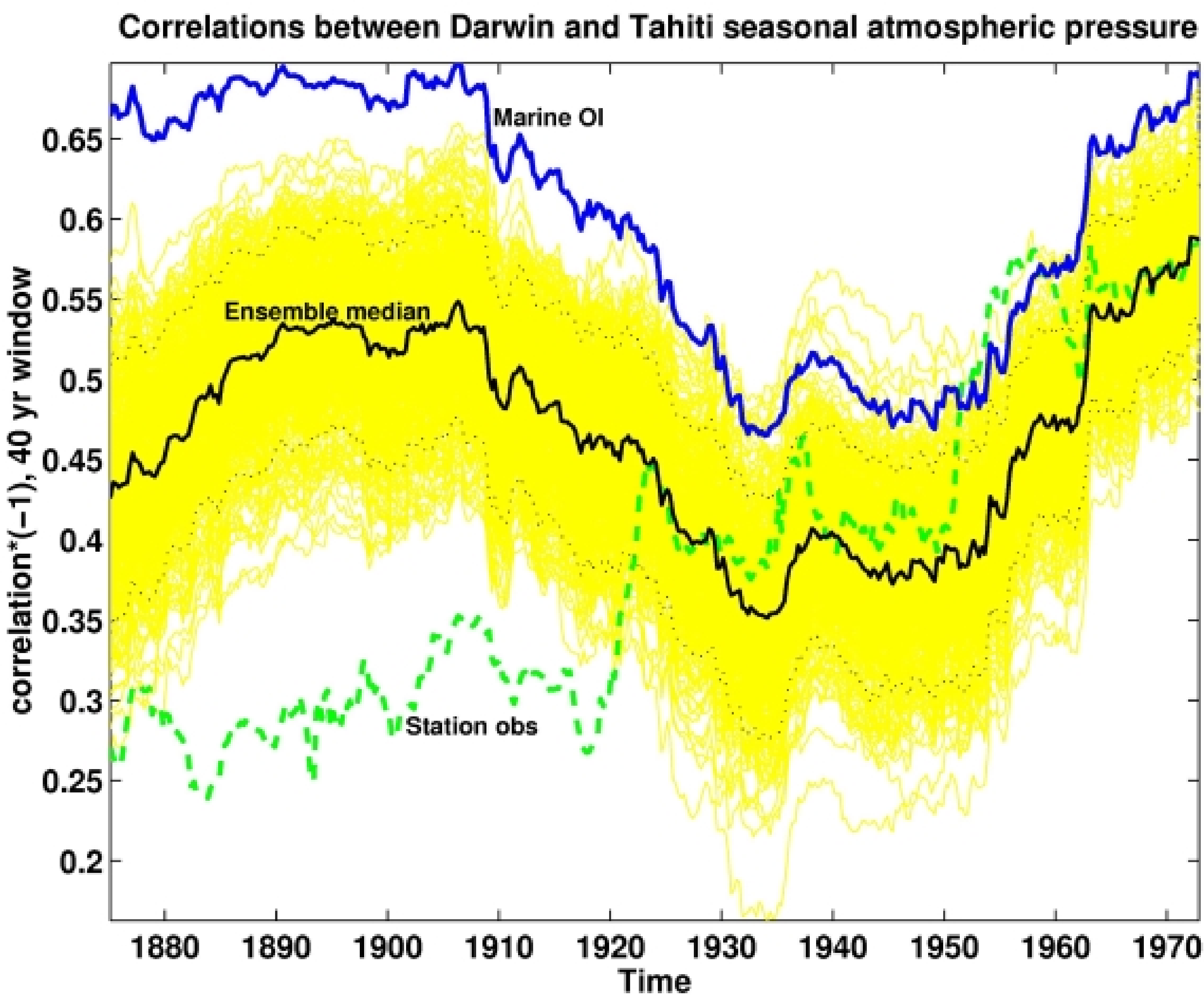
Figure 3: Panel (a) shows a hypothetical observation field for the true field (b). Panel (c) shows the RS part of the solution ($E\hat{\alpha}$ in terms of equation ()). Panel (d) has the result of approximate Bayesian kriging ($\Delta\hat{T}$) added

December 1986



Small-scale SST variability: spatial autocorrelations




Correlations between Darwin and Tahiti seasonal atmospheric pressure
Marine OI
Ensemble median
Station obs
correlation*(−1), 40 yr window
0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
Time

Take home points

- Spagetti-western properties of least-squares estimates of spectrally red signals: (good) can be approximated by a few modes, (bad) have less variance than the true signal, and (ugly) redder than the true signal.
- Since the effect of these properties is stronger for poor data, and the data quality generally improves with time, use of least-squares analyses at face value, as if they were the truth, poses a threat of misinterpretation.
- A possible way out (however expensive): use of ensembles drawn from the posterior distributions rather than a single ensemble mean.

Further Work

Scale separation approach allows to work towards conceptually uniform globally-complete high-resolution objective analyses of SST according to the following scheme:

(a) start from the reduced-space analysis (with an assumption of stationary mean and covariance);

(b) small-scale analysis of observational residuals;

(c) recomputing non-stationary mean and covariance;

(d)adding high-resolution corrections, globalization patches, and sea-ice analyzed fields by the same scheme: large-scale prediction from the SST fields + local-scale corrections.