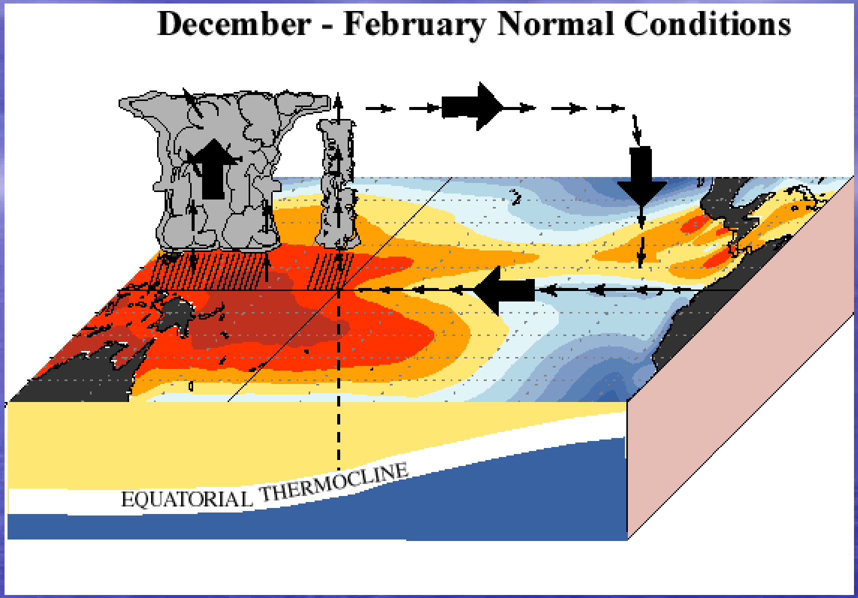
Utility of state-space reduction in ocean and climate inverse problems

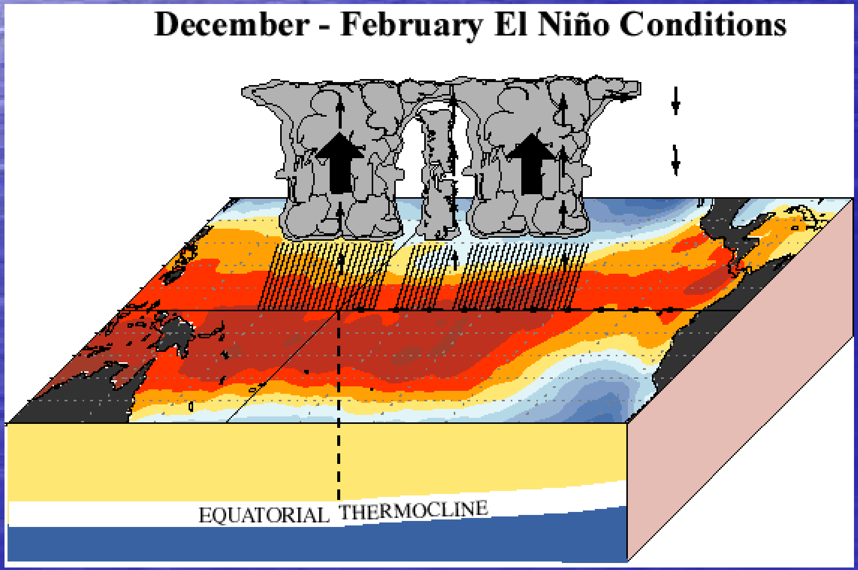
Alexey Kaplan

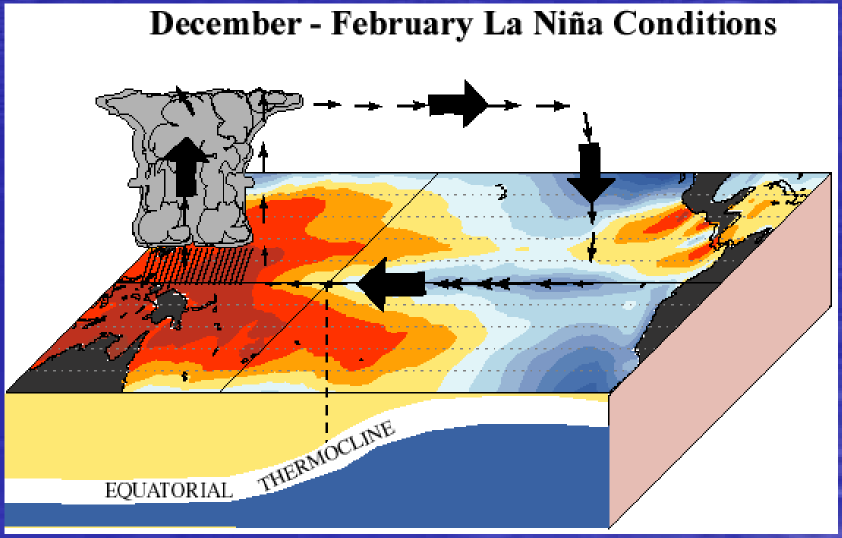
Lamont-Doherty Earth Observatory (LDEO) of Columbia University

In collaboration with: Mark Cane, Yochanan Kushnir, Benno Blumenthal (all LDEO), and Mike Evans (University of Arizona)

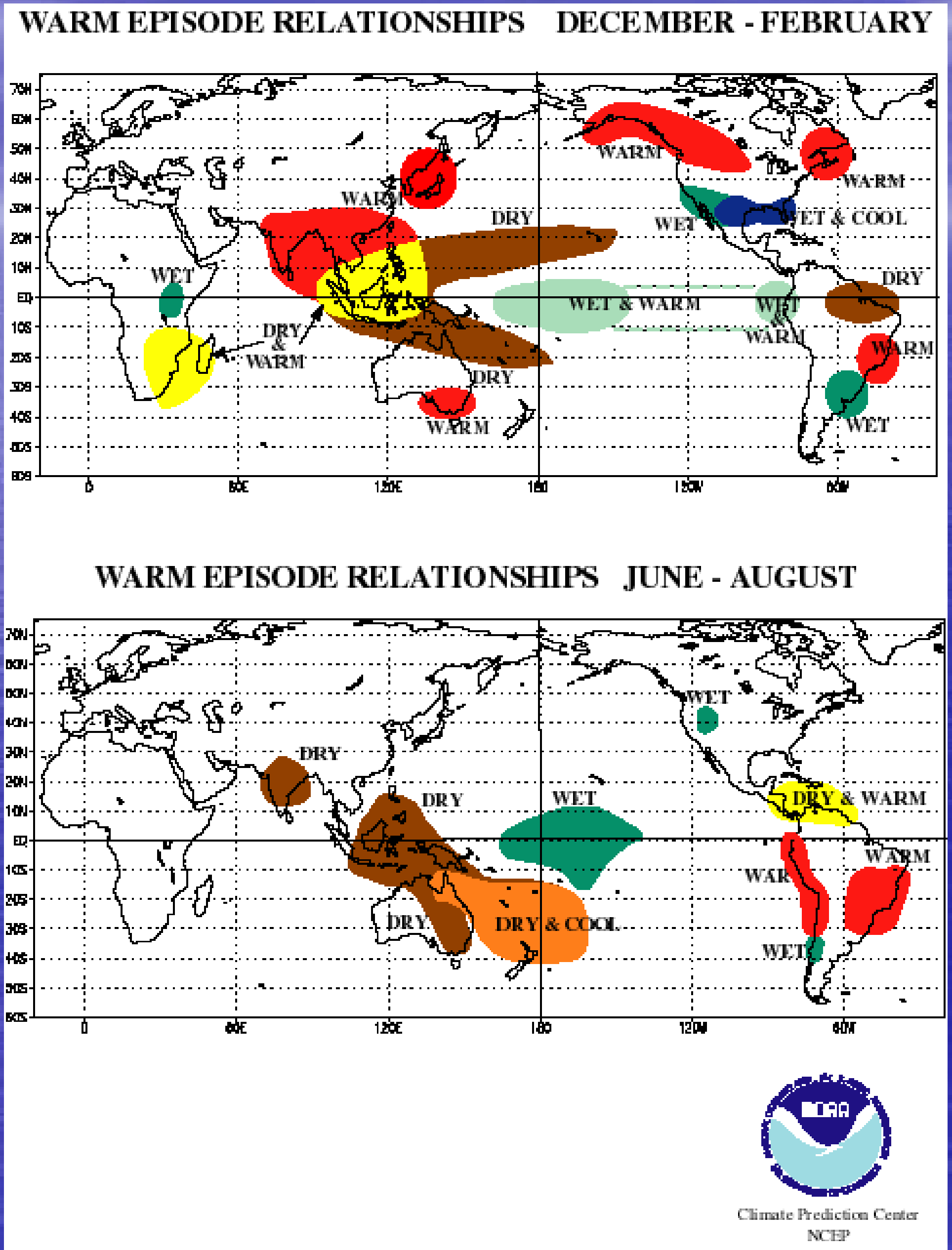
Dynamics of El Niño – Southern Oscillation



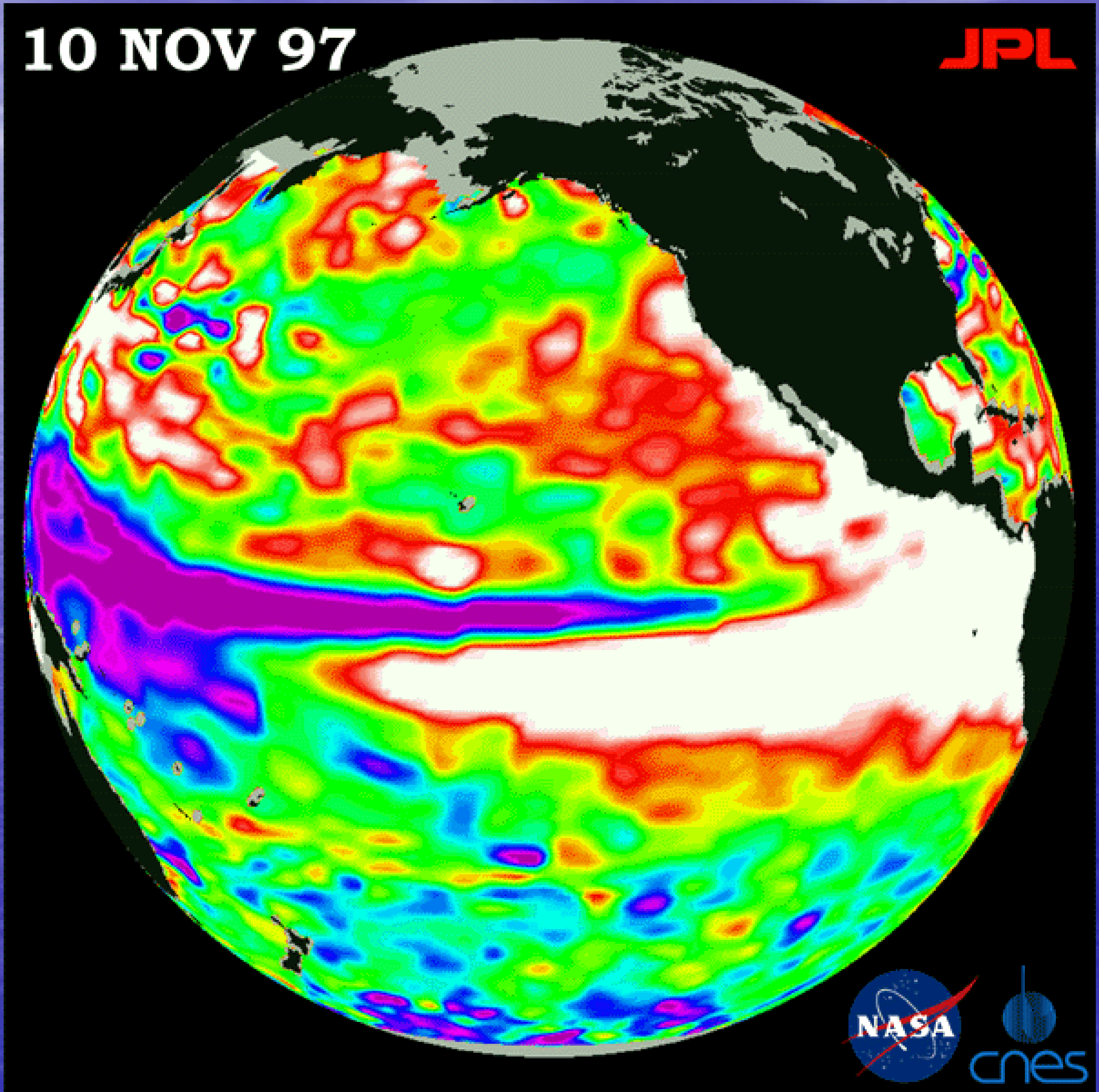


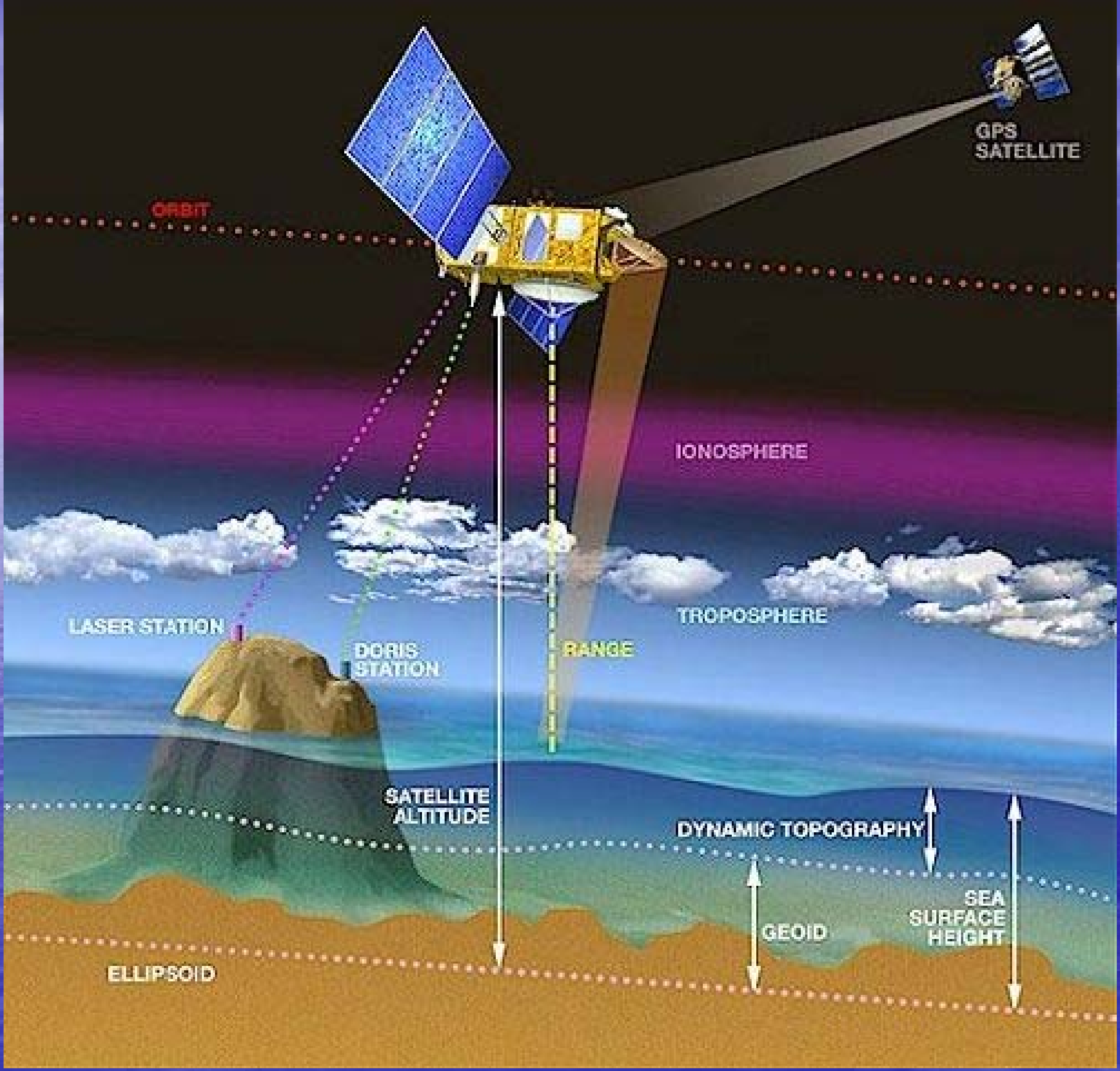


Global Impacts of El Niño



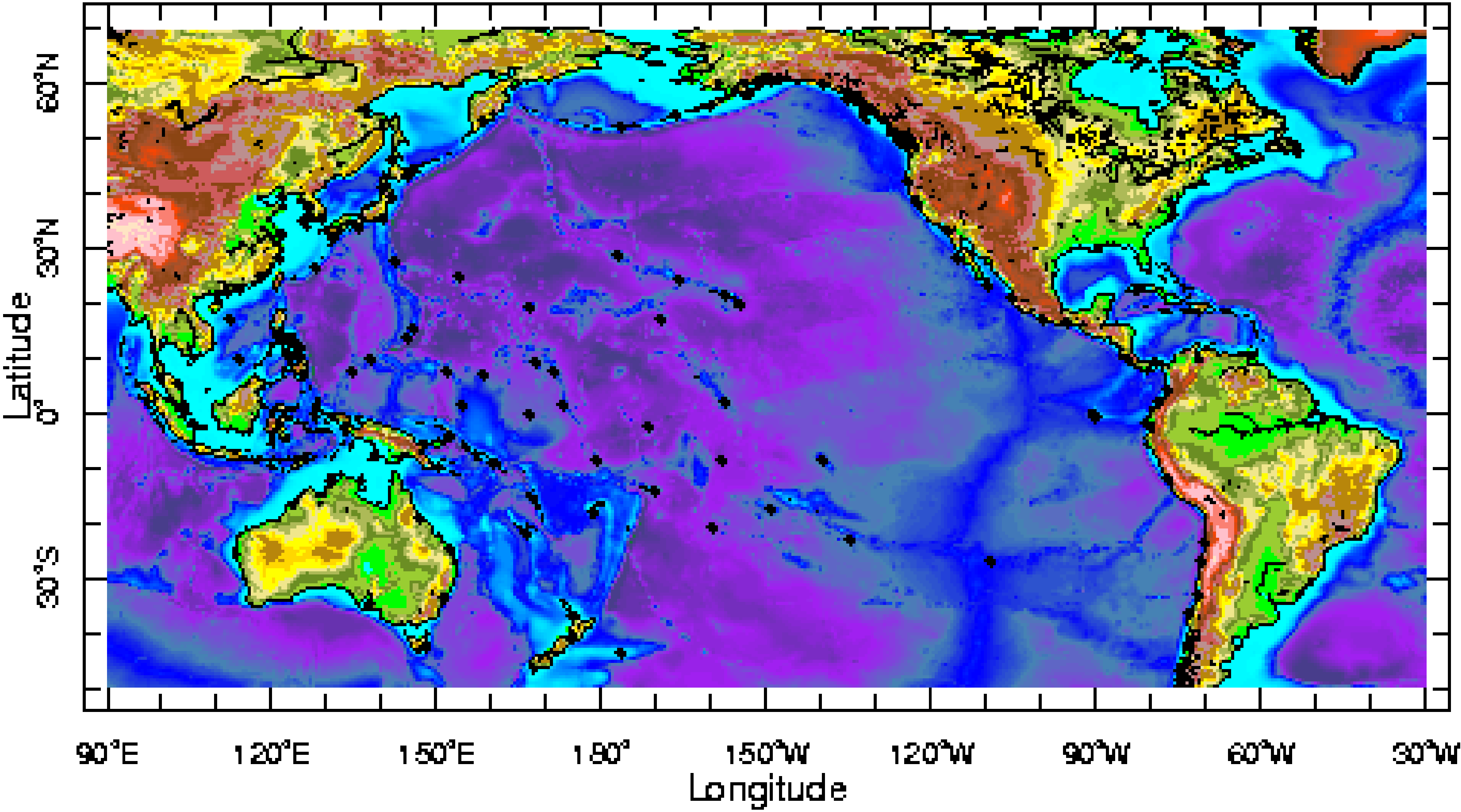
Sea Surface Height Anomaly



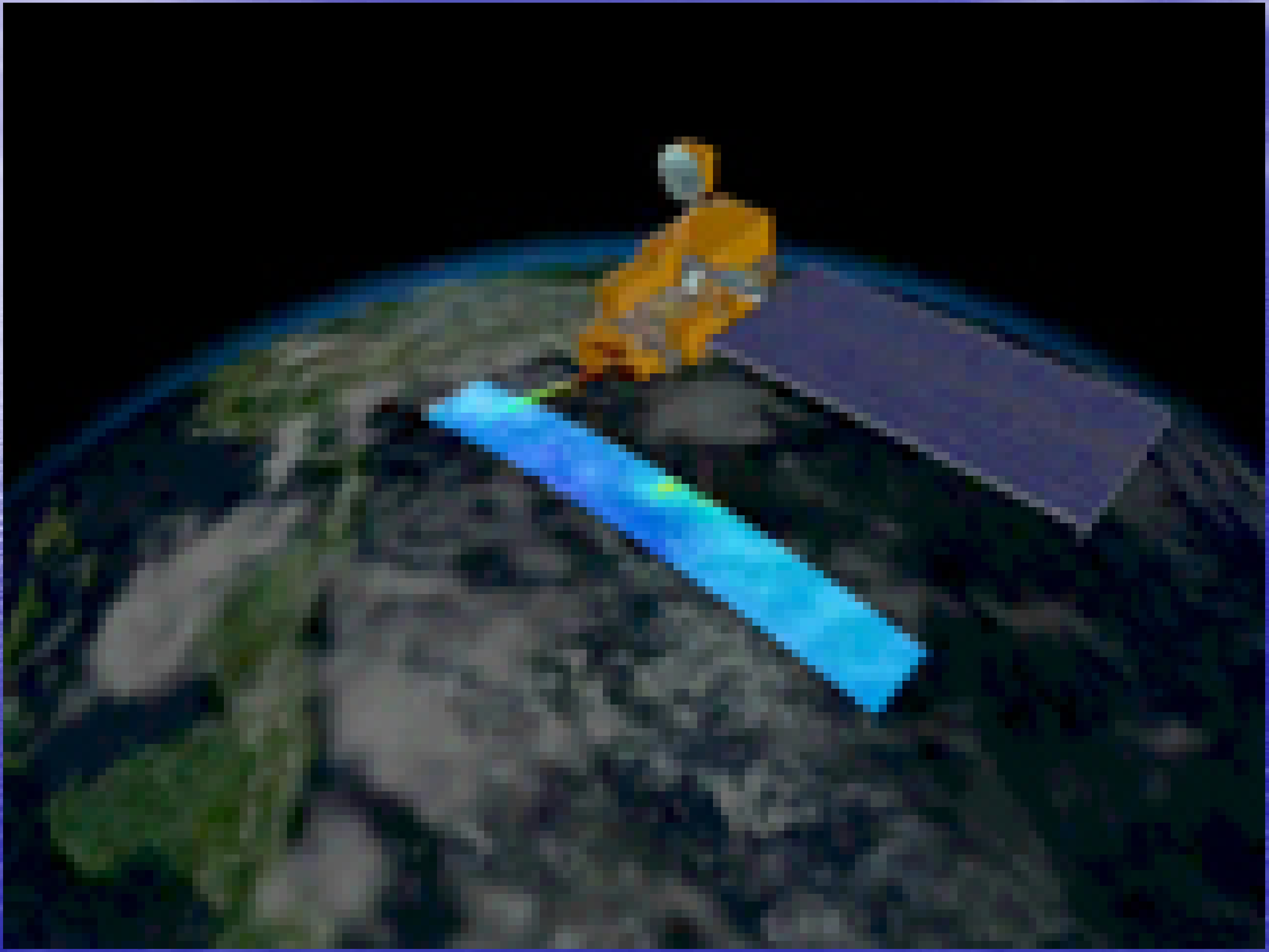


Altimetry principle from http://www.jason.oceanobs.com/html/alti/principe_uk.html

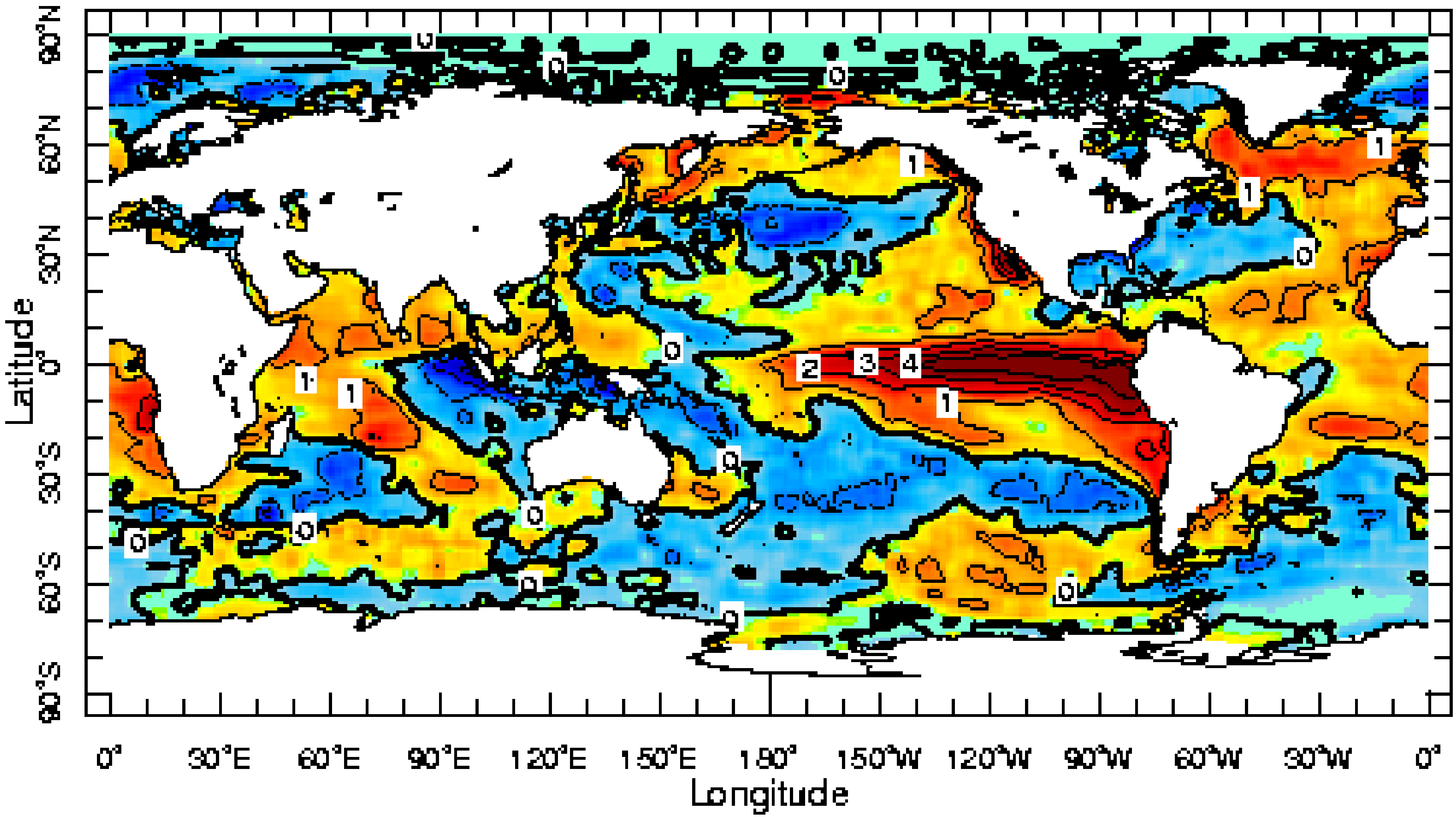
Pacific Ocean tide gauge stations network (from University of Hawaii sea level center)



MODIS Scanning Swath



Sea Surface Temperature Anomaly



9-15 Nov 1997

ABSTRACT LOG

RECOMMENDED BY THE

MARITIME CONFERENCE OF BRUSSELS,

Kept on board the U. S.

Commander, during the years 185

Indexed

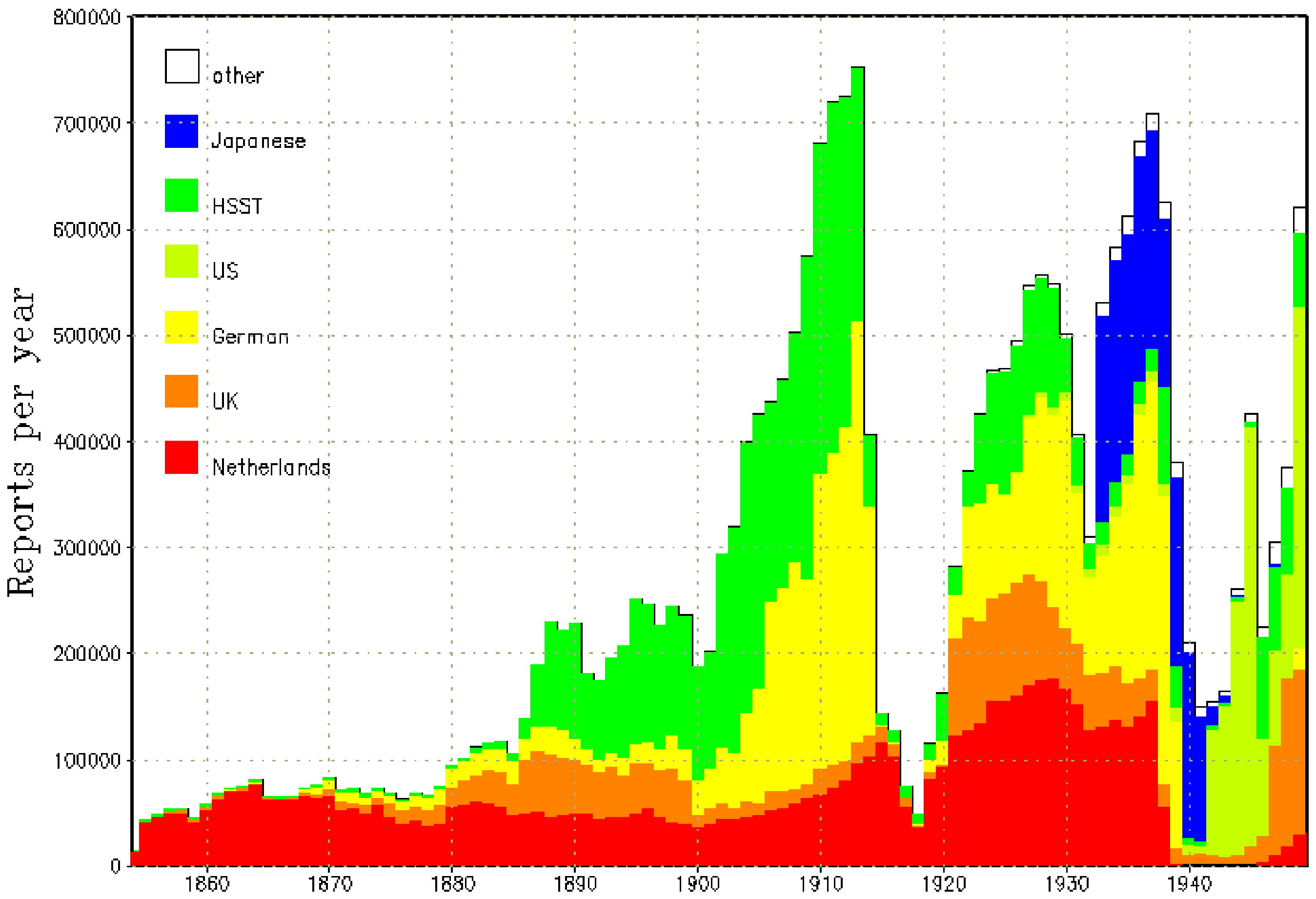
GENERAL ORDER.

Navy Department,

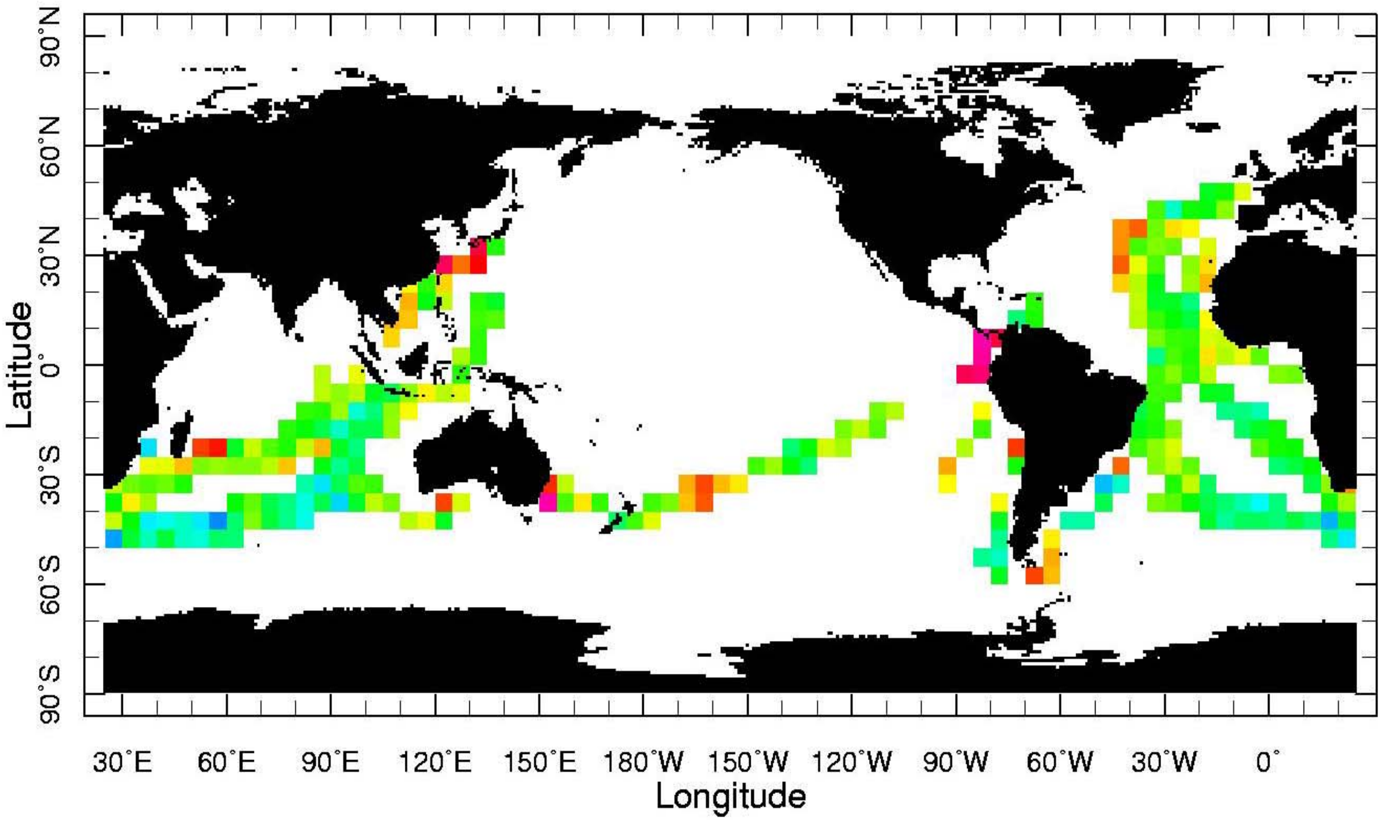
Signed, J. C. DOBBIN,

Secretary of the Navy.

Number of observations in COADS

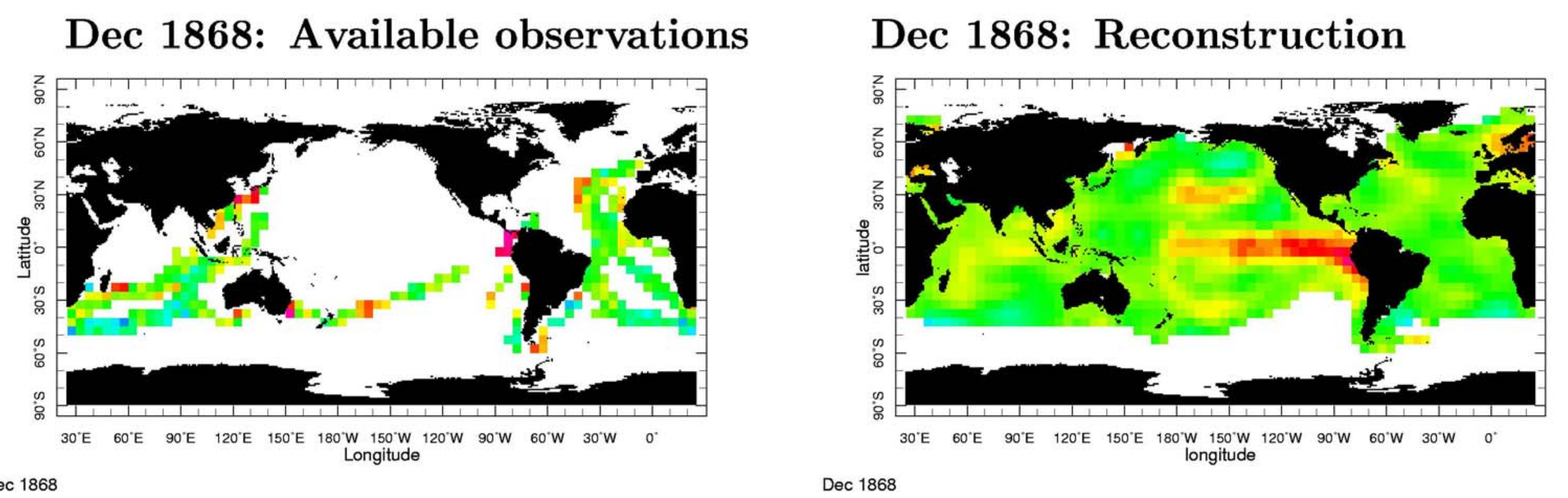


Dec 1868: Available observations

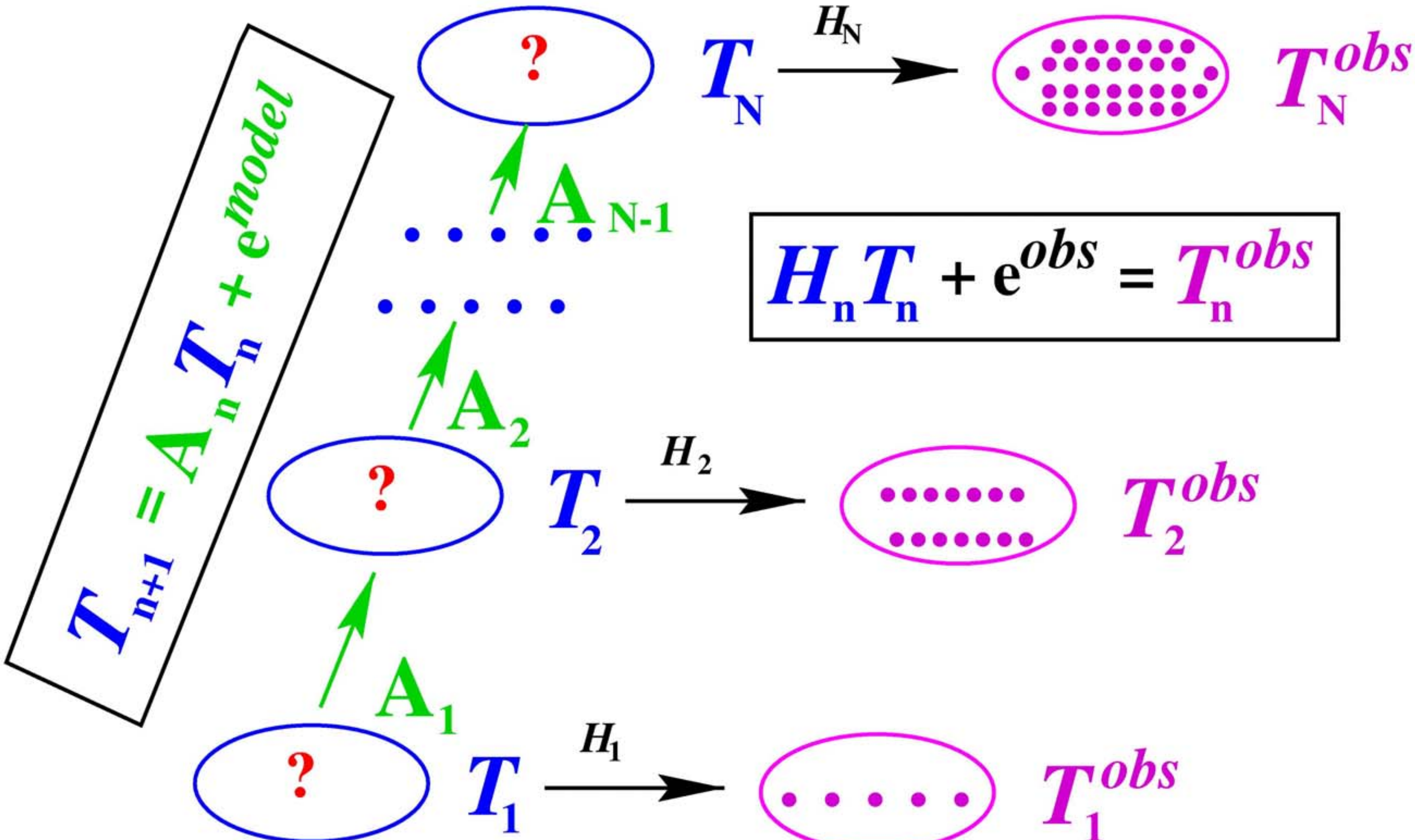


Dec 1868

Given a choice, climatologists in general would rather use the righthand panel below than the lefthand one



Generic problem of the analysis of time-evolving fields



Carl Friedrich Gauss

(1777–1855)

Discovery of least-squares estimation method: 1795

Gauss–Markov Theorem

If $\mathcal{T}^o = H\mathcal{T} + \varepsilon$,
$\langle \varepsilon \rangle = 0, \quad \langle \varepsilon\varepsilon^T \rangle = R, \quad \langle \varepsilon \mathcal{T}^T \rangle = 0,$
then the Least Squares Estimate (LSE)

$$\hat{\mathcal{T}} = (H^T R^{-1} H)^{-1} H^T R^{-1} \mathcal{T}^o$$

minimizes

$$S[\mathcal{T}] = (H\mathcal{T} - \mathcal{T}^o)^T R^{-1} (H\mathcal{T} - \mathcal{T}^o)$$

and is the **Best**[1] **Linear Unbiased Estimate (BLUE)** with error covariance

$$P \stackrel{\text{def}}{=} \langle (\mathcal{T} - \hat{\mathcal{T}})(\mathcal{T} - \hat{\mathcal{T}})^T \rangle = (H^T R^{-1} H)^{-1}.$$

ε is normal $\Longrightarrow$ $\mathcal{T}$ is a **Maximum Likelihood Estimate (MLE)**
ε and $\mathcal{T}$ are normal $\Longrightarrow$ $\mathcal{T}$ is the best among **all** (not necessarily linear) estimates.

---

[1] $\|\mathcal{T} - \hat{\mathcal{T}}\|_S^2 = \langle (\mathcal{T} - \hat{\mathcal{T}})^T S (\mathcal{T} - \hat{\mathcal{T}}) \rangle \longrightarrow \min \quad \forall S \Rightarrow$ minimal variance

To combine various sorts of data:
$T=(P^{-1}+R^{-1})^{-1}(P^{-1}M+R^{-1}D)$

Data Assimilation: Optimal Interpolation (OI), Kalman Filter (KF), Optimal Smoother (OS)

Right on the money

GENERAL PROBLEM OF RECONCILING MODELS WITH DATA

$$\mathcal{T}_{n+1} = A_n \mathcal{T}_n + \varepsilon_n^{\mathrm{m}}, \quad n = 1, \ldots, N-1$$
$$\mathcal{T}_n^o = H_n \mathcal{T}_n + \varepsilon_n^o, \quad n = 1, \ldots, N.$$

$$\langle \varepsilon_n^o \rangle = 0, \qquad \langle \varepsilon_n^o \varepsilon_n^{o\,T} \rangle = R_n, \quad n = 1, \ldots, N$$
$$\langle \varepsilon_{n_1}^o \varepsilon_{n_2}^{o\,T} \rangle = 0, \quad n_1 \neq n_2, \quad \langle \varepsilon_{n_1}^o \mathcal{T}_{n_2}^T \rangle = 0, \quad n_1, n_2 = 1, \ldots, N,$$
$$\langle \varepsilon_n^{\mathrm{m}} \rangle = 0, \qquad \langle \varepsilon_n^{\mathrm{m}} \varepsilon_n^{\mathrm{m}\,T} \rangle = Q_n, \quad n = 1, \ldots, N-1$$
$$\langle \varepsilon_{n_1}^{\mathrm{m}} \varepsilon_{n_2}^{\mathrm{m}\,T} \rangle = 0, \quad n_1 \neq n_2, \quad \langle \varepsilon_{n_1}^{\mathrm{m}} \mathcal{T}_{n_2}^T \rangle = 0, \quad n_1, n_2 = 1, \ldots, N-1$$
$$\langle \varepsilon_{n_1}^o \varepsilon_{n_2}^{\mathrm{m}\,T} \rangle = 0, \quad n_1 = 1, \ldots, N, \qquad n_2 = 1, \ldots, N-1$$

MINIMIZATION OF THE FULL COST FUNCTION:

$$\mathbf{S}[\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N] = \Sigma_{n=1}^{N} (H_n \mathcal{T}_n - \mathcal{T}_n^o)^T R_n^{-1} (H_n \mathcal{T}_n - \mathcal{T}_n^o) +$$
$$\Sigma_{n=1}^{N-1} (\mathcal{T}_{n+1} - A_n \mathcal{T}_n)^T Q_n^{-1} (\mathcal{T}_{n+1} - A_n \mathcal{T}_n)$$

OPTIMAL SMOOTHER (OS) and KALMAN FILTER (KF)

"Sweep up" – **KF**:

$$\hat{\mathcal{T}}_n^a = \hat{\mathcal{T}}_n^f + K_n \left(\hat{\mathcal{T}}_n^o - H_n \hat{\mathcal{T}}_n^f \right),$$
$$\hat{\mathcal{T}}_n^f = A_n \hat{\mathcal{T}}_{n-1}^a,$$
$$K_n = P_n^f H_n^T \left(H_n P_n^f H_n^T + R_n \right)^{-1}$$
$$P_n^a = (I_M - K_n H_n) P_n^f$$
$$P_n^f = A_{n-1} P_{n-1}^a A_{n-1}^T + Q_{n-1}, \qquad n = 2, 3, \ldots, N$$

"Sweep down" – **OS**:

$$\hat{\mathcal{T}}_n^s = \hat{\mathcal{T}}_n^a + G_n \left(\hat{\mathcal{T}}_{n+1}^s - A_n \hat{\mathcal{T}}_n^a \right), \qquad G_n = P_n^a A_n^T (P_{n+1}^f)^{-1},$$
$$P_n^s = P_n^a + G_n \left(P_{n+1}^s - P_{n+1}^f \right) G_n^T, \qquad n = N-1, \ldots, 2, 1$$

Example of Optimal Interpolation

$$T=T_B+e_B$$

$$HT=T_o+e_o$$

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

$$<e_B e_B^T>=C$$ **Hard to know in detail!**

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

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 of *C* (EOFs)

$C = EDE^T$

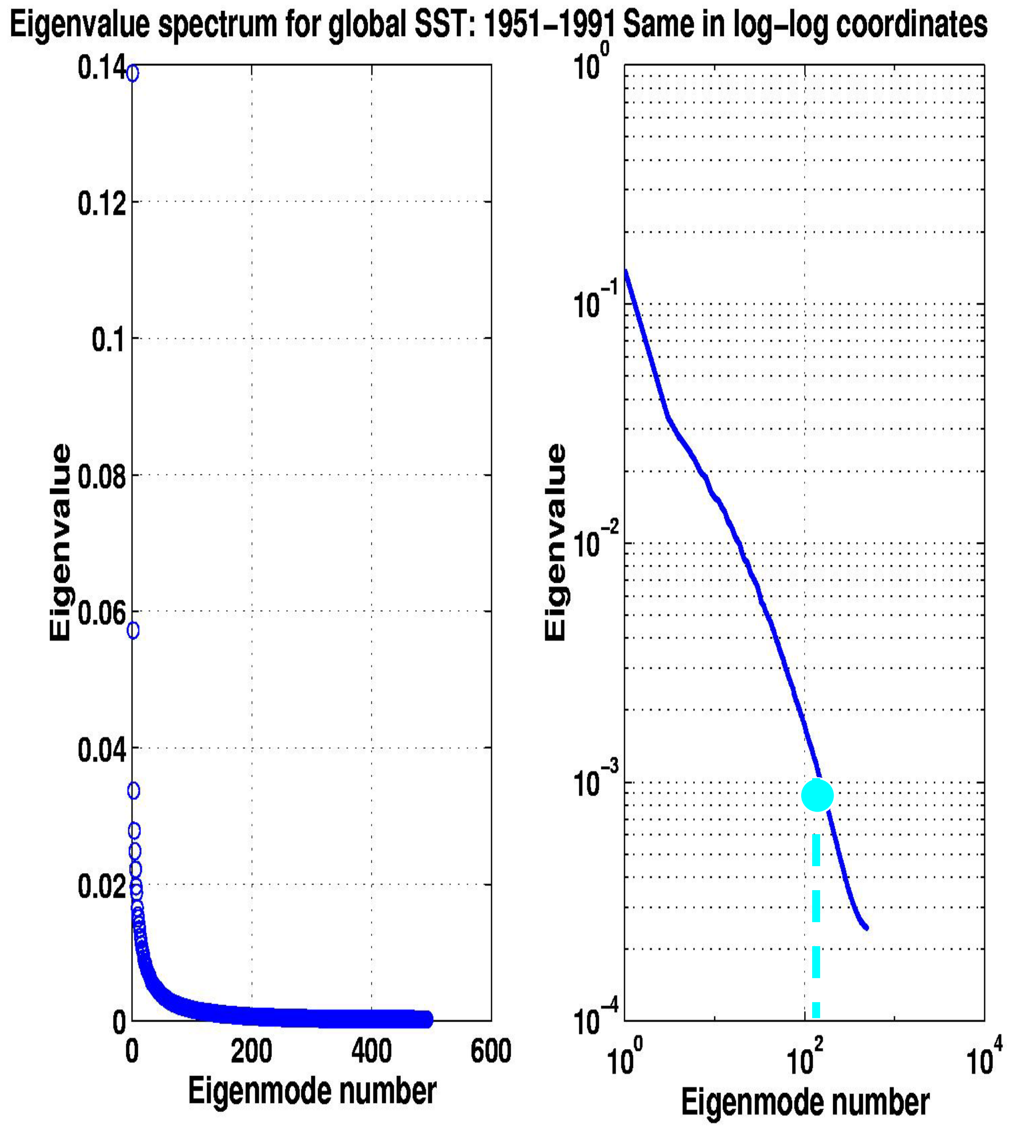
$T = Ea$

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

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

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

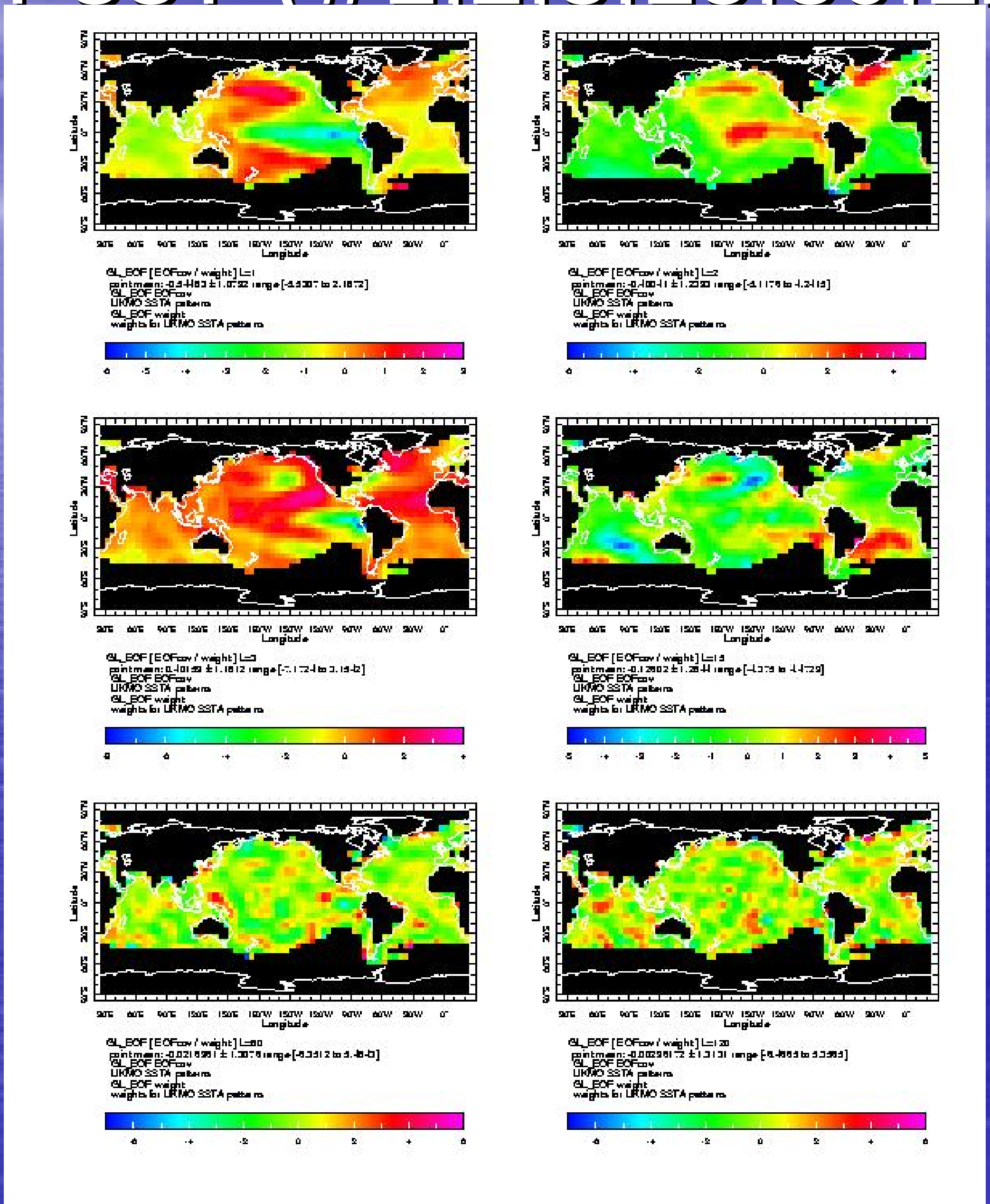
In many applications (for spectrally red signals) diagonal elements of this matrix decrease from ~1 to ~0. In effect, the solution is constrained to the subspace spanned by the patterns with $d_i >> r$.



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

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

EOF 1
14%



EOF 15
1%

EOF 80
0.1%

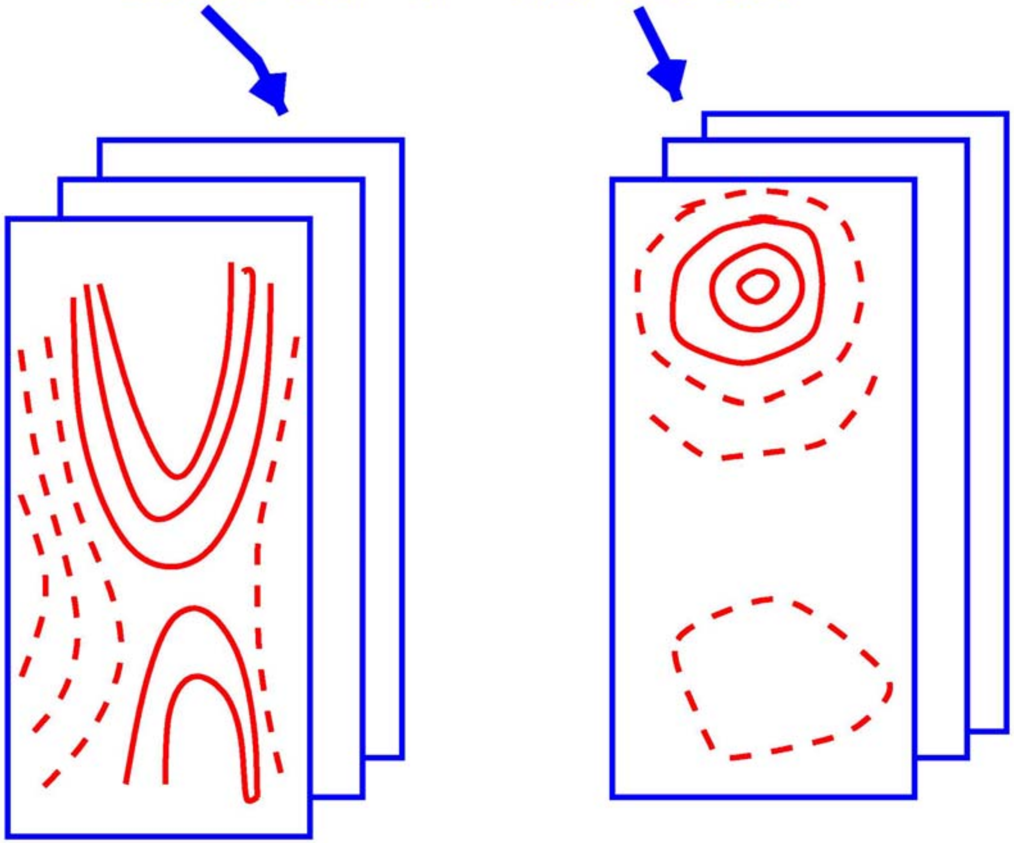
EOF 120
0.02%

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

SPACE REDUCTION

$$C = E\Lambda E^T + E'\Lambda' E'^{\,T}$$

$$\mathcal{T}_n = E\alpha_n + \varepsilon^r_n, \;\; n = 1, \ldots, N$$

ESTIMATION PROBLEM IN THE REDUCED SPACE

$$\mathcal{T}^o_n = H_n E\alpha_n + (H_n\varepsilon^r_n + \varepsilon^o_n) \stackrel{\text{def}}{=} \mathcal{H}_n\alpha_n + \check{\varepsilon}^o_n, \quad n = 1, \ldots, N,$$

$$\alpha_{n+1} = \mathcal{A}_n\alpha_n + E^T\varepsilon^{\mathrm{m}}_n \stackrel{\text{def}}{=} \mathcal{A}_n\alpha_n + \check{\varepsilon}^{\mathrm{m}}_n, \quad n = 1, \ldots, N-1.$$

$$\mathcal{Q}_n = \langle \check{\varepsilon}^{\mathrm{m}}_n \check{\varepsilon}^{\mathrm{m}\,T}_n \rangle = E^T \langle \varepsilon^{\mathrm{m}}_n \varepsilon^{\mathrm{m}\,T}_n \rangle E = E^T Q_n E$$

$$\mathcal{R}_n = \langle \check{\varepsilon}^o_n \check{\varepsilon}^{o\,T}_n \rangle = \langle (H_n\varepsilon^r_n + \varepsilon^o_n)(H_n\varepsilon^r_n + \varepsilon^o_n)^T \rangle =$$
$$\langle \varepsilon^o_n \varepsilon^{o\,T}_n \rangle + H_n \langle \varepsilon^r_n \varepsilon^{r\,T}_n \rangle H^T_n \stackrel{\text{def}}{=} R_n + H_n Q^r H^T_n \stackrel{\text{def}}{=} R_n + R'_n.$$

REDUCED SPACE OPTIMAL ANALYSIS

Cost function:

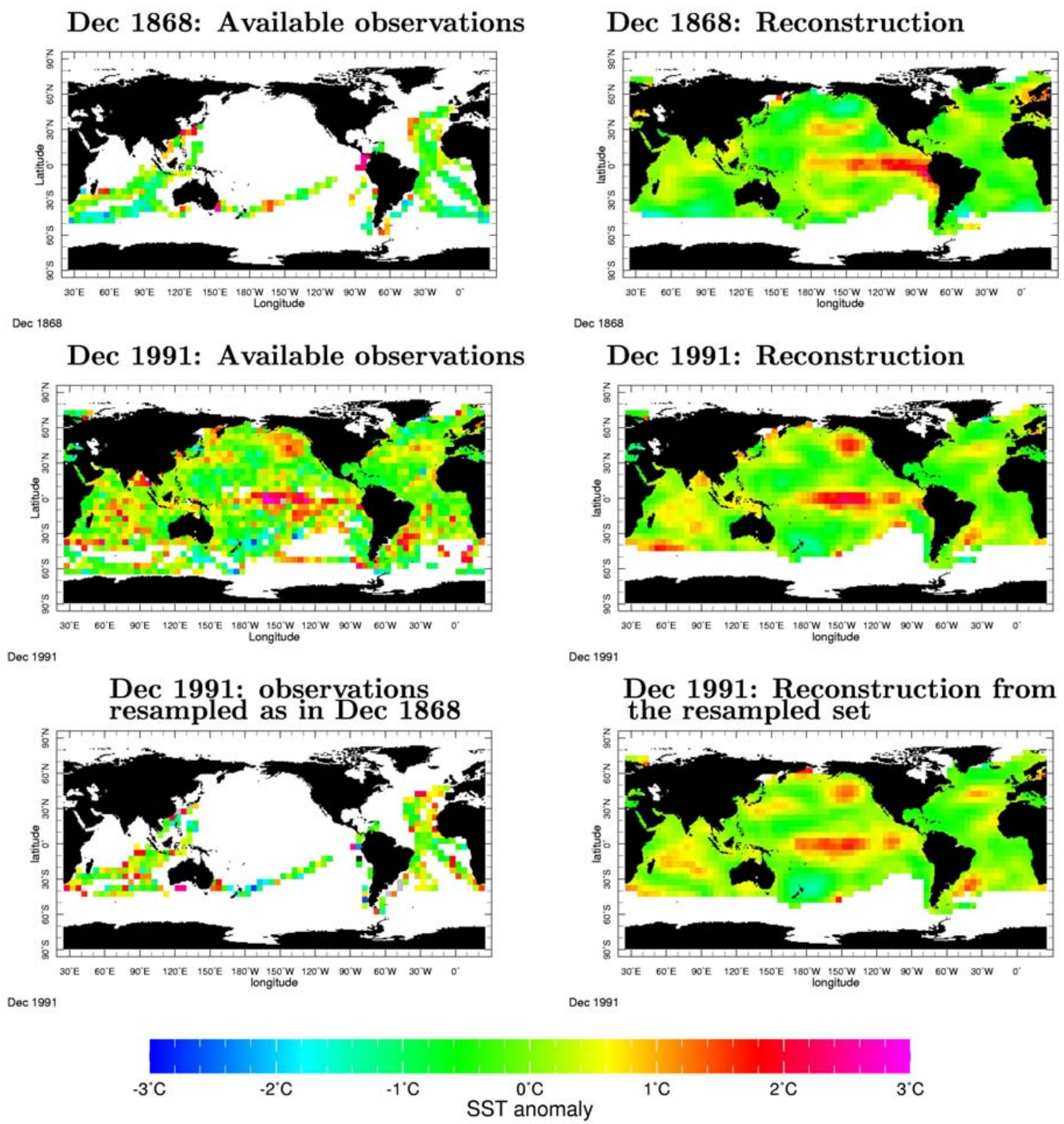
$$\mathcal{S}[\alpha_1, \alpha_2, \ldots, \alpha_N] = \Sigma_{n=1}^{N}(\mathcal{H}\alpha_n - \mathcal{T}_n^o)^T \mathcal{R}_n^{-1}(\mathcal{H}\alpha_n - \mathcal{T}_n^o) + \Sigma_{n=1}^{N-1}(\alpha_{n+1} - \mathcal{A}_n\alpha_n)^T \mathcal{Q}_n^{-1}(\alpha_{n+1} - \mathcal{A}_n\alpha_n).$$

KF:

$$\begin{aligned}
\alpha_n^a &= \alpha_n^f + \mathcal{K}_n\left(\mathcal{T}_n^o - \mathcal{H}_n\alpha_n^f\right), \\
\alpha_n^f &= \mathcal{A}_n\alpha_{n-1}^a, \\
\mathcal{K}_n &= \left(\mathcal{H}_n^T\mathcal{R}_n^{-1}\mathcal{H}_n + \mathcal{P}_n^{f\,-1}\right)^{-1}\mathcal{H}_n^T\mathcal{R}_n^{-1} \\
\mathcal{P}_n^a &= (I_L - \mathcal{K}_n\mathcal{H}_n)\mathcal{P}_n^f \\
\mathcal{P}_n^f &= \mathcal{A}_{n-1}\mathcal{P}_{n-1}^a\mathcal{A}_{n-1}^T + \mathcal{Q}_{n-1}, \qquad n = 2, 3, \ldots, N
\end{aligned}$$

OS:

$$\begin{aligned}
\alpha_n^s &= \alpha_n^a + G_n\left(\alpha_{n+1}^s - \mathcal{A}_n\alpha_n^a\right), \\
G_n &= \mathcal{P}_n^a\mathcal{A}_n^T(\mathcal{P}_{n+1}^f)^{-1}, \\
\mathcal{P}_n^s &= \mathcal{P}_n^a + G_n\left(\mathcal{P}_{n+1}^s - \mathcal{P}_{n+1}^f\right)G_n^T, \qquad n = N-1, \ldots, 2, 1
\end{aligned}$$


Dec 1868: Available observations
Dec 1868: Reconstruction
Dec 1991: Available observations
Dec 1991: Reconstruction
Dec 1991: observations resampled as in Dec 1868
Dec 1991: Reconstruction from the resampled set
Latitude
Longitude
90°N
60°N
30°N
0°
30°S
60°S
90°S
30°E
60°E
90°E
120°E
150°E
180°W
150°W
120°W
90°W
60°W
30°W
0°
Dec 1868
Dec 1991
-3°C
-2°C
-1°C
0°C
1°C
2°C
3°C
SST anomaly

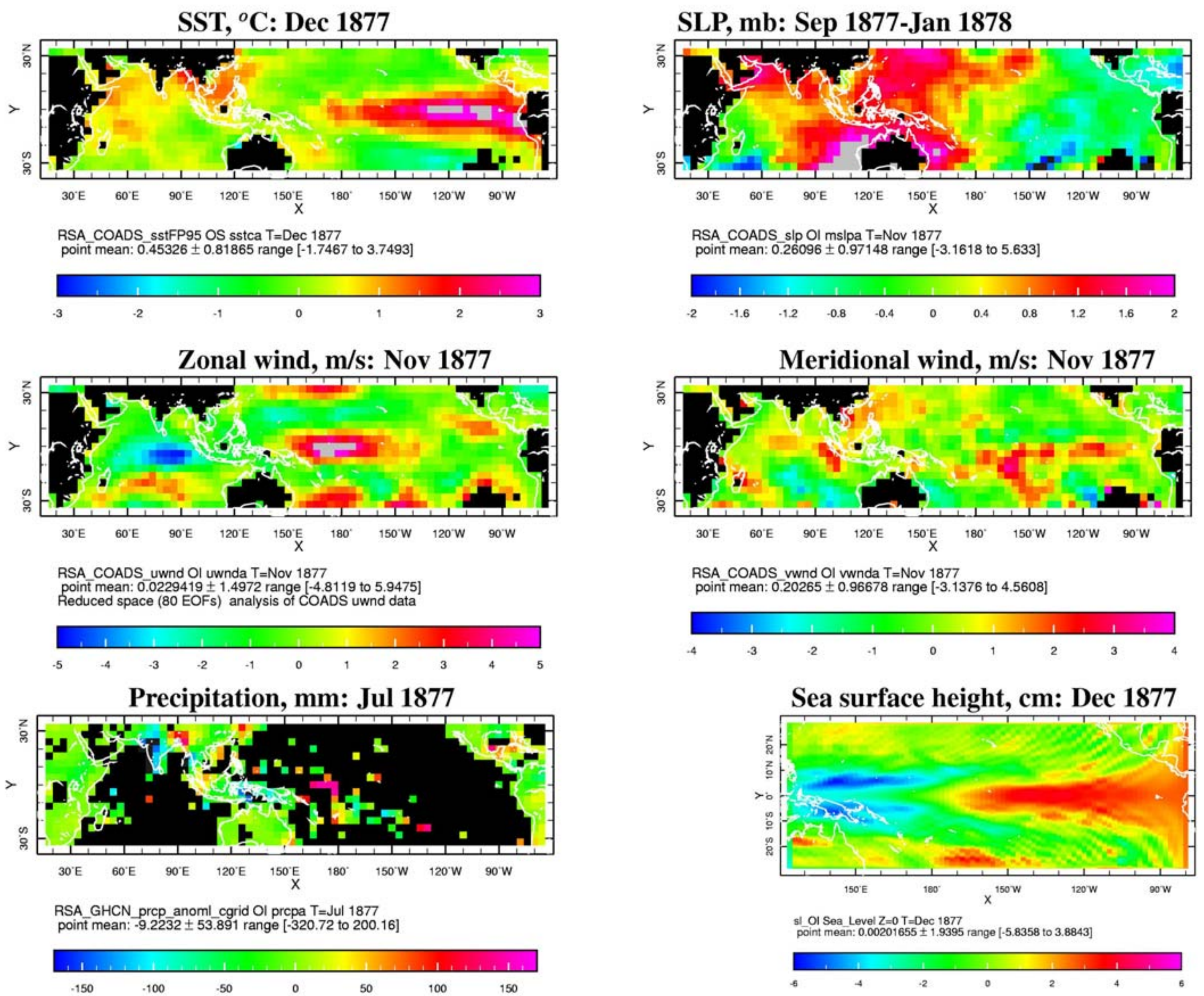
Kaplan, A., Y. Kushnir, M. Cane, and M. Blumenthal, 1997: Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures, *J. Geophys. Res*., 102, 27835--27860.

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B.Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991, *J. Geophys. Res*.,103, 18567-18589.

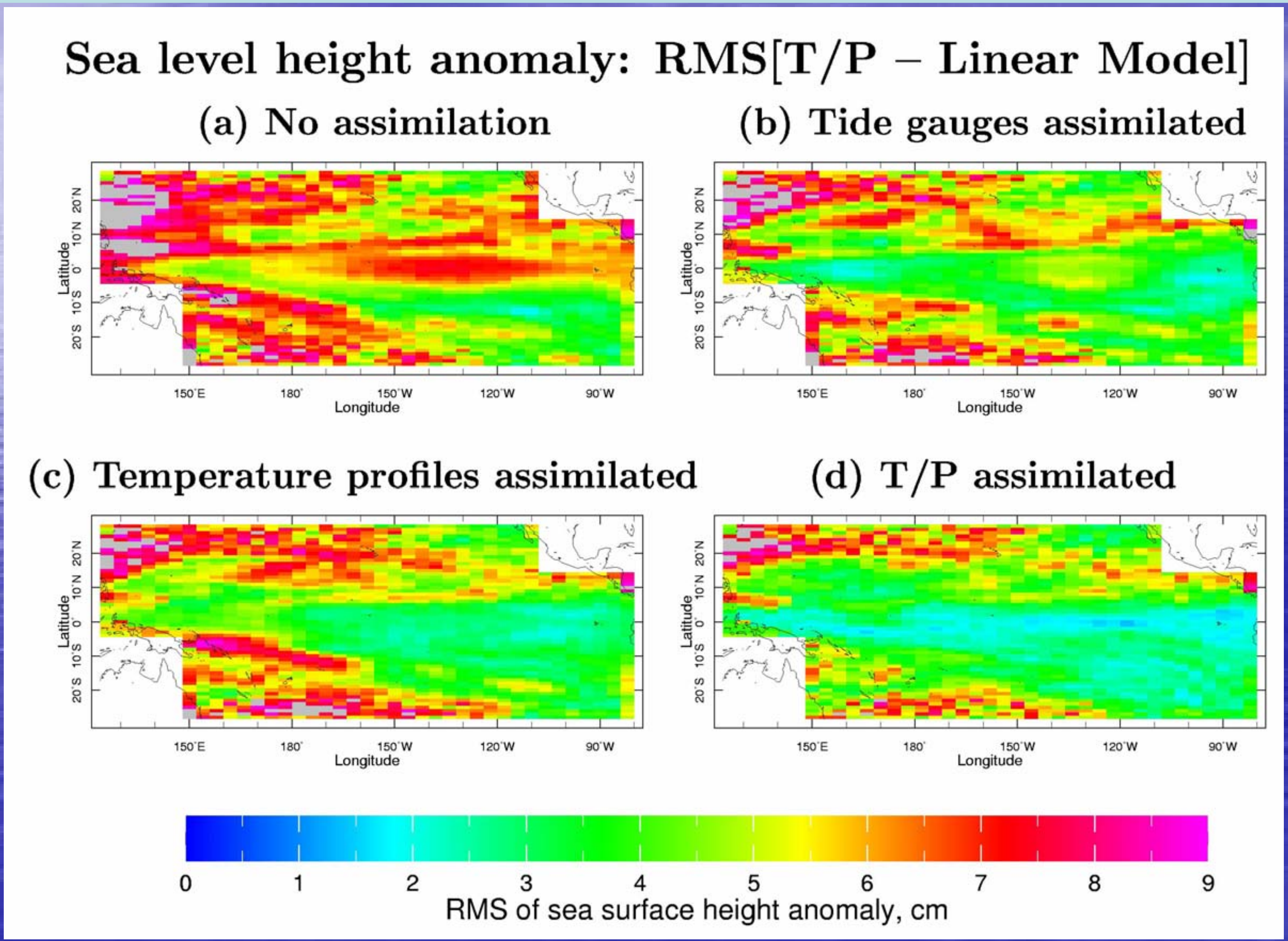
Kaplan A., Y. Kushnir, M.A. Cane, 2000: Reduced space optimal interpolation of historical marine sea level pressure: 1854-1992, *J. Climate*, 13, 2987-3002.

Kaplan A., M.A. Cane, and Y. Kushnir, 2003: Reduced space approach to the optimal analysis interpolation of historical marine observations: Accomplishments, difficulties, and prospects, in *Advances in the Applications of Marine Climatology: The Dynamic Part of the WMO Guide to the Applications of Marine Climatology*, WMO/TD-1081, pp. 199-216

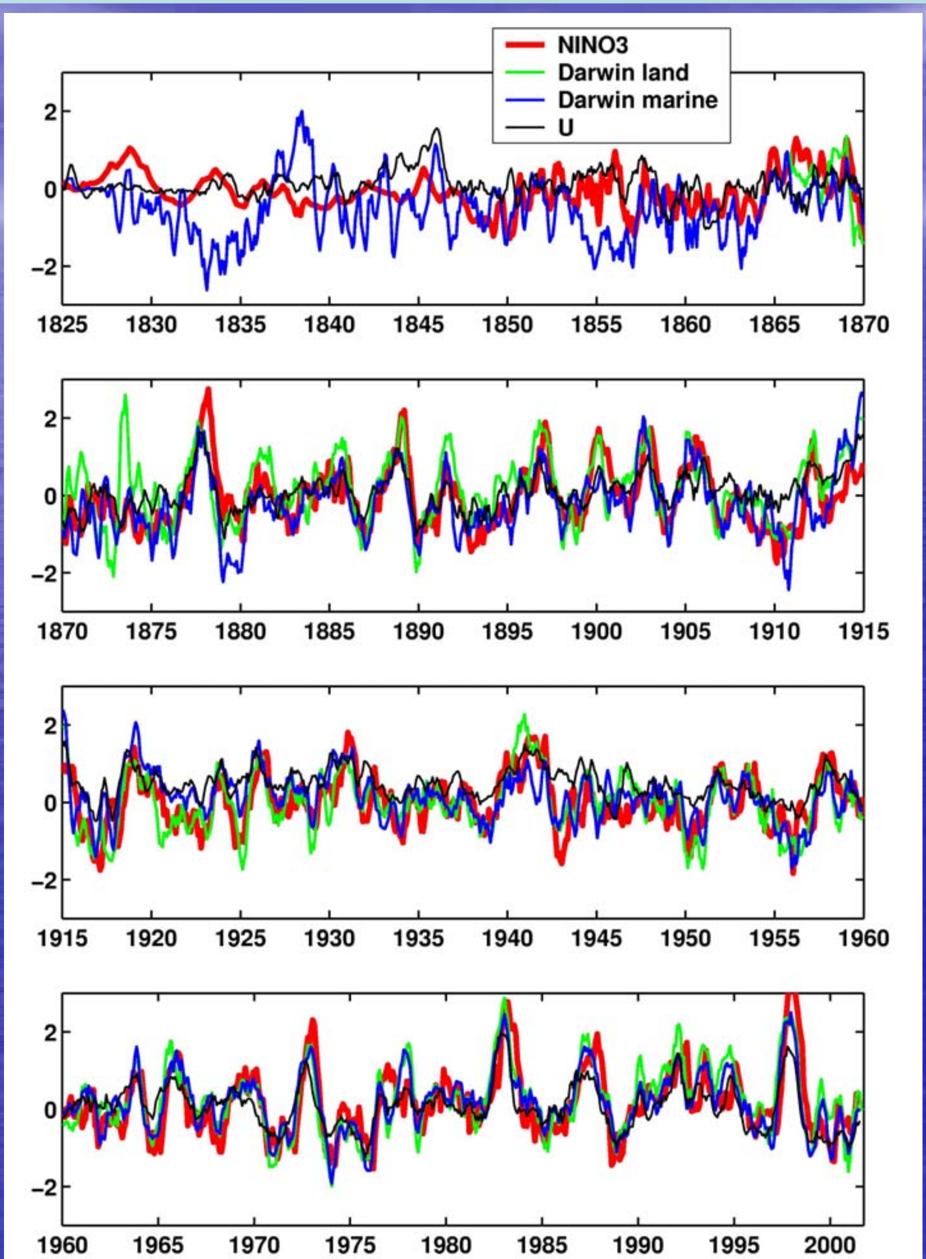
El Niño of 1877-1878 in analyzed anomalies



Cane, M.A., **A. Kaplan**, R.N. Miller, B. Tang, E.C. Hackert, and A.J.Busalacchi, 1996: Mapping tropical Pacific sea level: data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.*, **101**, 22599--22617.
Reverdin, G., **A. Kaplan**, and M. Cane, 1996: Sea level from temperature profiles in the tropical Pacific Ocean 1975--1992. *J. Geophys. Res.*, **101**, 18105-18119.
Kaplan, A., M.A. Cane, D. Chen, D.L. Witter, and R.E. Cheney, 2004: Small-scale variability and model error in tropical Pacific sea level, *J. Geophys. Res.*, **109**, C02001, doi:10.1029/2002JC001743.



Evans, M.N. and A. Kaplan, 2004: The Pacific sector Hadley and Walker Circulations in historical marine wind analyses, in *The Hadley Circulation: Present, Past and Future*. H.F.Diaz and R.S. Bradley (Eds.), Kluwer Academic Publishers, Netherland, 239-258.



3 corollaries:

[of $D(D+R)^{-1}$=diag[$d_i/(d_i+r)$] factor]

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

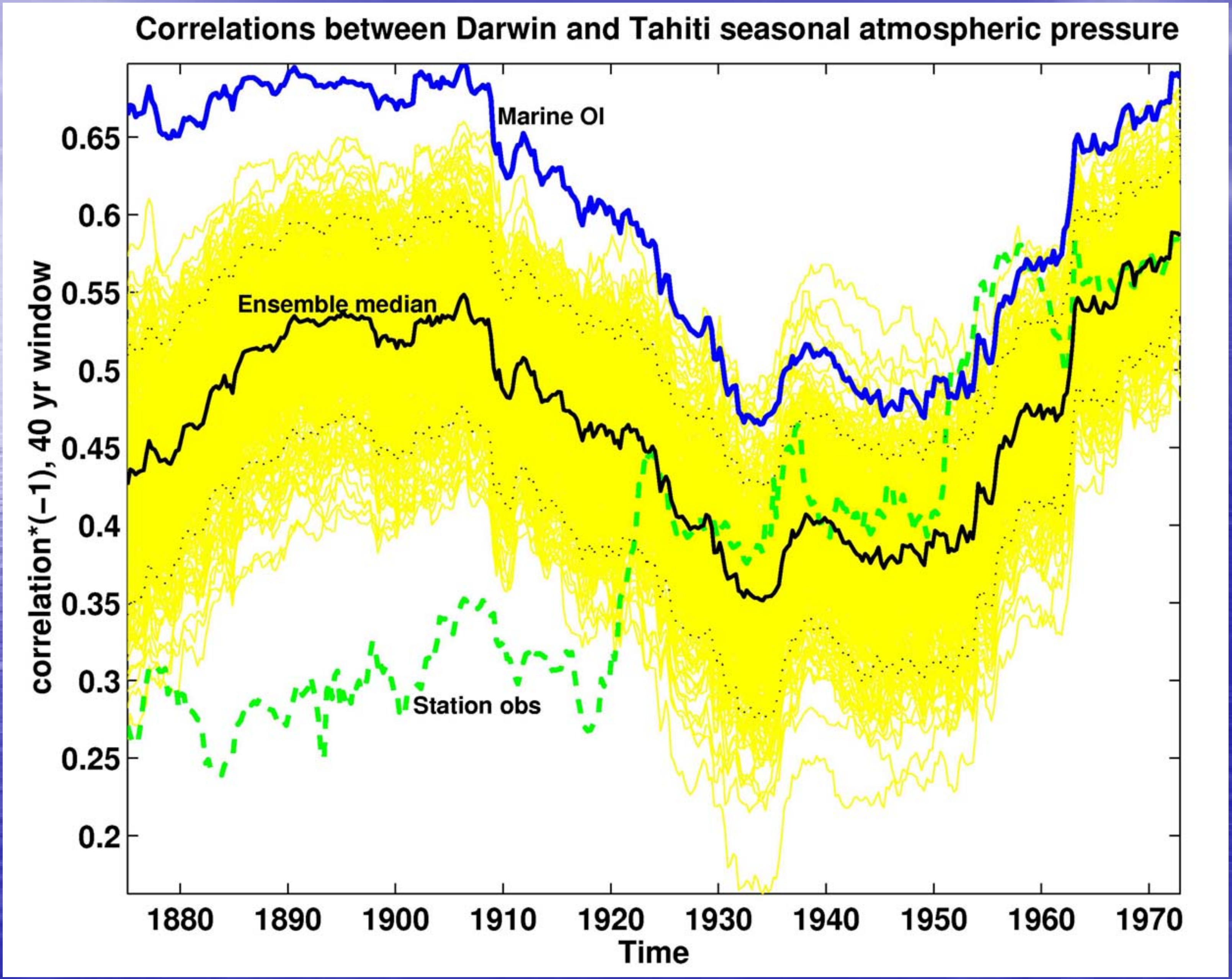
3 corollaries:

[of $D(D+R)^{-1}$=diag[$d_i/(d_i+r)$] factor]

- 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

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


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.
- These properties can be used for making analyses of sparse climate data cheaper and less ambiguous in their setup.
- 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.