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

December - February Normal Conditions

December - February El Niño Conditions

December - February La Niña Conditions
Global Impacts of El Niño
Sea Surface Height Anomaly
Pacific Ocean tide gauge stations network (from University of Hawaii sea level center)
MODIS Scanning Swath
Sea Surface Temperature Anomaly

9-15 Nov 1997
Number of observations in COADS
Dec 1868: Available observations
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

\[ T_{n+1} = A_n T_n + e_{model} \]

\[ H_n T_n + e^{obs} = T^{obs}_n \]

\[ H_1 T_1 = T^{obs}_1 \]

\[ H_2 T_2 = T^{obs}_2 \]

\[ H_N T_N = T^{obs}_N \]
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)

---

**Discovery of least-squares estimation method: 1795**

**Gauss–Markov Theorem**

If \( T^o = HT + \varepsilon \),
\[ \langle \varepsilon \rangle = 0, \quad \langle \varepsilon \varepsilon^T \rangle = R, \quad \langle \varepsilon T^T \rangle = 0, \]

then the Least Squares Estimate (LSE)

\[ \hat{T} = (H^T R^{-1} H)^{-1} H^T R^{-1} T^o \]

minimizes

\[ S[T] = (HT - T^o)^T R^{-1} (HT - T^o) \]

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

\[ P \overset{\text{def}}{=} \langle (T - \hat{T})(T - \hat{T})^T \rangle = (H^T R^{-1} H)^{-1}. \]

\[ \varepsilon \text{ is normal } \implies T \text{ is a Maximum Likelihood Estimate (MLE)} \]

\[ \varepsilon \text{ and } T \text{ are normal } \implies T \text{ is the best among all (not necessarily linear) estimates.} \]

---

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

---

---

(1777–1855)
Right on the money
GENERAL PROBLEM OF RECONCILING MODELS WITH DATA

\[ \mathcal{T}_{n+1} = A_n \mathcal{T}_n + e_n^m, \quad n = 1, \ldots, N - 1 \]
\[ \mathcal{T}_n^o = H_n \mathcal{T}_n + e_n^o, \quad n = 1, \ldots, N. \]

\[ \langle e_n^o \rangle = 0, \quad \langle e_n^o e_n^o T \rangle = R_n, \quad n = 1, \ldots, N \]
\[ \langle e_{n_1}^o e_{n_2}^o T \rangle = 0, \quad n_1 \neq n_2, \quad \langle e_{n_1}^o \mathcal{T}_{n_2}^T \rangle = 0, \quad n_1, n_2 = 1, \ldots, N, \]

\[ \langle e_n^m \rangle = 0, \quad \langle e_n^m e_n^m T \rangle = Q_n, \quad n = 1, \ldots, N - 1 \]
\[ \langle e_{n_1}^m e_{n_2}^m T \rangle = 0, \quad n_1 \neq n_2, \quad \langle e_{n_1}^m \mathcal{T}_{n_2}^T \rangle = 0, \quad n_1, n_2 = 1, \ldots, N - 1 \]

\[ \langle e_{n_1}^o e_{n_2}^m T \rangle = 0, \quad n_1 = 1, \ldots, N, \quad n_2 = 1, \ldots, N - 1 \]

MINIMIZATION OF THE FULL COST FUNCTION:

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

OPTIMAL SMOOTHER (OS) and KALMAN FILTER (KF)

“Sweep up” – KF:

\[ \hat{\mathcal{T}}_n^a = \hat{\mathcal{T}}_n^f + K_n (\mathcal{T}_n^o - H_n \hat{\mathcal{T}}_n^f), \]
\[ \hat{\mathcal{T}}_n^f = A_n \hat{\mathcal{T}}_{n-1}^a, \]
\[ K_n = P_n^f H_n^T (H_n P_n^f H_n^T + R_n)^{-1} \]
\[ P_n^a = (I_n - K_n H_n) P_n^f \]
\[ P_n^f = A_{n-1} P_{n-1}^a A_{n-1}^T + Q_{n-1}, \quad n = 2, 3, \ldots, N \]

“Sweep down” – OS:

\[ \hat{\mathcal{T}}_n^a = \hat{\mathcal{T}}_n^o + G_n (\mathcal{T}_{n+1}^o - A_n \hat{\mathcal{T}}_n^a), \quad G_n = P_n^o A_n^T (P_{n+1}^f)^{-1}, \]
\[ P_n^a = P_n^o + G_n (P_{n+1}^o - P_{n+1}^f) G_n^T, \quad n = N - 1, \ldots, 2, 1 \]
Example of Optimal Interpolation

\[ T = T_B + e_B \]
\[ HT = T_o + e_o \]
\[ <e_B> = <e_o> = <e_B e_o^T> = 0 \]
\[ <e_B e_B^T> = C \]
\[ <e_o e_o^T> = R \]

Solution minimizes the cost function
\[ S[T] = (HT - T_o)^T R^{-1} (HT - T_o) + (T - T_B)^T C^1 (T - T_B) \]

\[ T = (H^T R^{-1} H + C^1)^{-1} (H^T R^{-1} T_o + C^1 T_B) \]
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 $a=D(D+R)^{-1}E^TT_o$

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

In many applications (for spectrally red signals) diagonal elements of this matrix decrease from $\sim1$ to $\sim0$. In effect, the solution is constrained to the subspace spanned by the patterns with $d_i\gg r$. 
Eigenvalue spectrum for global SST: 1951–1991 Same in log–log coordinates
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.
$C = EE^T + E'E'^T$

Reduced space optimal analysis

Successive corrections; Kriging
SPACE REDUCTION

\[ C = E \Lambda E^T + E' \Lambda' E'^T \]
\[ \mathcal{T}_n = E \alpha_n + \varepsilon^r_n, \quad n = 1, \ldots, N \]

ESTIMATION PROBLEM IN THE REDUCED SPACE

\[ \mathcal{T}_n^o = H_n E \alpha_n + (H_n \varepsilon^r_n + \varepsilon^o_n) \overset{\text{def}}{=} \mathcal{H}_n \alpha_n + \tilde{\varepsilon}^o_n, \quad n = 1, \ldots, N, \]

\[ \alpha_{n+1} = A_n \alpha_n + E^T \varepsilon^m_n \overset{\text{def}}{=} A_n \alpha_n + \tilde{\varepsilon}^m_n, \quad n = 1, \ldots, N - 1. \]

\[ Q_n = \langle \tilde{\varepsilon}^m_n \tilde{\varepsilon}^m_n^T \rangle = E^T \langle \varepsilon^m_n \varepsilon^m_n^T \rangle E = E^T Q_n E \]

\[ R_n = \langle \tilde{\varepsilon}^o_n \tilde{\varepsilon}^o_n^T \rangle = \langle (H_n \varepsilon^r_n + \varepsilon^o_n)(H_n \varepsilon^r_n + \varepsilon^o_n)^T \rangle = \langle \varepsilon^o_n \varepsilon^o_n^T \rangle + H_n \langle \varepsilon^r_n \varepsilon^r_n^T \rangle H_n^T \overset{\text{def}}{=} R_n + H_n Q^r H_n^T \overset{\text{def}}{=} R_n + R'_n. \]
REDUCED SPACE OPTIMAL ANALYSIS

Cost function:

\[ S[\alpha_1, \alpha_2, \ldots, \alpha_N] = \sum_{n=1}^{N} (H\alpha_n - T_n^o)^T R_n^{-1} (H\alpha_n - T_n^o) + \sum_{n=1}^{N-1} (\alpha_{n+1} - A_n\alpha_n)^T Q_n^{-1} (\alpha_{n+1} - A_n\alpha_n). \]

KF:

\[ \alpha_n^a = \alpha_n^f + K_n (T_n^o - H_n\alpha_n^f), \]
\[ \alpha_n^f = A_n\alpha_{n-1}, \]
\[ K_n = (H_n^T R_n^{-1} H_n + P_n^{f-1})^{-1} H_n^T R_n^{-1} \]
\[ P_n^a = (I_L - K_n H_n)P_n^f \]
\[ P_n^f = A_{n-1}P_{n-1}^a A_{n-1}^T + Q_{n-1}, \quad n = 2, 3, \ldots, N \]

OS:

\[ \alpha_n^s = \alpha_n^a + G_n (\alpha_{n+1}^s - A_n\alpha_n^a), \]
\[ G_n = P_n^a A_n^T (P_{n+1}^f)^{-1}, \]
\[ P_n^s = P_n^a + G_n (P_{n+1}^s - P_{n+1}^f) G_n^T, \quad n = N-1, \ldots, 2, 1 \]


---

**El Niño of 1877-1878 in analyzed anomalies**

**SST, °C: Dec 1877**

**SLP, mb: Sep 1877-Jan 1878**

**Zonal wind, m/s: Nov 1877**

**Meridional wind, m/s: Nov 1877**

**Precipitation, mm: Jul 1877**

**Sea surface height, cm: Dec 1877**


---

**Sea level height anomaly: RMS[T/P -- Linear Model]**

(a) No assimilation  
(b) Tide gauges assimilated  
(c) Temperature profiles assimilated  
(d) T/P assimilated

![Sea level height anomaly](image-url)
3 corollaries:

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

- 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:

\[
\text{of } D(D+R)^{-1} = \text{diag}[d/(d+r)] \text{ factor}
\]

- 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

Time:
- 1880
- 1890
- 1900
- 1910
- 1920
- 1930
- 1940
- 1950
- 1960
- 1970

Correlation (−1), 40 yr window:
- 0.65
- 0.6
- 0.55
- 0.5
- 0.45
- 0.5
- 0.4
- 0.35
- 0.3
- 0.25
- 0.2
- 0.15
- 0.1
- 0.05
- 0

Graph shows the correlation over time between Darwin and Tahiti atmospheric pressure, highlighting trends and variability.
Take home points

- Spaghetti-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.