Extracting the dynamical essence of geophysical timeseries

Julien Emile-Geay
University of Southern California, Los Angeles
Part III:

Reconstructing Climate from Multiple Proxies
ENSO over the past millennium

Collaborators: Kim Cobb (GaTech) Michael Mann (Penn State) Tapio Schneider (Caltech) Andrew Wittenberg (GFDL) Scott Rutherford (RWU) Diana Sima (Katholieke Universiteit Leuven)

Acknowledgments: Kevin Anchukaitis (LDEO)

Outline:

1. Motivations
2. Methodology
3. ENSO reconstruction
4. Implications for ENSO sensitivity

NOAA grant NA06OAR4310116
**The Hockey Stick graph**

**Mann, Bradley & Hughes [1998,1999]**

“Northern Hemisphere mean annual temperatures for three of the past eight years are warmer than any other year since (at least) a millennium.”

- Widely used in IPCC 2001. Emblem of global warming
- Media Firestorm. Main target of climate denialists.
- Blog wars: climateaudit.org vs realclimate.org
- Triggers 2 congressional inquiries! (Wegman, NAS reports)

Julien Emile-Geay
10 YEARS AFTER: IMPROVEMENTS

- Much more data, improved methodologies
- No longer addicted to tree-rings
- Goes beyond 1000 AD
- Regional information
- Code/data are public
- Pseudoproxy and frozen network validation
- Uncertainties are reported... But not always transparent

Mann et al., PNAS, 2008
Evidence of La Niña-like conditions from ~900-1250AD:
Mono Lake lowstands (Stine et al., 1994)
high Warm Pool temperatures (Mg/Ca MD81, Stott et al., 2004)
decreased Peru runoff (lithic counts, Rein et al., 2004)
cool Santa Barbara basin temperatures (G. bull. $\delta^{18}$O, Field et al., in prep)

Data from :
Cobb lab, Georgia Institute of Technology

1) Discontinuous record
2) indeterminate
The value of multiple proxies

Ice cores

Sediment cores

Tree Rings

Speleothems

Corals

Julien Emile-Geay
Multiproxy network

Tropical Proxy database for NINO3 reconstruction: 82 records

Almost no data from NINO3 region: rely on stationary teleconnections

From Emile-Geay, Cobb, Mann, Rutherford & Wittenberg, J.Clim in prep
Multiproxy network (2)

Proxy representation by age (82 records)

From Emile-Geay, Cobb, Mann, Rutherford & Wittenberg, *in prep*
TARGET: MODERN SST VARIABILITY

Unfiltered NINO3

NINO3 (with 10 year lowpass filter)
Methodology

- **Hybrid reconstruction** [Rutherford et al, 2005]
  - Split at 10-year period (high-freq vs low-freq)
- **Screening for significant correlations w/ NINO3**
  - 10% level: knocks out 41 out of 82 proxies (no speleo)
- **Two-pronged approach**
  - Composite Plus Scale (CPS)
  - Regularized Expectation Maximization (RegEM) [Schneider, 2001]
    - Regularization by truncated total least squares (TTLS)
    - Nearest-neighbor diachronic covariability [AR(1)]
- **Jackknife sensitivity analysis** [Efron, 1981]
  - “leave-one-out” ⇒ 41 reconstructions
  - ensemble standard deviation ⇒ confidence intervals
KISS: Composite Plus Scale

NINO3 CPS Reconstruction, Historical Interval

NINO3 reconstructions

Error grows as proxies wane

Acceptable verification scores

Can we do better?
RegEM: a regression with error estimate

\[ \mathbf{x}_m = \mu_m + (\mathbf{x}_a - \mu_a) \mathbf{B} + \mathbf{e}. \]

\[ \mathbf{X} = \begin{pmatrix} \mathbf{PC}_1(2000) & \mathbf{PC}_2(2000) & \mathbf{PC}_3(2000) \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \text{Known} \\ \vdots \end{pmatrix} \begin{pmatrix} \mathbf{Proxy}_1(2000) & \cdots & \mathbf{Proxy}_N(2000) \\ \mathbf{Proxy}_1(1856) & \cdots & \mathbf{Proxy}_N(1856) \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \text{Calibration} \\ \vdots \end{pmatrix} \begin{pmatrix} \mathbf{PC}_1(1000) & \mathbf{PC}_2(1000) & \mathbf{PC}_3(1000) \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \text{Reconstruction} \quad (= \text{Regression}) \\ \vdots \end{pmatrix} \begin{pmatrix} \mathbf{Proxy}_1(1000) & \cdots & \mathbf{Proxy}_N(1000) \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \text{Proxies} \\ \text{SST} \end{pmatrix} \]
Regularized Expectation Maximization in a nutshell

\[ \{\mu_0, \Sigma_0\} \]

Gaussian

“Sufficient statistics”

\[ \mu = \text{mean} \]

\[ \Sigma = \text{covariance matrix} \]

Dempster Laird & Rubin, 1977

Schneider, J. Clim. 2001

Julien Emile-Geay, USC 2009
Regularized Expectation Maximization in a nutshell

\[ \{ \mu_0, \Sigma_0 \} \xrightarrow{\text{Filter } \Sigma} \]

\( \text{(Regularization)} \)

Schneider, J. Clim. 2001

Gaussian
“Sufficient statistics”
\[ \mu = \text{mean} \]
\[ \Sigma = \text{covariance matrix} \]

Julien Emile-Geay, USC 2009
**Regularized Expectation Maximization in a nutshell**

\[ \{ \mu_0, \Sigma_0 \} \]

**Filter** \[ \Sigma \]  
*(Regularization)*

**Compute regression coefficients**

\[ \hat{B} = \hat{\Sigma}_a^{-1} \hat{\Sigma}_{am} \]

Gaussian  
“Sufficient statistics”

\[ \mu = \text{mean} \]

\[ \Sigma = \text{covariance matrix} \]

Julien Emile-Geay, USC 2009
Regularized Expectation Maximization in a nutshell

\[
\{\mu_0, \Sigma_0\}
\]

Filter \( \Sigma \)

(Regularization)

\[ \text{Compute regression coefficients} \]

\[ \hat{\mathbf{B}} = \hat{\Sigma}_{aa}^{-1} \hat{\Sigma}_{am} \]

\[ \mathbf{x}_m = \mu_m + (\mathbf{x}_a - \mu_a) \mathbf{B} + \mathbf{e}. \]

Gaussian

“Sufficient statistics”

\( \mu = \text{mean} \)

\( \Sigma = \text{covariance matrix} \)

Estimate unknown values from the available ones

Julien Emile-Geay, USC 2009
Regularized Expectation Maximization in a nutshell

\[ \{ \mu_0, \Sigma_0 \} \]

Filter \( \sum \)

(Regularization)

\[ \text{Compute} \{ \mu, \Sigma \} \]

\[ \text{Compute regression coefficients} \]

\[ \hat{B} = \hat{\Sigma}_{aa} \hat{\Sigma}_{am} \]

\[ x_m = \mu_m + (x_a - \mu_a)B + e. \]

Gaussian

“Sufficient statistics”

\( \mu = \text{mean} \)

\( \Sigma = \text{covariance matrix} \)

Estimate unknown values from the available ones

Julien Emile-Geay, USC 2009
Regularized Expectation Maximization in a nutshell

\[ \{\mu_0, \Sigma_0\} \]

Compute \( \{\mu, \Sigma\} \)

Iterate until convergence

\[ x_m = \mu_m + (x_a - \mu_a)B + e. \]

Compute regression coefficients

\[ \hat{B} = \hat{\Sigma}_{aa}^{-1}\hat{\Sigma}_{am} \]

Gaussian

“Sufficient statistics”

\( \mu = \text{mean} \)

\( \Sigma = \text{covariance matrix} \)

Schneider, J. Clim. 2001

Estimate unknown values from the available ones

Julien Emile-Geay, USC 2009
Key assumptions of RegEM

- EM algorithm assumes the data is ‘missing at random’
  i.e. \( P(\text{missing value}) \) is independent of the value.
- Data are normally distributed
- Linear relationship between proxies and climate field
Putting the Reg in RegEM

(1) Maximum Likelihood Estimation using $L_2$-penalized likelihood and Expectation-Maximization [Dempster, Laird and Rubin, 1977]

$$L_2(\Sigma, h) = \frac{n}{2} \text{tr}(\Sigma^{-1} \bar{S}) - \frac{n}{2} |\Sigma^{-1}| + h \sum_{i<j} ||\sigma^{ij}||_2^2$$

(2) Regularization:

Schneider [2001]

$$\hat{\Sigma}_{aa} = V \Lambda^2 V^T$$

$$F \equiv \Lambda^\dagger V^T \hat{\Sigma}_{am}$$ (Fourier coefficients)

So one can estimate $B$ as:

$$\hat{B} = V \text{Diag}(f_j) \Lambda^\dagger F$$

Tikhonov Regularization (“Ridge regression”)

$$f_j = \lambda_j^2 / (\lambda_j^2 + h^2)$$

with $f_j$ the filter factors
Verification RE scores

Early Verification. High-Frequency, TTLS, RE=+0.70

Late Verification. High-Frequency, TTLS, RE=+0.60

Early Verification. Low-Frequency, TTLS, RE=+0.10

Late Verification. Low-Frequency, TTLS, RE=+0.48

Early Verification. Total, RE=+0.62

Late Verification. Total, RE=+0.57
A virtual climate laboratory

SSTA. Related indices, such as the Niño-3.4 average and the first principal component of tropical Pacific SSTAs correlate with Niño-3 at levels exceeding 0.95 for monthly values in the CM2 models and Extended Reconstructed SST version 2 (ER.v2) observations. One may therefore regress onto any of these indices and obtain very similar spatial patterns. We select Niño-3 because in the observations it has a high correlation with the first principal component of tropical Pacific SSTAs (0.98 correlation for monthly ER.v2 SSTAs, 1954–2003), making it a key target for coupled simulations. Niño-3 also sits in the overlap region of the observed and simulated SST variability, where the shallow thermocline exerts a strong influence on SST. Finally, the Niño-3 index is simple to compute, and permits straightforward comparisons with existing results in the literature.

Figure 15 shows the tropical Pacific precipitation regressed onto Niño-3 SSTAs. The observations indicate wet conditions along the equator in the central and eastern Pacific during warm events, with peak rainfall anomalies just east of the date line. Meanwhile, drier-than-normal conditions prevail away from the equator and west of 155°E. In the eastern Pacific, the observed rainfall response is meridionally asymmetric, with wet conditions north of the equator but much less of a change in the south.

While the models do show increased rainfall over the central equatorial Pacific during warm events, there are clear differences with the observations. The precipitation response is too far west—consistent with the equatorial cold bias and the westward displacement of the annual-mean convection in the models (Figs. 1–3). In the east Pacific, the rainfall anomalies are too symmetric about the equator, probably due to the south-equatorial climatological warm bias and double ITCZ. There are also clear differences between the models themselves: in the west the peak rainfall response is stronger in CM2.0 than in CM2.1, while in the central/eastern Pacific the equatorial rainfall response is stronger in CM2.1.

The zonal wind stress anomaly ($\frac{\xi}{H}$) response to

Fig. 14. Std dev of interannual SST anomalies (°C). The anomalies are filtered via two applications of a 4-month running mean, transmitting 25% and 75% of the spectral amplitude at periods of 6.6 and 14 months, respectively. Observations correspond to the ER.v2 reconstruction. The dashed box in each panel indicates the Niño-3 region (5°S–5°N, 150°–90°W).

Fig. 15. Precipitation anomalies regressed onto Niño-3-averaged SST anomalies, all months included (mm day$^{-1}$ °C$^{-1}$). Observations correspond to the CMAP.v2 precipitation anomalies regressed onto the ER.v2 SSTAs for 1979–2003.

GFDL CM2.1 Specs:
- Coupled General Circulation Model
- Very good ENSO simulation
- Realistic “surrogate climate”
Pseudoproxy Tests

The hybrid RegEM methodology is tested on synthetic proxies derived from NINO3 in a 1000-year long, simulations of 2 CGCMS: NCAR CSM1.4 (volcanic and solar) and GFDL CM2.1 (unforced)

\[ P_p(\mathbf{x}, t) = \text{SNR} \cdot \text{NINO3}(t) + \xi(\mathbf{x}, t) \]

where \( \xi \) is a standard Gaussian white noise process

Signal to noise ratio:
\[ \text{SNR} = \frac{\rho}{\sqrt{1 - \rho^2}} \]

Test Statistic:
\[ \text{RE} = 1 - \frac{\sum_i (y_i - \bar{y})^2}{\sum_i (y_i - \mu_v)^2}; \]

<table>
<thead>
<tr>
<th>Case</th>
<th>SNR = 1</th>
<th>SNR = 0.4</th>
<th>SNR = 0.25</th>
<th>“realistic”</th>
</tr>
</thead>
</table>

Pseudoproxy Tests w/ CSM1.4

NINO3 RegEM-lagD reconstruction, varying SNR, CSM1.4

- Target NINO3
- SNR = 1.0, RE= +0.05; +0.92
- SNR = 0.4, RE= +0.27; +0.65
- SNR = 0.25, RE= -0.13; +0.46
- Observed SNR, RE= +0.09; +0.49

NINO3 RE scores for 200 ensemble members

- SNR = 1.0
- SNR = 0.4
- SNR = 0.25
- Observed SNR
- AR(1) benchmark

Same (low-frequency)

- SNR = 1.0
- SNR = 0.4
- SNR = 0.25
- Observed SNR
- AR(1) benchmark
Part 3: 

ENSO reconstruction
Comparison to radiative forcing

Volcanic and Solar Forcing over the Tropics

Solar forcing: Bard et al. (2007); Volcanic loading from Gao et al. (2008)

NINO3 reconstructions

- NINO3 CPS (20-year lowpass)
- NINO3 Hybrid TTLS (unfiltered)
- NINO3 Hybrid TTLS (20-year lowpass)
- Jackknife 95% C.I.
**Spectral Content**

Morlet Wavelet Power Spectral Density

**NINO3**

- **Non-stationary timeseries**
- **2-8 year band preserved throughout record**
- **Peak around 205-year variability before 1600 AD**
Response to solar forcing

Cross Wavelet Transform: Solar Forcing-NINO3

Wavelet Transform Coherency: Solar Forcing-NINO3

Supports the notion of a thermostat response to solar forcing
Clement et al. [1996], Mann et al. [2005], Emile-Geay et al. [2007]
Who does the talking?

1850 A.D.

Who does the talking?
Who does the talking?

1750 A.D.
Who does the talking?

1650 A.D.
Who does the talking?

1550 A.D.
Who does the talking?

1450 A.D.
Who does the talking?

1350 A.D.
Who does the talking?

1250 A.D.
Who does the talking?

1150 A.D.
Who does the talking?

1050 A.D.
Comparison to existing reconstructions

Reconstructions of ENSO Indices

- Wilson COA
- Wilson TEL
- TEXMEX
- Mann00
- Mann09
- This Study, CPS
- This Study, RegEM TTLS (4,3)

NINO3 or NINO34 (C)

1200 1300 1400 1500 1600 1700 1800 1900 2000

-1.5 -1 -0.5 0 0.5 1
Conclusions

- **CPS and hybrid RegEM converge to a similar answer**
  - Convergence is surprising given vast difference in complexity
  - Means either of 2 things:
    - Both are wrong (pseudoproxy study suggests otherwise)
    - Neither method significantly distorts the message conveyed by the proxies

- **Error analysis suggests:**
  - High skill in interannual band (but dating uncertainties limit use)
  - Jackknife 95% confidence intervals <1C until ~1350 A.D.
  - Acceptable verification scores until ~1400 A.D.

- **The result is a rather homoskedastic NINO3 series**
  - Nothing outstanding about 20\textsuperscript{th} century
  - But supports thermostat response to solar forcing
  - Radiation-dependent ENSO sensitivity?
Grand Conclusion for short course

Many tools are available to crack the paleoecology coconut

- Univariate (amplitude, phase)
- Multivariate

They are now readily available to the paleoecologist

- Matlab, R, Python
- “Clean hands are not learning hands” (after lunch)
- (Sharing data and code is the only way forward)

But advanced tools have advanced caveats

- Know your assumptions
- Know your data
- Know the methods’ pros and cons
- Don’t be afraid to ask....

julieneg@usc.edu
Thank you!

http://climdyn.usc.edu

http://www.flickr.com/photos/visbeek/3839267305/sizes/l/