Modeling North Pacific Climate Time Series

Don Percival

Applied Physics Lab, University of Washington

overheads for talk available at

`http://staff.washington.edu/dbp/talks.html`

joint work with Jim Overland & Hal Mofjeld (PMEL/NOAA)
(partially based on Dec. 2001 *Journal of Climate* paper)

Introduction

- goal: investigate nature of interdecadal variability in climate time series
- Fig. 1: average Nov–Mar Aleutian low sea level pressure field (North Pacific index (NPI))
- shortness of series poses major difficulties
- one approach is through modeling
  - stochastic
  - oscillator
  - other possibilities: nonlinear dynamics, SSA, ...
- models have different implications for extrapolations (e.g., nature of regime shifts)
- will fit/assess/compare three models
  - short memory stochastic model
  - long memory stochastic model
  - 'signal + noise' model: square wave oscillator & white noise

Overview of Remainder of Talk

- describe short & long memory stochastic models
- describe rationale for square wave oscillator model (picked using matching pursuit)
- discuss estimation of model parameters
- look at fitted models for NPI
- discuss use of goodness of fit tests to assess models (will find that all 3 models fit equally well)
- discuss how well we can expect to discriminate amongst models
- look at implications of models (regime shifts)
- state conclusions

Short & Long Memory Models

- will consider two Gaussian stationary models for data
  - first order autoregressive process (AR(1))
  - fractionally differenced (FD) process
- both processes fully specified by 3 parameters (and hence both are ‘equally simple’)
  1. process mean
  2. parameter that controls process variance
  3. parameter controlling shape of both
     - autocovariance sequence (ACVS) and
     - spectral density function (SDF)
- essential difference between processes
  - AR(1) ACVS dies down quickly (exponentially), so process said to have ‘short memory’
  - FD ACVS dies down slowly (hyperbolically), so process said to have ‘long memory’ (LM)

Short Memory Stochastic Model

- regard data as realization of portion $X_0, X_1, \ldots, X_{N-1}$ of stationary Gaussian AR(1) process:

$$X_t - \mu_X = \phi(X_{t-1} - \mu_X) + \epsilon_t = \sum_{k=0}^{\infty} \phi^k \epsilon_{t-k}$$

  where

  1. $\mu_X = E\{X_t\}$ is process mean
  2. ϵ_t is white noise with mean zero and variance σ_ϵ^2
  3. $|\phi| < 1$ (if $\phi = 0$, then X_t is white noise)

- ACVS and SDF given by

$$s_{X,\tau} \equiv \operatorname{cov}\{X_t, X_{t+\tau}\} = \frac{\sigma_\epsilon^2 \phi^{|\tau|}}{1-\phi^2} \;\&\; S_X(f) = \frac{\sigma_\epsilon^2}{1+\phi^2 - 2\phi\cos(2\pi f)},$$

  where τ is an integer & $|f| \leq \frac{1}{2}$

- related to discretized 1st order differential equation (has single damping constant (related to ϕ))

- can define measure of decorrelation (or integral time scale):

$$\tau_D \equiv 1 + 2\sum_{\tau=1}^{\infty} \frac{s_{X,\tau}}{s_{X,0}} = \frac{1+\phi}{1-\phi};$$

  i.e., subseries $X_{n\lceil \tau_D \rceil}$, $n = \ldots, -1, 0, 1, \ldots$ is close to white noise

Long Memory Stochastic Model

- regard data as realization of portion $Y_0, Y_1, \ldots, Y_{N-1}$ of stationary Gaussian FD process:

$$\begin{aligned} Y_t - \mu_Y &= \sum_{k=0}^{\infty} \frac{\Gamma(1+\delta)}{\Gamma(k+1)\Gamma(1+\delta-k)}(-1)^k (Y_{t-k} - \mu_Y) \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(1-\delta)}{\Gamma(k+1)\Gamma(1-\delta-k)}(-1)^k \varepsilon_{t-k} \end{aligned}$$

  where

  1. $\mu_Y = E\{Y_t\}$ is process mean
  2. ε_t is white noise with mean zero and variance σ_ε^2
  3. $|\delta| < \frac{1}{2}$ (if $\delta = 0$, Y_t is white noise; LM if $\delta > 0$)

- ACVS and SDF given by

$$s_{Y,\tau} = \frac{\sigma_\varepsilon^2 \sin(\pi\delta)\Gamma(1-2\delta)\Gamma(\tau+\delta)}{\pi\Gamma(\tau+1-\delta)} \;\;\&\;\; S_Y(f) = \frac{\sigma_\varepsilon^2}{|2\sin(\pi f)|^{2\delta}}$$

- for $\tau \geq 1$ and approximately for large τ & small f,

$$s_{Y,\tau} = s_{Y,\tau-1}\frac{\tau+\delta-1}{\tau-\delta} \propto |\tau|^{2\delta-1} \;\text{ and }\; S_Y(f) \propto \frac{1}{|f|^{2\delta}}$$

- related to aggregation of 1st order differential equation involving many different damping constants

- integral time scale τ_D is infinite

Square Wave Oscillation Model: I

- Minobe (1999): NPI contains ‘regime’ shifts
- regime is time interval over which series is essentially either $>$ or $<$ its long term average value
- Fig. 1: plot of NPI and 5 year running mean
  - data for 1901–23 are essentially $>$ sample mean (exceptions are 1905 & 1919)
  - called positive regime with duration of 23 years
  - clearly identified in 5 year running mean
  - latter is essentially $<$ sample mean for 1924–46 (but not strictly so)
- Minobe (1999): regimes characterized by
  - 20 & 50 year oscillations
  - rapid transitions that ‘cannot be attributed to a single sinusoidal-wavelike variability’

  (cf. Figure 1 from Minobe, 1999)
- matching pursuit supports Minobe’s notions

Matching Pursuit: Basics

- idea: approximate time series $\mathbf{Z} \equiv [Z_0, \ldots, Z_{N-1}]^T$ using small # of vectors selected from a large set
- let $\mathcal{D} \equiv \{\mathbf{d}_n : n = 0, \ldots, N_{\mathcal{D}} - 1\}$ be 'dictionary' containing $N_{\mathcal{D}}$ different vectors
  - $\mathbf{d}_n = [d_{n,0}, d_{n,1}, \ldots, d_{n,N-1}]^T$
  - each vector normalized to have unit norm:
  $$\|\mathbf{d}_n\|^2 = \sum_{t=0}^{N-1} |d_{n,t}|^2 = 1$$
  - $\mathbf{d}_n$ can be real- or complex-valued
  - assume $\mathcal{D}$ to be highly redundant (allows us to find $\mathbf{d}_n$ well matched to $\mathbf{Z}$)
- matching pursuit successively approximates $\mathbf{Z}$ with orthogonal projections onto elements of $\mathcal{D}$

Matching Pursuit Algorithm: I

- for $\mathbf{d}_{n_0} \in \mathcal{D}$, form $\langle \mathbf{Z}, \mathbf{d}_{n_0} \rangle \mathbf{d}_{n_0}$
- define residual vector: $\mathbf{R}^{(1)} \equiv \mathbf{Z} - \langle \mathbf{Z}, \mathbf{d}_{n_0} \rangle \mathbf{d}_{n_0}$
  so that $\mathbf{Z} = \langle \mathbf{Z}, \mathbf{d}_{n_0} \rangle \mathbf{d}_{n_0} + \mathbf{R}^{(1)}$
- $\mathbf{d}_{n_0}$ and $\mathbf{R}^{(1)}$ are orthogonal: $\langle \mathbf{d}_{n_0}, \mathbf{R}^{(1)} \rangle = 0$
- hence have

  $$\|\mathbf{Z}\|^2 = \|\langle \mathbf{Z}, \mathbf{d}_{n_0} \rangle \mathbf{d}_{n_0}\|^2 + \|\mathbf{R}^{(1)}\|^2 = |\langle \mathbf{Z}, \mathbf{d}_{n_0} \rangle|^2 + \|\mathbf{R}^{(1)}\|^2$$

- choose $\mathbf{d}_{n_0}$ such that

  $$|\langle \mathbf{Z}, \mathbf{d}_{n_0} \rangle| = \max_{\mathbf{d}_n \in \mathcal{D}} |\langle \mathbf{Z}, \mathbf{d}_n \rangle|$$

- above is first step of algorithm; second step is

  $$\mathbf{R}^{(1)} = \langle \mathbf{R}^{(1)}, \mathbf{d}_{n_1} \rangle \mathbf{d}_{n_1} + \mathbf{R}^{(2)}$$

  with $\mathbf{d}_{n_1}$ picked such that

  $$\left|\langle \mathbf{R}^{(1)}, \mathbf{d}_{n_1} \rangle\right| = \max_{\mathbf{d}_n \in \mathcal{D}} \left|\langle \mathbf{R}^{(1)}, \mathbf{d}_n \rangle\right|$$

Matching Pursuit Algorithm: II

- after m such steps, have additive decomposition:

$$\mathbf{Z} = \sum_{k=0}^{m-1} \langle \mathbf{R}^{(k)}, \mathbf{d}_{n_k} \rangle \mathbf{d}_{n_k} + \mathbf{R}^{(m)}$$

(letting $\mathbf{R}^{(0)} \equiv \mathbf{Z}$) and energy decomposition:

$$\begin{aligned}
\|\mathbf{Z}\|^2 &= \sum_{k=0}^{m-1} \|\langle \mathbf{R}^{(k)}, \mathbf{d}_{n_k} \rangle \mathbf{d}_{n_k}\|^2 + \|\mathbf{R}^{(m)}\|^2 \\
&= \sum_{k=0}^{m-1} |\langle \mathbf{R}^{(k)}, \mathbf{d}_{n_k} \rangle|^2 + \|\mathbf{R}^{(m)}\|^2
\end{aligned}$$

- note: as m increases, $\|\mathbf{R}^{(m)}\|^2$ must decrease (must reach zero under certain conditions)

Square Wave Oscillation Model: II

- idea: construct $\mathcal{D}$ containing
  1. real- & complex-valued vectors from orthonormal discrete Fourier transform (ODFT)
  2. square wave oscillations (SWOs) with periods of $2, \ldots, N$ and all relevant shifts
- note: if complex-valued ODFT vector picked, will also pick its complex conjugate (to handle phases)
- Fig. 2: result of applying matching pursuit to NPI (after subtraction of sample mean)
  - 1st vector picked is SWO with period of 50 years
  - 2nd vector picked is SWO with period of 20 years
  - 5th, 6th & 10th vectors picked are from ODFT
- will consider simple SWO model for NPI time series:

$$Z_t = \mu_Z + \beta d_{n_0,t} + e_t$$

  - μ_Z & β are parameters (if $\beta = 0$, Z_t is white noise)
  - $d_{n_0,t}$ part of 1st vector picked by matching pursuit
  - e_t is Gaussian white noise with mean zero and variance σ_e^2

Estimation of Model Parameters: I

- AR(1) process X_t parameterized by μ_X, ϕ & σ_ϵ^2
- FD process Y_t parameterized by μ_Y, δ & σ_ε^2
- SWO process Z_t parameterized by μ_Z, β & σ_e^2
- can estimate μ_X, μ_Y & μ_Z via sample means:

  $$\hat{\mu}_X = \frac{1}{N}\sum_{t=0}^{N-1} X_t,\ \ \hat{\mu}_Y = \frac{1}{N}\sum_{t=0}^{N-1} Y_t\ \ \&\ \ \hat{\mu}_Z = \frac{1}{N}\sum_{t=0}^{N-1} Z_t$$

  (might be suboptimal, but little practical loss)
- form recentered series:

  $$\widetilde{X}_t \equiv X_t - \hat{\mu}_X,\ \ \widetilde{Y}_t \equiv Y_t - \hat{\mu}_Y\ \ \&\ \ \widetilde{Z}_t \equiv Z_t - \hat{\mu}_Z$$
- regard $\widetilde{X}_t$, $\widetilde{Y}_t$ & $\widetilde{Z}_t$ as AR(1), FD & SWO processes with $\mu_X = \mu_Y = \mu_Z = 0$
- can estimate ϕ & σ_ϵ^2, δ & σ_ε^2 or β & σ_e^2 via maximum likelihood (ML) method

Estimation of Model Parameters: II

- large sample theory on ML estimators says
  - $\hat{\phi}$ & $\hat{\sigma}_\epsilon^2$ are approximately normally distributed with means ϕ & σ_ϵ^2 and variances $\frac{1-\phi^2}{N}$ & $\frac{2\sigma_\epsilon^4}{N}$
  - $\hat{\delta}$ & $\hat{\sigma}_\varepsilon^2$ are approximately normally distributed with means δ & σ_ε^2 and variances $\frac{6}{\pi^2 N}$ & $\frac{2\sigma_\varepsilon^4}{N}$
  - $\hat{\beta}$ & $\hat{\sigma}_e^2$ are approximately normally distributed with means β & σ_e^2 and variances σ_e^2 & $\frac{2\sigma_e^4}{N}$
- Monte Carlo experiments: above valid for $N \geq 100$
- can use ML theory to form 95% confidence intervals (CIs) for unknown parameters
- can form residuals $\hat{\epsilon}_t$, $\hat{\varepsilon}_t$ and $\hat{e}_t$
- can use residuals to test adequacy of model (if adequate, residuals should resemble white noise)

Fitted Models for NPI

- Tab. 1: parameter estimates & CIs for NPI series
- all 3 models significantly different from white noise (i.e., $\phi \neq 0$, $\delta \neq 0$ & $\beta \neq 0$)
- SWO model has smallest estimated residual variation
- Fig. 3: estimated autocorrelation sequence (ACS) and estimated SDF (periodogram) for NPI, i.e.,

$$\hat{\rho}_\tau \equiv \frac{\hat{s}_{X,\tau}}{\hat{s}_{X,0}} = \frac{\sum_{t=0}^{N-\tau-1} \widetilde{X}_t \widetilde{X}_{t+\tau}}{\sum_{t=0}^{N-1} \widetilde{X}_t^2} \ \& \ \hat{S}(f_k) \equiv \frac{1}{N} \left| \sum_{t=0}^{N-1} \widetilde{X}_t e^{-i2\pi f_k t} \right|^2,$$

  along with ACSs & SDFs from fitted models (for SWO, SDF taken to be $E\{\hat{S}(f_k)\}$)
- qualitatively, all 3 models seem reasonable (arguably AR(1) ACS poorest match to $\hat{\rho}_\tau$)
- can use goodness of fit tests for quantitative assessment of models

Goodness of Fit Tests: I

1. compare fitted SDF to periodogram:

$$T_1 \equiv \frac{NA}{4\pi B^2}, \text{ where } A \equiv \sum_{k=1}^{\lfloor \frac{N-1}{2} \rfloor} \left(\frac{\hat{S}(f_k)}{S(f_k; \hat{\theta})} \right)^2; \;\; B \equiv \sum_{k=1}^{\lfloor \frac{N-1}{2} \rfloor} \frac{\hat{S}(f_k)}{S(f_k; \hat{\theta})};$$

   $S(f_k; \hat{\theta})$ is theoretical SDF depending on $\hat{\theta}$; & either $\hat{\theta} = [\hat{\phi}, \hat{\sigma}_\epsilon^2]^T$ or $\hat{\theta} = [\hat{\delta}, \hat{\sigma}_\varepsilon^2]^T$ (can't use with SWO)

2. cumulative periodogram test statistic:

$$T_2 = \max \left\{ \max_l \left(\frac{l}{\lfloor \frac{N-1}{2} \rfloor - 1} - \mathcal{P}_l \right), \max_l \left(\mathcal{P}_l - \frac{l-1}{\lfloor \frac{N-1}{2} \rfloor - 1} \right) \right\},$$

   where $\mathcal{P}_l$ is the normalized cumulative periodogram for $\hat{\epsilon}_t$ (likewise for $\hat{\varepsilon}_t$ & $\hat{e}_t$):

$$\mathcal{P}_l \equiv \frac{\sum_{k=1}^{l} \hat{S}_{\hat{\epsilon}_t}(f_k)}{\sum_{k=1}^{\lfloor \frac{N-1}{2} \rfloor} \hat{S}_{\hat{\epsilon}_t}(f_k)}$$

3. Box–Pierce portmanteau test statistic:

$$T_3 = N \sum_{\tau=1}^{K} \hat{\rho}^2_{\hat{\epsilon}_t, \tau}$$

   where $\rho_{\hat{\epsilon}_t, \tau}$ is estimated ACS for $\hat{\epsilon}_t$ (same for $\hat{\varepsilon}_t$ & $\hat{e}_t$)

4. Ljung–Box–Pierce portmanteau test statistic:

$$T_4 = N(N+2) \sum_{\tau=1}^{K} \frac{\hat{\rho}^2_{\hat{\epsilon}_t, \tau}}{N - \tau}$$

Goodness of Fit Tests: II

- if T_j 'too big,' reject 'model is adequate' hypothesis
- can determine what is 'too big' under null hypothesis that model is correct
- Tab. 2: model goodness of fit tests for NPI
  - can reject white noise model
  - cannot reject any of the 3 models for NPI
- Q: can we really expect to distinguish amongst 3 models given just $N = 100$ values for NPI?

Model Discrimination

- to address question, consider following experiment
- assume FD model with observed $\hat{\delta}$ is correct for NPI
- simulate time series of length N' from FD model
- fit AR(1) model to simulated FD series
- evaluate fitted AR(1) model using each T_j
- repeat above large # of times (2500)
- can estimate probability that T_j will (correctly) reject null hypothesis that AR(1) model is correct
- gives power of T_j in saying AR(1) model is incorrect
- repeat above for variety of sample sizes N'
- can repeat all of the above with different combinations of AR(1), FD & SWO processes
- Fig. 4: power of various test statistics vs. N'
  - at best, 30% chance of rejecting null hypothesis
  - need $N' \approx 500$ to have 50% chance of discriminating between AR(1) & FD models
  - no one test uniformly better than others

Model Implications: I

- no statistical reason to one model over other two
- all three models depend on 3 parameters & hence are equally simple (ignoring matching pursuit step)
- even though all match NPI equally well, models can have different & potentially important implications
- Fig. 5: examples of 1000 year simulations
- Q: how well do models support notion of regimes?

Model Implications: II

- to address question, consider following experiment
- generate deviate $\tilde{\delta}$ from normal distribution with mean $\hat{\delta}$ from NPI and variance $\frac{6}{\pi^2 N} = \frac{6}{\pi^2 100}$
- assume FD model with $\tilde{\delta}$ is correct for NPI
- simulate time series of length 1024 from FD model
- tabulate sizes of observed regimes in
  1. simulated series
  2. five year running mean of series
- repeat above 1000 times
- also repeat using fitted AR(1) and SWO models
- Fig. 6: plots of empirically determined probabilities of regime sizes being $\geq$ specified sizes
- intermediate regime sizes most likely under SWO
- large regime sizes most likely under FD
- regime size $\geq$ 23 is 4 times more likely under FD model than under AR(1)

Conclusions

- AR(1), FD & SWO models equally adequate for NPI
- cannot realistically hope to distinguish between three models given available sample sizes
- all 3 models include white noise as special case (all 3 lead to rejection of hypothesis of white noise)
- AR(1) model has most rapid drop off of ACS
- FD model has long tail of small positive correlations
- SWO model has oscillating ACS
- loose physical considerations might favor FD model (aggregation of first order differential equations)
- FD model more supportive of regimes than AR(1)
- FD model more supportive of long regimes than SWO
- estimated δ compatible with notion of regimes, but NPI does not have strong long memory