Modeling Atmospheric Circulation Changes over the North Pacific

Don Percival

Applied Physics Laboratory, University of Washington
Seattle, Washington, USA

overheads for talk available at

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

joint work with Jim Overland & Hal Mofjeld (Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, Washington, USA)
Introduction

• goal: investigate nature of changes in atmospheric circulation over North Pacific

• will concentrate on two atmospheric time series
  – Fig. 1: average Nov–Mar Aleutian low sea level pressure field (North Pacific index (NPI))
  – Fig. 2: Sitka, Alaska, air temperatures

• shortness of both series (100 and 146 points) is major difficulty

• one approach is through modeling
  – pure stochastic
  – deterministic signal + stochastic noise
  – other possibilities (nonlinear dynamics, SSA, ...)

• models have different implications for extrapolations

• will fit/assess/compare three models
  – short memory stochastic model
  – long memory stochastic model
  – signal + noise model: square wave oscillator (SWO) & white noise
Overview of Remainder of Talk

• describe short & long memory stochastic models

• describe rationale for SWO model (matching pursuit)

• discuss estimation of model parameters

• look at fitted models

• discuss goodness of fit tests used 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

• state conclusions
Short & Long Memory Models

• will consider two Gaussian stationary models
  – 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. $\epsilon_t$ is white noise with mean zero and variance $\sigma_{\epsilon}^2$
3. $|\phi| < 1$ (if $\phi = 0$, then $X_t$ is white noise)

- ACVS and SDF given by

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

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

- related to discretized 1st order differential equation (has single damping constant dictated by $\phi$)

- can define integral time scale (decorrelation measure):

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

implies subseries $X_{n[\tau_D]}, n = \ldots, -1, 0, 1, \ldots$, 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:

\[
Y_t - \mu_Y = \sum_{k=0}^{\infty} \frac{\Gamma(1 + \delta)}{\Gamma(k + 1)\Gamma(1 + \delta - k)} (-1)^k (Y_{t-k} - \mu_Y)
\]

\[
= \sum_{k=0}^{\infty} \frac{\Gamma(1 - \delta)}{\Gamma(k + 1)\Gamma(1 - \delta - k)} (-1)^k \varepsilon_{t-k}
\]

where

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

• ACVS and SDF given by

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

• for \( \tau \geq 1 \) and approximately for large \( \tau \) & small \( f \),

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

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

• integral time scale \( \tau_D \) 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’

• can use matching pursuit to assess Minobe’s claim
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 $\mathcal{D} \equiv \{\mathbf{D}_k : k = 0, \ldots, K - 1\}$ be ‘dictionary’ containing $K$ different vectors
  
  $\mathbf{D}_k = [D_{k,0}, D_{k,1}, \ldots, D_{k,N-1}]^T$
  
  - vectors normalized to have unit norm (‘energy’):
    
    $$\|\mathbf{D}_k\|^2 = \sum_{t=0}^{N-1} |D_{k,t}|^2 = 1$$

  $\mathbf{D}_k$ can be real- or complex-valued

  - assume $\mathcal{D}$ to be highly redundant in order to find $\mathbf{D}_k$ well matched to $\mathbf{Z}$

• matching pursuit successively approximates $\mathbf{Z}$ with orthogonal projections onto elements of $\mathcal{D}$
Matching Pursuit Algorithm: I

- for each $D_k \in \mathcal{D}$, form approximation $A_k \equiv \langle Z, D_k \rangle D_k$, where
  $$\langle Z, D_k \rangle \equiv \sum_{t=0}^{N-1} Z_t D_{k,t}$$
  (assumes $D_k$ real-valued; can adjust if not so)
- define residuals $R_k \equiv Z - A_k$ so that $Z = A_k + R_k$
- $A_k$ and $R_k$ are orthogonal; i.e., $\langle A_k, R_k \rangle = 0$
- hence $\|Z\|^2 = \|A_k\|^2 + \|R_k\|^2 = |\langle Z, D_k \rangle|^2 + \|R_k\|^2$
- to minimize $\|R_k\|^2$, select $k^{(1)}$ such that
  $$|\langle Z, D_{k^{(1)}} \rangle| = \max_{D_k \in \mathcal{D}} |\langle Z, D_k \rangle|$$
- let $A^{(1)}$ & $R^{(1)}$ be approximation and residuals
- 1st stage of algorithm thus yields $Z = A^{(1)} + R^{(1)}$
- 2nd stage: use $R^{(1)}$ rather than $Z$ in above
- yields $R^{(1)} = A^{(2)} + R^{(2)}$ with $k^{(2)}$ picked such that
  $$|\langle R^{(1)}, D_{k^{(2)}} \rangle| = \max_{D_k \in \mathcal{D}} |\langle R^{(1)}, D_k \rangle|$$
Matching Pursuit Algorithm: II

• after $m$ such steps, have additive decomposition

$$Z = \sum_{n=1}^{m} A^{(n)} + R^{(m)} \equiv \hat{Z}^{(m)} + R^{(m)} ,$$

where $\hat{Z}^{(m)}$ is $m$th order approximation to $Z$

• also have ‘energy’ decomposition

$$\|Z\|^2 = \sum_{n=1}^{m} \|A^{(n)}\|^2 + \|R^{(m)}\|^2$$

$$= \sum_{n=1}^{m} \left| \langle R^{(n-1)}, D_{k(n)} \rangle \right|^2 + \|R^{(m)}\|^2 ,$$

where $R^{(0)} \equiv Z$

• note: as $m$ increases, $\|R^{(m)}\|^2$ must decrease (must reach zero under certain conditions)
Square Wave Oscillation Model: II

- Fig. 3: construct $\mathcal{D}$ containing
  1. vectors from discrete Fourier transform (sinusoids)
  2. SWOs with periods of $2,\ldots,N$ & all shifts
  3. single cycles from SWOs (Haar wavelet vectors)
  4. half cycles from SWOs (Haar scaling vectors)

- Fig. 4: result of applying matching pursuit to NPI (after subtraction of sample mean)
  - 1st vector picked is SWO with period of 50 years
  - 2nd to 4th vectors are Haar wavelet vectors
  - 5th vector is sinusoid

- Fig. 5: result of applying matching pursuit to Sitka
  - 1st vector picked is SWO with period of 54 years
    (location of transitions match up well with NPI’s)

- results lend support for Minobe’s hypothesis
Square Wave Oscillation Model: III

- will consider simple SWO model:

\[ Z_t = \mu_Z + \beta D_{k(0),t} + e_t \]

- \(\mu_Z\) & \(\beta\) are parameters (if \(\beta = 0\), \(Z_t\) is white noise)
- \(D_{k(0),t}\) part of 1st vector from matching pursuit
- \(e_t\) is Gaussian white noise with mean zero and variance \(\sigma_e^2\)
Estimation of Model Parameters: I

• AR(1) process $X_t$ parameterized by $\mu_X$, $\phi$ & $\sigma^2_\epsilon$
• FD process $Y_t$ parameterized by $\mu_Y$, $\delta$ & $\sigma^2_\epsilon$
• SWO process $Z_t$ parameterized by $\mu_Z$, $\beta$ & $\sigma^2_\epsilon$
• can estimate $\mu_X$, $\mu_Y$ & $\mu_Z$ via sample means:
  \[
  \hat{\mu}_X = \frac{1}{N} \sum_{t=0}^{N-1} X_t, \quad \hat{\mu}_Y = \frac{1}{N} \sum_{t=0}^{N-1} Y_t \quad \text{&} \quad \hat{\mu}_Z = \frac{1}{N} \sum_{t=0}^{N-1} Z_t
  \]
  (might be suboptimal, but little practical loss)
• form recentered series:
  \[
  \tilde{X}_t \equiv X_t - \hat{\mu}_X, \quad \tilde{Y}_t \equiv Y_t - \hat{\mu}_Y \quad \text{&} \quad \tilde{Z}_t \equiv Z_t - \hat{\mu}_Z
  \]
• regard $\tilde{X}_t$, $\tilde{Y}_t$ & $\tilde{Z}_t$ as AR(1), FD & SWO processes with $\mu_X = \mu_Y = \mu_Z = 0$
• can estimate $\phi$ & $\sigma^2_\epsilon$, $\delta$ & $\sigma^2_\epsilon$ or $\beta$ & $\sigma^2_\epsilon$ via maximum likelihood (ML) method
Estimation of Model Parameters: II

- large sample theory on ML estimators says
  - \( \hat{\phi} \) & \( \hat{\sigma}^2 \) are approximately normally distributed
    with means \( \phi \) & \( \sigma^2 \) and variances \( \frac{1-\phi^2}{N} \) & \( \frac{2\sigma^4}{N} \)
  - \( \hat{\delta} \) & \( \hat{\sigma}^2 \) are approximately normally distributed
    with means \( \delta \) & \( \sigma^2 \) and variances \( \frac{6}{\pi^2 N} \) & \( \frac{2\sigma^4}{N} \)
  - \( \hat{\beta} \) & \( \hat{\sigma}^2 \) are approximately normally distributed
    with means \( \beta \) & \( \sigma^2 \) and variances \( \sigma^2 \) & \( \frac{2\sigma^4}{N} \)

- 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 \( \hat{\epsilon}_t \), \( \hat{\epsilon}_t \) 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

• 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. 6: estimated autocorrelation sequence (ACS) and estimated SDF (periodogram) for NPI, i.e.,

$$\hat{\rho}_\tau \equiv \frac{\hat{s}_{X,\tau}}{\hat{s}_{X,0}} = \frac{\sum_{t=0}^{N-\tau-1} \tilde{X}_t \tilde{X}_{t+\tau}}{\sum_{t=0}^{N-1} \tilde{X}_t^2}$$

$$\hat{S}(f_k) \equiv \frac{1}{N} \left| \sum_{t=0}^{N-1} \tilde{X}_t e^{-i2\pi f_k t} \right|^2,$$

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 $\hat{\rho}_\tau$)

• found similar results for Sitka air temperatures

• can use goodness of fit tests for quantitative assessment of models
Goodness of Fit Tests: I

1. compare fitted SDF to periodogram:

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

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

2. cumulative periodogram test statistic:

\[ T_2 = \max \left\{ \max_l \left( \frac{l}{\lceil N/2 \rceil - 1} - \mathcal{P}_l \right), \max_l \left( \mathcal{P}_l - \frac{l - 1}{\lceil N/2 \rceil - 1} \right) \right\}, \]

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

\[ \mathcal{P}_l \equiv \frac{\sum_{k=1}^{\lfloor N-1 \rfloor} \hat{S}_t(f_k)}{\sum_{k=1}^{\lfloor N/2 \rfloor} \hat{S}_t(f_k)} \]

3. Box–Pierce portmanteau test statistic:

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

where \( \hat{\rho}_{\hat{\epsilon}_t,\tau} \) is estimated ACS for \( \hat{\epsilon}_t \) (same for \( \hat{\epsilon}_t \) & \( \hat{\epsilon}_t \))
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. 7: 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. 8: 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 \( \frac{6}{\pi^2N} = \frac{6}{\pi^2100} \)
• 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. 9: 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 and Sitka air temperatures

• SWO models picked out by matching pursuit & offer some support for Minobe’s hypothesis

• 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 $\delta$ compatible with notion of regimes, but neither NPI nor Sitka exhibit strong long memory