Interpretation of North Pacific Variability as a Short and Long Memory Process

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)
(forthcoming paper in *Journal of Climate*)

Introduction

- goal: investigate nature of interdecadal variability in climate time series
- shortness of series poses major difficulties
- one approach: fit & compare various models
  - oscillator
  - nonlinear dynamics (chaos)
  - stochastic
- models have different implications
  (e.g., nature of regime shifts)
- will investigate influence of choice of stochastic models on representing North Pacific atmospheric data
  - short vs. long memory stochastic models
  - two different atmospheric data sets
    * Fig. 1: average (Nov–Mar) Aleutian low sea level pressure field (North Pacific index (NPI))
    * Fig. 2: temperature record at Sitka, Alaska

Overview of Remainder of Talk

- describe short & long memory stochastic models
- discuss maximum likelihood (ML) estimation of model parameters
- look at fitted models for NPI & Sitka
- discuss use of goodness of fit tests to assess models (will find that both models fit equally well)
- discuss how well we can hope to discriminate between short & long memory models
- look at implications of short & long memory models with regard to regime shifts
- consider interpretation of long memory models
- 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

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

  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

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
- given data that can be regarded as realization of $X_0, \ldots, X_{N-1}$ or $Y_0, \ldots, Y_{N-1}$, can estimate process mean via sample means:

$$\hat{\mu}_X = \frac{1}{N} \sum_{t=0}^{N-1} X_t \text{ and } \hat{\mu}_Y = \frac{1}{N} \sum_{t=0}^{N-1} Y_t$$

- can form recentered series:

$$\widetilde{X}_t \equiv X_t - \hat{\mu}_X \text{ and } \widetilde{Y}_t \equiv Y_t - \hat{\mu}_Y$$

- regard $\widetilde{X}_t$ & $\widetilde{Y}_t$ as zero mean AR(1) & FD processes
- can use to estimate ϕ & σ_ϵ^2 or δ & σ_ε^2 via maximum likelihood (ML) method
- 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 $\frac{1-\phi^2}{N}$ & $\frac{\sigma_\epsilon^4}{2N}$
  - $\hat{\delta}$ & $\hat{\sigma}_\varepsilon^2$ are approximately normally distributed with means δ & σ_ε^2 and variances $\frac{6}{\pi^2 N}$ & $\frac{\sigma_\varepsilon^4}{2N}$
- Monte Carlo experiments: above valid for $N \geq 100$

Estimation of Model Parameters: II

- can use ML theory to form 95% confidence intervals (CIs) for unknown parameters
- can adjust ML procedure to handle time series with missing values (no need to use interpolation)
- can form residuals $\hat{\epsilon}_t$ and $\hat{\varepsilon}_t$
- can use residuals to test adequacy of model
  (if adequate, residuals should resemble white noise)

Fitted Models for NPI and Sitka

- Tab. 1: parameter estimates & CIs for NPI & Sitka
  - AR(1) & FD models both significantly different from white noise (i.e., $\phi \neq 0$ and $\delta \neq 0$)
  - $\hat{\phi}$'s similar for NPI & Sitka (as are $\hat{\delta}$'s)
  - interpolation increases estimated $\hat{\phi}$ & $\hat{\delta}$ for Sitka
- Fig. 3: estimated autocorrelation sequence (ACS) and estimated SDF (periodogram) for NPI, i.e.,

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

  along with ACSs and SDFs from fitted models
  - qualitatively, both models seems reasonable (arguably FD ACS better match to $\hat{\rho}_\tau$ than AR(1))
  - get similar results for Sitka
- 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 $\hat{\theta}$; and either $\hat{\theta} = [\hat{\phi}, \hat{\sigma}_\epsilon^2]^T$ or $\hat{\theta} = [\hat{\delta}, \hat{\sigma}_\varepsilon^2]^T$

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 $\hat{\varepsilon}_t$):

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

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 $\hat{\epsilon}_t$ (likewise for $\hat{\varepsilon}_t$)

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 either AR(1) or FD model for NPI (some *very* weak hint that FD is better)
  - similar results obtained for Sitka
- Q: can we really expect to distinguish between AR(1) and FD models given just $N = 100$ values for NPI?

AR(1) & FD 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
- evaluated 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 roles of FD & AR(1) processes interchanged
- Fig. 4: power of various test statistics vs. N'
  - in best case, need $N' \approx 500$ to have 50% chance of discriminating between models
  - portmanteau tests to be preferred over T_1 and T_2

AR(1) & FD Model Implications: I

- no statistical reason to prefer AR(1) over FD model for NPI (or *vice versa*)
- both AR(1) & FD models depend on 3 parameters & hence are equally simple (i.e., cannot appeal to Occram's razor here)
- even though both describe NPI equally well, models can have potentially important implications if one is selected in favor of the other
- as example, will consider extent to which models support notion of 'regimes' in NPI
- 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)

AR(1) & FD Model Implications: II

- Q: how do fitted AR(1) & FD models impact distribution of regime sizes?
- 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) model for NPI
- Fig. 5: plots of empirically determined probabilities of regime sizes being $\geq$ specified sizes
- regime size $\geq$ 23 is 4 times more likely under FD model than under AR(1)
- similarly, regime size $\geq$ 35 is 10 times more likely

Long Memory Model Interpretation

- for both NPI & Sitka, fitted FD model has $\hat{\delta} \approx 0.2$
- allowable range of δ for stationary FD models with long memory is $0 < \delta < \frac{1}{2}$
- as $\delta \to 0$, FD process $\to$ white noise ('no memory')
- as $\delta \to \frac{1}{2}$, FD process has strong long memory effect
- Fig. 6: gives some idea how to interpret δ
  1. $\delta = 0.02$ is lower end of 95% CI for δ in NPI case
  2. $\delta = 0.17$ is estimated value for NPI
  3. $\delta = 0.32$ is upper end of 95% CI
  4. $\delta = 0.45$ corresponds to strong LM effect
- simulated series in a given column constructed using same white noise sequences
- as δ increases, average regime sizes increase
- increase not linear in δ: cases $\delta = 0.32$ & 0.17 similar & substantially different from $\delta = 0.02$ case
- can thus interpret δ as indicator of how much regime-like structure there is in a time series

Conclusions

- AR(1) & FD models equally adequate for time series considered here (NPI & Sitka)
- cannot realistically hope to distinguish between AR(1) & FD processes given available sample sizes
- both models include white noise as special case (both lead to rejection of hypothesis of white noise)
- AR(1) model has rapid drop off of ACVS
- FD model has long tail of small positive correlations
- loose physical considerations might favor FD model (aggregation of first order differential equations)
- FD model more supportive of regime-like behavior than AR(1)
- can use δ as indicator of regime-like behavior
- for NPI & Sitka, estimated δ compatible with notion of regimes, but neither series has strong long memory