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: study changes in North Pacific atmospheric circulation
- will concentrate on two atmospheric time series
  - Fig. 1: average Nov–Mar Aleutian low sea level pressure field (North Pacific index (NPI))
  - Fig. 2: air temperatures at Sitka, Alaska
- 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
  1. short memory stochastic model
  2. long memory stochastic model
  3. signal + noise model:
     square wave oscillator (SWO) & white noise

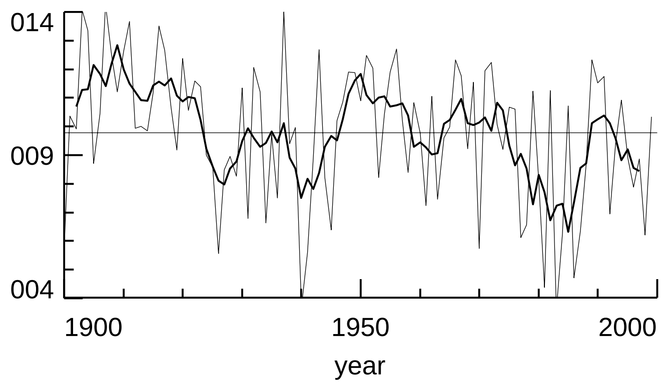

Figure 1: Plot of the NP index (thin curve) and a five year running average of the index (thick). The thin horizontal line depicts the sample mean (1009.8) for the index.

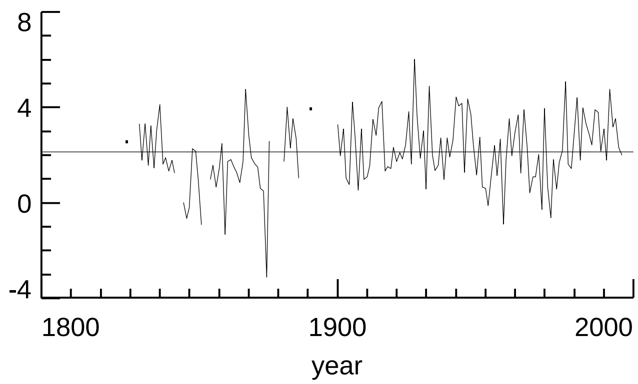

Figure 2: Plot of Sitka winter air temperatures (broken curve). The thin horizontal line depicts the sample mean (2.13) for the series.

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
- brief discussion on combining model together (model averaging)
- conclusions

Short & Long Memory Models

- will consider two Gaussian stationary models (stationarity reasonable; Gaussianity open to question)
  1. first order autoregressive process (AR(1))
  2. 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

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

  where $\tau \in \mathbb{Z}$ (set of all integers) & $|f| \leq 1/2$

- default model for correlated time series in climatology
- related to discretized 1st order differential equation (has single damping constant dictated by ϕ)
- 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 $X_{n\lceil \tau_D \rceil}$, $n \in \mathbb{Z}$, 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:

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

  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. $|\delta| < 1/2$ (if $\delta = 0$, then 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 large positive τ and small positive f, have approximately

$$
s_{Y,\tau} \propto \tau^{2\delta-1} \ \ \& \ \ S_Y(f) \propto 1/f^{2\delta}
$$

- related to aggregation of 1st order differential equations 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'
- 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 ('dictionary')
- 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 $\mathbf{Z}$ using orthogonal projections onto elements of $\mathcal{D}$ (similar in spirit to step-wise regression)

Matching Pursuit Algorithm: I

- for each $\mathbf{D}_k \in \mathcal{D}$, form approximation $\mathbf{A}_k \equiv \langle \mathbf{Z}, \mathbf{D}_k \rangle \mathbf{D}_k$, where

$$\langle \mathbf{Z}, \mathbf{D}_k \rangle \equiv \sum_{t=0}^{N-1} Z_t D_{k,t}$$

  (assumes $\mathbf{D}_k$ real-valued; can adjust if not so)

- define residuals $\mathbf{R}_k \equiv \mathbf{Z} - \mathbf{A}_k$ so that $\mathbf{Z} = \mathbf{A}_k + \mathbf{R}_k$
- $\mathbf{A}_k$ and $\mathbf{R}_k$ are orthogonal; i.e., $\langle \mathbf{A}_k, \mathbf{R}_k \rangle = 0$
- hence $\|\mathbf{Z}\|^2 = \|\mathbf{A}_k\|^2 + \|\mathbf{R}_k\|^2 = \left|\langle \mathbf{Z}, \mathbf{D}_k \rangle\right|^2 + \|\mathbf{R}_k\|^2$
- to minimize $\|\mathbf{R}_k\|^2$, select $k^{(1)}$ such that

$$\left|\langle \mathbf{Z}, \mathbf{D}_{k^{(1)}} \rangle\right| = \max_{\mathbf{D}_k \in \mathcal{D}} |\langle \mathbf{Z}, \mathbf{D}_k \rangle|$$

- let $\mathbf{A}^{(1)}$ & $\mathbf{R}^{(1)}$ be approximation and residuals
- 1st stage of algorithm thus yields $\mathbf{Z} = \mathbf{A}^{(1)} + \mathbf{R}^{(1)}$
- 2nd stage: use $\mathbf{R}^{(1)}$ rather than $\mathbf{Z}$ in above
- yields $\mathbf{R}^{(1)} = \mathbf{A}^{(2)} + \mathbf{R}^{(2)}$ with $k^{(2)}$ picked such that

$$\left|\langle \mathbf{R}^{(1)}, \mathbf{D}_{k^{(2)}} \rangle\right| = \max_{\mathbf{D}_k \in \mathcal{D}} \left|\langle \mathbf{R}^{(1)}, \mathbf{D}_k \rangle\right|$$

Matching Pursuit Algorithm: II

- after m such steps, have additive decomposition

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

  where $\widehat{\mathbf{Z}}^{(m)}$ is mth order approximation to $\mathbf{Z}$

- also have 'energy' decomposition

$$\begin{aligned} \|\mathbf{Z}\|^2 &= \sum_{n=1}^{m} \|\mathbf{A}^{(n)}\|^2 + \|\mathbf{R}^{(m)}\|^2 \\ &= \sum_{n=1}^{m} \left| \langle \mathbf{R}^{(n-1)}, \mathbf{D}_{k(n)} \rangle \right|^2 + \|\mathbf{R}^{(m)}\|^2, \end{aligned}$$

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

- note: as m increases, $\|\mathbf{R}^{(m)}\|^2$ 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 possible 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

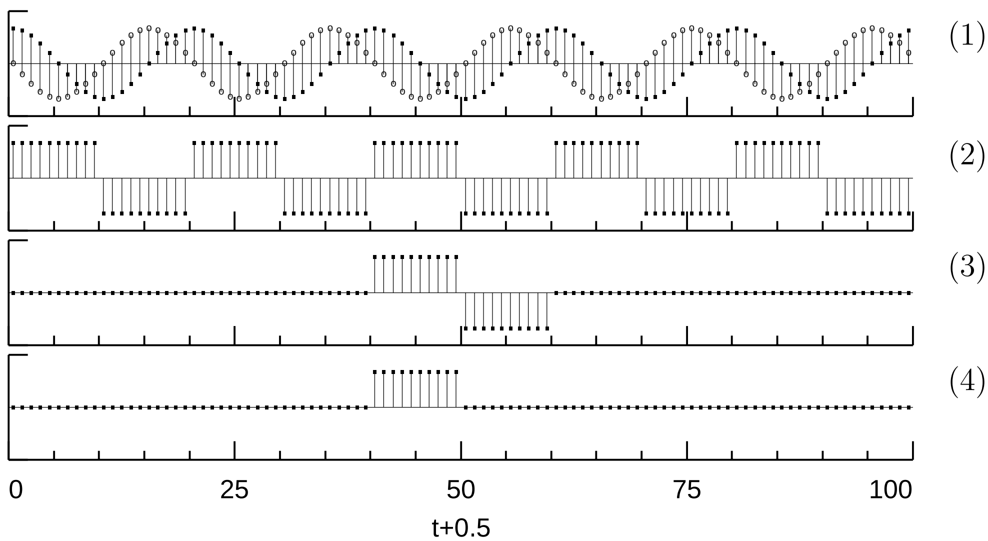

Figure 3: Examples of dictionary vectors $\mathbf{D}_k$ used in various matching pursuits of the NP index. The elements $D_{k,t}$, $t = 0, \ldots, 99$, for each $\mathbf{D}_k$ are plotted versus $t + 0.5$. The vector in (1) is a complex-valued vector from an orthonormal discrete Fourier transform (the real and imaginary parts are indicated by, respectively, solid dots and open circles). The period associated with this vector is twenty. In (2), the vector contains a square wave oscillation, also with a period of twenty. In (3), the vector is created from a discretized Haar wavelet function associated with changes on a scale of ten, while (4) shows one from a corresponding Haar scaling function.

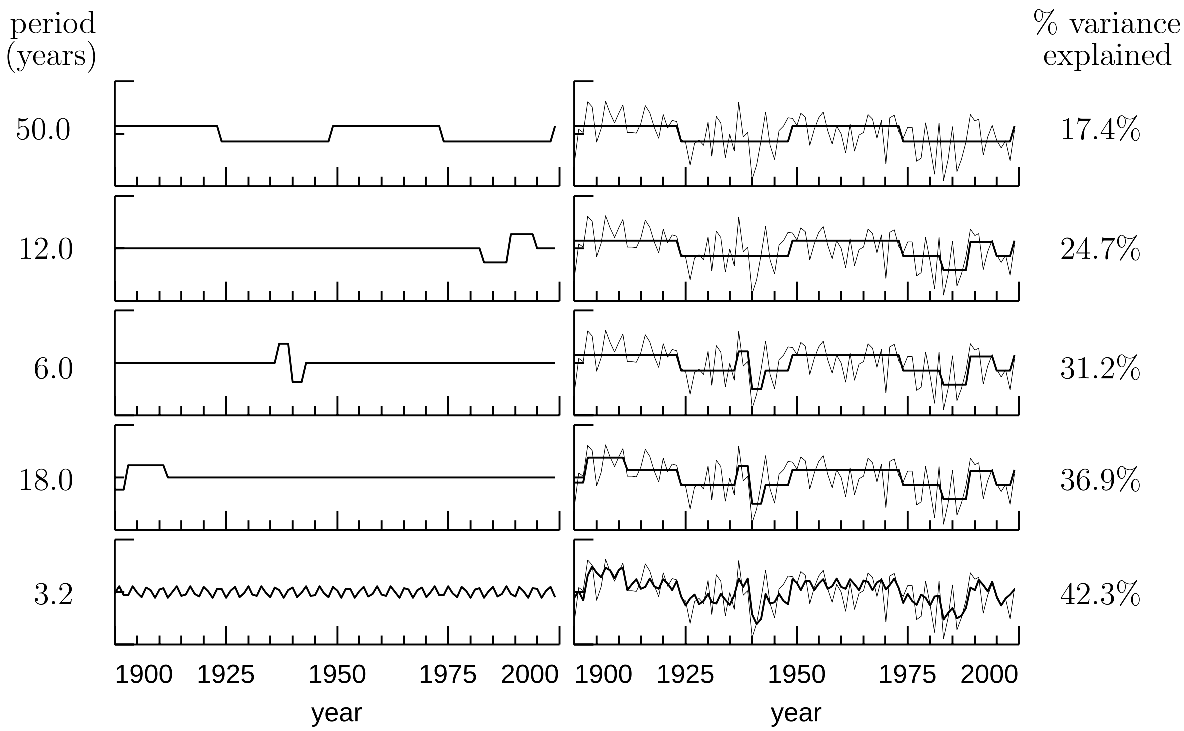

Figure 4: Matching pursuit of NP index using dictionary consisting of sinusoids, square wave oscillations, Haar wavelet vectors and Haar scaling vectors. The thin jagged curve in each right-hand plot shows the NP index $\mathbf{Z}$. The thick curves in the left-hand plots depict the vector that was selected in steps $m = 1, \ldots, 5$ (top to bottom, respectively). The thick curves in the right-hand plots show the corresponding approximation $\widehat{\mathbf{Z}}^{(m)}$. The period associated with each vector is stated in the left-hand margin, while the right-hand margin lists the percentage of the variance that is explained by $\widehat{\mathbf{Z}}^{(m)}$ (by definition, this is $(\|\mathbf{Z}\|^2 - \|\mathbf{R}^{(m)}\|^2)/\|\mathbf{Z}\|^2 \times 100\%$, where $\mathbf{R}^{(m)} = \mathbf{Z} - \widehat{\mathbf{Z}}^{(m)}$).

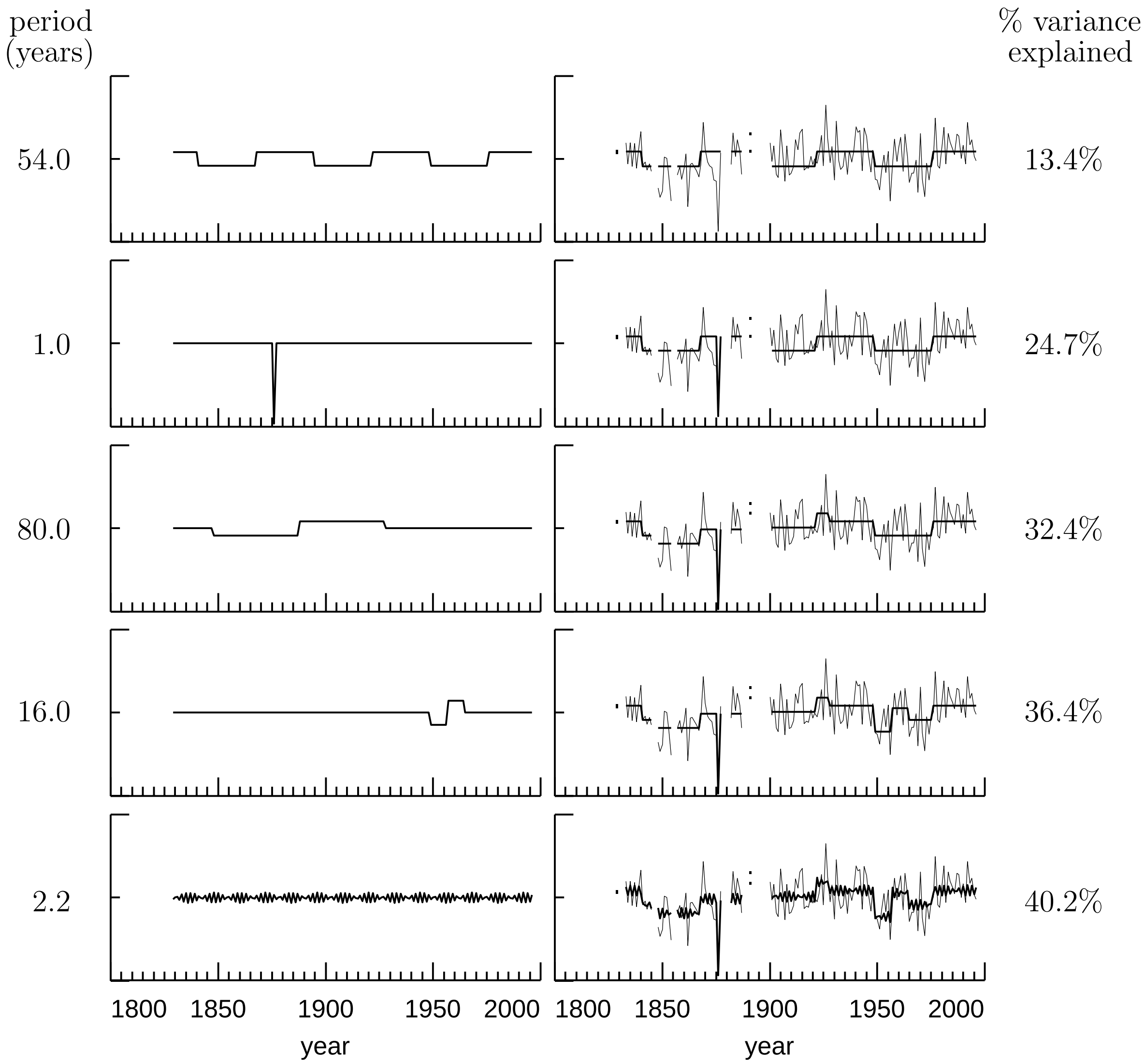

Figure 5: As in Figure 4, but now using the Sitka air temperatures.

Square Wave Oscillation Model: III

- will consider simple SWO model:

$$Z_t = \mu_Z + \beta D_{k(1),t} + e_t$$

  - μ_Z & β are parameters (if $\beta = 0$, Z_t is white noise)
  - $D_{k(1),t}$ part of 1st vector from 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, β & σ_e^2 via maximum likelihood (ML)

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 $(1-\phi^2)/N$ & $2\sigma_\epsilon^4/N$
  - $\hat{\delta}$ & $\hat{\sigma}_\varepsilon^2$ are approximately normally distributed with means δ & σ_ε^2 and variances $6/(\pi^2 N)$ & $2\sigma_\varepsilon^4/N$
  - $\hat{\beta}$ & $\hat{\sigma}_e^2$ are approximately normally distributed with means β & σ_e^2 and variances σ_e^2 & $2\sigma_e^4/N$
- 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 Gaussian 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: plots of residuals and histograms (latter calls into question Gaussian assumption)
- Fig. 7: 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 $\hat{\rho}_\tau$)
- found similar results for Sitka air temperatures
- can use goodness of fit tests for quantitative assessment of models

model	parameter	95% CI	σ	95% CI
AR	$\hat{\phi} = 0.21$	$[0.02, 0.40]$	$\hat{\sigma}_\epsilon = 2.37$	$[2.01, 2.67]$
FD	$\hat{\delta} = 0.17$	$[0.02, 0.32]$	$\hat{\sigma}_\varepsilon = 2.35$	$[2.00, 2.66]$
SWO	$\hat{\beta} = -10.09$	$[-14.51, -5.67]$	$\hat{\sigma}_e = 2.21$	$[1.88, 2.50]$

Table 1: Autoregressive (AR), fractionally differenced (FD) and square wave oscillator (SWO) process parameter estimates for the NP index.

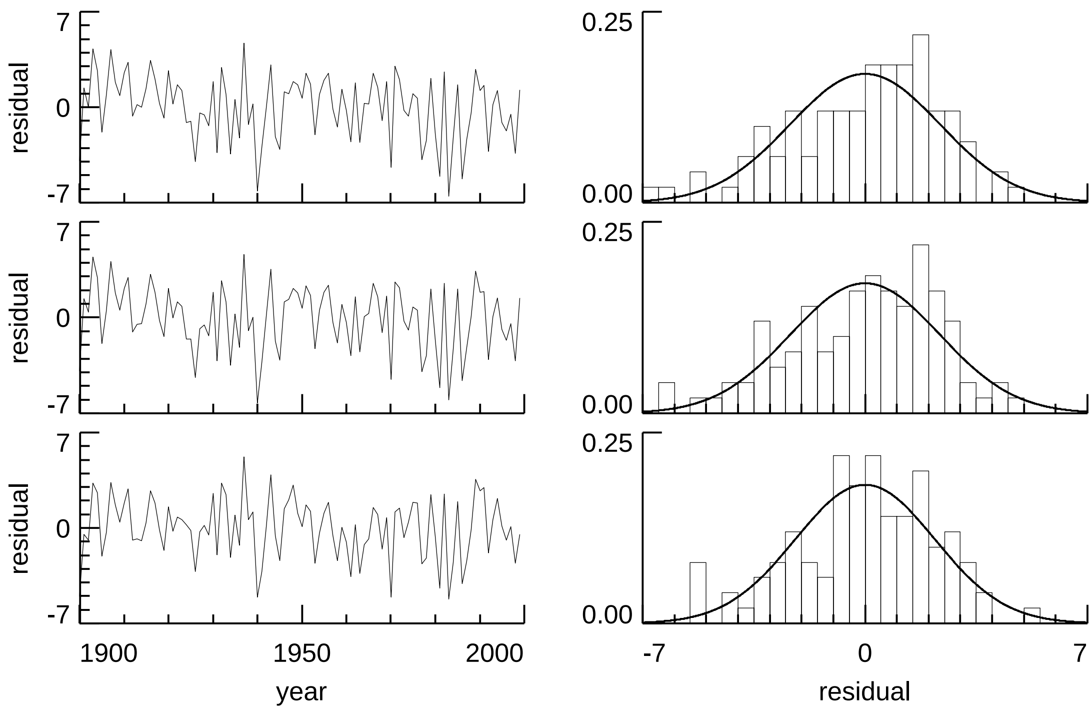

Figure 6: Residuals from AR, NP and SWO fits to NP index (top to bottom, respectively, on left-hand plots), along with histograms and Gaussian probability density functions (right-hand plots) with same mean and variances as the residuals.

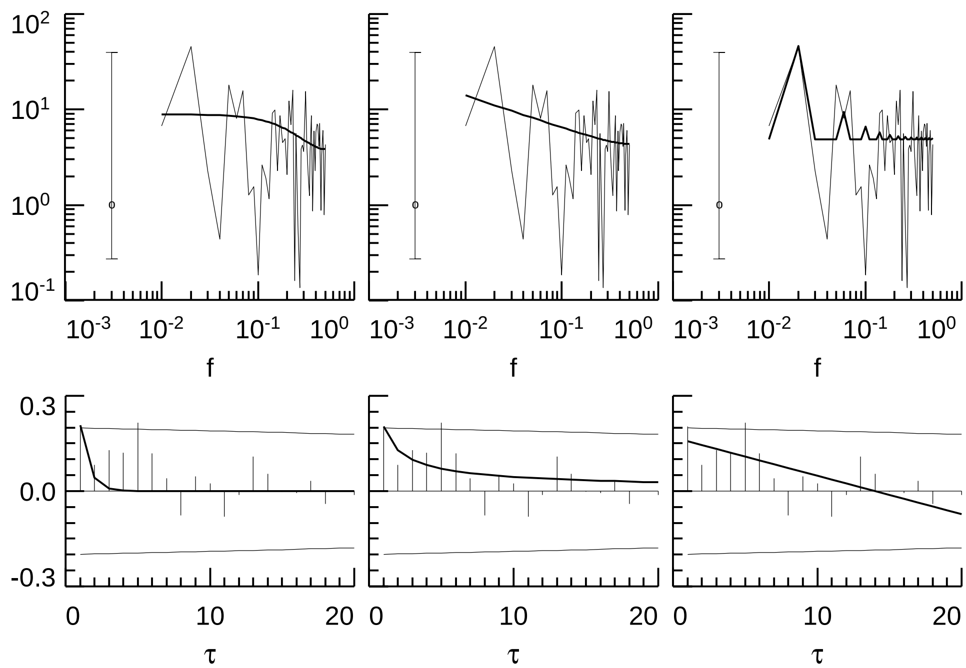

Figure 7: Sample autocorrelation sequence (ACS) and periodogram for the NP index, along with theoretical ACSs and spectral density functions (SDFs) for fitted AR, FD and SWO models (left, middle and right plots, respectively).

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:

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

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

Goodness of Fit Tests: II

- if T_j 'too big,' reject 'model is adequate' hypothesis
- can determine what is 'too big' under null hypothesis (i.e., 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?

j	model	T_j	$Q_j(0.90)$	$Q_j(0.95)$	$Q_j(0.99)$	$\alpha = 0.05$ test result	$\hat{\alpha}$
1	AR	0.30	0.38	0.39	0.42	fail to reject	0.67
	FD	0.28	"	"	"	fail to reject	0.78
	WN	0.39	"	"	"	reject	0.05
2	AR	0.10	0.17	0.19	0.23	fail to reject	$\gg 0.1$
	FD	0.07	"	"	"	fail to reject	$\gg 0.1$
	SWO	0.10	"	"	"	fail to reject	$\gg 0.1$
	WN	0.21	"	"	"	reject	≈ 0.03
3	AR	4.65	7.74	9.45	13.31	fail to reject	0.32
	FD	3.12	"	"	"	fail to reject	0.54
	SWO	2.83	"	"	"	fail to reject	0.59
	WN	12.63	"	"	"	reject	0.01

Table 2: Model goodness of fit tests for the NP index.

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. 8: power of various test statistics vs. N'
  - at best, 30% chance of rejecting null hypothesis for $N' = 100$
  - need $N' \approx 500$ to have 50% chance of discriminating between AR(1) & FD models
  - no one test uniformly better than others

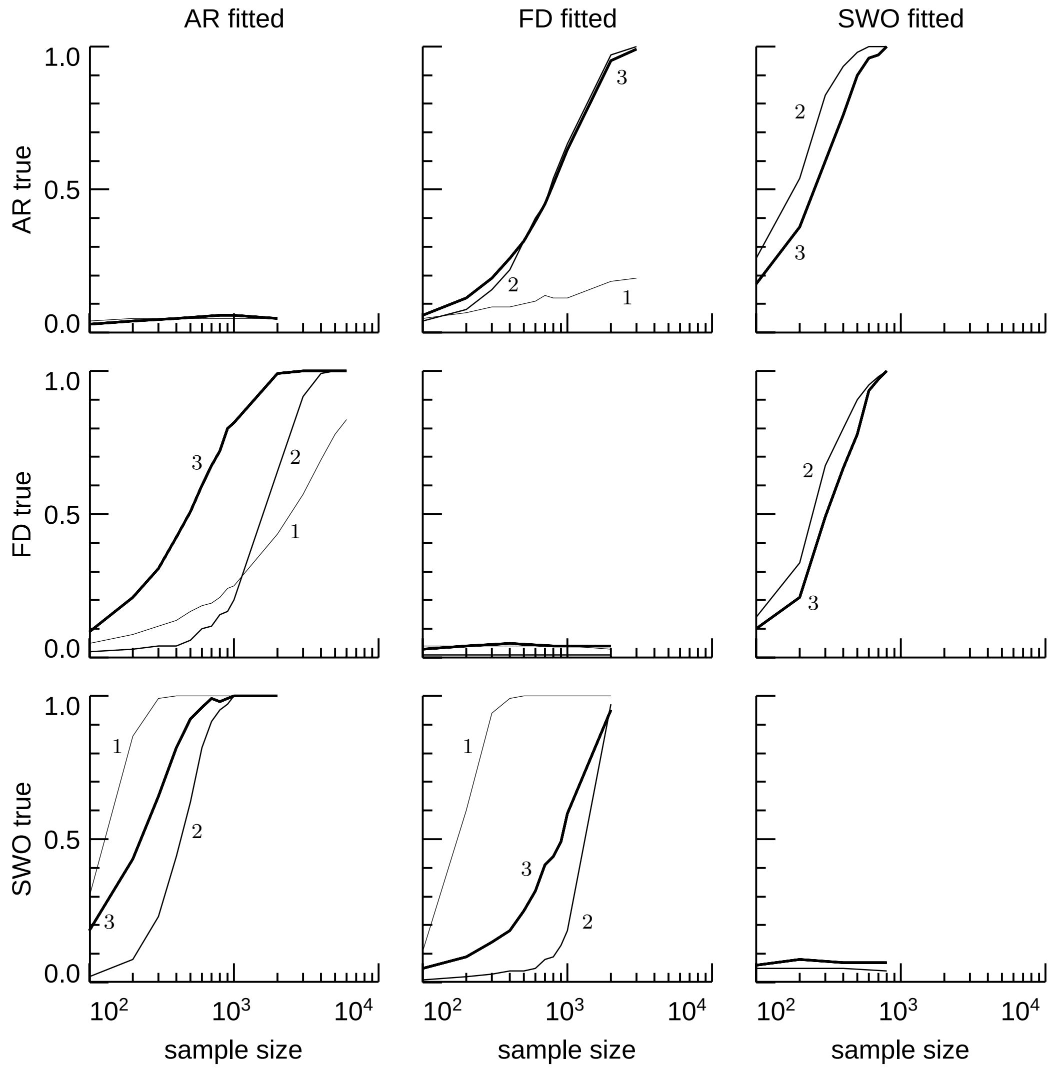

Figure 8: Probability (as a function of sample size) of rejecting the null hypothesis at a 0.05 level of significance that a fitted model A is adequate for a realization of a process B when using the test statistics T_1, T_2 and T_3. For the plots in the left- to right-hand columns, the fitted models A is, respectively, an FD, AR(1) and SWO model. The same ordering is used for the process B for the plots in the top to bottom rows.

Model Implications: I

- no statistical reason to prefer 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. 9: examples of 1000 year simulations
- Q: how well do models support notion of regimes?

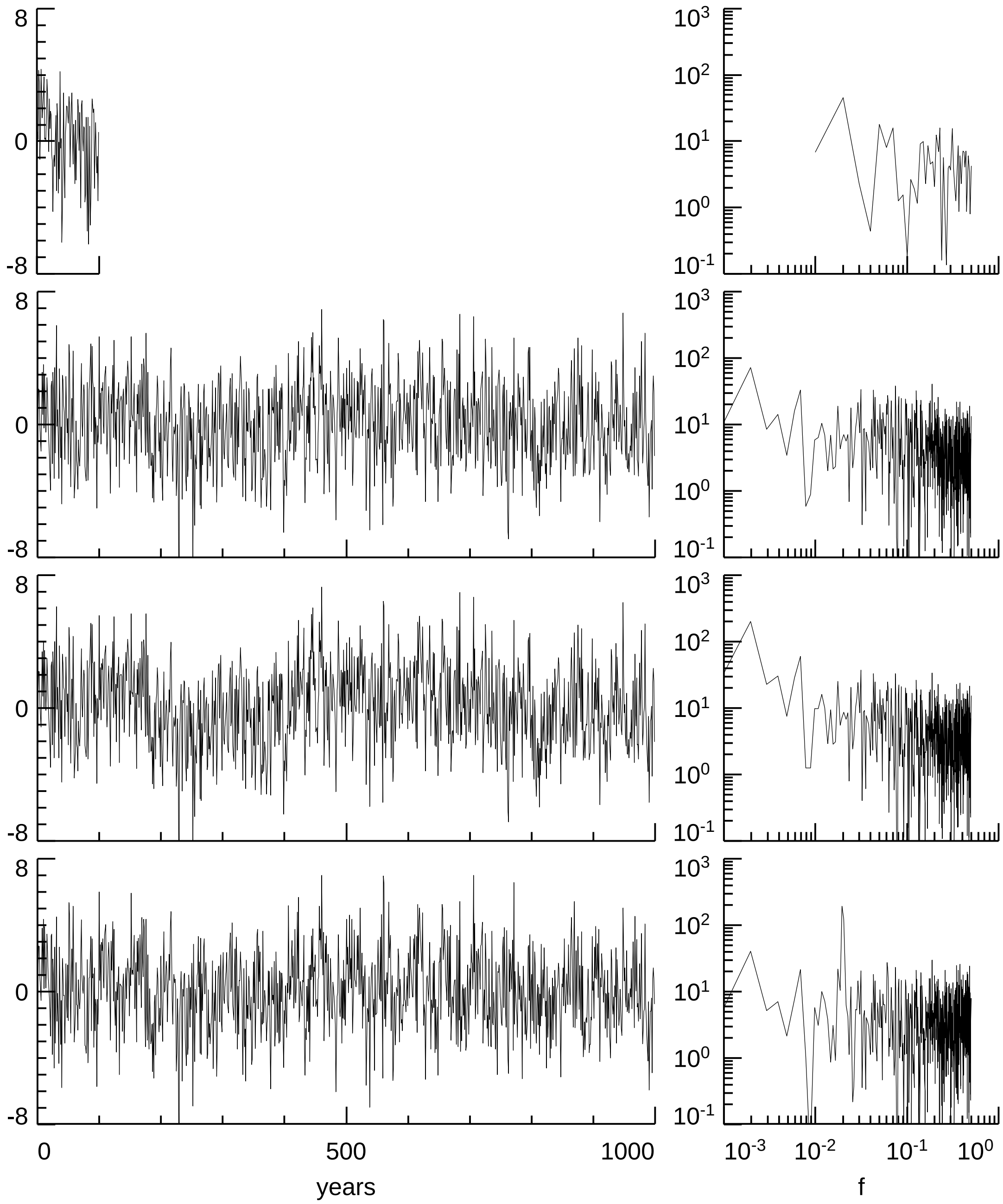

Figure 9: Simulated realizations (left-hand column, bottom three rows) of AR(1), FD and SWO processes (second, third and fourth rows, respectively) with model parameters set to values estimated for the NP index, along with associated periodograms (right-hand column). The actual NP index and its periodogram are shown in the top row. Each realization was created using an exact circulant embedding method. This method converts $2N$ uncorrelated standard Gaussian deviates into the desired realization of length N. For each series, the same $2N$ deviates were used to make it easier to compare realizations from the different models.

Model Implications: II

- to address question, consider following experiment
- generate deviate $\tilde{\delta}$ from normal distribution with mean $\hat{\delta}$ from NPI and variance $6/(\pi^2 N) = 6(\pi^2 100)$
- assume FD model with $\tilde{\delta}$ is correct for NPI
- simulate time series of length 1024 from FD model
- tabulate sizes of observed regimes (i.e., run lengths) 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. 10: 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 ≥ 23 is 4 times more likely under FD model than under AR(1)

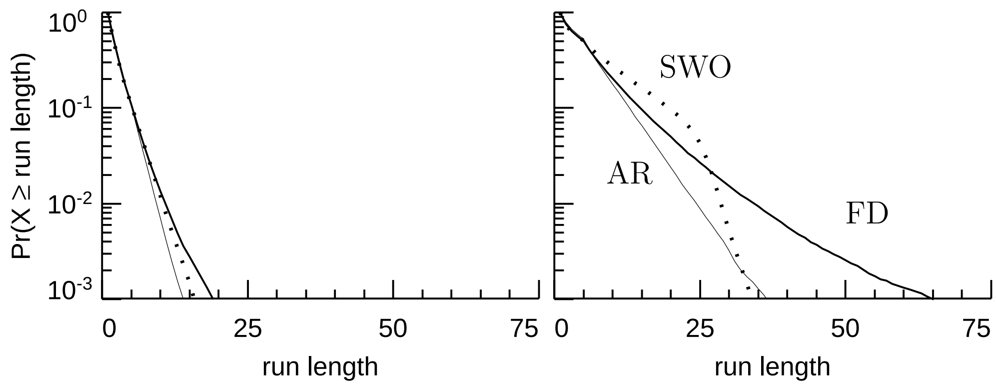

Figure 10: Probability of observing a run that is greater than or equal to a specified run length. The thin, thick and dotted curves denote the AR, FD and SWO processes. The left-hand plot is for processes without smoothing, whereas the right-hand plot is for processes subjected to a five year running average.

Combining Models

- since AR(1), FD & SWO models equally adequate, will consider combined model:

$$U_t \equiv \mu + \left(\widetilde{X}_t + \widetilde{Y}_t + \widetilde{Z}_t\right)/\sqrt{3}$$

  - $\widetilde{X}_t$ is zero mean version of X_t, etc.,
  - $\widetilde{X}_t$, $\widetilde{Y}_t$ and $\widetilde{Z}_t$ are independent of each another

- Fig. 11: 'regime size' experiment redone using combined model
- variations on the above
  - sample from standardized observed residuals to create U_t (little change to Fig. 11, but does affect other statistics)
  - assign random weights to $\widetilde{X}_t$, $\widetilde{Y}_t$ and $\widetilde{Z}_t$ (again, little change to Fig. 11)
- presumably can formulate in terms of 'Bayesian model averaging' (Hoeting, Madigan, Raftery and Volinsky, 1999)

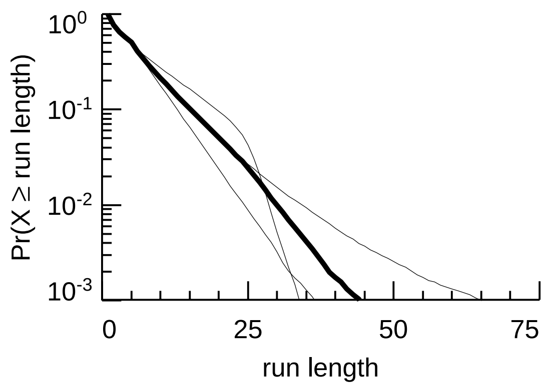

Figure 11: As in right-hand plot of Figure 10, but now showing combined model (thick curve) in comparison to AR, FD and SWO processes (thin curves).

Conclusions

- AR(1), FD & SWO models equally adequate for NPI and Sitka air temperatures
- all 3 models include white noise as special case
  (all 3 lead to rejection of hypothesis of white noise)
- SWO models picked out by matching pursuit – offers some support for Minobe's hypothesis
- cannot realistically hope to distinguish between three models given available sample sizes
- 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
- combining models has considerable appeal (focus of current efforts)