Wavelet-Based Analysis for Multispectral Fractal Processes

Don Percival
MathSoft, Inc., Seattle, WA
Applied Physics Lab, Univ. of Washington (UW)

overheads for talk available at

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

joint work with:

- Jim Bassingthwaighte, Dept. of Bioeng, UW
- Bill Constantine, MathSoft (PI)
- Peter Craigmile, Dept. of Statistics, UW
- Peter Guttorp, Dept. of Statistics, UW
- Jim Pitton, Applied Physics Lab, UW
- Per Reinhall, Dept. of Mech. Eng., UW (PI)
Introduction and Overview

• motivation: ABL aerothermal turbulence data
  – Fig. 1: 7.5 million points (100 point averages)
  – spatial resolution \( \approx 1.83 \) cm

• will model using time-varying stochastic process

• basic idea: combine wavelets with stochastic fractals
  – wavelets give time/scale decomposition
    (yields multiscale approach to modeling)
  – fractals describe connections across scales
    (will use fractionally differenced processes)
Outline of Talk

- overview of discrete wavelet transform (DWT)
- overview of fractionally differenced (FD) processes
- basic properties of DWT of an FD process
  (DWT acts as decorrelator of FD processes)
- DWT-based estimation of parameters for FD process
  – maximum likelihood and least squares estimators
- application to ABL data
- future work
Overview of DWT: I

• let $X = [X_0, X_1, \ldots, X_{N-1}]^T$ be observed time series (for convenience, assume $N$ integer multiple of $2^{J_0}$)
• let $W$ be $N \times N$ orthonormal DWT matrix
• $W = WX$ is vector of DWT coefficients
• orthonormality says $X = W^TW$, so $X \Leftrightarrow W$
• can partition $W$ as follows:

$$W = \begin{bmatrix}
W_1 \\
\vdots \\
W_{J_0} \\
V_{J_0}
\end{bmatrix}$$

• $W_j$ contains $N_j = N/2^j$ wavelet coefficients
  – related to changes of averages at scale $\tau_j = 2^{j-1}$ ($\tau_j$ is $j$th ‘dyadic’ scale)
  – related to times spaced $2^j$ units apart
• $V_{J_0}$ contains $N_{J_0} = N/2^{J_0}$ scaling coefficients
  – related to averages at scale $\lambda_{J_0} = 2^{J_0}$
  – related to times spaced $2^{J_0}$ units apart
• Fig. 2: DWT of small segment of ABL data
Overview of DWT: II

• obtain DWT via filtering with subsampling

• filter \( X_0, X_1, \ldots, X_{N-1} \) to obtain

\[
2^{j/2} \hat{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \mod N}, \quad t = 0, 1, \ldots, N - 1
\]

- \( h_{j,l} \) is \( j \)th level wavelet filter
- width of \( h_{1,l} \) is \( L_j = (2^j - 1)(L - 1) + 1 \)
- \( \hat{W}_{j,t} \) part of ‘maximal overlap’ DWT (MODWT)

• subsample to obtain DWT wavelet coefficients:

\[
W_{j,t} = 2^{j/2} \hat{W}_{j,2^j(t+1)-1}, \quad t = 0, 1, \ldots, N_j - 1,
\]

where \( W_{j,t} \) is \( t \)th element of \( W_j \)

• Fig. 3: Haar & ‘least asymmetric’ (LA) wavelet filters

• \( j \)th filter is band-pass with pass-band \( [2^{j-1}, 1] \)

• similarly, scaling filters yield \( V_{J_0} \)

• Fig. 3: Haar & LA(8) scaling filters

• \( J_0 \)th filter is low-pass with pass-band \( [0, 1] \)
Overview of FD Processes

• $X_t$ called fractionally differenced (FD) process if it has a spectral density function (SDF) given by

$$S_X(f) = \frac{\sigma^2}{|2\sin(\pi f)|^{2\delta}},$$

where $\sigma^2 > 0$ and $-\infty < \delta < \infty$

• Fig. 4: for small $f$, have $S_X(f) \approx C/|f|^{2\delta}$; i.e., ‘$1/f$ type,’ ‘power law’ or ‘fractal’ process

• also called ARFIMA(0,δ,0) process

• special cases
  – stationary if $\delta < \frac{1}{2}$
    * white noise if $\delta = 0$
    * has ‘long memory’ if $0 < \delta < \frac{1}{2}$
      · autocorrelation sequence $s_{X,\tau} \approx C_s \tau^{-1+2\delta}$
      · quite similar to fractional Gaussian noise
  – has stationary increments if $\delta \geq \frac{1}{2}$
    * random walk if $\delta = 1$
    * like fractional Brownian motion if $\frac{1}{2} < \delta < \frac{3}{2}$
    * like $-\frac{5}{3}$ power law (Kolmogorov) if $\delta = \frac{5}{6}$
DWT of FD Processes

• Fig. 5: DWT of realization of FD process ($\delta = 0.4$)
• sample ACSs suggest random variables (RVs) in $W_j$ are approximately uncorrelated
• ignoring ‘boundary’ coefficients, $W_j$ is stationary
• Fig. 6: SDFs for $W_j$, $j = 1, 2, 3, 4$
  – quite close to white noise
  – remaining structure close to SDF for first or second order autoregressive process
• $W_j \& W_{j'}$, $j \neq j'$, approximately uncorrelated (can improve approximation by increasing $L$)
• DWT acts as a whitening transform (basis for wavelet-based maximum likelihood scheme)
• have $\nu^2_{X}(\tau_j) \equiv \text{var}\{\widetilde{W}_{j,t}\} \propto \tau_j^{2\delta-1}$ approximately
  – implies $\log(\nu^2_{X}(\tau_j)) \approx \zeta + (2\delta - 1) \log(\tau_j)$
  – $\nu^2_{X}(\tau_j)$ called wavelet variance (note: based on MODWT $\tilde{W}_{j,t}$ rather than DWT $W_{j,t}$)
  – basis for wavelet-based least squares scheme
ML Estimation for FD Processes: I

- suppose we are given $U_0, \ldots, U_{N-1}$ such that
  \[ U_t = T_t + X_t \]

  where $T_t \equiv \sum_{j=0}^{r} a_j t^j$ is polynomial trend & $X_t$ is FD process

- width $L$ wavelet filter has embedded differencing operation of order $L/2$

- if $\frac{L}{2} \geq r + 1$, reduces polynomial trend to 0

- can partition DWT coefficients as
  \[ W = W_s + W_b + W_w \]

  where
  - $W_s$ has scaling coefficients and 0s elsewhere
  - $W_s$ has boundary-dependent wavelet coefficients
  - $W_w$ has boundary-independent wavelet coefficients
ML Estimation for FD Processes: II

• since $U = \mathcal{W}^T W$, can write
  $$U = \mathcal{W}^T (W_s + W_b) + \mathcal{W}^T W_w \equiv \bar{T} + \bar{X}$$

• can use values in $W_w$ to form likelihood:
  $$L(\delta, \sigma^2_\epsilon) \equiv \prod_{j=1}^{J_0} \prod_{t=1}^{N'_j} \frac{1}{(2\pi \sigma^2_j)^{1/2}} e^{-W_{j,t}^2 + L'_j - 1/(2\sigma^2_j)}$$

where
  $$\sigma^2_j \equiv \int_{-1/2}^{1/2} \mathcal{H}_j(f) \frac{\sigma^2_\epsilon}{|2\sin(\pi f)|^{2\delta}} df,$$

and $\mathcal{H}_j(f)$ is squared gain for $h_{j,l}$

• leads to maximum likelihood estimator $\hat{\delta}^{(ml)}$ for $\delta$

• $\hat{\delta}^{(ml)}$ asymptotically normal with mean $\delta$ and
  $$\text{var} \{\hat{\delta}^{(ml)}\} = 2\left[ \sum_{j=1}^{J_0} N'_j \gamma_j^2 - \frac{1}{N'} (\sum_{j=1}^{J_0} N'_j \gamma_j)^2 \right]^{-1},$$

where $N' \equiv \sum_{j=1}^{J_0} N'_j$ and

$$\gamma_j \equiv \frac{d\text{var} \{W_{j,t}\}}{d\delta} \left[ \frac{4\sigma^2_\epsilon}{\text{var} \{W_{j,t}\}} \int_0^{1/2} \mathcal{H}_j(f) \frac{\log (2\sin(\pi f))}{[2\sin(\pi f)]^{2\delta}} df \right]$$

• works well in Monte Carlo simulations
LS Estimation for FD Processes: I

- define unbiased estimator of wavelet variance $\nu_X^2(\tau_j)$:
  \[
  \hat{\nu}_X^2(\tau_j) \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{W}_{j,t}^2, \quad \text{where } M_j \equiv N - L_j + 1
  \]

- $\hat{\nu}_X^2(\tau_j)$ is approximately distribution as $\nu_X^2(\tau_j)\chi_{\eta_j}^2/\eta_j$, where
  - $\chi_{\eta_j}^2$ is chi-square RV with $\eta_j$ degrees of freedom
  - can approximate $\eta_j$ by $\max\{M_j/2^j, 1\}$

- using $\log (\nu_X^2(\tau_j)) \approx \zeta + \beta \log (\tau_j)$ with $\beta \equiv 2\delta - 1$, can formulate regression model; with
  \[
  Y(\tau_j) \equiv \log (\hat{\nu}_X^2(\tau_j)) - \psi(\frac{\eta_j}{2}) + \log (\frac{\eta_j}{2}),
  \]
  have $Y(\tau_j) = \zeta + \beta \log (\tau_j) + e_j$, where
  \[
  e_j \equiv \log \left( \frac{\hat{\nu}_X^2(\tau_j)}{\nu_X^2(\tau_j)} \right) - \psi(\frac{\eta_j}{2}) + \log (\frac{\eta_j}{2})
  \]
  has distribution $\log (\chi_{\eta_j}^2) - \psi(\frac{\eta_j}{2}) - \log (2)$

- have $E\{e_j\} = 0$ and $\text{var} \{e_j\} = \psi'(\frac{\eta_j}{2})$, where $\psi'(\cdot)$ is trigamma function

- $e_j$ approximately Gaussian if $\eta_j \geq 10$
LS Estimation for FD Processes: II

• weighted least squares (LS) estimator for $\beta$:

$$
\hat{\beta}^{(wls)} = \frac{\sum w_j \sum w_j \log (\tau_j) Y(\tau_j) - \sum w_j \log (\tau_j) \sum w_j Y(\tau_j)}{\sum w_j \sum w_j \log^2(\tau_j) - (\sum w_j \log (\tau_j))^2},
$$

where $w_j \equiv 1/\psi'(\eta_j^2)$

• have

$$
\text{var} \{ \hat{\delta}^{(wls)} \} = \frac{\sum w_j}{\sum w_j \sum w_j \log^2(\tau_j) - (\sum w_j \log (\tau_j))^2}
$$

• use $\delta = \frac{1}{2}(\beta + 1)$ to get $\hat{\delta}^{(wls)} \equiv \frac{1}{2}(\hat{\beta}^{(wls)} + 1)$ with

$$
\text{var} \{ \hat{\delta}^{(wls)} \} = \frac{1}{4} \text{var} \{ \hat{\beta}^{(wls)} \}
$$

• works well in Monte Carlo simulations
Analysis of ABL Data: I

• initial approach: divide into nonoverlapping blocks
  – each block has 10,000 points
  – blocks are contiguous
  – allows analysis out to $\tau_{10} = 9.37$ meters

• Fig. 7: wavelet variance estimates for ‘typical’ block
  – based upon LA(8) wavelet filter
  – single $\delta$ (i.e., power law) inadequate
  – will combine 3 adjacent scales via separate FD models

• Fig. 8, lower left-hand portion: scatter plots for $\log(\hat{\nu}_{X,b}^2(\tau_j))$
  – $b$ is block index
  – $\log(\hat{\nu}_{X,b}^2(\tau_j))$ versus $\log(\hat{\nu}_{X,b}^2(\tau_k))$ for different $j, k$
  – lines shows expected pattern if $\sigma^2$ held fixed, but $\delta$ is changing across blocks
  – reasonable agreement at higher scales when $k = j \pm 1$ or $k = j \pm 2$
Analysis of ABL Data: II

- Fig. 8, upper right-hand portion: ‘slope differential’ plots
  - plot for \((j, j + 1)\) with \((k, k + 1)\) defined as
    \[
    \frac{\log (\hat{\nu}_X^2 X, b(\tau_{j+1})) - \log (\hat{\nu}_X^2 X, b(\tau_j))}{\log (\hat{\nu}_X^2 X, b(\tau_{k+1})) - \log (\hat{\nu}_X^2 X, b(\tau_k))} - 1 \text{ versus } b
    \]
  - above is zero if estimated slopes are identical
  - box plots assess significance of deviations from 0

- conclusion: reasonable to combine scales as suggested by ‘typical’ block

- Fig. 9: blocked WLS estimates of power law exponent \(\alpha \equiv -2\delta\)
  - scale \(\tau_4\) has periodic burst (artifact?)
  - scales \(\tau_5, \tau_6, \tau_7\) swing from \(\alpha = 0\) to \(-\frac{5}{3}\)
  - 95% confidence intervals say variations in \(\alpha\) are significant
Analysis of ABL Data: III

• Fig. 10: comparison of WLS estimates for scales $\tau_5, \tau_6, \tau_7$ and $\tau_8, \tau_9, \tau_{10}$
  
  - two groups do not track each other
  
  - largest scales generally consistent with $-\frac{5}{3}$ power law, but show significant deviations at times

• Figs. 11–2: corresponding plots for ML estimates
  
  - very good agreement with WLS estimates (except for $\tau_1, \tau_2, \tau_3$ – not surprising)
  
  - 95% confidence intervals similar to those for WLS (but now block dependent)
Future Work

• ‘instantaneous’ LS and ML estimates
  – designed to get away from block dependence
  – Fig. 13: use MODWT coefficients co-located across scales (one coefficient per scale)
  – easy to modify LS and ML estimators
  – Fig. 14: preliminary LS results for scales $\tau_5, \tau_6, \tau_7$
    * individual estimates very noisy, so have smoothed
    * good agreement with blocked estimates
  – need to study distributional properties of instantaneous estimates
  – need to study ways to denoise instantaneous estimates (waveshrink)

• need to study ways to model evolution of $\alpha$

• need to study ways of combining multiscale models
Papers, Thesis and Book


http://www.staff.washington.edu/dbp/wmtsa.html