Wavelet methods
Peter Craigmile

• Wavelet transformations
  – Some useful references

• The discrete wavelet transform (DWT)
  – Defining the wavelet and scaling filter
  – Daubechies wavelet filters
  – The pyramid algorithm
  – Partitioning the DWT coefficients
  – DWT plot of the Malindi coral series

• Decomposing trend using wavelets

• Estimating the parameters of an FD process with trend

• Returning to the Malindi example

• Limitations of the DWT

• Decomposing the variance using wavelets

• Wavelet-based analysis of variance for the daily Swedish temperatures
Wavelet transformations

• We previously used spectral analysis to view a space-time process in terms of a sum of sinusoids at different frequencies.
  – In many such transformations we lose all information about time.

• A wavelet transformation uses wavelets (“small waves”) to decompose a time series into:
  1. averages over a specified (typically long) time scale, and
  2. changes of averages over a range of shorter time scales.

• It is a time-scale (some say frequency) decomposition.

• By extension, wavelet transforms are a natural tool for investigating the scaling relationships inherent in many space-time processes.
  – Example: Certain climate relationships may be significant over decades, but not visible at shorter times scales [e.g., Madden, 1986]. These relationships can vary spatially.
Some useful references

- For time series: Vidakovic [1998] and Percival and Walden [2000].

- For gridded spatial processes (images): Gonzalez and Woods [2001], Mondal and Percival [2012], Geilhufe et al. [2013]

- Berliner et al. [1999], Wikle et al. [2001], and Craigmile and Guttorp [2011] use hierarchical Bayesian space-time wavelet-based models.

- Examples from the analysis of longitudinal data: Brown et al. [2001], Morris et al. [2003], Vannucci et al. [2003], Morris and Carroll [2006], Morris et al. [2006], Ramirez and Vidakovic [2007].
The discrete wavelet transform (DWT)

- Again we demonstrate with time series.

- Suppose \( \{X_t\} \) is a time series, observed from time 0 to \( N - 1 \) (slight change of notation).

Let \( \mathbf{X} = (X_0 \ldots X_{N-1})^T \).

(assume \( N \) divisible by \( 2^J \), for some \( J \)).

- Most simply, the **DWT coefficients** are

\[
W = \mathcal{W} \mathbf{X},
\]

where \( \mathcal{W} = N \times N \) is an **orthonormal** DWT matrix.

- In practice use **pyramid algorithm** [Mallat, 1989] to calculate \( W \).
Defining the wavelet and scaling filter

• We define two filters:

\{h_l\} is a wavelet filter of even length \( L \), i.e., \( \sum_l h_l = 0, \sum_l h_l h_{l+2k} = 1_{[k=0]} \).

\( g_l = (-1)^l h_{L-1-l} \) is the scaling filter.

• There are many choices of wavelet filter we can use in practice.

  – Each filter has different properties.

• A popular choice is of filters is the Daubechies class

  [Daubechies, 1992, Sect. 6.2], which includes the extremal phase \( \text{(D)} \) and least asymmetric \( \text{(LA)} \) sets of wavelet filters.

  – This class of orthogonal filters are characterized by the maximum number of vanishing moments.

  – The wavelet filter \( \{h_l\} \), of even width \( L \), passes high frequencies but filters out low ones.

  – It has \( L/2 \) vanishing moments.
Daubechies wavelet filters
The pyramid algorithm

• Let $V_{0,t} = X_t$ for each $t$.

• For each $j = 1, \ldots, J$ calculate

$$W_{j,t} = \sum_{l=0}^{L-1} h_l \ V_{j-1,2t+1-l \mod N_j},$$

$$V_{j,t} = \sum_{l=0}^{L-1} g_l \ V_{j-1,2t+1-l \mod N_j},$$

for $N_j = N/2^j$ and $t = 0, \ldots, N_j - 1$. 
Partitioning the DWT coefficients

- **Decomposition** of $W$:

  \[
  W = (W_1, W_2, \ldots, W_J, V_J).
  \]

- $W_j$ are the $N_j$ level $j$ **wavelet** coefficients:
  - associated with changes in averages on scale $2^{j-1}$;
  - relates to times spaced $2^j$ units
  - is associated with the frequency interval

  \[
  \lambda_j = [-1/2^j, -1/2^{j+1}) \cup (1/2^{j+1}, 1/2^j]
  \]

  (this follows by examining Fourier transforms of the wavelet and scaling filters).

- $V_J$ are the $N_J$ **scaling** coefficients:
  - associated with averages on scale $2^J$;
  - relates to times spaced $2^J$ units apart.
Decomposing trend using wavelets

- Can also decompose $W$ as

$$W = W_s + W_b + W_{nb}.$$  

$W_s$: scaling coefficients and zeros elsewhere.

$W_b$: wavelet coef. affected by boundaries.

$W_{nb}$: wavelet coef. not affected by boundaries.

- Since $X = \mathcal{W}^T W$

$$X = \mathcal{W}^T (W_s + W_b) + \mathcal{W}^T (W_{nb})$$

$$= \tilde{\mu} + \tilde{\epsilon},$$

where

$\tilde{\mu}$: an estimate of trend;

$\tilde{\epsilon}$: tapered estimate of error.
Estimating the parameters of the FD process

- Suppose that a model for the Malindi oxygen isotope series is

\[ X_t = \mu_t + \epsilon_t, \]

where \( \mu_t \) is a polynomial of degree \( K \), and \( \epsilon_t \) is a mean zero stationary Gaussian error process.

- Again suppose \( \{\epsilon_t\} \) is a long memory process – the fractionally differenced (FD) process with parameters \( \delta \) and \( \sigma^2 \).

- We use a filter of length \( L/2 \geq \max\{K + 1, \lfloor \delta + 1/2 \rfloor \} \).

- Since the wavelet filter acts as a differencing filter:
  - \( W_{nb} \) does not contain the trend component;
  - \( E(W_{nb}) = 0 \);
  - Can use \( W_{nb} \) to estimate the FD process parameters (e.g., via least squares or maximum likelihood).

- Can employ approximate schemes to estimate the process parameters [e.g., McCoy and Walden, 1996, Craigmile et al., 2005, Craigmile and Mondal, 2013].
Returning to the Malindi example

• Using an approximate ML estimate:

\[ \hat{\delta} = 0.359 \ (95\% \ CI \ of \ [0.143, 0.597]) \]
\[ \hat{\sigma} = 0.0667. \]

• **Strong** evidence of long range dependence.
  
  – But cannot discriminate between a stationary \((\delta < 1/2)\)
    or nonstationary FD process \((\delta \geq 1/2)\).

• Evidence that the oxygen concentration is above average for the latter years. (Monte Carlo test for trend p-value < 0.001).

• Can extend to nonpolynomial \(\{\mu_t\}\) using wavelet shrinkage methods [e.g., Donoho and Johnstone, 1994, Percival and Walden, 2000].
Limitations of the DWT

- Not shift invariant [e.g., Percival and Mofjeld, 1997].

- We lose time resolution by downsampling (the use of the ’2t’ in the filtering operation).

- Instead, we can use the maximum overlap discrete wavelet transform (MODWT).

- Let $\tilde{g}_l = g_l / \sqrt{2}$ and $\tilde{h}_l = h_l / \sqrt{2}$. Then the pyramid algorithm that yields the MODWT scaling and wavelet coefficient on level $j$ is given by:

$$\tilde{V}_{j,t} = \sum_l \tilde{g}_l \, V_{j-1,t-l \mod N};$$
$$\tilde{W}_{j,t} = \sum_l \tilde{h}_l \, V_{j-1,t-l \mod N};$$

respectively.

- This transformation is shift invariant. On each level $j$, the wavelet or scaling coefficient can be assigned to a specific time point (not $t$ but a circular shift of $t$).
MODWT plot of the Malindi coral series
Decomposing the variance using wavelets

- Suppose we perform the MODWT (ignoring circularity) on the entire process, \( \{X_t\} \), analyzing to all possible scales.

- Let \( \tilde{H}_j(\cdot) \) denote the transfer function (Fourier transform) of the wavelet filter that would yield the level \( j \) wavelet coefficients directly from \( \{X_t\} \).

- Then the wavelet variance decomposes the variance of \( X_t \):

\[
\text{var}(X_t) = \int_{-1/2}^{1/2} S_X(f) df = \int_{-1/2}^{1/2} \left[ \sum_{j=1}^{\infty} |\tilde{H}_j(f)|^2 \right] S_X(f) df \\
= \sum_{j=1}^{\infty} \int_{-1/2}^{1/2} |\tilde{H}_j(f)|^2 S_X(f) df \\
= \sum_{j=1}^{\infty} \text{var}(\tilde{W}_{j,t}).
\]

- This is a time-scale (frequency) decomposition of the variance since each wavelet coefficient is associated with a certain time point and scale, \( \lambda_j \).
Decomposing the variance continued

• Given data, simple to estimate \( \text{var}(\tilde{W}_{j,t}) \) using averages of the squared wavelet coefficients.

• Properties of these estimators for stationary (and some non-stationary) processes are well known [Percival, 1995, Serroukh et al., 2000]

• Can also extend to gappy time series [Mondal and Percival, 2010, Craigmile and Mondal, 2013], and to images [Mondal and Percival, 2012, Geilhufe et al., 2013].

• This extends to the decomposition of

\[
\text{cov}(X_t, X_{t+h}) \text{ and } \text{cov}(X_t, Y_{t+h})
\]

(and associated correlations) for certain processes \( \{X_t\} \) and \( \{Y_t\} \) [Serroukh and Walden, 2000a,b, Whitcher et al., 2000].

• Currently working on spatio-temporal extensions!
Wavelet-based analysis of variance for the daily Swedish temperatures

6: Sundsvalls Flygplats
Daily average temperature

10: Films Kyrkby A
Daily average temperature

16: Kilsbergen–Suttarboda
Daily average temperature

6: Sundsvalls Flygplats
Deseasonalized and standardized temp.

10: Films Kyrkby A
Deseasonalized and standardized temp.

16: Kilsbergen–Suttarboda
Deseasonalized and standardized temp.

(a)

(b)
References


