Modeling in Transformed Domains: Overview and Generalizations

Debashis Mondal[†] and Don Percival[‡]

[†]Department of Statistics
Oregon State University, Corvallis

[‡]Applied Physics Laboratory
Department of Statistics
University of Washington, Seattle

Outline

- overview: data transforms used in statistics
- data transforms used by Guttorp and co-authors
- questions to motivate breakout session

Overview: Data Transforms Used in Statistics

- extend applicability of linear regression (transform predictors and/or response; see, e.g., Weisberg, 2014, Chapter 8[†])
- force non-Gaussian data into Gaussian shoe
- stabilize variances – e.g., square root transform does so for
  - Poisson-distributed point processes
  - kernel-based estimates of probability density functions
- for time series and spatial series,
  - facilitate modeling
  - facilitate characterization of correlation
  - compensate for correlation
  - help extract signal in presence of noise
  - handle nonstationarities by forcing data into stationary shoe

[†]references Brillinger (1982)

Data Transforms Used by Guttorp and Co-Authors

- Sampson & Guttorp (1991): looked at how power transforms alter interaction effects that exist in pretransformed data
- Sampson & Guttorp (1992): advocated transforms to model nonstationary spatial covariances
- Guttorp & co-authors (10 articles, 1994–2012): used wavelet transforms for, e.g., trend extraction

Power Transforms (Sampson & Guttorp, 1991): I

- problem of interest: look at interaction effect comparing measured pollutant levels before and after closure of copper smelter between regions presumed affected and unaffected by smelter
- for ANOVA analysis, need to apply a transform such that model residuals are approximately normally distributed and of constant variance
- cube root transform yielded residuals nicely satisfying distributional requirements
- alas, after transformation, magnitude of apparent interaction effect smaller
  - in other problems, power transforms often advocated as way to eliminate interactions when null hypothesis is false

Power Transforms (Sampson & Guttorp, 1991): II

- solution: devise test for interaction in original data using transformed data in conjunction with second-order Taylor series expansion (asymptotically equivalent to an approximate likelihood ratio test)
- Monte Carlo simulations verified efficacy of proposed test
- application to copper smelter data indicated closure of smelter did indeed reduce sulfate deposition in a near-downwind region from smelter, but no significant reduction in region further away
- particular lesson: dangerous to assess interactions after a transformation with intent of interpreting assessment directly in terms on interaction on original (raw) data
- general lesson: use of transform to solve one problem can induce a new problem – **no free lunch**!

Another Costly Lunch (Rothrock et al., 2008): I

- problem of interest: assess decline in arctic sea-ice thickness using submarine data collected over a quarter of a century and over different arctic regions
- knowledge of sonar-based recording system allows assessment of amount of variance in original data due to measurement errors
- multiple linear regression used to model annual variations, spatial variations and interannual changes with additive measurement errors
- reasonable to assume normality of errors, but assumption of independence spatially within a given year dicey – evidence for long-range dependence

Another Costly Lunch (Rothrock et al., 2008): II

- while OLS estimates of regression coefficients are unbiased under long-range dependence, statistical theory would advocate use of generalized least square (GLS)
- GLS can be interpreted as OLS after application of a decorrelating transform
- alas, decorrelating transform does not preserve variance – hence can't properly assess effect of measurement errors
- solution: use OLS rather than GLS because standard deviations of OLS-estimated parameters are only 5% greater on the average than GLS-estimated parameters (i.e., although spatial correlation has long-range dependence, overall effect on multiple regression coefficients is small)

Another Costly Lunch (Rothrock et al., 2008): III

- research question (unexplored): does there exist an orthonormal transform (hence variance-preserving) that approximates decorrelating transform well enough to offer an improvement over OLS?

Spatial Transforms (Sampson & Guttorp, 1992): I

- problem of interest: get around common assumption that spatial covariances are stationary – unreasonable due to, e.g., effect of landscape on air pollution or rainfall
- data taken from random function $Z(x, t)$ observed at locations x_i in two-dimensional plane and times t_i
- solution: assume temporal stationarity and model spatial dispersions

  $$D^2(x_i, x_j) = \text{var}\,\{Z(x_i, t) - Z(x_j, t)\} = g(|f(x_i) - f(x_j)|)$$

  as a general smooth function of station pairs (x_i, x_j), where function in question is composition of

  – f, a multidimensional scaling (MDS) mapping

  – g, a monotone function

Spatial Transforms (Sampson & Guttorp, 1992): II

- MDS mapping f implements nonstationary covariances by transforming them into stationary covariances
- careful choice of g yields valid covariances, i.e., ones satisfying condition of nonpositive definiteness
- spatial data consists of measurements of $Z(x,t)$ recorded at locations x and times t, but Sampson–Guttorp deformation method transforms just x's to allow using stationary models to handle certain nonstationarities
- deformation method quite successful – inspiration for a lot of subsequent research

Wavelet Transforms: I

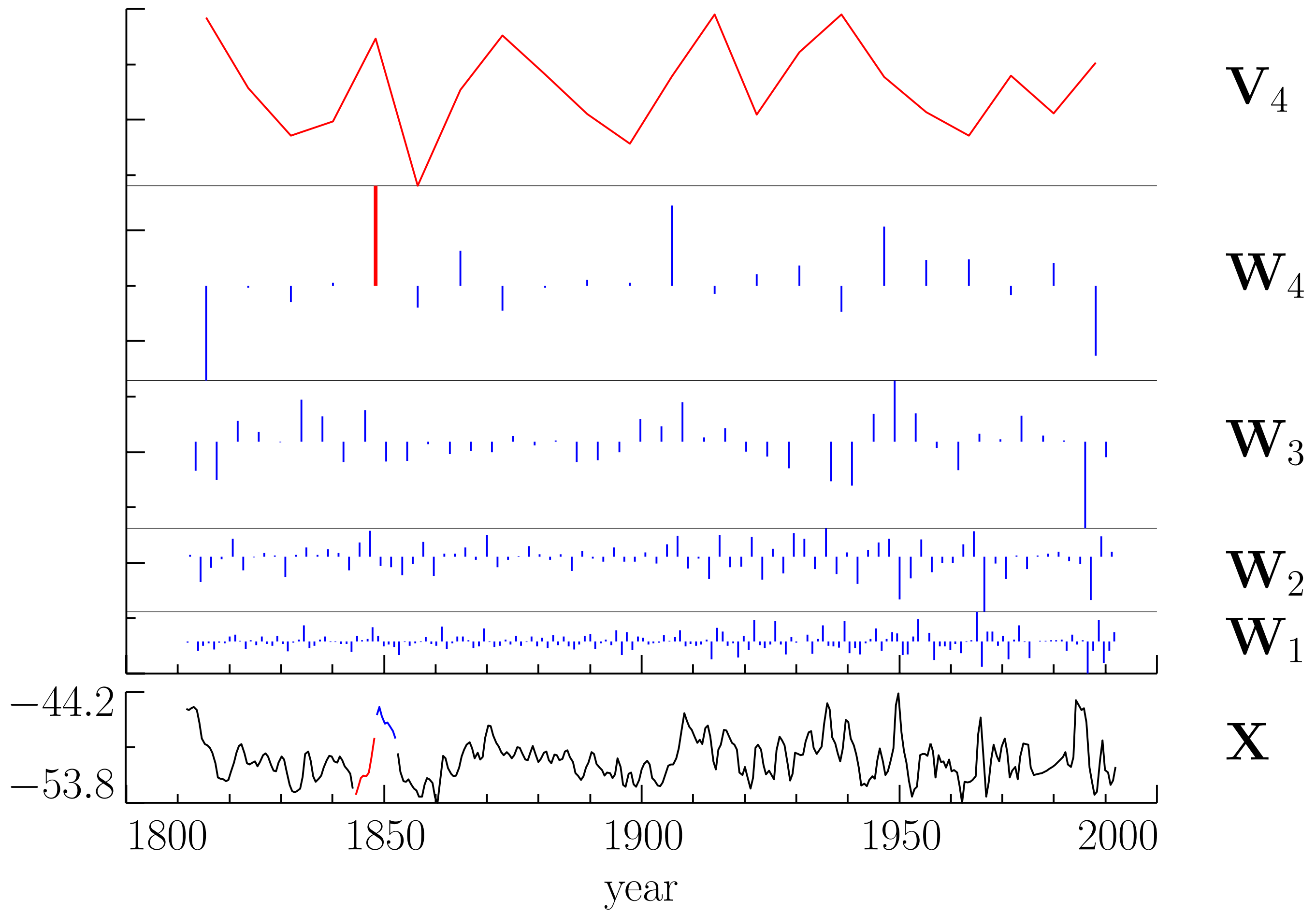
- wavelets are analysis tools for time series and images (primarily)
- interest in wavelet tranforms started in geophysics in early 1980s and then migrated to other fields
- wavelet and Fourier transforms often billed as alternatives
- two transforms have some properties in common, including:
  - transforms fully equivalent to original data (inverse transforms exist to recover data from transform coefficients)
  - transforms preserves variance of original data
  - both act as a decorrelating transform (approximately)
  - manipulation of tranform coefficients – in conjunction with inverse transform – can lead to useful signal extraction
  - transform coefficients attached to physically meaningful variables (frequency or time/scale)

Wavelet Transforms: II

- two transforms differ in important aspects, including
  - for stationary processes, arguably Fourier transform better at decorrelating processes with short-range dependence, while wavelet transform better for long-range dependence
  - Fourier transforms are better at capturing global aspects, while wavelet transforms are better with local aspects
  - types of signals for which two transforms are well adapted quite different – for signals of practical interest, wavelet transform often promotes sparsity better
- see Guttorp et al. (2007) for a comprehensive comparison of two transforms for analyzing space-time processes

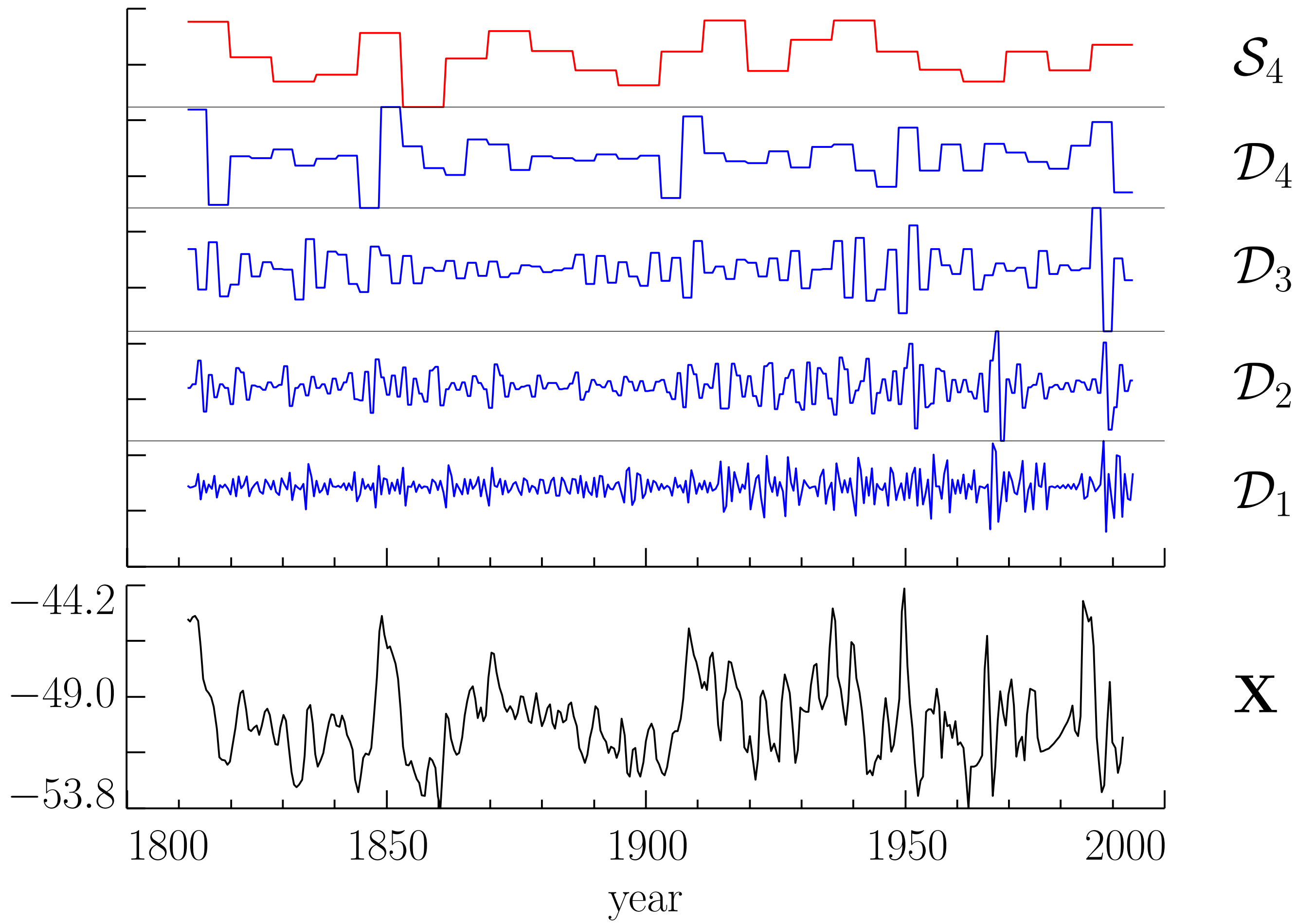
Example of Haar DWT (Four Levels)

- oxygen isotope records $\mathbf{X}$ from Antarctic ice core ($N = 352$)



Multiresolution Analysis

- oxygen isotope records **X** from Antarctic ice core



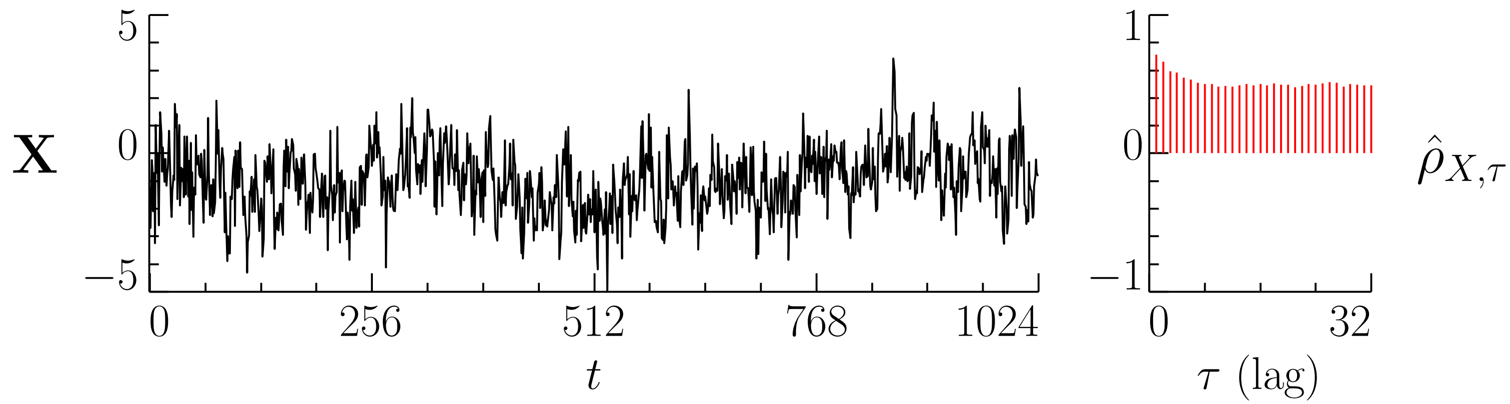
Scale-based Analysis of Variance

- decomposition of sample variance

$$\hat{\sigma}_X^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}\left(X_t - \overline{X}\right)^2 = \sum_{j=1}^{4}\frac{1}{N}\|\mathbf{W}_j\|^2 + \frac{1}{N}\|\mathbf{V}_4\|^2 - \overline{X}^2$$

- Haar-based example for oxygen isotope records
  - 0.5 year changes: $\frac{1}{N}\|\mathbf{W}_1\|^2 \doteq 0.295$ ($\doteq$ 9.2% of $\hat{\sigma}_X^2$)
  - 1.0 years changes: $\frac{1}{N}\|\mathbf{W}_2\|^2 \doteq 0.464$ ($\doteq$ 14.5%)
  - 2.0 years changes: $\frac{1}{N}\|\mathbf{W}_3\|^2 \doteq 0.652$ ($\doteq$ 20.4%)
  - 4.0 years changes: $\frac{1}{N}\|\mathbf{W}_4\|^2 \doteq 0.846$ ($\doteq$ 26.4%)
  - 8.0 years averages: $\frac{1}{N}\|\mathbf{V}_4\|^2 - \overline{X}^2 \doteq 0.947$ ($\doteq$ 29.5%)
  - sample variance: $\hat{\sigma}_X^2 \doteq 3.204$

Discrete Wavelet Transform as a Decorrelator: I



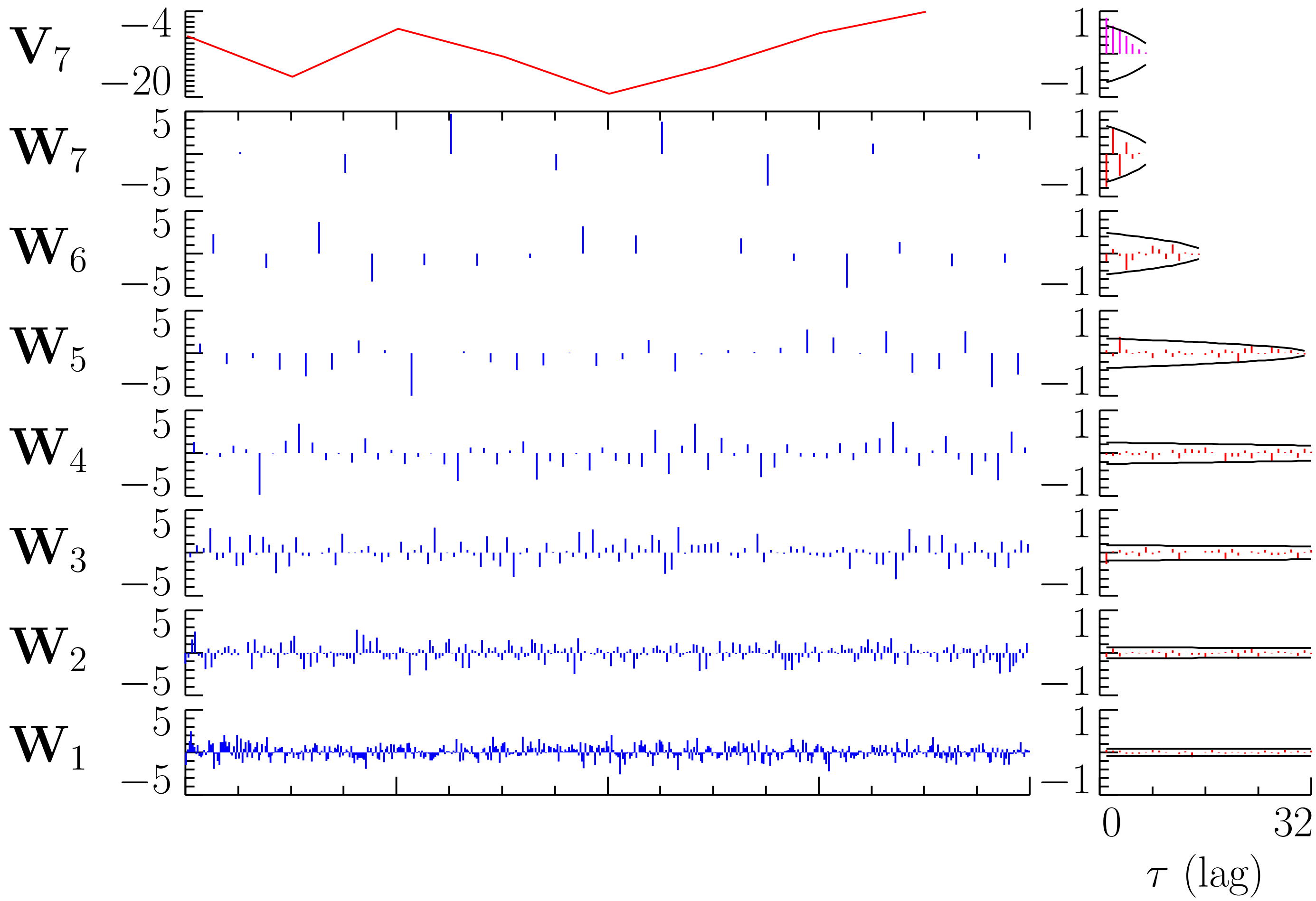
- realization of time series $\mathbf{X}$ with long-range dependence along with its sample autocorrelation sequence (ACS): for $\tau \geq 0$,

$$\hat{\rho}_{X,\tau} = \frac{\frac{1}{N}\sum_{t=0}^{N-1-\tau} X_t X_{t+\tau}}{\frac{1}{N}\sum_{t=0}^{N-1} X_t^2} = \frac{\sum_{t=0}^{N-1-\tau} X_t X_{t+\tau}}{\sum_{t=0}^{N-1} X_t^2}$$

(assumes time series has known mean or has been centered)

- note that ACS dies down slowly (typical for series with long-range dependence)

Discrete Wavelet Transform as a Decorrelator: II



- DWT of $\mathbf{X}$ and sample ACSs for its components $\mathbf{W}_j$ & $\mathbf{V}_7$, along with 95% confidence intervals for white noise

Wavelet Analysis and Wavelet-Based Modeling: I

- long-range dependence, Allan variance & wavelets (Percival & Guttorp, 1994)
  - Allan variance used to characterize frequency instability of atomic clocks (Allan, 1966)
  - Flandrin (1992) briefly noted connection between Allan variance and variance based upon Haar wavelet coefficients (Haar wavelet variance)
  - paper assessed advantages and disadvantages of Haar wavelet variance vs. wavelet variances based upon Daubechies wavelet transforms (latter can handle intrinsically stationary processes of various orders)

Wavelet Analysis and Wavelet-Based Modeling: II

– proposed use of nonorthogonal (but variance preserving) version of discrete wavelet transform (DWT) known as maximal-overlap DWT (MODWT)
– outlined efficient algorithm for computing MODWT (same order of computational complexity as FFT algorithm)
– analyzed time series of vertical shear measurements from the ocean, demonstrating value of time-dependent multiresolution analysis and fact that Allan variance leads to misleading analysis as compared to one based on Daubechies wavelet

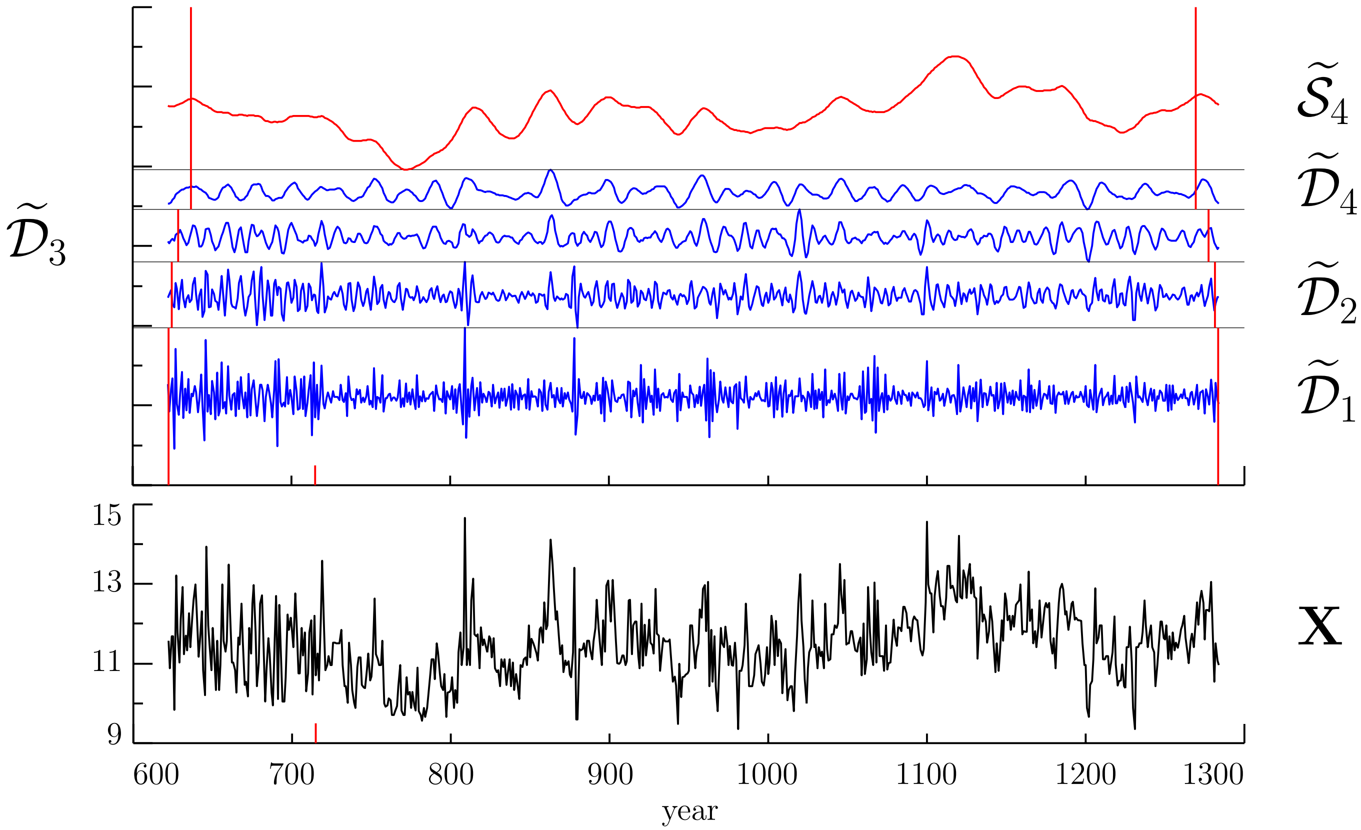
Wavelet Analysis and Wavelet-Based Modeling: III

- wavelet-based covariance analysis (Whitcher, Guttorp & Percival, 2000a)
  - introduced multiscale analysis of covariance between two time series
  - defined MODWT-based wavelet covariance and wavelet correlation as alternative to cross-spectrum analysis
  - defined wavelet cross covariance and wavelet cross correlation to investigate scale-based lead/lag relationships
  - looked at Madden–Julian Oscillation as manifested in the bivariate relationship between the Southern Oscillation Index and pressure series at Truk Island

Wavelet Analysis and Wavelet-Based Modeling: IV

– localized nature of wavelet transform allows scale-based subseries to be partitioned into seasonal periods (winter or summer) and according to state of El Niño–Southern Oscillation (ENSO)
  * found statistically significant increased correlations and increased variances in boreal winter over scales associated with periods of 16 to 128 days
  * also found reduced variance and reduced correlation during warm ENSO episodes over scales associated with periods of 8 to 512 days

Minimum Annual Water Levels $\mathbf{X}$ of Nile River



- data from ≈ 715 to 1284 recorded at Roda gauge near Cairo
- method(s) used to record data from 622 to ≈ 715 source of speculation

Wavelet Analysis and Wavelet-Based Modeling: V

- testing for homogeneity of variance in a time series exhibiting long-range dependence (Whitcher, Byers, Guttorp & Percival, 2002)
  - DWT of time series with long-range dependence yields wavelet coefficients that are approximately white noise across a given scale
  - under a Gaussian assumption, can assess null hypothesis of homogeneity of variance on a scale-by-scale basis by using a test based on cumulative sum of squares of wavelet coefficients (test designed originally for white noise)
  - if null hypothesis is rejected, can use MODWT to locate change points

Wavelet Analysis and Wavelet-Based Modeling: VI

- for Nile River, can reject the null hypothesis at two smallest scales (1 and 2 years), but not at higher scales
- MODWT-based change point detector picked out a change point at 720, which is consistent with change of of measurement method
- alternative explanation of change in long-range dependence behaviour at 715 harder to justify physically than change due to new measurement method suggested by wavelet analysis (new method decreased small scale noise)
- analysis in transform domain here allowed use of existing test for homogeneity of variance that cannot be used with untransformed data

Wavelet Analysis and Wavelet-Based Modeling: VII

- assessing existence of trend in a time series exhibiting long-range dependence (Craigmile, Guttorp & Percival, 2004)
  - no commonly accepted precise definition for trend, but to quote Kendall (1973):

    "the essential idea of trend is that it shall be smooth"
  - assume time series X_t can be modeled as $X_t = T_t + Y_t$, where
    - T_t is a nonstochastic trend component
    - Y_t is stochastic: either a stationary process with long-range dependence or an intrinsically stationary process (nonstationary, but stationary after suitable differencing)
  - trend assessment challenging: Y_t has significant low frequency components hard to distinguish from smoothly varying T_t

Wavelet Analysis and Wavelet-Based Modeling: VIII

– conservative approach to trend assessment: will not falsely declare a significant trend in X_t if in fact Y_t is reasonably capable of generating observed low frequency variations
– assuming T_t well approximated at least locally by a low order polynomial, DWT based on Daubechies wavelet filter can transform X_t into wavelet coefficients that are invariant with respect to T_t and scaling coefficients that trap T_t
– ability of DWT to cleanly separate X_t into components trapping Y_t and T_t is key to proposed methodology for
  * estimating T_t
  * testing for significance of trend
  * constructing confidence bands for unknown trend
– methodology worked well in assessing trend in a climatological time series

Wavelet Analysis and Wavelet-Based Modeling: IX

- space-time modeling of trends (Craigmile & Guttorp, 2011)
  - goal: characterize temperature trends jointly over space-time
  - approach: build wavelet-based space-time hierarchical Bayesian models to simultaneously model trend, seasonality and error, with error component accounting for long-range dependence
  - as motivation, use five decades of daily temperature series collected at 17 locations in central Sweden
  - site-by-site analysis elicited key common characteristics including seasonal dependent variability (handled by expressing log of standard deviation as a two-term harmonic model)
  - spatial structure trapped in scaling coefficients, for which a separable space-time model is entertained
  - inference for hierarchical model done in wavelet domain

Questions to Motivate Breakout Session: I

- adaptation of existing transforms (Fourier, wavelet etc.) to handle new problems arising in environmental data analysis: what are interesting directions to pursue?
- are there other transforms of interest for analysis of environmental data that have yet to be fully explored?
  - empirical mode decomposition (Huang et al, 1998)
  - synchrosqueezed wavelet transforms (billed as an empirical mode decomposition-like tool, but more amenable to mathematical analysis; Daubechies et al., 2011)
  - dynamic mode decomposition (Schmid, 2010)
  - multiresolution dynamic mode decomposition (Kutz et al., 2016)

Questions to Motivate Breakout Session: II

- ideas for entirely new transforms?
- what is the best way to deal with transforms that help satisfy distributional problems, but then mess up correlations?
- for transforms that do not preserve variance, are there ways in which we can do at least a quasi-ANOVA?

References: I

- D.W. Allan (1966), 'Statistics of atomic frequency clocks,' *Proceedings of the IEEE*, vol. 31, pp. 221–30.
- V. Berrocal, P.F. Craigmile and P. Guttorp (2012), 'Regional climate model assessment using statistical upscaling and downscaling techniques,' *Environmetrics*, vol. 23, pp. 482–92.
- D.R. Brillinger (1982), 'A generalized linear model with "Gaussian" regressor variables,' *A Festschrift for Erich L. Lehmann in Honor of His Sixty-Fifth Birthday*, edited by P.J. Bickel, K.A. Doksum and J.L. Hodges, Jr., Wadsworth International Group, Belmont, CA, pp. 971–114.
- P.F. Craigmile and P. Guttorp (2011), 'Space-time modeling of trends in temperature series,' *Journal of Time Series Analysis*, vol. 32, pp. 378–95.
- P.F. Craigmile, P. Guttorp and D.B. Percival (2004), 'Trend assessment in a long memory dependence model using the discrete wavelet transform,' *Environmetrics*, vol. 15, no. 4, pp. 313–35.
- P.F. Craigmile, P. Guttorp and D.B. Percival (2005), 'Wavelet-based parameter estimation for polynomial contaminated fractionally differenced processes,' *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3151–61.
- P.F. Craigmile, D.B. Percival and P. Guttorp (2001), 'The impact of wavelet coefficient correlations on fractionally differenced process estimation,' *European Congress of Mathematics (Barcelona, July 10-14, 2000), Volume II*, edited by C. Casacuberta, R.M. Miro-Roig, J. Verdera and S. Xambo-Descamps. Basel, Switzerland: Birkhauser Verlag, pp. 591–9.
- I. Daubechies, J. Lu, H.-T. Wu (2011), 'Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool,' *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243–61

References: II

- P. Flandrin (1992), 'Wavelet analysis and synthesis of fractional Brownian motion,' *IEEE Transactions on Information Theory*, vol. 38, pp. 910–7.
- P. Guttorp, M. Fuentes, and P.D. Sampson (2007), 'Using transforms to analyze space-time processes,' in *Statistics of Spatio-Temporal Systems*, edited by V. Isham, B. Finkelstadt and W. Härdle. Boca Raton: Taylor and Francis, pp. 77–150.
- N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung and H.H. Liu (1998), 'The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,' *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences*, vol. 454, no. 1971, pp. 903–95.
- M. Kendall, *Time Series*, Charles Griffin: 1973.
- J. N. Kutz, X. Fu, and S. L. Brunton (2016), 'Multiresolution dynamic mode decomposition,' *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 2, pp. 713–35.
- D.B. Percival and P. Guttorp (1994), 'Long-memory processes, the Allan variance and wavelets,' in *Wavelets in Geophysics*, edited by E. Foufoula-Georgiou and P. Kumar, a volume in the series *Wavelet Analysis and Its Applications* (series edited by C. Chui). New York: Academic Press, pp. 325–44.
- D.A. Rothrock, D.B. Percival and M. Wensnahan (2008), 'The decline in arctic sea-ice thickness: Separating the spatial, annual, and interannual variability in a quarter century of submarine data,' *Journal of Geophysical Research – Oceans*, vol. 113, C05003.
- P.D. Sampson and P. Guttorp (1991), 'Power transformations and tests of environmental impact as interaction effects,' *American Statistician*, vol. 45, pp. 83–9.

References: III

- P.D. Sampson and P. Guttorp (1992), 'Nonparametric estimation of nonstationary spatial covariance structure,' *Journal of the American Statistical Association*, vol. 87, pp. 108–19.
- P.J. Schmid (2010), 'Dynamic mode decomposition of numerical and experimental data,' *Journal of Fluid Mechanics*, vol. 656, no. 11, pp. 5–28
- S. Weisberg, *Applied Linear Regression* (Fourth Edition), Wiley: 2014.
- B.J Whitcher, S.D. Byers, P. Guttorp and D.B. Percival (2002), 'Testing for homogeneity of variance in time series: long memory, wavelets and the Nile River,' *Water Resources Research*, vol. 38, no. 5, 10.1029/2001WR000509.
- B.J Whitcher, P. Guttorp and D.B. Percival (2000a), 'Wavelet analysis of covariance with application to atmospheric time series,' *Journal of Geophysical Research*, vol. 105, no. D11, pp. 14,941–62.
- B.J Whitcher, P. Guttorp and D.B. Percival (2000b), 'Multiscale detection and location of multiple variance changes in the presence of long memory,' *Journal of Statistical Computation and Simulation*, vol. 68, no. 1, pp. 65–88.

Thanks to:

- Peter Craigmile et al. for organizing a stimulating workshop in an amazing location
- Peter Guttorp for being a mentor and colleague since 1980 (37 years . . . and counting!)