Recent Improvements in Water Flows Time Series Analysis

A. Ian McLeod, aimcleod@uwo.ca
University of Western Ontario

This presentation and additonal resources is available at:

http://www.stats.uwo.ca/faculty/aim/2006/CEPEL/

Cepel Road Map Workshop, May 20, 2006

Online Textbook

Time Series Modelling of Water Resources and Environmental Systems by Keith W. Hipel & A. Ian McLeod. This book previously published in 1994 by Elsevier is now freely available online:

http://www.stats.uwo.ca/faculty/aim/1994Book/default.htm

High Level Quantitative Programming Environments

Much better computing environments than Fortran now exist. These new computing environments allow for much more rapid development and are ideal for many sorts of mathematical and statistical research.

- *Mathematica*
- MatLab
- R and S-Plus

"tools for thought"

Outline of Topics

- Trend Analysis
- New time series model for riverflow, GAR and GARP
- Daily time series simulation and forecasting

Trend Analysis

- Testing for monotonic (nondecreasing or nonincreasting) trend
- Visualizing trend and seasonality using STL
- Visualizing trend and nonlinear features using Multiresolution Analysis
- Jump or changepoint detection using wavelets
- Intervention analysis for modelling environmental impacts

Testing for Monotonic Trend

- Mann-Kendall Test using block bootstrap
- R library: Kendall

Illustrative Example:
Monthly average river height of Rio Negro at Manaus, January 1903 until December 1992

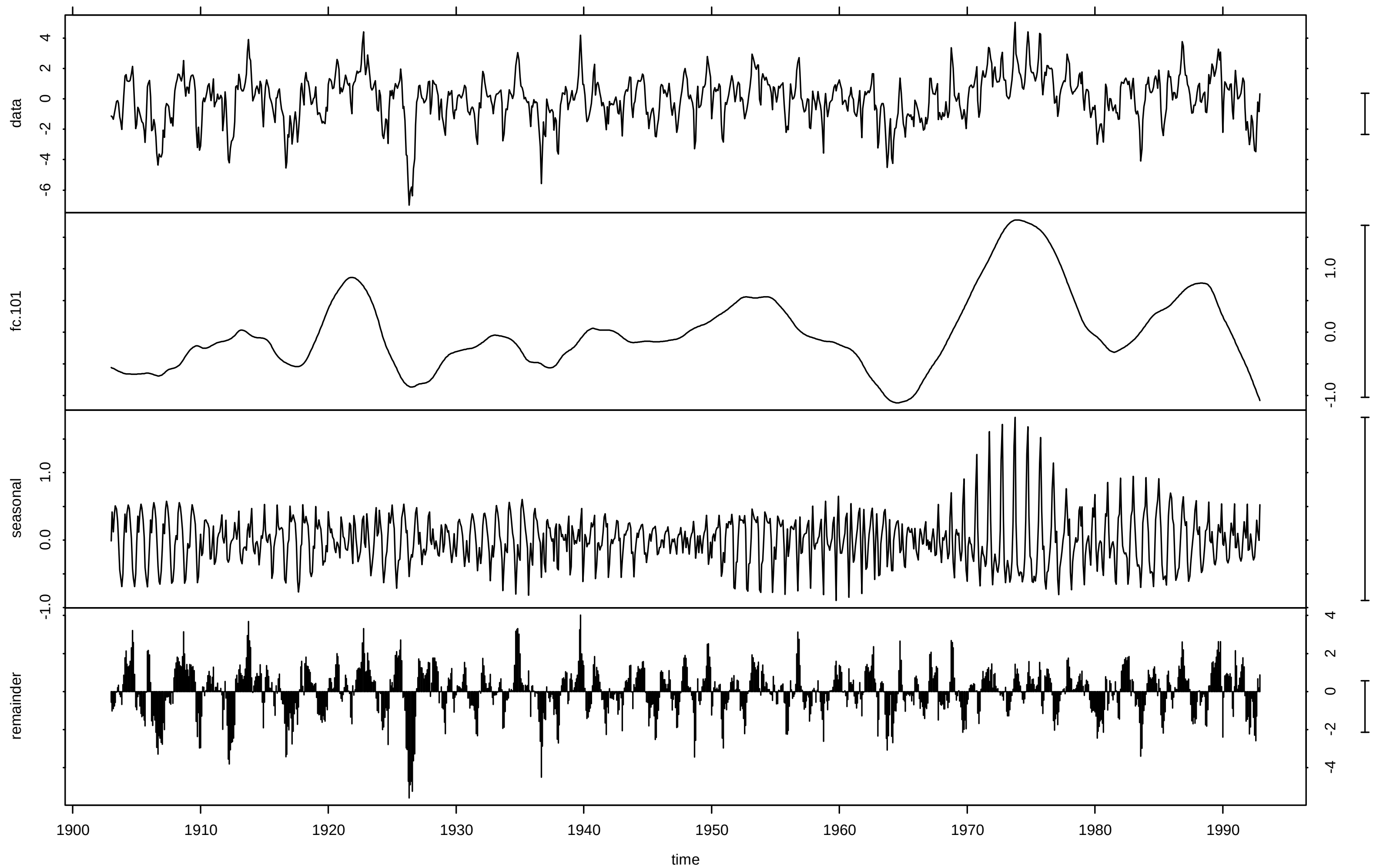
Deseasonalize by subtracting monthly means. Regular Mann-Kendall trend test P-value = 10^{-5} but using block bootstrap to allow for autocorrelation we obtain a revised P-value = 4.8%.

This is in agreement with the result obtained by Brillinger (1989, Biometrika) of 4.1% based on the daily averages.

Seasonal Trend Loess

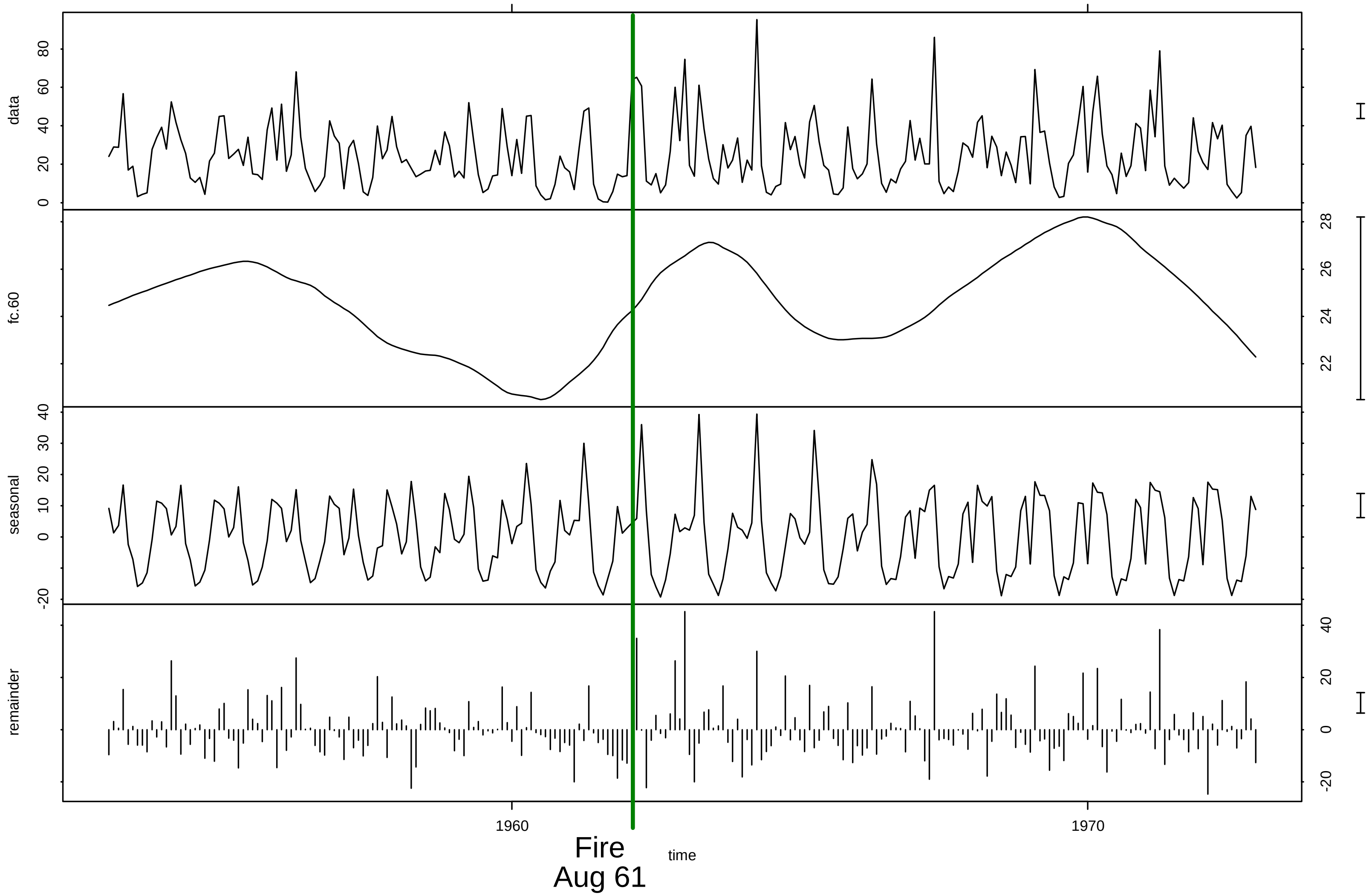
- Cleveland, W.S. (1994) Visualizing Data. Hobart Press.
- R. B. Cleveland, W. S. Cleveland, J.E. McRae, and I. Terpenning (1990) STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, **6**, 3–73.
- Availability: R and S-Plus

Rio Negro River Height at Manaus



STL Analysis

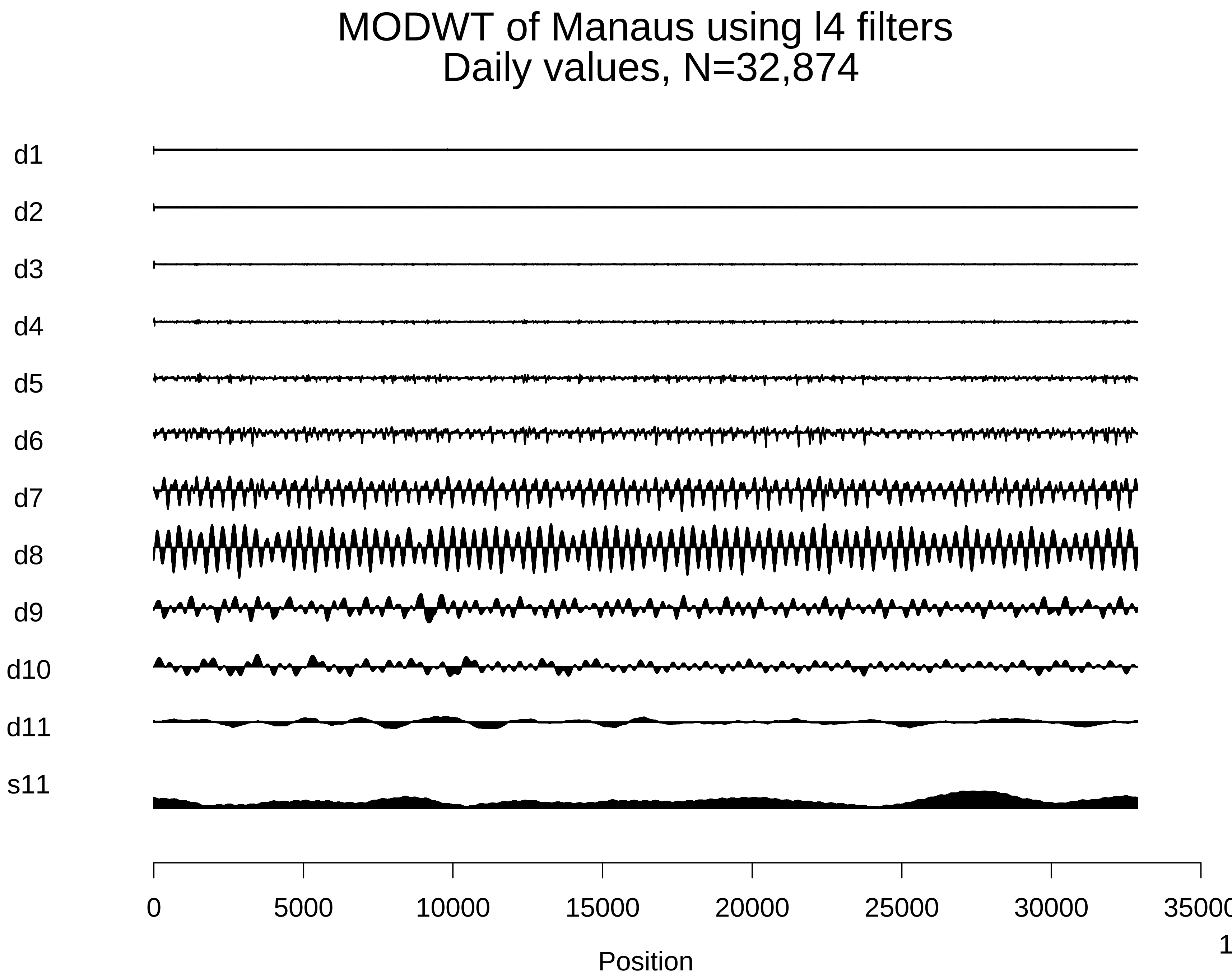
Piper's Hole River, 1953 to 1981

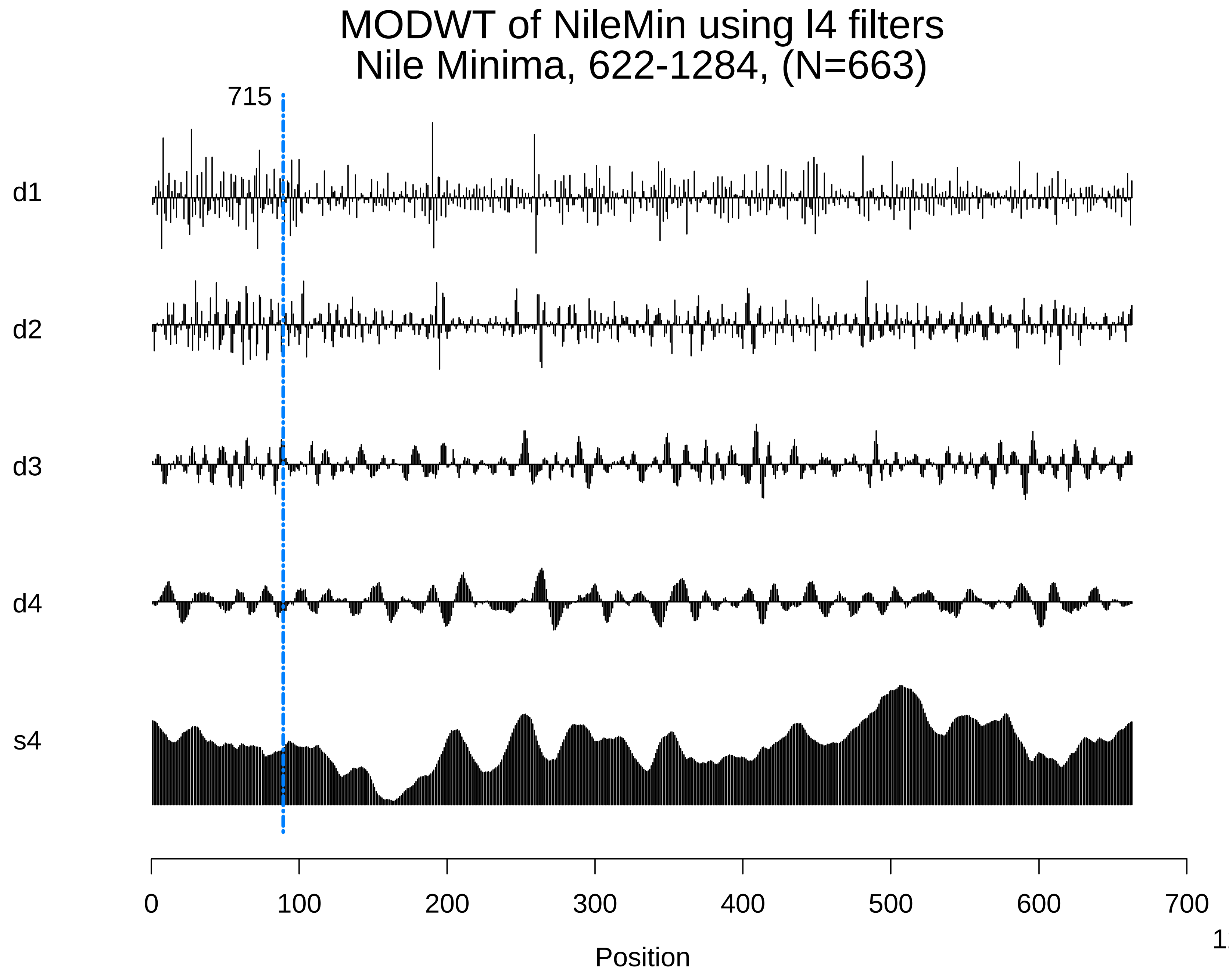


Multiresolution Analysis (MRA)

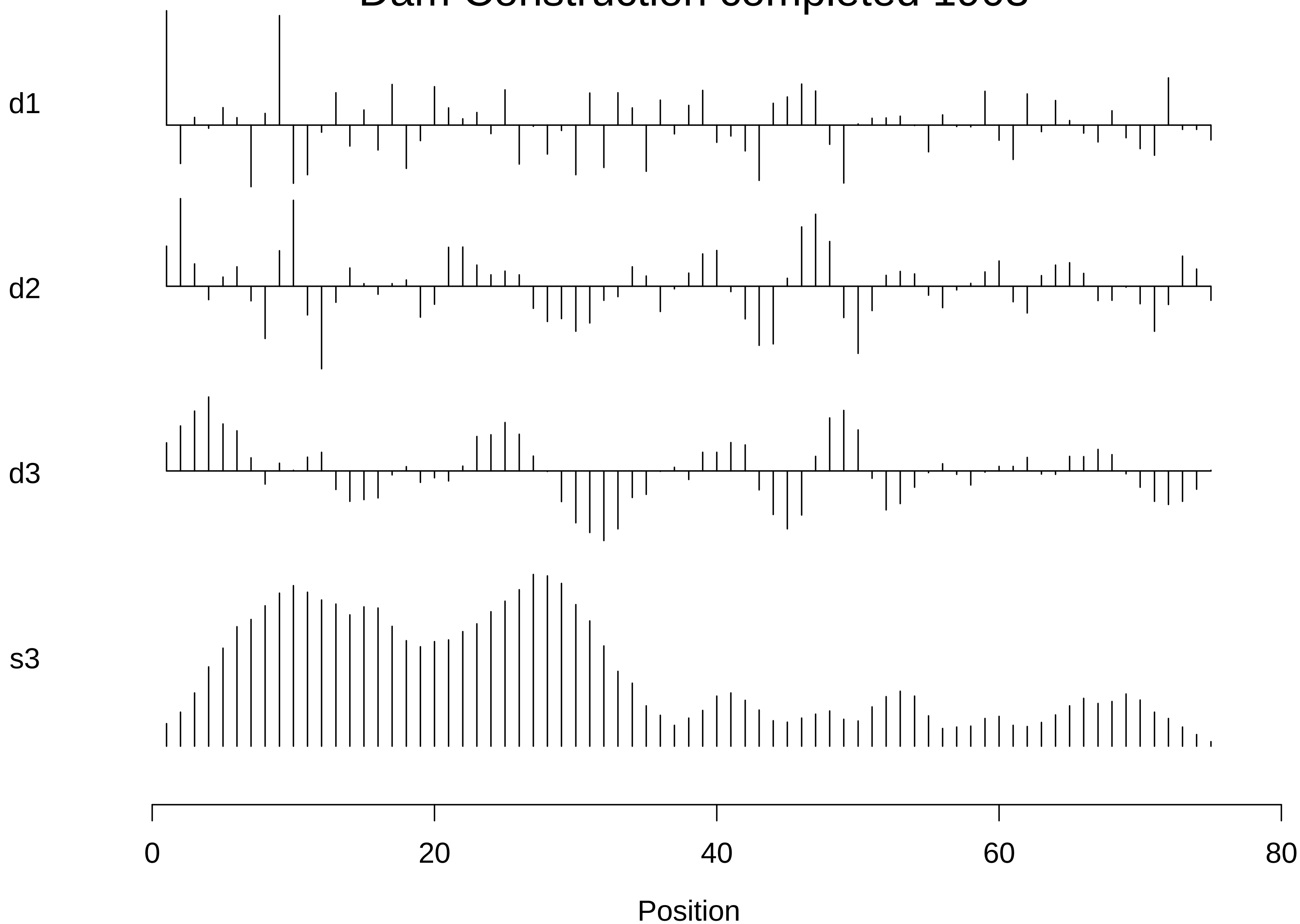
- decomposition of a time series into Details, D_j, $j=1,\ldots,J$ and Smooth S_J
- D_j indicates changes at scale 2^{j-1}
- S_J indicates mean level at scale 2^{J-1}

Percival, D.B. & Walden, A.T. (2000). *Wavelet Methods for Time Series Analysis.* Cambridge University Press.

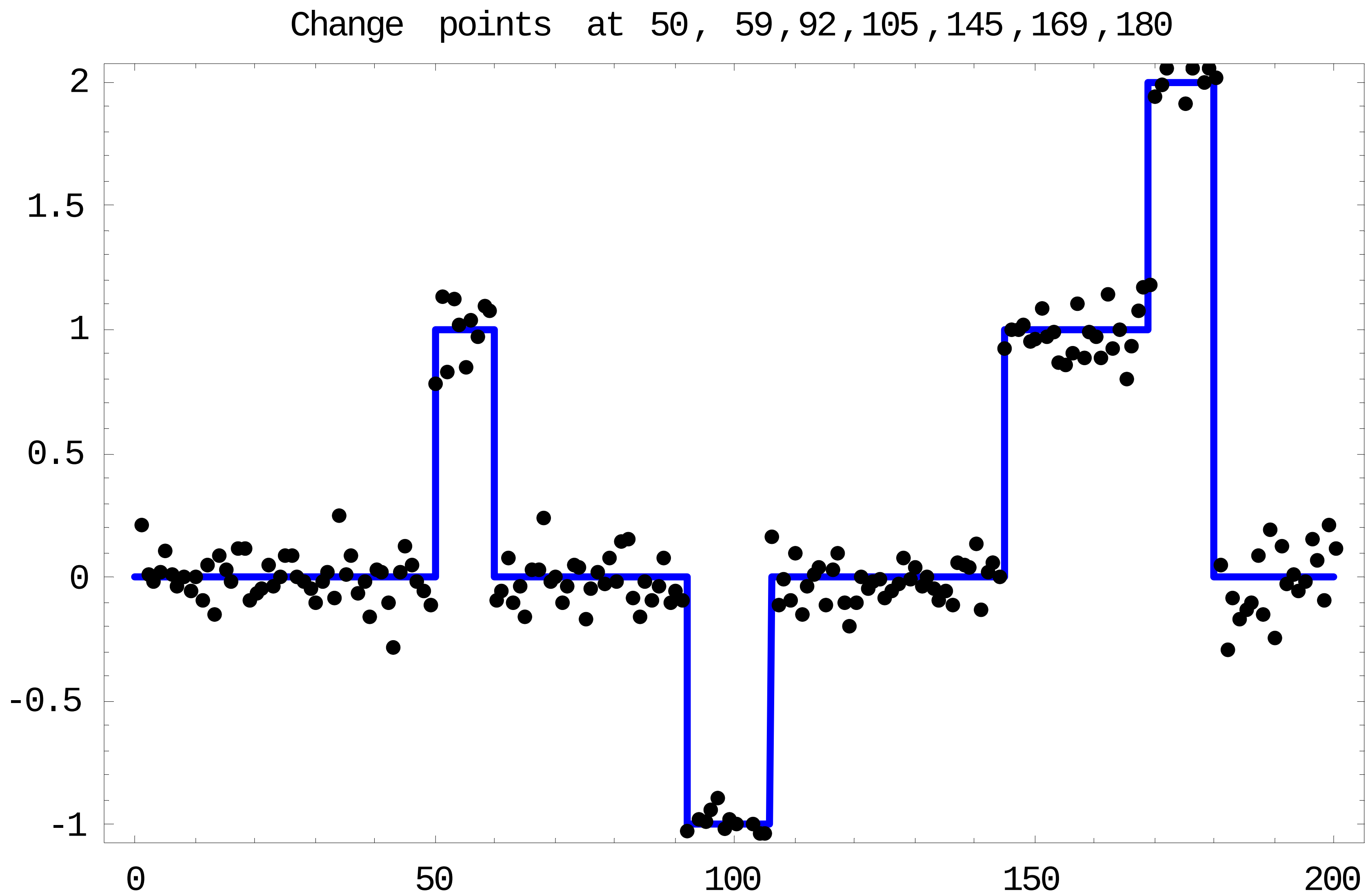

MODWT of Manaus using l4 filters
Daily values, N=32,874
d1
d2
d3
d4
d5
d6
d7
d8
d9
d10
d11
s11
0
5000
10000
15000
20000
25000
30000
35000
Position


MODWT of NileMin using l4 filters
Nile Minima, 622-1284, (N=663)
715
d1
d2
d3
d4
s4
0
100
200
300
400
500
600
700
Position

MODWT of Nile using Haar filters
Annual Nile Flows at Aswan, 1870-1945
Dam Construction completed 1903



Changepoint Detection via Wavelets



Partial DWT

The partial DWT is a special orthonormal transformation,

$$(z_0,\ \cdots,\ z_{N-1}) \longleftrightarrow (W_1,\ \cdots,\ W_J,\ V_J),$$

W_j is a vector of length $N_j = N / 2^j$

$W = \mathcal{W}\, X$

$$\mathcal{W} = \begin{pmatrix} \mathcal{W}_1 \\ \cdots \\ \mathcal{W}_J \\ \mathcal{V}_J \end{pmatrix}$$

$\mathcal{W}_j$ is $N_j \times N$, $j = 1,\ \cdots,\ J$

$\mathcal{V}_J$ is $N_J \times N$

Daubechies D4 Wavelet Filter

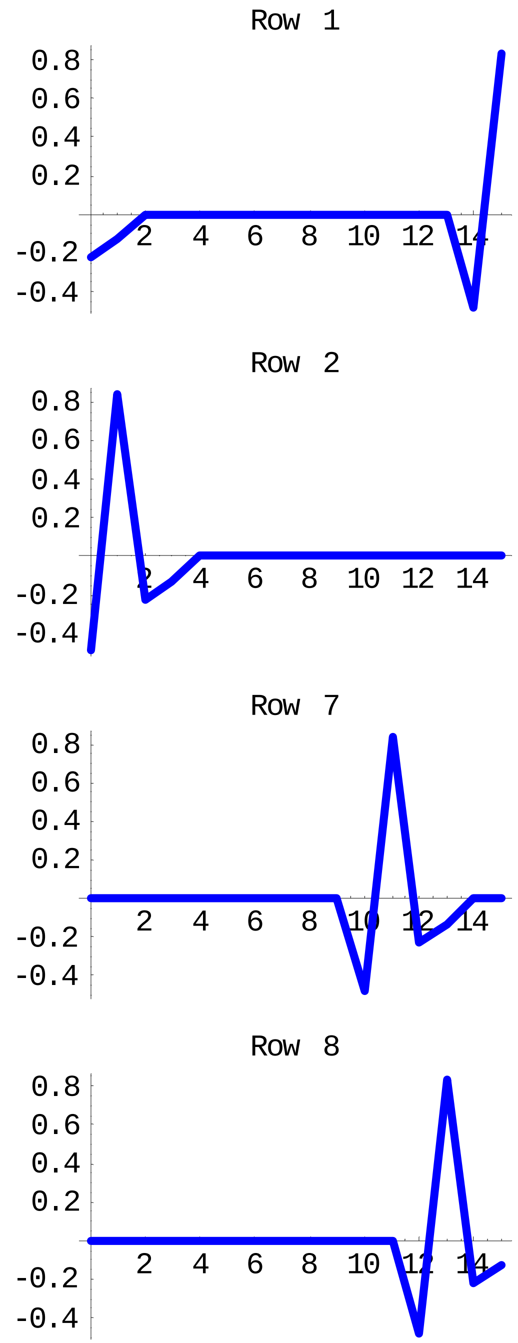
$$(a_0,\, a_1,\, a_2,\; a_3) = (\,-0.12941 \quad -0.224144 \quad 0.836516 \quad -0.482963\,)$$

$$\mathcal{W}_1 = \begin{pmatrix}
a_1 & a_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_3 & a_2 \\
a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_3 & a_2 & a_1 & a_0
\end{pmatrix}$$

N=16

The rows of $\mathcal{W}_1$ behave like first differences

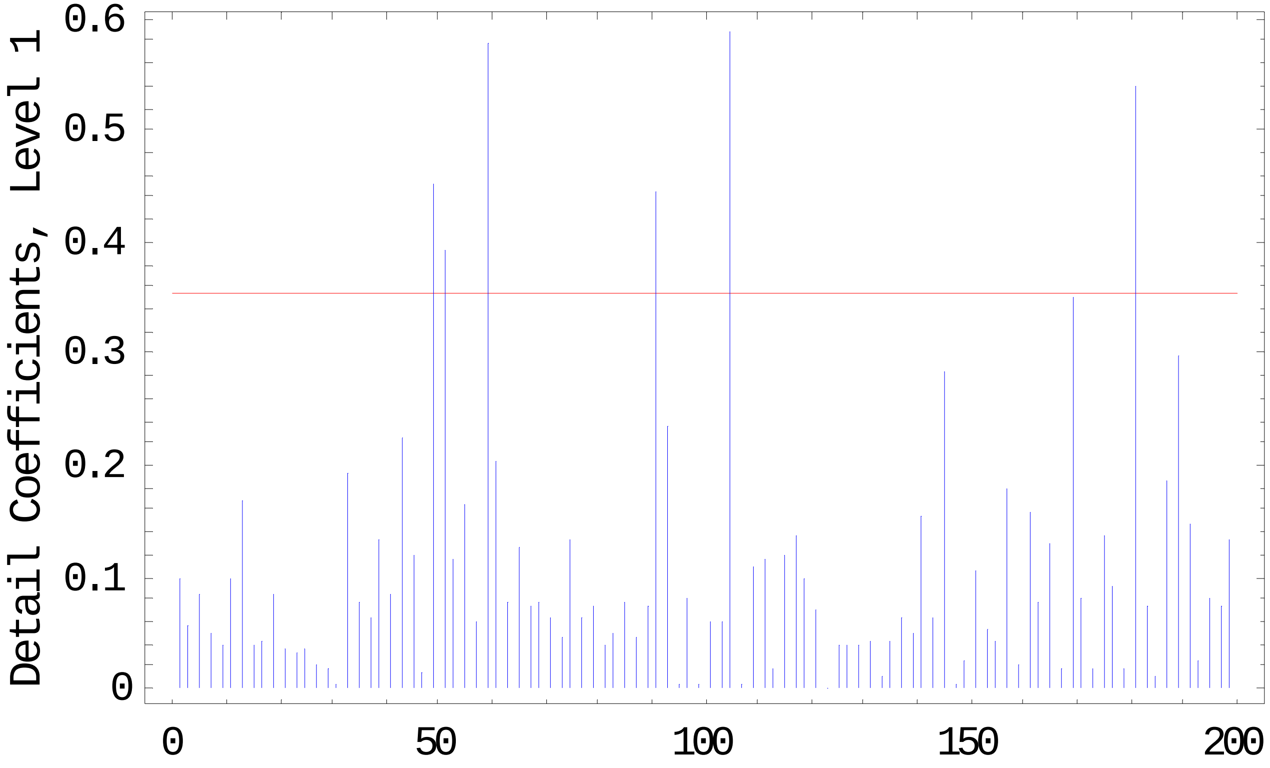
(except for Row 1, due to boundary effect)



Wang, Yazhen (1995). Jump and Sharp Cusp Detection by Wavelets. *Biometrika*, 82, 385-397.

$$\text{threshold:} \quad \sigma\sqrt{2\log(n)}$$

Change points at 50, 59,92,105,145,169,180



MRA (Derivation)

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

$\mathcal{W}_1, \; \ldots, \; \mathcal{W}_J$ and $\mathcal{V}_J$ also orthonormal

$$X = \sum_{j=1}^{J} \mathcal{W}_j^T W_j + \mathcal{V}_J^T V_J$$

$$D_j = \mathcal{W}_j^T W_j \qquad\qquad S_J = \mathcal{V}_{J_0}^T V_J$$

Intervention Analysis (IA)

$$z_t = \mu + \omega\, B^b \;\; P_t(T) + u_t\,,$$

$t = 1,\, 2,\;\; \ldots\,,\;\, n,$

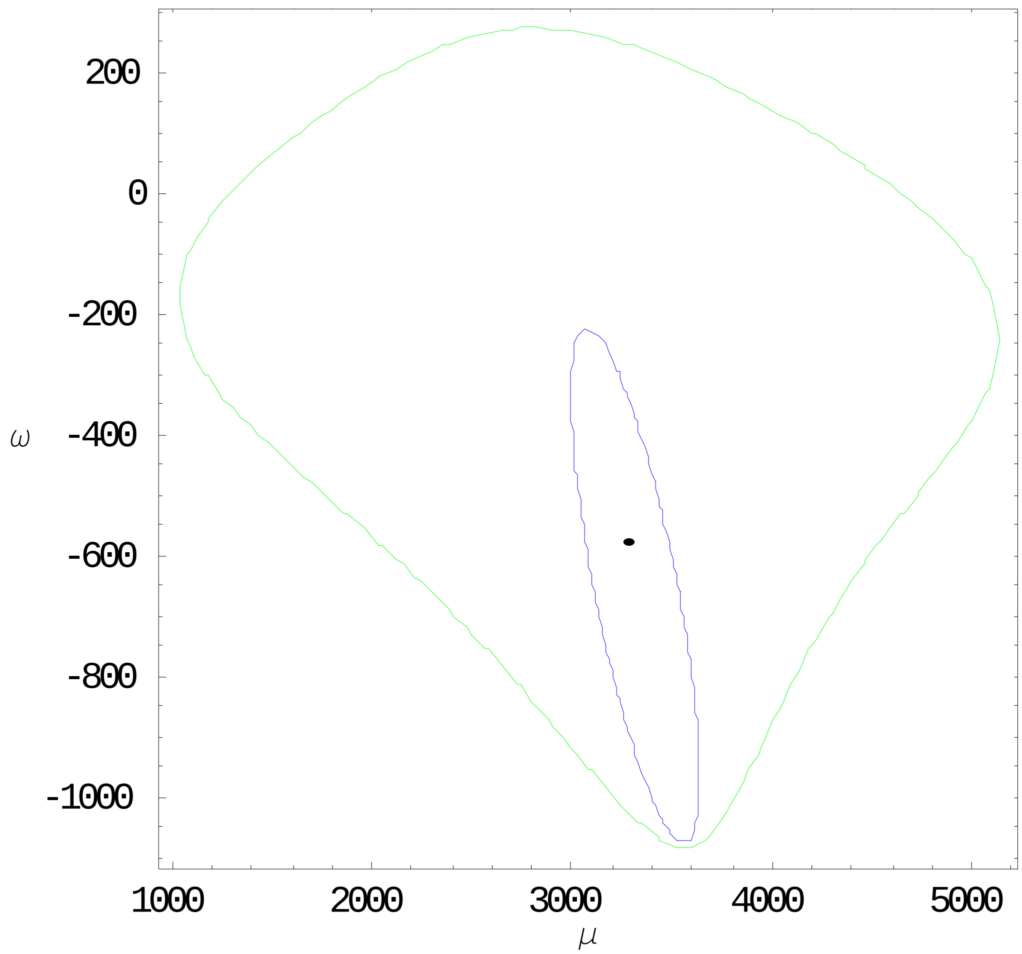
B backshift operator

$$P_t\,(T) = \begin{cases} 0 & t \le T \\ 1 & t > T \end{cases}$$

New Developments in IA

a) Power Computation (reprint available)
b) S-Plus library **`tfm`** for IA (forthcoming)
c) A fundamental limitation to statistical inference (working paper available)

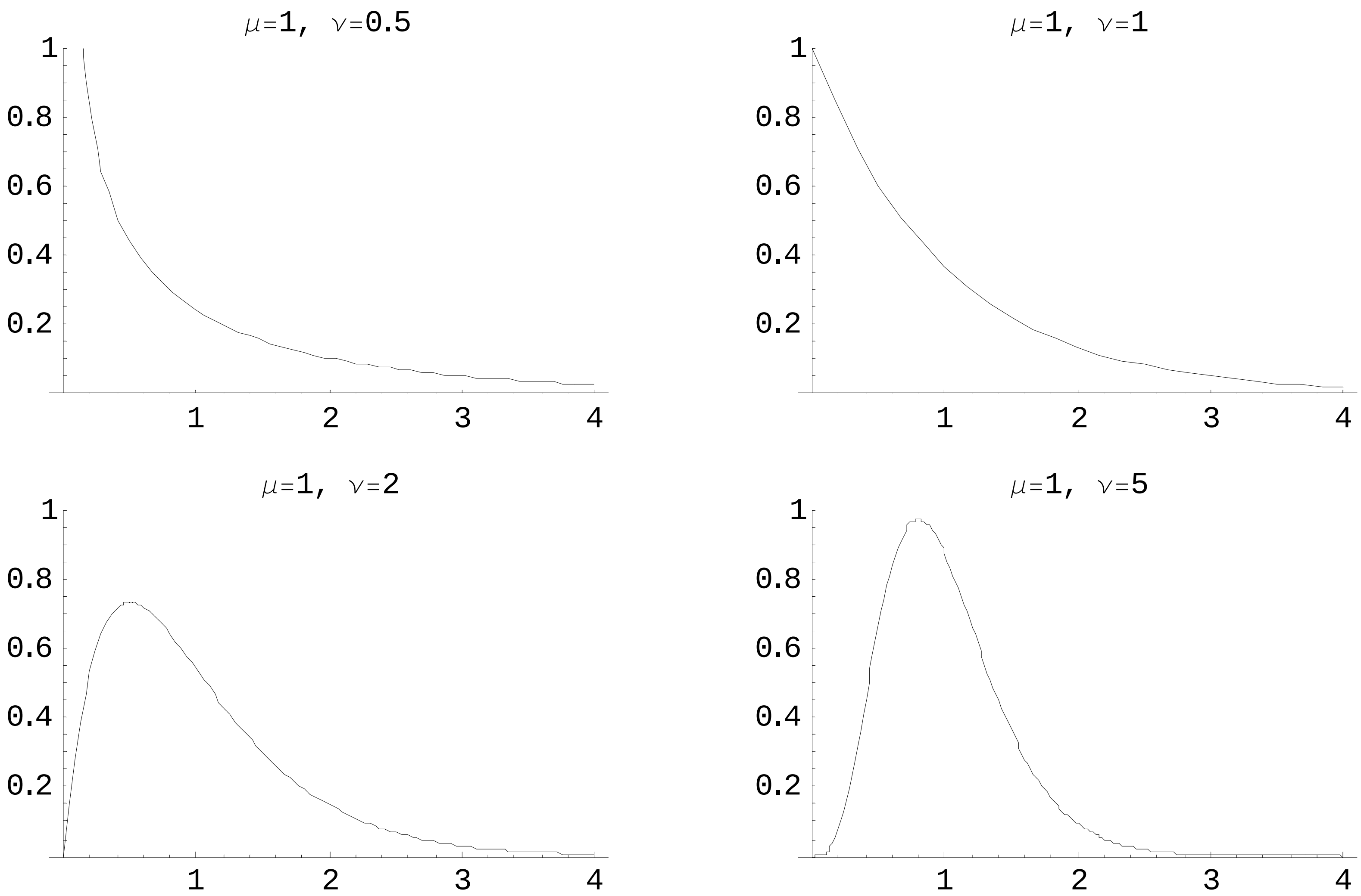
Illustration of the lack of robustness of the 95% confidence intervals due to different models for autocorrelated error



Gamma Distribution for Riverflow Modelling

- Normal distribution can lead to negative flow values in simulations
- Using log transformation produces a model not in the original domain
- Gamma distribution always >0 and may fit better than normal or lognormal
- May be fit using glm software
- May be extended to multisite and seasonal

Shapes of the gamma distribution



GAR(p) Model

$$\frac{1}{\Gamma(\nu)}\left(\frac{\nu\, y}{\mu}\right)^{\nu}\exp\left(-\frac{\nu\, y}{\mu}\right)d(\log y)$$

The distribution is denoted by $G(\mu, \vee)$.

$$z_t \sim G(\xi + \phi_1 z_{t-1} + \cdots + z_{t-p},\ \nu)$$

Two New Methods for Daily Riverflow Forecasting & Simulation

- Feedforward neural nets provide a powerful method for forecasting nonlinear time series. Reprint of paper on webpage.
- Wavelets. Since the lower level detail coefficients behave like white noise we can simply simulate them and use the other coefficients to synthesize the time series.