Detecting changes in second order structure within oceanographic time series

Rebecca Killick

Joint work with Idris Eckley and Phil Jonathan

Lancaster University

March 26, 2012

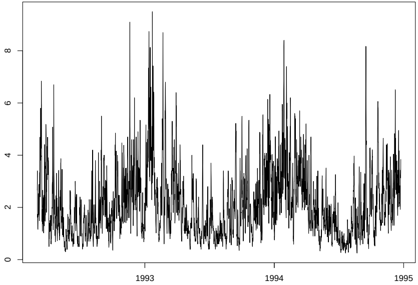
Summary

- Motivation
- Changepoint recap
- Locally Stationary Wavelet (LSW) model
- Wavelet Likelihood Method for changes in autocovariance
- Simulation Study
- Oceanographic Example

Motivation

Wave Heights

- Interested in detecting start and end of storm season
- Unknown dependence structure changing over time
- Cyclic mean

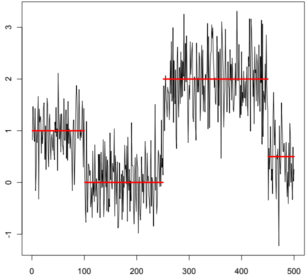


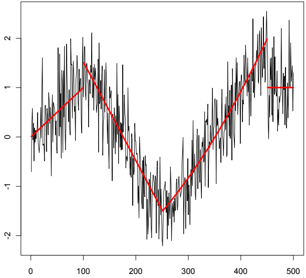
Changepoint recap

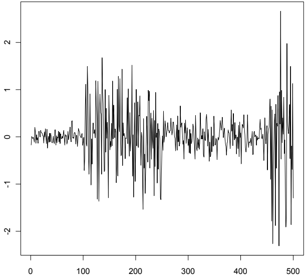
Changepoints

For data $y_1, \ldots, y_n$, if a changepoint exists at τ, then $y_1, \ldots, y_\tau$ differ from $y_{\tau+1}, \ldots, y_n$ in some way.

There are many different types of changes.







Change in second order structure

The traditional hypothesis, for some lag, v,

$$\begin{aligned}
\mathbf{H_0} &: \operatorname{cov}(X_0, X_{0-v}) = \operatorname{cov}(X_1, X_{1-v}) = \ldots = \operatorname{cov}(X_{n-1}, X_{n-1-v}) = \rho_{0,v} \\
\mathbf{H_1} &: \rho_{1,v} = \operatorname{cov}(X_0, X_{0-v}) = \ldots = \operatorname{cov}(X_\tau, X_{\tau-v}) \\
&\qquad \neq \operatorname{cov}(X_{\tau+1}, X_{\tau+1-v}) = \ldots = \operatorname{cov}(X_{n-1}, X_{n-1-v}) = \rho_{n,v}.
\end{aligned}$$

Change in second order structure

The traditional hypothesis, for some lag, v,

$$
\begin{aligned}
\mathbf{H_0} &: \operatorname{cov}(X_0, X_{0-v}) = \operatorname{cov}(X_1, X_{1-v}) = \ldots = \operatorname{cov}(X_{n-1}, X_{n-1-v}) = \rho_{0,v} \\
\mathbf{H_1} &: \rho_{1,v} = \operatorname{cov}(X_0, X_{0-v}) = \ldots = \operatorname{cov}(X_\tau, X_{\tau-v}) \\
&\qquad \neq \operatorname{cov}(X_{\tau+1}, X_{\tau+1-v}) = \ldots = \operatorname{cov}(X_{n-1}, X_{n-1-v}) = \rho_{n,v}.
\end{aligned}
$$

Likelihood-Based test statistic

$$
\lambda = \max_\tau \frac{L(y_{1:n}, \tau | \rho_{1,v}, \rho_{n,v}, \theta)}{L(y_{1:n} | \rho_{0,v}, \theta)}.
$$

Methods exist that make structural assumptions about the covariance to get $L(\cdot)$, e.g. piecewise AR.

Locally Stationary Wavelet Model

The Model

Locally Stationary Wavelet Model

$$X_{t,n} = \sum_j \sum_k W_j\left(\frac{k}{n}\right)\psi_{j,k-t}\xi_{jk}.$$

Relevant assumptions:

- $\mathbb{E}\xi_{jk} = 0$, $\mathrm{cov}(\xi_{jk}, \xi_{lm}) = \delta_{jl}\delta_{km}$
- We assume that the $\xi_{jk} \sim \mathrm{Normal}$
- Hence $\mathbb{E}X_t = 0$, i.e. remove any mean or trend before analysis
- Bounded total variation condition on the $W_j^2(\cdot)$.

The Model

Locally Stationary Wavelet Model

$$X_{t,n} = \sum_j \sum_k W_j\left(\frac{k}{n}\right)\psi_{j,k-t}\xi_{jk}.$$

Reasons for using LSW:

- Models the local structure as it changes over time
- Has a more general covariance structure than other time series models
- Piecewise constant covariance structure $\implies$ p.c. spectrum

Wavelet Likelihood Method

The Hypothesis

The traditional hypothesis, for some lag, v,

$$\mathbf{H_0}: \operatorname{cov}(X_0, X_{0-v}) = \operatorname{cov}(X_1, X_{1-v}) = \ldots = \operatorname{cov}(X_{n-1}, X_{n-1-v}) = \rho_{0,v}$$
$$\mathbf{H_1}: \rho_{1,v} = \operatorname{cov}(X_0, X_{0-v}) = \ldots = \operatorname{cov}(X_\tau, X_{\tau-v})$$
$$\neq \operatorname{cov}(X_{\tau+1}, X_{\tau+1-v}) = \ldots = \operatorname{cov}(X_{n-1}, X_{n-1-v}) = \rho_{n,v}.$$

is equivalent to

$$\mathbf{H_0}: W_j^2\left(\tfrac{0}{n}\right) = W_j^2\left(\tfrac{1}{n}\right) = \ldots = W_j^2\left(\tfrac{n-1}{n}\right) = \gamma_{0,j} \qquad \forall j$$
$$\mathbf{H_1}: \gamma_{1,j} = W_j^2\left(\tfrac{0}{n}\right) = \ldots = W_j^2\left(\tfrac{\tau}{n}\right) \neq W_j^2\left(\tfrac{\tau+1}{n}\right) = \ldots = W_j^2\left(\tfrac{n-1}{n}\right) = \gamma_{n,j},$$

for some $j \in \{1, 2, \ldots, J = \log_2 n\}$.

The Likelihood

$$X_{t,n} = \sum_j \sum_k W_j\left(\tfrac{k}{n}\right)\psi_{j,k-t}\xi_{jk}.$$

If ξ are Gaussian then,

$$\ell(W|\mathbf{x}) = \tfrac{n}{2}\log 2\pi - \tfrac{1}{2}\log|\Sigma_W| - \tfrac{1}{2}\mathbf{x}'\Sigma_W^{-1}\mathbf{x},$$

where the variance-covariance matrix, Σ_W, has the following form:

$$\Sigma_W(k,k') = \operatorname{cov}(X_k, X_{k'}) = \sum_l \sum_m W_l^2\left(\tfrac{m}{n}\right)\psi_{l,m-k}\psi_{l,m-k'}.$$

Changepoint test statistic

$$\lambda_\tau = \max_{J<\tau<n-J} \left\{ \log \left|\hat{\Sigma}_0\right| + \mathbf{x}'\hat{\Sigma}_0^{-1}\mathbf{x} - \log \left|\hat{\Sigma}_1\right| - \mathbf{x}'\hat{\Sigma}_1^{-1}\mathbf{x} \right\}.$$

Here $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$ have elements,

$$\hat{\Sigma}_0(k,k') = \sum_l \sum_m \hat{\gamma}_{0,l}\psi_{l,m-k}\psi_{l,m-k'},$$

$$\hat{\Sigma}_1(k,k') = \sum_l \sum_{m\leq\tau} \hat{\gamma}_{1,l}\psi_{l,m-k}\psi_{l,m-k'} + \sum_{m>\tau} \hat{\gamma}_{n,l}\psi_{l,m-k}\psi_{l,m-k'},$$

Simulation Study

We compare with

- AutoPARM (AP) - Davis, Lee, Rodriguez, Yam (2006)
  - Piecewise AR models
  - Likelihood ratio test statistic
  - Minimum Description Length Penalty
- CF - Cho, Fryzlewicz (2011)
  - LSW model with Gaussian innovations
  - Designed to stabilize variance prior to changepoint estimation
  - Non-parametric test statistic

Simulations

Important measures,

- Number of changepoints identified
- Location of changepoints

Simulation Results

Stationary AR Model

$$X_t = aX_{t-1} + \epsilon_t \qquad \text{for } 1 \leq t \leq 1024.$$

a	-0.7			-0.1			0.4			0.7		
no. cpts	WL	CF	AP	WL	CF	AP	WL	CF	AP	WL	CF	AP
0	**100**	**71**	**100**	**100**	**89**	**100**	**100**	**94**	**100**	**91**	**92**	**100**
1	0	24	0	0	11	0	0	5	0	9	7	0
≥ 2	0	5	0	0	0	0	0	1	0	0	1	0

Piecewise AR Models

$$\text{B} \qquad X_t = \begin{cases} 0.9X_{t-1} + \epsilon_t & \text{if } 1 \leq t \leq 512, \\ 1.68X_{t-1} - 0.81X_{t-2} + \epsilon_t & \text{if } 513 \leq t \leq 768, \\ 1.32X_{t-1} - 0.81X_{t-2} + \epsilon_t & \text{if } 769 \leq t \leq 1024, \end{cases}$$

$$\text{C} \qquad X_t = \begin{cases} 0.4X_{t-1} + \epsilon_t & \text{if } 1 \leq t \leq 400, \\ -0.6X_{t-1} + \epsilon_t & \text{if } 401 \leq t \leq 612, \\ 0.5X_{t-1} + \epsilon_t & \text{if } 613 \leq t \leq 1024, \end{cases}$$

$$\text{D} \qquad X_t = \begin{cases} 0.75X_{t-1} + \epsilon_t & \text{if } 1 \leq t \leq 50, \\ -0.5X_{t-1} + \epsilon_t & \text{if } 51 \leq t \leq 1024. \end{cases}$$

no. of	Model B			Model C			Model D		
cpts	WL	CF	AP	WL	CF	AP	WL	CF	AP
0	0	0	0	0	0	0	4	2	0
1	0	0	0	0	0	0	**94**	**83**	**100**
2	**98**	**70**	**94**	**94**	**76**	**100**	2	15	0
3	2	27	6	6	22	0	0	0	0
≥ 4	0	3	0	1	2	0	0	0	0

Simulation Results

Models

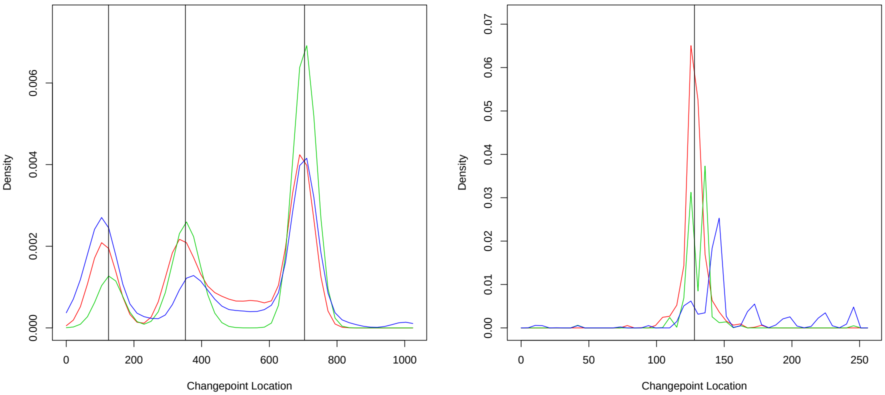
$$\text{E} \qquad X_t = \begin{cases} 1.399X_{t-1} - 0.4X_{t-2} + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, 0.8^2) & \text{if } 1 \leq t \leq 400, \\ 0.999X_{t-1} + \epsilon_t, & \epsilon_t \sim \mathcal{N}(0, 1.2^2) & \text{if } 401 \leq t \leq 750, \\ 0.699X_{t-1} + 0.3X_{t-2} + \epsilon_t & \epsilon_t \sim \mathcal{N}(0, 1) & \text{if } 751 \leq t \leq 1024. \end{cases}$$

$$\text{F} \qquad X_t = \begin{cases} 0.7X_{t-1} + \epsilon_t + 0.6\epsilon_{t-1} & \text{if } 1 \leq t \leq 125, \\ 0.3X_{t-1} + \epsilon_t + 0.3\epsilon_{t-1} & \text{if } 126 \leq t \leq 352, \\ 0.9X_{t-1} + \epsilon_t & \text{if } 353 \leq t \leq 704, \\ 0.1X_{t-1} + \epsilon_t - 0.5\epsilon_{t-1} & \text{if } 705 \leq t \leq 1024. \end{cases}$$

$$\text{G} \qquad X_t = \begin{cases} \epsilon_t + 0.8\epsilon_{t-1} & \text{if } 1 \leq t \leq 128, \\ \epsilon_t + 1.68\epsilon_{t-1} - 0.81\epsilon_{t-2} & \text{if } 129 \leq t \leq 256, \end{cases}$$

no. of cpts	Model E WL	Model E CF	Model E AP	Model F WL	Model F CF	Model F AP	Model G WL	Model G CF	Model G AP
0	0	0	0	0	0	0	0	0	0
1	26	9	9	20	12	51	**99**	**85**	**100**
2	**45**	**75**	**33**	22	36	33	1	15	0
3	26	15	31	**35**	**45**	**16**	0	0	0
≥ 4	3	1	27	23	7	0	0	0	0

Simulation Results



Model F

Model G

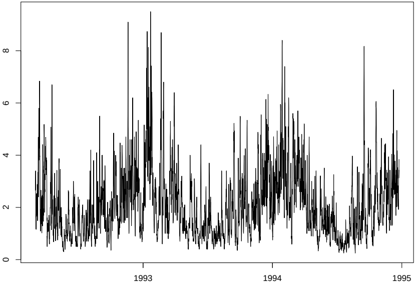
Red- WL

Green - AP

Blue - CF

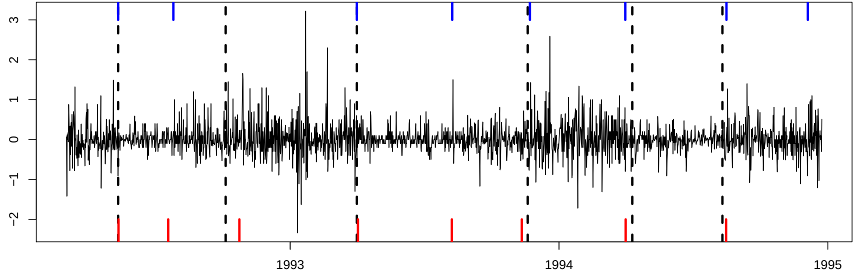
Oceanographic Example

Application to North Sea Wave Heights



- Significant Wave Heights at a Central North Sea location
- March 1992 - December 1994
- First difference to remove the mean

Application to North Sea Wave Heights



Black - WL Red - AP Blue - CF

Ocean Engineers when presented with the results preferred the WL estimates.

Summary

Summary

- Demonstrated why changes in second order structure are important
- Developed a method that:
  - has less restrictive assumptions than existing likelihood-based methods
  - performs on par with existing likelihood-based methods for AR models
  - is useful in practice when the covariance structure is unknown

References

- Cho, H. and Fryzlewicz, P. (2011) *Multiscale and multilevel technique for consistent segmentation of nonstationary time series* Statistica Sinica, 22:207229
- Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006) *Structural break estimation for nonstationary time series models* JASA, 101:223239.
- Killick, R., Eckley, I. A., Jonathan, P. (2012) *Detecting changes in second order structure within oceanographic time series* In Submission.