The Uncertainty of Storm Season Changes: Quantifying the Uncertainty of Autocovariance Changepoints

Supplementary Material

Christopher F. H. Nam[1], John A. D. Aston[1,2], Idris A. Eckley[3], Rebecca Killick[3]

[1]Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

[2]Statistical Laboratory, DPMMS, University of Cambridge, Cambridge, CB3 0WB, UK

[3]Department of Mathematics and Statistics, University of Lancaster, Lancaster, LA1 4YF, UK

{c.f.h.nam | j.a.d.aston }@warwick.ac.uk

{i.eckley | r.killick }@lancaster.ac.uk

February 13, 2014

1 Joint density of $\mathbf{I}_k = (I_{1,k}, I_{2,k}, \ldots, I_{J,k})$

The following section considers the generalised version of computing the density of a transformed random vector. X and Y denote standard random vectors here with no connections to the HMM or wavelet setup. This material is from Grimmett and Stirzaker (2001). As $(Y_1 = X_1^2, Y_2 = X_2^2) = T(X_1, X_2)$ is a many-to-one mapping, direct application of a standard change of variable via the Jacobian argument (Grimmett and Stirzaker, 2001, p. 109) is not permissible. In the one-dimensional case, the following proposition is proposed.

Proposition 1.1. *Let $I_1, I_2, \ldots, I_n$ be intervals which partition $\mathbb{R}^2$, and suppose that $Y = g(x)$ where g is strictly monotone and continuously differentiable on every I_i. For each i, the function $g : I_i \to \mathbb{R}$ is invertible on $g(I_i)$ with the inverse function h_i. Then*

$$f_Y(y) = \sum_{i=1}^{n} f_X(h_i(y))|h_i'(y)|,$$

with the convention that the ith summand is 0 if h_i is not defined at y, and $h_i'(\cdot)$ is the first derivative of $h_i(\cdot)$.

Proof. See page 112 of Grimmett and Stirzaker (2001). □

Therefore,

Proposition 1.2. *For* $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) = (X_1^2, X_2^2, \ldots, X_n^2)$

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}}(y_1, \ldots, y_n) = \frac{1}{2^n \prod_{i=1}^n |x_i|} \sum_{a_1,\ldots,a_n \in \{+,-\}} f_{\mathbf{X}}\left(a_1|x_1|, \ldots, a_n|x_n|\right).$$

Proof. Applications of Propositions 1.1 and Corollary 4 (Grimmett and Stirzaker, 2001, p. 109).

□

2 Computing Σ_k^D, the covariance structure of $\mathbf{D}_k$

This section outlines how the covariance structure of $\mathbf{D}_k = (D_{1,j} \ldots, D_{J,k})$, can be computed from the Evolutionary Wavelet Spectrum $W_j^2(\frac{k}{n})$.

Proposition 2.1. *The autocovariance structure for the observation process,* Y_t*, can be characterised by the Evolutionary Wavelet Spectrum as follows:*

$$\mathrm{Cov}(Y_t, Y_{t-v}) = \sum_l \sum_m W_l^2 \left(\frac{m}{n}\right) \psi_{l,m-t} \psi_{l,m-t+v}.$$

Proof. See proof of Proposition 1 in Nason et al. (2000). □

Proof of Proposition 2 in paper. As LSW processes are assumed to have mean zero, $\mathbb{E}[Y_t] = 0$, then it follows that the wavelet coefficients are mean zero themselves since they can be seen as

a linear combination of Gaussian observations. Thus $\mathbb{E}[D_{j,k}] = \mathbb{E}[D_{j',k'}] = 0$. Then

$$\begin{aligned}
\mathrm{Cov}(D_{j,k}, D_{j',k'}) &= \mathbb{E}[D_{j,k}D_{j',k'}] - \mathbb{E}[D_{j,k}]\mathbb{E}[D_{j',k'}] = \mathbb{E}[D_{j,k}D_{j',k'}] \\
&= \mathbb{E}\left[\left(\sum_t Y_t \psi_{j,k-t}\right)\left(\sum_s Y_s \psi_{j',k'-s}\right)\right] \\
&= \mathbb{E}\left[\sum_t \left(\sum_l \sum_m W_l\left(\frac{m}{n}\right)\psi_{l,m-t}\xi_{l,m}\right)\psi_{j,k-t} \sum_s \left(\sum_p \sum_q W_p\left(\frac{q}{n}\right)\psi_{p,q-s}\xi_{p,q}\right)\psi_{j',k'-s}\right] \\
&= \sum_{t,l,m,s,p,q} W_l\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{j,k-t}W_p\left(\frac{q}{n}\right)\psi_{p,q-s}\psi_{j',k'-s}\mathbb{E}[\xi_{l,m}\xi_{p,q}]/
\end{aligned}$$

$$\text{By definition,} \qquad \mathbb{E}[\xi_{l,m}\xi_{p,q}] = \begin{cases} 1, & \text{iff } l = p,\ m = q; \\ 0, & \text{otherwise.} \end{cases}$$

Thus,

$$\begin{aligned}
\mathrm{Cov}(D_{j,k}, D_{j',k'}) &= \sum_{t,l,s,m} W_l^2\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{l,m-s}\psi_{j,k-t}\psi_{j',k'-s} \\
&= \sum_t \psi_{j,k-t} \sum_s \psi_{j',k'-s} \sum_l \sum_m W_l^2\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{l,m-s}.
\end{aligned}$$

Let $s = t - v$, then

$$\begin{aligned}
\mathrm{Cov}(D_{j,k}, D_{j',k'}) &= \sum_t \psi_{j,k-t} \sum_{t+v} \psi_{j',k'-t+v} \sum_l \sum_m W_l^2\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{l,m-t+v} \\
&= \sum_t \psi_{j,k-t} \sum_v \psi_{j',k'-t+v} \sum_l \sum_m W_l^2\left(\frac{m}{n}\right)\psi_{l,m-t}\psi_{l,m-t+v} \\
&= \sum_t \sum_v \psi_{j,k-t}\psi_{j',k'-t+v}\mathrm{Cov}(Y_t, Y_{t-v}).
\end{aligned}$$

Thus,

$$\mathrm{Cov}(D_{j,k}, D_{j',k'}) = \sum_t \sum_v \psi_{j,k-t}\psi_{j',k-t+v}\mathrm{Cov}(Y_t, Y_{t-v}). \tag{2.1}$$

□

3 Determining how much of the EWS one needs to know to compute Σ_k^D

In determining how much of the EWS needs to be known when computing the covariance structure at location k, we consider the following lines of logic. Let $\mathcal{L}_j$ denote the support for the wavelet at scale j (number of non-zero filter coefficients in ψ_j). The number of non-zero

product filtering coefficients, $\psi_{l,m-t}\psi_{l,m-t+v}$, is greatest when we consider the variance of the wavelet coefficients or observations process and no lag is present ($v = 0$). We thus consider $\text{Var}(D_{j,k})$ and $\text{Var}(Y_t)$. In addition, the number of non-zero product terms will be greatest for the coarsest scale considered, J, with corresponding support $\mathcal{L}_J$

$\text{Var}(D_{j,k})$ will be dependent on observations $Y_k, \ldots, Y_{k-(\mathcal{L}_j-1)}$ for any scale $j = 1, \ldots, J$. Thus for the coarsest scale $\text{Var}(D_{J,k})$ will be dependent on observations $Y_k, \ldots, Y_{k-(\mathcal{L}_J-1)}$. The variance for the most distant observation $Y_{k-(\mathcal{L}_J-1)}$ is dependent on the power from the following locations: $k-(\mathcal{L}_j-1)-(\mathcal{L}_j-1), \ldots, k-(\mathcal{L}_j-1)$, for scale j. The coarsest scale requires the most power feeding into it: $W_J^2\left(\frac{k-2(\mathcal{L}_J-1)}{n}\right), \ldots, W_J^2\left(\frac{k-(\mathcal{L}_J-1)}{n}\right)$. For the most recent observation Y_k at the coarsest scale, the following power needs to be known $W_J^2\left(\frac{k-(\mathcal{L}_J-1)}{n}\right), \ldots, W_J^2\left(\frac{k}{n}\right)$.

Thus to compute Σ_k^D, the covariance structure of the wavelet coefficients at location k, we must record the power from the locations $k - 2(\mathcal{L}_j - 1), \ldots, k$ for scale $j = 1, \ldots, J$.

4 Order of HMM with respect to analysing wavelet and J

We briefly comment on the behaviour of the order of the HMM as we consider more scales and different choices in analysing wavelet. Recall that the order of the HMM is associated with the analysing wavelet considered and J, the number of scales considered. More specifically, the HMM order is $2\mathcal{L}_J - 1$.

For the case of the Haar wavelet, where $\mathcal{L}_j = 2, 4, 8, 16$ for $j = 1, 2, 3, 4$, the corresponding order of the induced HMM is $3, 7, 15, 31$ for $J^* = 1, 2, 3, 4$. Similarly, Daubechies Extrexmal Phase wavelets with two vanishing moment has the following supports $\mathcal{L}_j = 4, 10, 22, 46$ for $j = 1, 2, 3, 4$. The induced order of HMM is thus $7, 19, 43, 91$ for $J^* = 1, 2, 3, 4$ scale processes respectively. Thus by considering coarser scales and smoother analysing wavelets, the order of the induced HMM grows exponentially which causes computational problems eventually. The use of a Haar wavelet and only considering a few finer scale processes is thus advocated.

5 SMC samplers example implementation

This section describes more explicitly the SMC samplers implementation described in Section 3.2. Defining $l(\mathbf{d}_{1:n}^2|\theta, H)$ as the likelihood, and $p(\theta|H)$ as the prior of the model parameters,

we can define the following sequence of distributions,

$$\pi_b(\theta) \propto l(\mathbf{d}^2_{1:n}|\theta, H)^{\gamma_b} p(\theta|H) \qquad b = 1, \ldots, B, \tag{5.1}$$

where $\{\gamma_b\}_{b=1}^B$ is a non-decreasing tempering schedule such that $\gamma_1 = 0$ and $\gamma_B = 1$. We could therefore sample from the sequence of distribution $\{\pi_b\}_{b=1}^B$ as follows:

Initialisation, Sampling from $\pi_1 = p(\theta|H)$: Assume independence between the transition probability matrix, $\mathbf{P}$ and the state dependent power, W^2.

$$p(\theta|H) = p(\mathbf{P}|H)p(W^2|H). \tag{5.2}$$

Transition Probability matrix, P: Sample each of the H transition probability rows $p_r = (p_{r1}, \ldots, p_{rH}), r = 1, \ldots, H$ independently from a Dirichlet prior distribution. As HMMs are typically associated with persistent behaviour in the same underlying state, asymmetric priors encouraging persistent behaviour are generally implemented. That is,

$$p_r \overset{\text{iid}}{\sim} \text{Dir}(\alpha_r) \qquad r = 1, \ldots, H$$
$$p(\mathbf{P}|H) = \prod_{r=1}^{H} p(p_r|H),$$

where α_r is the associated hyperparameter encouraging persistency.

State Dependent Power, W^2: Sample each of the state dependent inverse power for each scale independently from a Gamma distribution. That is,

$$\lambda_{j,r} = \frac{1}{w^2_{j,r}} \overset{\text{iid}}{\sim} \text{Gamma}(\alpha_\lambda, \beta_\lambda) \qquad j = 1, \ldots, J, r = 1, \ldots, H$$
$$p(\Lambda = \frac{1}{W^2}|H) = \prod_{j=1}^{J}\prod_{r=1}^{H} p(\frac{1}{w^2_{j,r}}|H),$$

where α_λ and β_λ are associated shape and scale hyperparameters.

Mutation and Reweighting, approximating π_b from π_{b-1}: We consider Random Walk Metropolis Hastings proposal kernels on different domains given the constraints of the parameters; $\mathbf{P}$ is a stochastic matrix, $w^2_{j,r}$ are non-negative. We consider mutating and updating components of θ separately, using the most recent value of the components (akin to Gibbs sampling). In particular, we consider the following mutation strategies to move from θ^i_{b-1} to θ^i_b, for particle i at iteration b.

Transition Probability matrix, P: Consider each of the H transition probability rows p_r separately, and mutate on the logit scale. That is, we propose moving from p_r to p_r^P via:

$$\text{Define the current logits:} \qquad l_r = \left(l_{r1} = \log \frac{p_{r1}}{p_{rH}}, \ldots, l_{rH} = \log \frac{p_{rH}}{p_{rH}} = 0 \right), \tag{5.3}$$

$$\text{Proposal logits:} \qquad l_r^P = l_r + \epsilon_l \qquad \epsilon_l \sim \text{MVN}(0, \Sigma_l), \quad \text{with } l_{rH}^P = 0, \tag{5.4}$$

$$\text{Proposal probability vectors:} \quad p_r^P = \left(\frac{\exp l_{r1}^P}{\sum_{n=1}^H \exp l_{rn}^P}, \ldots, \frac{\exp l_{rH}^P}{\sum_{n=1}^H \exp l_{rn}^P} \right), \tag{5.5}$$

where Σ_l is a suitable $H \times H$ proposal covariance matrix.

State Dependent Power, W^2: Consider each of the state dependent inverse powers for each scale independently, and mutate on the log scale. That is we propose moving from $\lambda_{j,r}$ to $\lambda_{j,r}^P$ via:

$$\lambda_{j,r}^P = \exp(\log \lambda_{j,r} + \epsilon_\lambda) \qquad \epsilon_\lambda \sim \text{N}(0, \sigma_\lambda^2), j = 1, \ldots, J, r = 1, \ldots, H, \tag{5.6}$$

where σ_λ^2 is a suitable proposal variance.

Reweighting: One can show that under general conditions of SMC samplers, the reweighting formula for particle i to approximate π_b is:

$$U_b^i = \frac{U_{b-1}^i \tilde{u}_b(\theta_{b-1}^i, \theta_b^i)}{\sum_{i=1}^N U_{b-1}^i \tilde{u}_b(\theta_{b-1}^i, \theta_b^i)} \tag{5.7}$$

$$\text{with } \tilde{u}_b(\theta_{b-1}^i, \theta_b^i) = \frac{\pi_b(\theta_{b-1}^i)}{\pi_{b-1}(\theta_{b-1}^i)} = \frac{l(\mathbf{d}_{1:n}^2 | \theta_{b-1}^i, H)^{\gamma_b}}{l(\mathbf{d}_{1:n}^2 | \theta_{b-1}^i, H)^{\gamma_{b-1}}}. \tag{5.8}$$

Final Output: We have a weighted cloud of N particles approximating the parameter posterior:

$$p(\theta | \mathbf{d}_{1:n}^2, H) \approx \{\theta_B^i, U_B^i | H\}_{i=1}^N \equiv \{\theta^i, U^i | H\}_{i=1}^N. \tag{5.9}$$

References

Grimmett, G. and D. Stirzaker (2001). *Probability and Random Processes* (3rd ed.). Oxford university press.

Nason, G., R. Von Sachs, and G. Kroisandt (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B 62*(2), 271–292.