A Latent Structure Model for High River Flows

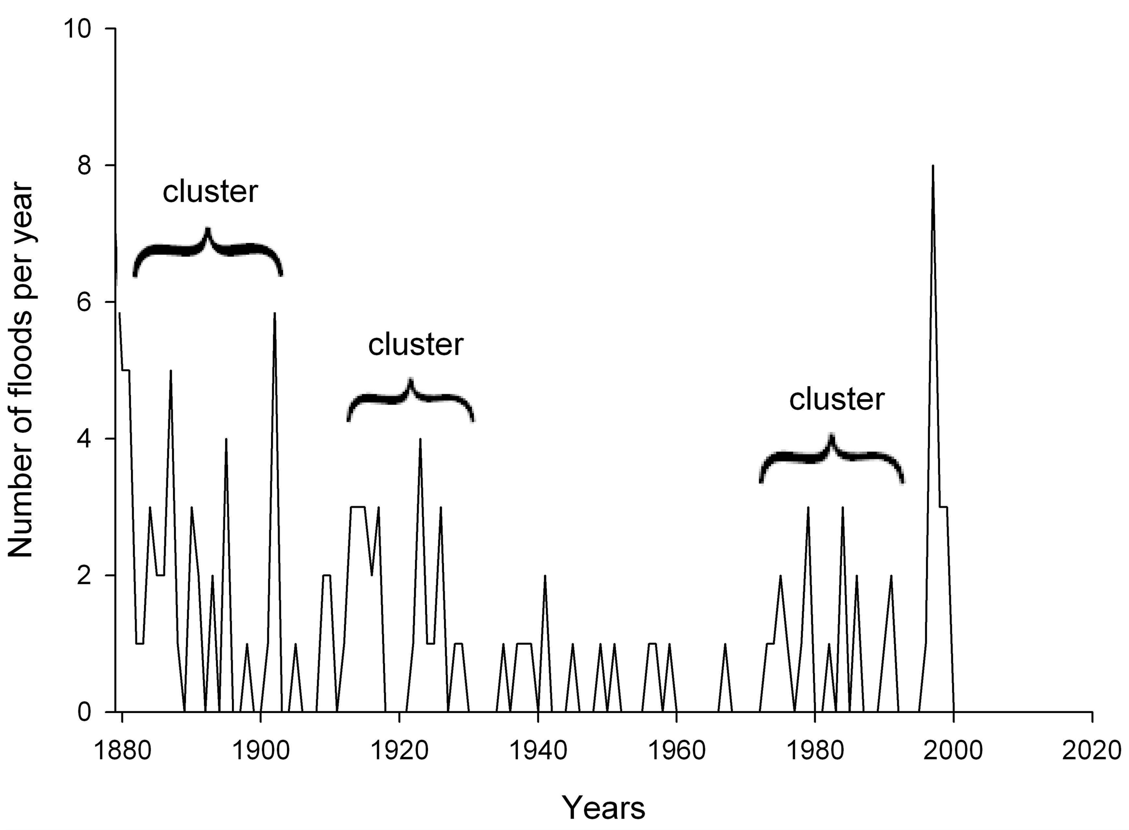
Theodoros Economou*
Renato Vitolo
Trevor Bailey
Emma Waterhouse
Zoran Kapelan

Introduction

Floods are natural catastrophes which may have a disastrous effect in economic terms. For example, UK floods in summer 2007 resulted in the largest flood-related aggregate insured loss in the UK (Lane, 2007).

Recent studies of UK catchments reveal significant *spatio-temporal clustering* in major floods (Robson, 2002; Lane, 2007) over decadal timescales and such non-stationary behaviour makes the problem of modelling high-flow river discharge particularly complex.

Threshold exceedances in discharge: River Lee, UK



One possible explanation for this clustering is low frequency climatic variability, but other unobserved processes (e.g. anthropogenic factors) may also be involved. It is therefore crucial to adopt models that can reflect both *climate trends* and latent structure. This work involves a *hidden semi-Markov model* which can account for features in the variability of flooding that may be driven by *unobserved processes*.

Model Formulation

Consider a data set which is a time series $\boldsymbol{y} = \{y_1, y_2, \ldots, y_N\}$ of yearly counts of single days for which a flood event was recorded. We assume a non-stationary Poisson model for y_t ($t = 1, \ldots, N$), where the mean $\Lambda(\boldsymbol{x}_t; S_t)$ may depend on *time dependent covariates* x_t and on the hidden state S_t of a semi-Markov chain at time t where $S_t \in \{1, 2, \ldots, S\}$ is the state space of the process. The mean, is:

$$\Lambda(\boldsymbol{x}_t; S_t) = \exp\{\theta_{S_t} + \boldsymbol{\beta}\boldsymbol{x}_t\}$$

so that given the state S_t, y_t is Poisson with a mean depending directly on time through $\boldsymbol{x}_t = (x_{1t}, x_{2t}, \ldots, x_{pt})$ and indirectly through the *state dependent intercept* θ_{S_t}. Note that through the hidden chain, some correlation structure is introduced in the counts y_t. To derive the likelihood of the HSMP consider first the likelihood of the conditional Poisson model over a period τ:

$$\ell(y_1, y_2, \ldots, y_\tau | s) = \prod_{i=1}^{\tau} \frac{e^{-\Lambda(\boldsymbol{x}_i; s)} \Lambda(\boldsymbol{x}_i; s)^{y_i}}{y_i!} \quad (1)$$

Second, consider a semi-Markov chain which is often defined by an initial state distribution $\boldsymbol{\pi} = (\pi(1), \pi(2), \ldots, \pi(S))$, a transition probability matrix $\boldsymbol{P} = \{p_{ij}\}$ where $p_{ii} = 0$, $\sum_j p_{ij} = 1$ and a vector of holding time distributions $\boldsymbol{h}(\tau) = \{h_i(\tau)\}$. The likelihood of a realisation $(\tau_{s_1}, \tau_{s_2}, \ldots, \tau_{s_n})$ of this chain involving n state changes is

$$\pi(s_1) h_{s_1}(\tau_{s_1}) \prod_{j=2}^{n} p_{s_{j-1}, s_j} h_{s_j}(\tau_{s_j}) \quad (2)$$

$h_i(\tau)$ can be any discrete distribution and if it is geometric then the chain is Markov and not semi-Markov. Now suppose that both the time series $\boldsymbol{y}$ and the semi-Markov chain have been observed. Then the joint likelihood $L(y_1, \ldots, y_N; \tau_{s_1}, \ldots, \tau_{s_n})$ is obtained by combining (1) and (2):

$$\pi(s_1) h_{s_1}(\tau_{s_1}) \ell(y_1, \ldots, y_{\tau_{s_1}} | s_1) p_{s_1, s_2} h_{s_2}(\tau_{s_2})$$
$$\times\ \ell(y_{\tau_{s_1}+1}, \ldots, y_{\tau_{s_1}+\tau_{s_2}} | s_2) p_{s_2, s_3} h_{s_3}(\tau_{s_3})$$
$$\times \text{ etc.}$$

The chain is not observed though, so we sum over all possible states $s_i \in \{1, 2, \ldots, S\}$ and all possible holding times $\tau_i \in \{1, 2, \ldots, \infty\}$ to obtain the marginal likelihood $L(y_1, \ldots, y_N)$ of the HSMP model as:

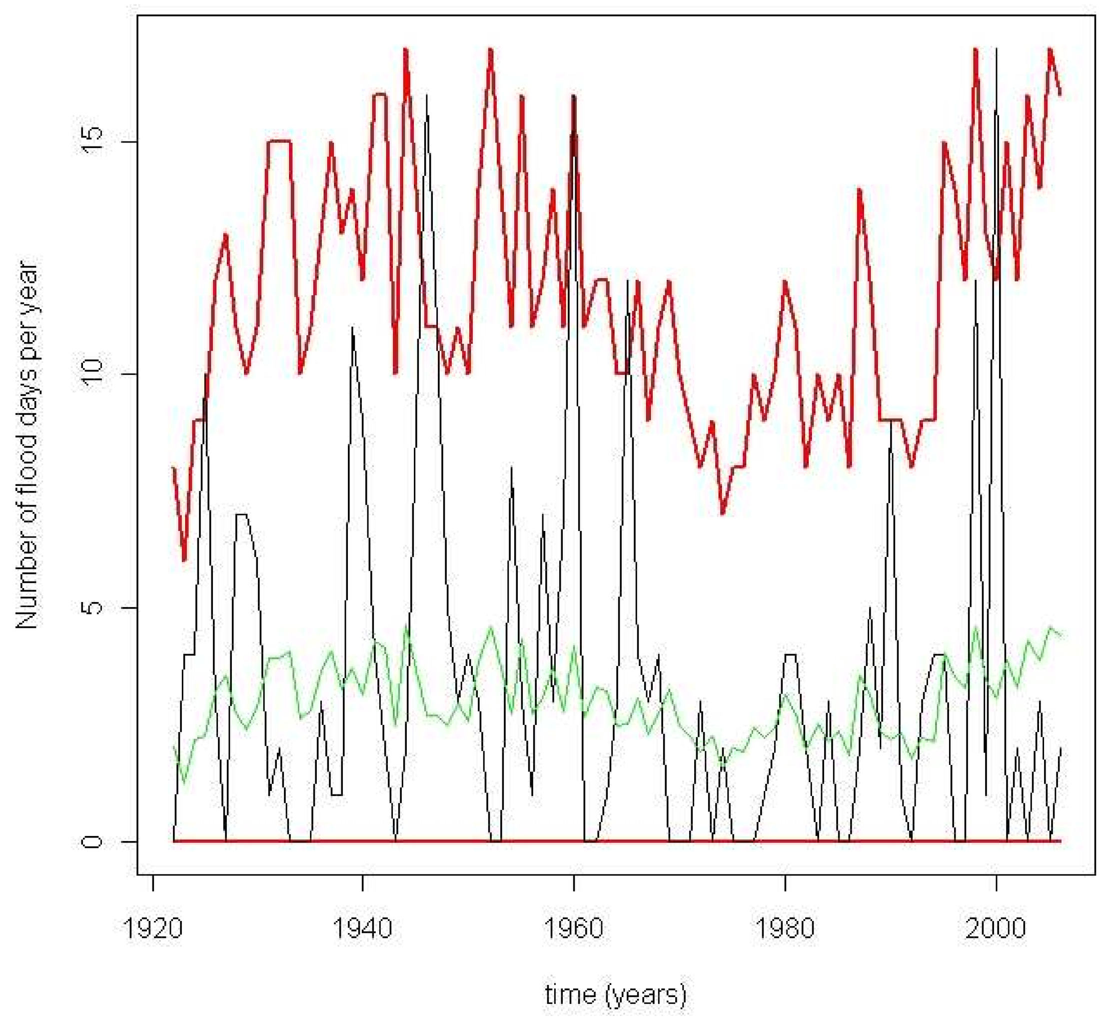
$$\sum_{\tau_{s_1}=1}^{\infty} \cdots \sum_{\tau_{s_n}=1}^{\infty} \sum_{s_1=1}^{S} \cdots \sum_{s_n=1}^{S} L(y_1, \ldots, y_N; \tau_{s_1}, \ldots, \tau_{s_n})$$

This likelihood is a function of the parameters in $\Lambda(\boldsymbol{x}(t); S_t)$, in $\boldsymbol{\pi}$, in $\boldsymbol{P}$ and in $h_i(\tau_i)$. Direct evaluation of the likelihood is prohibitively *computationally intensive* therefore we employ *recursive algorithms* used in the HMM literature (Rabiner, 1989) to efficiently evaluate it. Given the complexity of the model we adopt an *MCMC* approach and the likelihood is used in conjunction with Metropolis-Hastings to provide a computationally feasible estimation procedure. We use a combination of the random walk and independence Metropolis samplers.

Model Application

We examine daily discharge data for the river Severn at Bewdley (UK) in the 85 year period between 1922 and 2006. The response $\boldsymbol{y}$ is the number of days a flood has occurred in a year. Covariates $x_1(t)$ and $x_2(t)$ are used corresponding to the yearly averages of Atlantic multidecadal oscillation (AMO) and North Atlantic oscillation (NAO) indexes during that period. The possible presence of other, not explicit low-frequency processes is accounted for by the hidden semi-Markov chain. Specifically we assume two hidden states S_t in the chain where each has a Poisson holding time with a different mean. The model is then

$$y_t \sim \text{Pois}(\Lambda(x_{1t}, x_{2t}; S_t))$$
$$\Lambda(x_{1t}, x_{2t}; S_t) = \exp\{\theta_{S_t} + \beta_1 x_{1t} + \beta_2 x_{2t}\}$$



The figure shows the fitted (green line) and actual values (black line) of the response. Although the overall behaviour is mainly explained by the covariate AMO and not the hidden chain, the model did identify two hidden states, one with a very small prevalence in time but with a higher value of θ_{S_t} in the Poisson mean which is what the data is suggesting looking at the 'spikes' of large values in the observed counts. This is reflected in the sufficiently high 95% credible intervals (red lines).

References

Lane, S. N. (2007). The 2007 UK summer floods: a scientific perspective. *www.willisresearchnetwork.com/Lists/Publications/DispForm.aspx?ID=6*.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected application in speech recognition. *Proceedings of the IEEE*, **77**, 257-285.

Robson, A.J. (2002). Evidence for trends in UK flooding. *Philosophical transactions of the Royal Society A*, **360**, 1327.

International Workshop on Statistical Modelling, July 2009

corresponding author: t.economou@ex.ac.uk

UNIVERSITY OF EXETER
School of Engineering, Computing and Mathematics