ITT7: Attribution of changes in river flows in the UK

Benjamin Aslan, Aoibheann Brady, Nadeen Khaleel, Sean Longfield & Ilaria Prosdocimi

University of Bath, LSGNT & Environment Agency

February 2nd, 2018

What we said we'd do...

Aim

To both detect *and* attribute changes in peak river flows in the UK.

What we said we'd do...

Aim

To both detect *and* attribute changes in peak river flows in the UK.

Plan

- ~~**Changepoint analysis** to detect expected changes in flows with urbanisation.~~ Too many changepoints to be useful (i.e. basically everything)!

What we said we'd do...

Aim

To both detect *and* attribute changes in peak river flows in the UK.

Plan

- ~~**Changepoint analysis** to detect expected changes in flows with urbanisation.~~ Too many changepoints to be useful (i.e. basically everything)!
- Construct a series of **peaks over threshold** data for each catchment to investigate a **point process** approach (station-by-station).

What we actually did!

Data Processing (this took $\sim 90\%$ **of the time):**

- Constructed POT data for each of the 5 "urbanised" catchments with records from 31-50 years long.
- Added Q99 (the yearly 99th quantile) of daily catchment average rainfall and annual urbanisation data (linearly interpolated).

Modelling ($\sim 5\%$**)**:

What we actually did!

Data Processing (this took $\sim 90\%$ **of the time):**

- Constructed POT data for each of the 5 "urbanised" catchments with records from 31-50 years long.
- Added Q99 (the yearly 99th quantile) of daily catchment average rainfall and annual urbanisation data (linearly interpolated).

Modelling ($\sim 5\%$**)**:

- Investigate whether there are associations between peak flows and time, urbanisation and/or rainfall.

What we actually did!

Data Processing (this took $\sim$ 90% of the time):

- Constructed POT data for each of the 5 “urbanised” catchments with records from 31-50 years long.
- Added Q99 (the yearly 99th quantile) of daily catchment average rainfall and annual urbanisation data (linearly interpolated).

Modelling ($\sim$ 5%):

- Investigate whether there are associations between peak flows and time, urbanisation and/or rainfall.
- Fit generalised Pareto distributions to the *size of the peaks* over a specific threshold.

What we actually did!

Data Processing (this took $\sim 90\%$ of the time):

- Constructed POT data for each of the 5 "urbanised" catchments with records from 31-50 years long.
- Added Q99 (the yearly 99th quantile) of daily catchment average rainfall and annual urbanisation data (linearly interpolated).

Modelling ($\sim 5\%$):

- Investigate whether there are associations between peak flows and time, urbanisation and/or rainfall.
- Fit generalised Pareto distributions to the *size of the peaks* over a specific threshold.
- Fit Poisson regression models to the *counts of the peaks*.

What we actually did!

Data Processing (this took $\sim 90\%$ **of the time):**

- Constructed POT data for each of the 5 "urbanised" catchments with records from 31-50 years long.
- Added Q99 (the yearly 99th quantile) of daily catchment average rainfall and annual urbanisation data (linearly interpolated).

Modelling ($\sim 5\%$**)**:

- Investigate whether there are associations between peak flows and time, urbanisation and/or rainfall.
- Fit generalised Pareto distributions to the *size of the peaks* over a specific threshold.
- Fit Poisson regression models to the *counts of the peaks*.
- Fit a point process model to look at both!

GPD: Is the magnitude of the flow increasing with *insert covariate here*?

Given peak flow data X, for a large threshold u, the distribution of $(X - u)$ conditioned on $X > u$ may be approximated by:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{1/\xi}.$$

This function is defined on $\{y : y > 0 \;\&\; \xi y/\bar{\sigma} > 0\}$, and $\bar{\sigma} = \sigma + \xi(u - \mu)$.

This family of distributions is known as the **generalised Pareto family of distributions**. The *size* of threshold exceedances may be approximated by a member of this family.

Sadly this didn't work as we'd hoped...

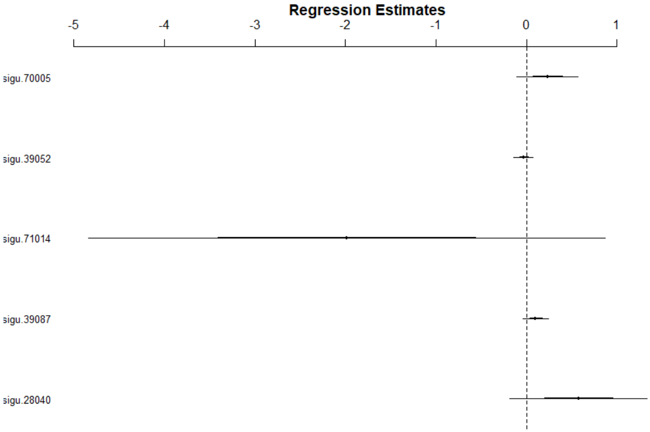

Figure: GPD model for urbanisation vs. magnitude of the POT data

...even when we account for rainfall along with it...

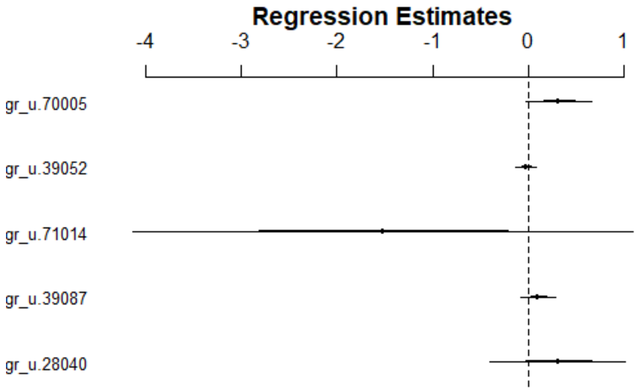

Figure: GPD model for urbanisation with rainfall vs. magnitude of the POT data

Poisson regression: Is the frequency increasing with *insert covariate here*?

We fit a **generalised linear model**. We have count data for the numbers of peaks over threshold for each year, so can assume a **Poisson distribution**.

$$\log(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k = x_i^T \beta$$

where $y_i \sim \text{Poisson}(\mu_i)$ and we use the natural log link $g(\mu) = \log(\mu)$.

Sadly this also didn't work as we'd hoped...

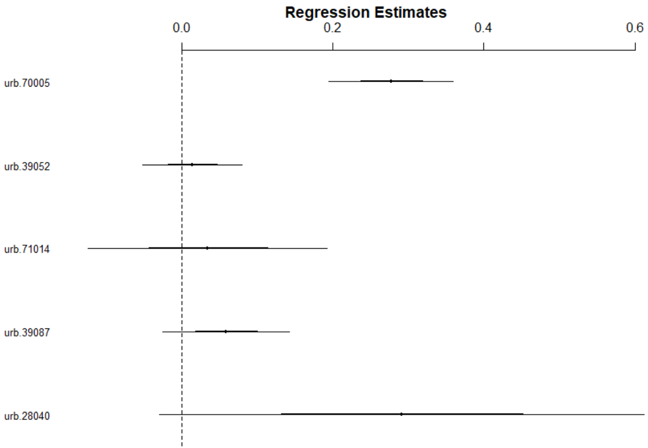

Figure: Poisson regression model for urbanisation vs. counts of the POT data

...but there may be another way! (point process representation)

We want something that looks both at the size and number of exceedances.

The **point process model** describes both the magnitude of threshold exceedances and the rate at which the threshold u is exceeded.

It is parameterised by three parameters – location, scale and shape.

...neither did this :'(

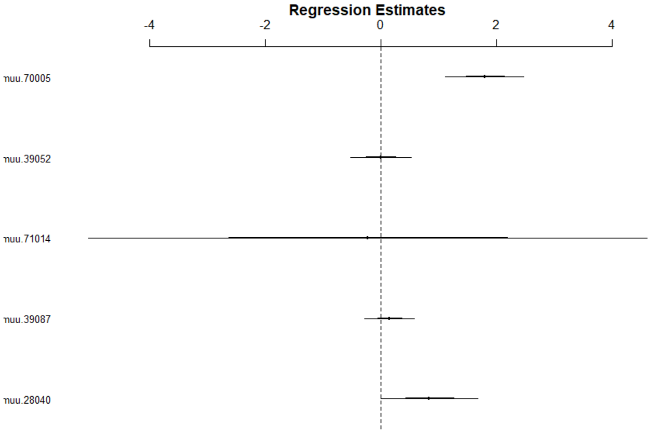

Figure: Point process model for urbanisation vs POT data

Future ideas (AKA definitely not my PhD)

Choice of covariates

- Urbanisation values are linearly interpolated between decades – may not be reliable.
- Quantification of the impact of urbanisation is difficult as authorities may be offsetting any increased risk.
- We may be looking at the wrong covariates – should investigate the effect of other climate drivers.

Causality

- Attribution is difficult! Initial approach: combining variables in regression models. In the future: ...?

Pooling of information/hierarchical model:

- At-site trend tests aren't very powerful & are sensitive to fluctuations. Use countrywide hierarchical model approach to "pool" information (#TeamDetection).