

Signature Methods to Cluster Flow-Data

Teo Deveney & Momchil Konstantinov

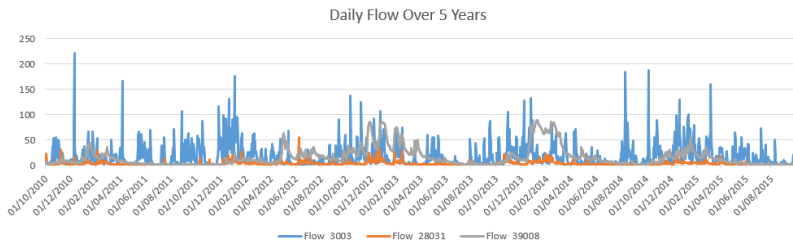
University of Bath

ITT Presentation

February 2, 2018

River flow data

Given the daily flow for each catchment

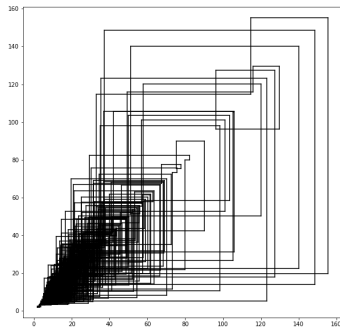
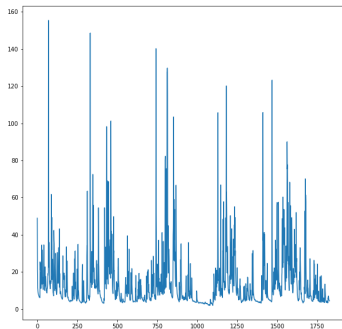


Can we cluster similar catchments based on their behaviour?

Idea:

Transform discrete time series data $\{Y_1, \dots, Y_N\}$ into a 2d path.
Can be done with a lead-lag transformation

$$\text{LeadLag} : Y = \{Y_1, \dots, Y_N\} \mapsto \begin{cases} Y_{\text{Lead}} = \{Y_0, Y_1, Y_1, Y_2, \dots, Y_N\} \\ Y_{\text{Lag}} = \{Y_0, Y_0, Y_1, Y_1, \dots, Y_N\} \end{cases}$$



Signatures Characterise the Structure

The signatures of a 2d path y up to 2nd order are

$$S = [S_1^0, S_1^1, S_2^1, S_{1,1}^2, S_{1,2}^2, S_{2,1}^2, S_{2,2}^2]$$

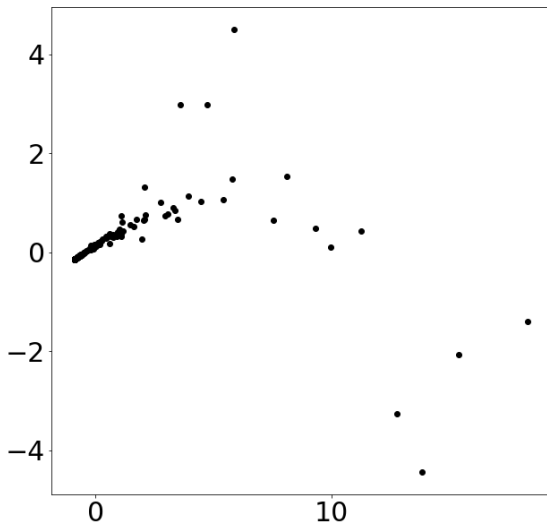
Where

$$S_{i_1, \dots, i_n}^n = \int \cdots \int_{t_1 \leq \dots \leq t_n} dy_{t_1} \cdots dy_{t_n}$$

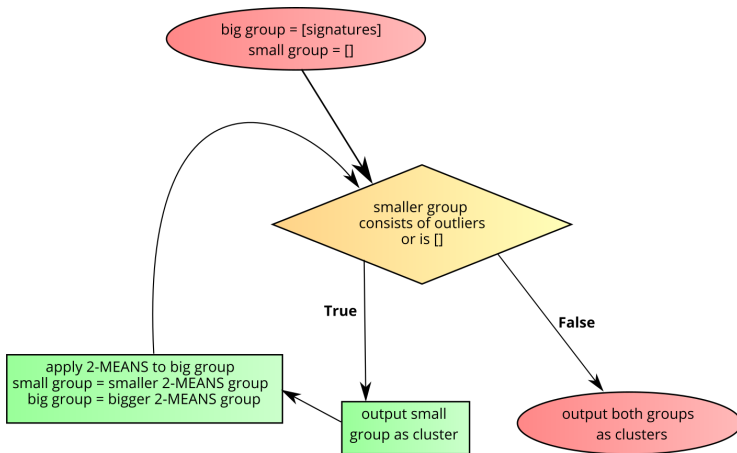
Features might be distinguished more clearly by considering the log signatures. Up to 2nd order these are

$$\log S = \left[S_1^2, S_2^1, \frac{1}{2}(S_{1,2}^2 - S_{2,1}^2) \right]$$

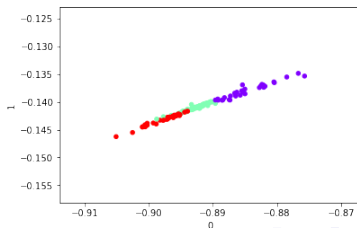
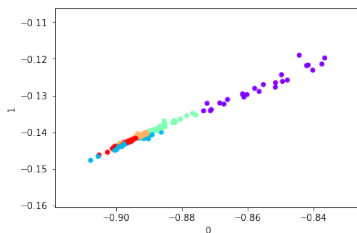
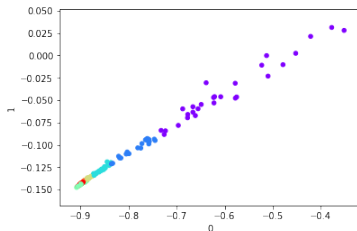
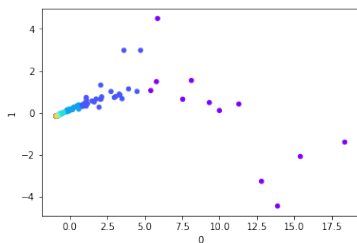
Visualising the Signatures of the Catchments



Clustering Catchments by Signatures



Results



Further Steps

- ▶ Implement hierarchical clustering algorithm which incorporates higher order signatures
- ▶ Further investigating into output. Are the clusters reasonable? Can we compare signature distances with similarity distance measures
- ▶ Outside Classification - Used for Regression
 - ▶ Gaussian Process took 2 minutes
 - ▶ Signatures took 3 seconds

