Stochastic modelling approach for future flood risk modelling

S. Patidar$^1$
H. Haynes$^1$
Ji Li$^1$
R. Haynes$^2$

$^1$Institute for Infrastructure and Environment, Heriot-Watt University, Edinburgh, Scotland, UK
$^2$Scottish Environmental Protection Agency, SEPA
Motivation

Climate change is playing a key role in the manifestation of the extreme climatic events, Eg. Flooding, hurricanes, cyclones, etc.

- For the UK one of the most challenging issue is:
  - increased risk of frequent and severe flooding in the future

- Mostly flood risk management project involves:
  - Use of single 1:N year extreme flow/rainfall event.

- This approach does not accounts for:
  - effect of flood clustering on channel cross-section
  - effect of change in channel capacity on flood risk

- Future climate change scenarios indicate that:
  - storm and flood frequency will increase ➡ the UK’s susceptibility to storm clusters will also be magnified.
Aim

To develop an efficient modelling approach that combines:

- *stakeholder knowledge* (Scottish Environment Protection Agency, SEPA) with cross-disciplinary academic expertise (statistical modelling, river process, climate change & flood modelling)

- *to refine mathematical modelling approaches* in flood risk.

Our approach includes:

- Statistical modelling
- sediment transport processes & related flood risk
- multi-event simulation
- flood sequence/cluster risk-recovery processes
- …*but, able to work within the constraints of limited data & run times*
Methodology

- **Stochastic modelling of 15 min mean flow**
  - Generate multiple (e.g.100) synthetic flow sequences
  - Hidden Markov Model + Generalised Pareto (HMM-GP)

- **1D/2D sediment transport modelling of flow sequences**
  - 100 future channel configurations
  - Define ‘worst-case’ channel

- **1D/2D flood hazard modelling using ‘worst-case’ channel**
  - Quantify change in inundation

---

Methodology

Previous Work:
- EPSRC’s FloodMEMORY: Multi-Event Modelling Of Risk & Recovery (EP/K013513/1) project as part of FCERM.net (EP/L000180/1) project has HMM-GP modelling framework for simulating synthetic daily flood sequences.

Present work:
- Investigated suitability of HMM-GP modelling approach for simulating streamflow at a much finer temporal resolution of 15 minutes.
- Validated for 14 distinct gauging stations across 4 cases study rivers Don, Nith, Dee and Tweed.
- Developed a systematic approach that exploits STL: a Seasonal-Trend decomposition procedure based on Loess for reframing HMM-GP modelling framework to allow integration of climate impact.
- Utilise 2D sediment transport and flood inundation modelling.
Why do we need stochastic model?

Quite often, **observed flow records are not long enough:**
- to extract reliable statistics to understand the variability and uncertainty in the flow patterns and the occurrences of all significant rare events (high/low flow).

Statistically significant **impact analyse of 1: N year return period extreme event requires a flow series of at least 4 * N years:**
- i.e., for conducting a statistically significant impact analyse of a 1:200 year return period extreme event a flow series of length at least 800 years is needed.

**Daily record skips several significant small duration intense flooding events:**
- capable of causing severe damages.

**Existing gauge data are not sufficiently extensive:**
- practically unobtainable at this scale (length up to 800 years and resolution of 15 minutes).
Why do we need stochastic model?

- Computationally efficient at generating multiple realisations for uncertainty analysis
- Provide realistic realisations of river flows
- Can be used for long-term modelling
- Allows estimation of sediment transport and loading on flood defences
- Ensure long-term sustainability of flood defence assets
- Easily applicable at multiple sites
- Limits error accumulation that occurs using rainfall and hydrological models
Catchment geographical characteristics

- The gauging stations were selected for the availability of 15-minutely data, the length of records and the minimal amount of missing data (less than 5%).

- All the selected river catchments have both the urban development areas (limited to less than 0.5%) and headwater agricultural areas.

- The average daily flow located furthermost upstream of the rivers are below 13 m³/s and the downstream gauging stations show a variability within a range from 28 m³/s to 81 m³/s.

The 15 minutely gauged dataset is provided by the Scottish Environmental Protection Agency (SEPA).
## Catchment geographical and Statistical characteristics

<table>
<thead>
<tr>
<th>Catchment geographical characteristics</th>
<th>Don (Houghton)</th>
<th>Nith (Friars Carse)</th>
<th>Dee (Woodend)</th>
<th>Tweed (Norham)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catchment area (km²)</td>
<td>1273</td>
<td>799</td>
<td>1844</td>
<td>4390</td>
</tr>
<tr>
<td>River slope (m/km)</td>
<td>111</td>
<td>157</td>
<td>169</td>
<td>136</td>
</tr>
<tr>
<td>Urban extent (%)</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>River mobility</td>
<td>High deposition – high erosion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gauging station statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (μ)</td>
<td>14.72</td>
<td>28.49</td>
<td>38.76</td>
<td>81.99</td>
</tr>
<tr>
<td>Standard Deviation (σ)</td>
<td>13.98</td>
<td>39.04</td>
<td>40.96</td>
<td>91.87</td>
</tr>
<tr>
<td>Minimum (Q0)</td>
<td>0.77</td>
<td>1.11</td>
<td>3.48</td>
<td>6.64</td>
</tr>
<tr>
<td>5th Percentile (Q5)</td>
<td>4.15</td>
<td>2.98</td>
<td>7.98</td>
<td>14.56</td>
</tr>
<tr>
<td>10th Percentile (Q10)</td>
<td>4.89</td>
<td>3.97</td>
<td>10.35</td>
<td>17.63</td>
</tr>
<tr>
<td>25th Percentile (Q25)</td>
<td>6.91</td>
<td>7.17</td>
<td>16.91</td>
<td>28.06</td>
</tr>
<tr>
<td>50th Percentile (Q50)</td>
<td>10.79</td>
<td>15.13</td>
<td>27.29</td>
<td>53.45</td>
</tr>
<tr>
<td>75th Percentile (Q75)</td>
<td>17.47</td>
<td>34.31</td>
<td>45.30</td>
<td>100.34</td>
</tr>
<tr>
<td>90th Percentile (Q90)</td>
<td>27.77</td>
<td>66.51</td>
<td>76.11</td>
<td>174.05</td>
</tr>
<tr>
<td>95th Percentile (Q95)</td>
<td>37.49</td>
<td>97.15</td>
<td>107.20</td>
<td>243.71</td>
</tr>
<tr>
<td>100th Percentile (Q100)</td>
<td>721.64</td>
<td>1119.27</td>
<td>1605.27</td>
<td>1602.16</td>
</tr>
</tbody>
</table>
Catchment statistical characteristics

Ordered monthly box plot illustrating five summary statistics (5th percentile, 25th percentile, 50th percentile, 75th percentile and 95th percentile; maximum values are specified at top of each box plot).

Case study rivers:
- Don at Haughton,
- Nith at Friars Carse,
- Dee at Woodend
- Tweed at Norham.
Stochastic Modelling of stream flow

- to produce desired number of the synthetic flow-series, our approach combines Hidden-Markov Model with extreme value distribution [Generalised Pareto (GP)].

Hidden Markov Model

- Based on the evolution of process/system in time from a given state to another state, i.e.
- Exploits the probability of the system to jump from one state to another
- Accounts for the transition of system through hidden states
Hidden Markov Model

A HMM model is comprised of five elements:

1. **Set of distinct observed states** - Percentile analysis of all observed values using increment of 10% to define eleven distinct observed states (A,B,C,D,E,F,G,H,I,J,K)

   - State A – flow between the minimum and 10th Percentile
   - State B – flow between the 10th and 20th percentile
   - ...
   - ...
   - So on.
Hidden Markov Model

2. **Set of unobserved (hidden) states within observed states** – account for all the discrete values with one decimal place within the range of each eleven distinct observed states.

**For example:**

- if state A corresponds to values between 0 to 1 then set of unobserved states corresponding to state A will be 0.1, 0.2, 0.3, ..., 0.9

- if state B corresponds to values between 1 to 2 then set of unobserved states corresponding to state B will be 1.1, 1.2, 1.3, 1.4, ..., 1.9

- ... so on
Hidden Markov Model

3. **State transitional probability matrix** – probability of transition between different observed states.

- Corresponding to eleven distinct states $\Rightarrow 11 \times 11$ matrix

\[
\begin{bmatrix}
  m_{aa} & m_{ab} & \cdots & \cdots & m_{ak} \\
  m_{ba} & m_{bb} & \cdots & \cdots & m_{bk} \\
  \vdots & \vdots & \ddots & \cdots & \vdots \\
  \vdots & \vdots & \cdots & \ddots & \vdots \\
  m_{ka} & m_{kb} & \cdots & \cdots & m_{kk}
\end{bmatrix}
\]
4. **Emission probability matrix** – emission probability of underlying (hidden) states from discrete observed states.

- Corresponding to eleven distinct states and nine underlying (hidden) state $\Rightarrow 11 \times 9$ matrix

\[
\begin{bmatrix}
m_{a1} & m_{a2} & \ldots & m_{a9} \\
m_{b1} & m_{b2} & \ldots & m_{b9} \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
m_{k1} & m_{k2} & \ldots & m_{k9}
\end{bmatrix}
\]

5. **Set of eleven initial probability of observed states** – obtained from the analysis of time series data
Extreme Value Distribution

Generalized Pareto (GP) distribution has been fitted to the data > 99th percentile to sample extreme values

Why Extreme Value Distribution?

➢ With pure HMM upper extreme values were controlled by measured data, e.g.

For example, if original data has $n$ discrete extreme values $x_1, x_2, ..., x_n$; then all synthetic series will have extreme values sampled from these $n$ values.

➢ Application of extreme value distribution provide an opportunity to sample extreme values from a continuous distribution rather than few discrete values.
Stochastic modelling framework

Synthetic simulate of 15 minutely streamflow sequences using HMM-GP methodology:

1. **Take the log of the time-series:**
   - transforms an additive time series into a multiplicative series.

2. **Apply the STL time-series deseasonalisation procedure** based on the Loess process to the log series:
   - segregate the log time-series into three: trend, seasonal, and a random component.

3. **Fit an HMM to the random component of the time-series:**
   - generate $N$ simulations of the random component of user-specified length.

4. **Add simulated random components to the corresponding seasonal and trend component.**

5. **Resamples extreme values from an extreme value distribution**
   - fitted to the extreme values over 99th percentile of the observed series.
Results: Comparing annual statistics

Comparing:
Probability Density Distribution (PDD);
Quantiles (from 0 to 98th percentile with a step size of 1); and
Auto-Correlation Function (ACF)

for the observed (solid thick black lines) and 100 synthetic streamflow profiles (solid brown lines).

Statistics are estimated for entire 15 minutely streamflow dataset available over the observation period.
Results: Comparing annual statistics

Comparing:
5th, 25th, 50th, 75th and 95th percentiles

for the observed (black dotted lines with solid circles) and 100 synthetic streamflow profiles (brown solid circles)

for the case study river Don at Haughton.

Percentiles are estimated for 15 minutely streamflow data for each year and over the period 1972 - 2015.
Results: Comparing extreme percentiles

Comparing 10 percentiles values from 99th to 100th percentile with a step size of 0.1

for the observed (black dotted lines with solid circles) and 100 synthetic streamflow profiles (brown solid circles)

for the case study river Don at Haughton, Nith at Friars Carse, Dee at Woodend and Tweed at Norham.

Percentiles are estimated for entire 15 minutely streamflow dataset available over the observation period.
Current uses within flood risk studies

Sediment transport modelling:

The two-dimensional hydro-morphological model (Telemac-2D with SYSYPHE) is used to reproduce the floodplain inundation depth, areas and drying time.

Flood cluster scenario design:

- Case 1: 200 year RP followed by a 2 year RP
- Case 2: 200 year RP followed by a 10 year RP
- Case 3: 200 year RP followed by a 200 year RP
Current uses within flood risk studies

The relationship between the **maximum inundation area** and the **time between two events**.
Current uses within flood risk studies

The relationship between the **maximum flow depth** and **the time between two events**.
Current uses within flood risk studies

The relationship between the drying time and the time between two events.
Conclusions

The methodology developed has several advantages, a few listed below:

- The HMM-GP model can be *utilised to generate* $N$ (sufficiently large integer value) *synthetic series* to serve a range of potential applications.

- The methodology *requires no additional information other than the historic observed data series* and *does not dependent on the length of the observed records*.

- All synthetic series generated with the model represents a *realistically plausible distinct scenario* and have *same (considerably close) statistical properties to the historical observed series*.
Future Work

Aim to improve & constrain methods towards practitioner needs ...

Objective 1: Climate change within the HMM-GP

✓ use of 15 minute gauge data
✓ which UKCIP scenarios?
✓ use of multiple gauges (pan-Scotland & downstream translation of sensitivity)

Objective 2: Constrain sequence → clusters of influence

✓ use of 15 minute gauge data to run sequences
✓ describe/design clusters better (POT thresholds, event duration)
✓ compare cluster-in-sequence (CIS) to equivalent cluster-from-benchmark (CFB)
✓ minimum timeframe to capture change (≥CFB?)
✓ compare hazard analysis

Objective 3: Consider downstream translation of sensitivity
Integrating precipitation with HMM-GP model for synthesising flows sequences

Percentile Distribution Model (PDM) to integrate precipitation information with the synthetic river flow series (generated by HMM-GP)

**Upper panel** - HMM-GP Model has been trained using 1960-1990 daily river flow and precipitation data.

Clearly HMM-GP model follows closely with original flow data

**Lower Panel** – Model trained on 1960 -1990 data has been used to synthesis synthetic flow series for 1990 – 2013.

Demonstrate capabilities of PDM-HMM-GP modelling approach in effectively incorporating influence of precipitation with flow series.
Application of PDM-HMM-GP Model to generate future flow series using daily probabilistic precipitation projections (available from UKCP09) for two future time periods:

**Upper Panel**
2030s – 2020 to 2049

**Lower Panel**
2050s – 2040 to 2069

**Some notes**
Method need to be rigorously refined and possibilities for including other influencing variables need to be assessed.
Many thanks for attention 😊

Journal papers


Conference papers


Contact: S.Patidar@hw.ac.uk