The role of emergent technologies in understanding the natural environment and the complex web of interconnected systems

Dr Ross Towe

Email: r.towe@lancaster.ac.uk  Twitter: @RPTowe

25 July 2017
What is *Ensemble*?

- Research programme led by Professor Gordon Blair
- Aim to explore how digital technology can help to address the grand challenges in environmental science
- Interdisciplinary team of researchers
- A number of industrial partners
- Focus on communication to stakeholders
Research Questions

- How can we manage the underlying intrinsic complexity?
Research Questions

- How can we manage the underlying intrinsic complexity?
- How can we reason and manage uncertainty across a wide range of data sources and environmental models?
Research Questions

- How can we manage the underlying intrinsic complexity?
- How can we reason and manage uncertainty across a wide range of data sources and environmental models?
- How can we raise the level of abstraction of technologies to help environmental scientists concentrate on science rather than computing infrastructure?
Research Questions

• How can we manage the underlying intrinsic complexity?

• How can we reason and manage uncertainty across a wide range of data sources and environmental models?

• How can we raise the level of abstraction of technologies to help environmental scientists concentrate on science rather than computing infrastructure?

• How do we understand the overall software architecture and technological building blocks to deploy such an approach in the cloud?
## Ensemble Five Year Plan

<table>
<thead>
<tr>
<th></th>
<th>JAN</th>
<th>FEB</th>
<th>MAR</th>
<th>APR</th>
<th>MAY</th>
<th>JUN</th>
<th>JUL</th>
<th>AUG</th>
<th>SEP</th>
<th>OCT</th>
<th>NOV</th>
<th>DEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sprint Cycle:**

**Colour Coding:**

- **Think:** 3 months
- **Act:** 8 months
- **Inform:** 9 months

**Dissemination**

<table>
<thead>
<tr>
<th>FLOOD</th>
<th>Bio-Diversity</th>
<th>Soil</th>
<th>Eco-System Services</th>
</tr>
</thead>
</table>
Flood Sprint

- Two day workshop
- Attendees from academia, government and industry
- Representatives from across the flood risk sector
- Identify industry, policy and technology needs
- What can we do differently to the status quo?
A Data-Centric View of Place
Data-Driven Decisions

- Flood Event
- Physical System
- States
- Duration of the Event
- Components
- Causality
- Probability of the Event
- Provenance of Data
- Statistical Dependencies
- Physical Drivers
- Hazard
- Hydrology
- Impact
- Natural World
- Weather
- Spatial Scale
- Citizen Science
- Observed
- Simulated
- Local
- Central
Two-Way Knowledge Sharing

Local Communities

Scientific Experts

Decision Makers
Cross-Cutting Themes

Abstraction, Adaptation, Complexity and Uncertainty

Examples include:

- Evidence to justify confidence in the accuracy of models
- Ability to reason and understand the provenance of a number of heterogeneous data sets
- Transparency in decision making
- How do we communicate, understand and manage uncertainty?
Cross-Cutting Themes

Abstraction, Adaptation, Complexity and Uncertainty

Examples include:

- Evidence to justify confidence in the accuracy of models
- Ability to reason and understand the provenance of a number of heterogeneous data sets
- Transparency in decision making
- How do we communicate, understand and manage uncertainty?
Integrated Flood Risk Modelling

Integrated Simulation Pathway

Global Weather
- Weather analysis and forecast data
- Natural climate variability + climate change

Local Weather
- Nested kilometre-scale local weather models
- Probabilistic representation of rainfall
- Coastal storm surges

Hydrology
- Integrated hydrological modelling:
  - Soil moisture
  - Ground water
  - Surface water
  - River flows

Inundation
- Hydraulic modelling:
  - Inundation -- flood depth, velocity and extents
  - River channel dynamics

Impacts
- Flood Hazard Impact Modelling
  - Hazard x Vulnerability x Exposure

Response
- Flood risk mitigation:
  - Engineered defences
  - Landscape and water management

Linked and Compound Hazards e.g. landslides

Statistical Modelling Pathway

Hydrology
- Statistical modelling:
  - River flows
  - Return periods for extreme peak flows

Inundation
- Hydraulic modelling:
  - Inundation -- flood depth and extents
  - Extreme Flood Outline

Impacts
- Flood Hazard Impact Modelling
  - Hazard x Vulnerability x Exposure

Response
- Flood risk mitigation:
  - Engineered defences
  - Landscape and water management

Challenges in Modelling Hydrology

- 916 gauge locations
- Need a flexible and scalable statistical model
- Spatial and temporal dependence
- Missing data
- Computationally expensive simulations
UK Flooding

Storm Desmond: Homes flooded and thousands without power

Credit: BBC News

Credit: goburrito
Integrated Risk Assessment
Integrated Risk Assessment

Pathway

Source

Hazard (H) → Flood Defence (S) → Damage (Z)
Calculation of Damage ($\mu_Z$)

\[ \mu_Z = \int_H f(H)dH \]
Calculation of Damage ($\mu_Z$)

\[ \mu_Z = \int_H \int_S P(S \mid H) dS \frac{f(H)}{dH} \]

Due to computational complexity

\[ \mu_Z = \int_H P(H) \frac{f(H)}{dH} \]
Calculation of Damage ($\mu_Z$)

\[ \mu_Z = \int_H \left( \int_S P(S \mid H) dS \right) f(H) dH \]

\[ = \int_H \int_S Z(S, H) P(S \mid H) dS f(H) dH \]
Calculation of Damage ($\mu_Z$)

$$\mu_Z = \int_H \int_S P(S | H) dS f(H) dH$$
$$= \int_H \int_S Z(S, H) P(S | H) dS f(H) dH$$

Due to computational complexity

$$\mu_Z = \sum_H \sum_S Z(S, H) P(S | H) f(H)$$
Existing Implementations - Approach 1

Determine 9 load levels and run a complex flood inundation model to calculate damages

\[
\mu_Z = \sum_{i=1}^{9} \sum_S Z(S, H_i) P(S | H_i) f(H_i)
\]
Existing Implementations - Approach 1

Determine 9 load levels and run a complex flood inundation model to calculate damages

\[ \mu_Z = \sum_{i=1}^{9} \sum_s Z(S, H_i) P(S \mid H_i) f(H_i) \]

- Loss of information
- Difficult to explore the uncertainty around the estimate
Existing Implementations - Approach 2

Determine 39 load levels and run a less detailed flood inundation model to calculate conditional damages ($c_i$)

$$\mu_Z = \sum_{i=1}^{39} c_i P(H \leq H_i)$$
Existing Implementations - Approach 2

Determine 39 load levels and run a less detailed flood inundation model to calculate conditional damages \( (c_i) \)

\[
\mu_Z = \sum_{i=1}^{39} c_i P(H \leq H_i)
\]

- Loss of information
- Inconsistencies in the ground datum
How can a Data-Driven Approach Help?

**Loss of information**
- Expose available information from metadata
- Use of semantic web technologies

**Computational cost**
- Cloud computing
- Alternative simulation techniques to Monte Carlo
How can a Data-Driven Approach Help?

Dealing with missing data
- Running models on-demand
- Comparing and combining multiple models

Single query
- Providing a multi-faceted view of risk
- Flexible time scales
- Computational efficiency vs accuracy of calculation
- Consider multiple damage functions
Open Challenges

- How do we couple process and statistical models?
- How can we use ontologies to reason over spatial and temporal scales?
- Can we use a domain specific language to consider a range of scenarios for decision planning?
- Can a data-centric approach enable us to do better science?
- Are these ideas transferable to other domains of environmental science?
Any Questions?

Email: r.towe@lancaster.ac.uk    Twitter: @RPTowe
R Shiny Application for Calculating Joint Probability