A first collective validation of global fluvial flood models for major floods in Nigeria and Mozambique

Mark Bernhofen  PhD Student - University of Leeds  cn13mvb@leeds.ac.uk


GFP Conference, Delft. Session 1 – 25 June, 2018
Background

• Global flood model intercomparison study by Trigg et al\(^1\) compared the flood hazard output of 6 global flood models in the continent of Africa.

• 6 models compared in the study were: CaMa-Flood, CIMA-UNEP, ECMWF, GLOFRIS, JRC, and SSBN (now Fathom Global).

• Only 30-40% agreement in flood extent.

• Identified the need for validation against observed events

\(^1\) Trigg et al 2016 The credibility challenge for global fluvial flood risk analysis ERL
Project Objectives

- Identify validation regions with large flood events and sufficient data for analysis
- Come up with a validation framework to consistently test all 6 models under
- Test and compare each individual model
- Test and compare the aggregated models
- Test a “combination” of the best individual models
Study Regions

- 3 study regions, 2 flood events
- 2012 flooding in Nigeria of the Niger and Benue rivers
- 2007 flooding in Mozambique of the Zambezi river
- Both events had roughly estimated return periods of 50 years
Performance Metrics

- Critical Success Index (CSI) – proportion correct [1 best, 0 worst]
- Hit Rate (HR) – proportion of observed captured by model [1 all, 0 none]
- Bias – measures bias towards under (-ve) or over prediction (+ve)

\[ CSI = \frac{F_m \cap F_o}{F_m \cup F_o} \]
\[ HR = \frac{F_m \cap F_o}{F_o} \]
\[ Bias = \frac{(F_m \cap F_o) + F_m}{(F_m \cap F_o) + F_o} - 1 \]
Individual Models

Models from left to right are ordered in descending order of resolution (approx. 1km, 1km, 500m, 500m, 90m, 90m)
Aggregated Models

25 year return period

Model Agreement vs. Average Performance Scores

-CSI
-HR
-BIAS

Mean of all three sites

All flood areas from all models

Only flood area where all models agree
Model Combination

- Would a composite of the best individual models improve performance by removing outliers?
- How to define “best” models?
- How many models to include?
- The Ensemble Score (ES) takes the average CSI across the three sites and penalizes for bias:

\[
ES = \text{Average CSI} - |0.2 \times \text{Average Bias}|
\]

- The best models according the ES were: CIMA-UNEP, JRC & SSBN
Composite Model Performance

Composite CSI Score Comparison

- Model: Composite, ≥2 MA, JRC

CSI Score:
- Lokoja
- Idah Region
- Chamba
Conclusions

- Performance between models shows wide range of under to over prediction – hence poor agreement in original intercomparison.

- Return period (input flow) has a significant effect on model performance.

- Hydraulic characteristics of the region play a role in model performance.

- Aggregate and composite models do not perform better than best individual model. But may be useful in regions where validation is not possible?

- No apparent link between model resolution and performance.

Limitations

- Some models may have been updated since this study.

- Observational data assumed to be correct. May not always be the case.
Outlook

This study has highlighted the usefulness of a common validation procedure, going forward:

- Validation should include cases to cover a wide range of locations, conditions and scales.

- Expand the validation to included not only flood footprint overlap but also a comparison of modelled vs. measured flow and modelled vs. measured inundated depth.

- Validation data to be shared between groups.
Thanks for listening
Any questions?