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

Mark V. Bernhofen[1], Charlie Whyman[1,2], Mark A. Trigg[1], P. Andrew Sleigh[1], Andrew M. Smith[3], Christopher C. Sampson[3], Dai Yamazaki[4], Philip J. Ward[5], Roberto Rudari[6], Florian Pappenberger[7], Francesco Dottori[8], Peter Salamon[8], Hessel C. Winsemius[9]

[1] School of Civil Engineering, University of Leeds, Leeds, LS2 9JT, UK
[2] Stantec, High Wycombe, HP11 1JU, UK
[3] Fathom Global, Engine Shed, Temple Meads, Bristol, BS1 6QH, UK
[4] Institute of Industrial Science, The University of Tokyo, Tokyo, 153-8505, Japan
[5] Institute for Environmental Studies(IVM), Vrije Universiteit Amsterdam, 1081HV Amsterdam, The Netherlands
[6] CIMA Research Foundation, I-17100, Savona, Italy
[7] European Centre for Medium-Range Weather Forecasts, Reading, RG2 9AX, UK
[8] European Commission, Joint Research Centre, I-21027 Ispra, Italy
[9] Deltares, 2629 HV Delft, The Netherlands

Abstract

Global flood models are becoming increasingly important for disaster risk management internationally. However, these models have had little validation against observed flood events, making it difficult to compare model performance. In this paper we introduce the first collective validation of Global Flood Models (GFMs). We identify three hydraulically diverse regions in Africa with large scale flood events and evaluate the output of 6 global flood models against satellite observations. The Critical Success Index of individual models across the three regions ranges from 0.45 to 0.7 and the percentage of flood captured ranges from 52% to 97%. Using the results of our analysis we create and validate a three-model composite to investigate the usefulness of using model combinations to improve flood hazard modelling. We find the composite model performs similarly to the best individual or aggregated models. In the three study regions, we found no correlation between performance and model resolution. The best models show a useful level of performance for these large rivers.

Introduction

Flooding is the most frequent and the most damaging of natural disasters globally [2]. From 1995-2015, floods affected 2.3 billion people, killing 157,000 [3]. Fluvial (river) flooding is the most common type of flood event and with over half of the world's population living within 3 km of a freshwater body it has truly global implications [5]. Flood impacts will continue to increase in severity, as the population exposed to fluvial flooding is expected to rise by 31% over the next 30 years. Certain vulnerable regions, such as Sub-Saharan Africa, are predicted to see an increase in exposed population by as much as 104%

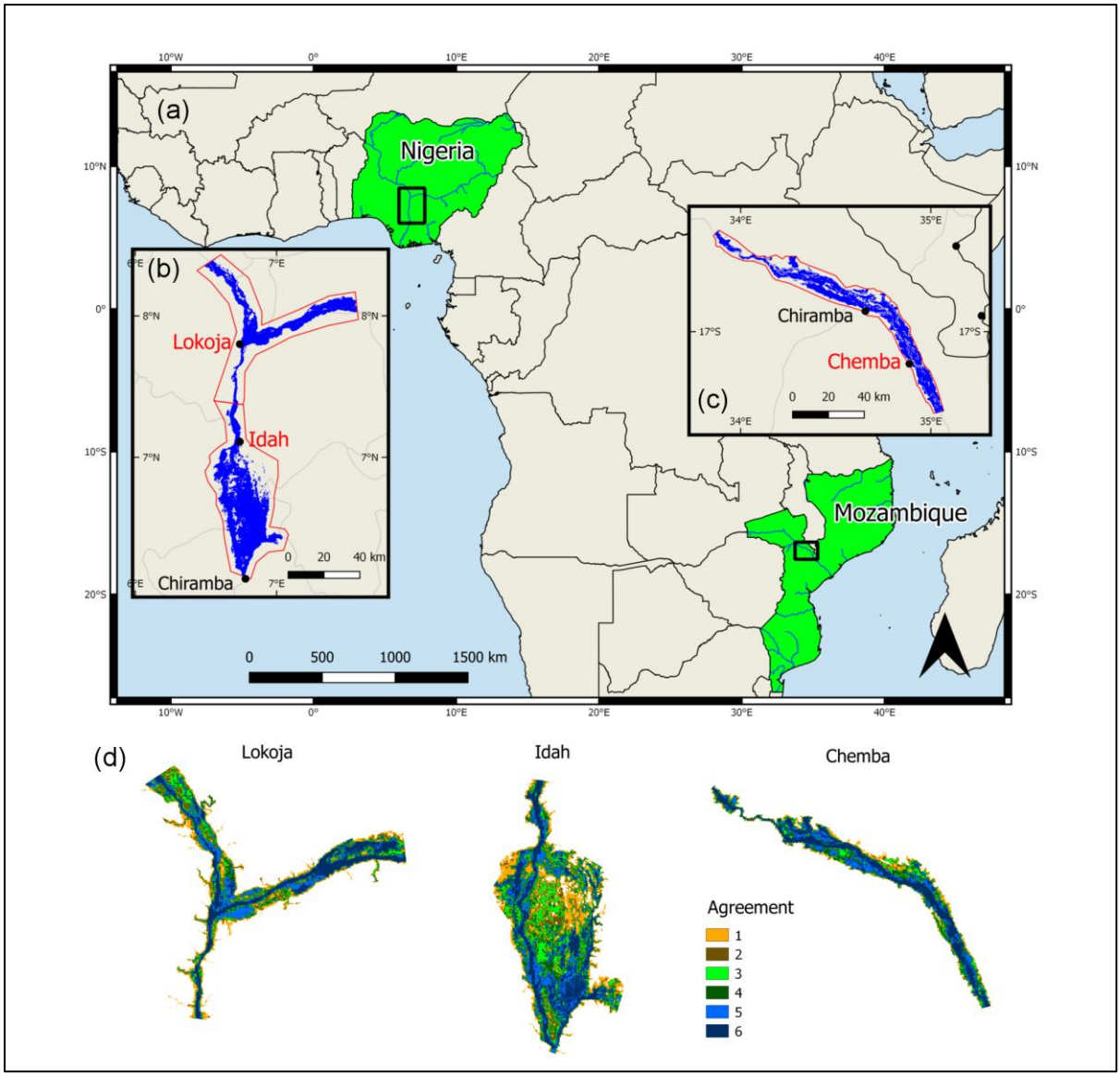

Figure 1. (a) Location of study regions in Africa. (b) Lokoja and Idah study regions with MODIS imagery of 2012 flooding [1]. (c) Chemba study region with MODIS imagery of 2007 flooding [1]. (d) GFM aggregated fluvial flood hazard (25 year return period) for each region [4].

[6]. Given current CO_2 emission trends, global temperatures could rise by up to 4º C by 2100 [7]. To put this into a river flooding context, a temperature rise of 4º C could result in 70% of the global population experiencing a 500% increase in flood risk [8]. Increased population exposure, coupled with the increased frequency and severity of flooding, means that reducing the risks associated with flooding is at the top of the agenda for the United Nations (UN). Reducing disaster vulnerability is a key target in goal 11 of the UN's Sustainable Development Goals [9]. Specific risk reduction targets, to be met by 2030, were introduced in the Sendai Framework for Disaster Risk Reduction [10].

Flood models are an integral tool for managing and reducing the risks associated with flooding. In the past decade, increased computing power and precision of remote sensing data sets has led to the development of Global Flood Models (GFMs) [11]. These GFMs are particularly useful in regions where, due to a lack of data and resources, there is no existing flood hazard information [12]. Despite

their extensive applicability, each flood model has only had limited, internal, validation against either observed events, existing regional models, or impact data [[13-20]]. The Global Flood Model Inter-comparison Project (GFMIP), undertaken by the Global Flood Partnership (https://gfp.jrc.ec.europa.eu/), was the first study to compare the output of six global flood models on the continent of Africa. Research from the GFMIP showed there was wide variation in the flood hazard output of the six GFMs [21]. The GFMIP identified the need for collective validation of the GFMs against observed flood extents.

This study is a continuation to the GFMIP, using its outputs and original GFM model output data to validate against observed flood events and expand on the testing of collective model output. It is the first study to collectively compare the performance of multiple GFMs against the same observed flood events. In doing so, it provides further insight into the reasons for model disagreement originally identified in the GFMIP [22].

Common validation should be seen as an integral stage in the science of GFMs. As the models improve and are used more extensively for disaster risk reduction, the need to compare model performance becomes increasingly apparent. GFMs differ in how they model flood extent; accordingly, different models may be better suited for different locations or uses. A common validation provides information pertinent to the potential applicability of GFMs.

Data and Methodology

The six global flood models compared in the GFMIP and in this study are CaMa-UT [19], CIMA-UNEP [16], ECMWF [15], GLOFRIS [17, 18], JRC [14], and SSBN (now Fathom Global Ltd.) [13]. The models use different techniques to predict flood extent and depth for a given return period flow. These range in complexity from 1D hydraulic modelling (CIMA-UNEP) and simple 2D flood re-distribution methods (GLOFRIS) to more complex 2D (ECMWF and CaMa-UT) and 2D hydrodynamic models (JRC and SSBN). Model output resolutions at the equator vary between ~90 m (SSBN, CIMA-UNEP), ~540 m (CaMa-UT, ECMWF), and ~900 m (GLOFRIS, JRC). Further information regarding model setup and the similarities and differences in model forcing and computational engine can be found in the GFMIP paper and the individual model literature [13-15, 17-19, 21]. The aggregated fluvial flood hazard (Figure 1d), an output of the GFMIP that shows the level of agreement in flood extent between all six models, was also validated in this study to assess the potential for using multiple model combinations for flood hazard prediction [4, 21].

Three hydraulically varied regions in Africa were chosen for validation: two in Nigeria and one in Mozambique (Figure 1). Nigeria and Mozambique were identified in the GFMIP as countries with high exposure to flooding [21]. An important factor in the choice of study regions was the size of the river. All the reaches contained rivers sufficiently large that they should be accurately represented in the

GFMs regardless of their model resolution. Smaller rivers would likely have been favourably biased towards finer resolution GFMs. In addition to this, delta regions were avoided for analysis to prevent issues associated with the demarcation of fluvial and coastal flooding, the latter of which is not currently represented in the GFMs, although recently CaMa-Flood was coupled with the results of a Global Tide and Surge Model [23] to simulate the influence of tide and surge on river levels [24]. The first region in Nigeria, referred to in this study as Lokoja, is at the confluence of the Niger and Benue rivers. It is a region with narrow, confined floodplains. The second region in Nigeria, located south of Lokoja between the cities of Idah and Onitsha, is referred to as Idah in this study. The Idah region is relatively flat and contains an extensive floodplain that has a number of smaller channels and streams within it. Downstream of the Idah floodplain is a tectonic constricted outlet. Located in central Mozambique, the final analysis region is referred to as Chemba and is located in the lower Zambezi basin, upstream of the delta. The Zambezi River in the Chemba region can be classified as anabranching (more than one channel) with a very wide valley floor trough [25].

The flood events used as the benchmark validation datasets were the floods of 2007 in Mozambique and of 2012 in Nigeria. Torrential rain between December 2006 and February 2007, coupled with the landfall of Cyclone Favio in February 2007, caused flooding in Mozambique that affected more than 130,000 people [26]. The 2012 flooding in Nigeria was even more devastating; affecting almost 4 million people [27]. The floods in Nigeria were caused by heavy rainfall between July and October 2012.

Flood imagery of both events was taken from the Dartmouth Flood Observatory (DFO) archive [1]. The DFO uses Moderate Resolution Image Spectroradiometer (MODIS) imagery to capture flood events globally, and stores them online in an open-access archive. None of the validation regions in this study are densely enough vegetated to cause underprediction of MODIS flood imagery [28]. The Chemba region is dominated by shrubbery and grasslands, and any woodland is sparse [29] and although there are forests in both regions in Nigeria, these have not detrimentally affected the observed MODIS flood imagery. For the 2012 event in Nigeria, 45 days of imagery (September 15 – October 29) were downloaded in the native vector GIS format and merged into one shapefile for analysis. Using over 6 weeks of data ensured that the entire event was captured (maximum extent) and any cloud obscuration was minimised. For the 2007 event in Mozambique, the flood extent was only available from the DFO as a JPEG file, which required georeferencing to allow for GIS analysis. The process of georeferencing the image for analysis is outlined in the Supplementary Material.

Both flood events had, very approximately, estimated return periods of around 50 years [30, 31]. The GFMIP compared the flood extent outputs of 6 different return periods: 25, 100, 250, 500, and 1000 years. To ensure that the validation results best represent the skill of the models, two return periods were tested in the individual analysis: 25 and 100 years. For the aggregated analysis only a 25 year

return period was used as it was sufficient to evaluate the usefulness of using a composite model. It should also be noted that the SSBN model had a return period of 20 years, not 25 years. Return period consistency does not need to be exact, as both modelled return periods and reported flood return periods are often based on limited historic data. Additionally, the GFMIP found that, apart from SSBN's GFM, modelled flood extent showed limited sensitivity to return period [21]. All the datasets used for validation in this study are open access, with the thought that the regions and events studied can be used for future GFM validation. The datasets are available from Research Data Leeds for academic research and education purposes (https://doi.org/10.5518/340).

The analysis in this study was done in QGIS (v2.18). Individual GFM outputs were converted from extents with pixels indicating depth of flooding to binary (wet/dry) water masks representing only flood extent. The modelled and observed extents were then overlapped in each of the study regions. The MODIS flood imagery used in this study was obtained in ~250 m resolution. To ensure continuity in the analysis, the MODIS imagery and all GFM outputs that were not previously of ~90 m resolution were resampled to ~90 m resolution. The degree of overlap between the modelled flood extents and the observed DFO MODIS extents was then calculated in terms of the number of pixels that showed flood agreement, model overprediction, and model underprediction. Maps visualizing this overlap were produced (Figure 3). The numerical data from these calculations was then used to calculate performance scores. The aggregated GFM output (Figure 1d) was extracted in 6 different model agreement levels. The extents ranged from largest to smallest: from any model agreement (≥1 models agree) to all model agreement (6 models agree). Each of the 6 model agreement levels was converted to a binary water mask and underwent the same analysis as the individual GFMs.

The performance metrics used in the analysis of the flood models are commonly used in flood model assessments and for forecast verification in the atmospheric sciences [32]. The scores were also used by a number of GFM providers for their own in-house validation [13, 14, 32-34]. The three performance scores were chosen as their results represent the most important aspects of model performance: model fit, model bias, and amount of flood captured. The first, and most comprehensive, score is the proportion correct or the Critical Success Index (CSI):

$$CSI = \frac{F_m \cap F_o}{F_m \cup F_o} \qquad (1)$$

where $F_m \cap F_o$ is the intersection of the modelled and observed flood extent, or number of correct forecasts, and $F_m \cup F_o$ is the union of modelled and observed extent. The CSI ranges from 1 (best) to 0 (worst). The CSI has been shown to favourably bias larger floods [35]. However, because the floods compared in this study have a similar return period and because model performance is being compared within the same flood, CSI was deemed appropriate. The second score, the Hit Rate (HR), measures the proportion of the observed flood that was captured by the model:

$$HR = \frac{F_m \cap F_o}{F_o} \quad (2)$$

where F_o is the total observed flood extent. The HR ranges from 1 (entire flood captured) to 0. The third score is the Bias score, which measures whether a forecast is biased towards under-prediction or over-prediction:

$$Bias = \frac{(F_m \cap F_o) + F_m}{(F_m \cap F_o) + F_o} - 1 \quad (3)$$

where F_m is the total modelled flood extent. A Bias score of 0 indicates an unbiased model. Positive and negative bias scores indicate bias towards overprediction and underprediction respectively. The further a bias score is from 0, the greater the bias in the respective direction.

The variation in flood hazard output between the GFMs identified in the GFMIP [21] raises the question of whether a composite model performs better than any individual flood model. Multiple model combinations have been used extensively in the atmospheric sciences in the form of model ensembles [36-41]. The composite model proposed in this study, although not as complex as an ensemble, aims to achieve the same goal: to reduce the uncertainty associated with individual models. The composite model will be made up of the best performing individual models determined by the following Composite Score (CS):

$$CS = Average\ CSI - |0.2 * Average\ Bias| \quad (4)$$

In order to have one common composite model output, the average of the 25 year return period performance scores across the regions was used to determine the CS. A Bias adjustment factor was added to the CS to penalise for any significant bias towards overprediction or underprediction. Excessive overprediction is especially detrimental to the composite model as it would be dominated by the model that tends towards overprediction. The number of individual models to include in the composite model was decided based on the results of the aggregated model validation.

Once the best individual models to include in the composite model had been determined, the composite model was created in QGIS (v2.18) by combining the flood extents of the individual models into one, binary, composite flood extent. The composite extent was then analysed using the same methodology as for the individual and aggregated models.

Results and Discussion

The performance scores are represented graphically in Figure 2 and the GFMs are arranged from left to right in order of resolution from coarsest to finest. The results indicate that there is a significant variation between the GFMs in their ability in modelling the flood events in each region. The average CSI of the GFMs range from 0.45 (GLOFRIS) to 0.70 (JRC) for a 25 year return period flow. To put these scores

into context, CSI scores from other flood validation literature, mostly from river reach scale models, range from 0.3-0.9 [13, 14, 34], with above 0.7 considered good and below 0.5 poor.

Lokoja stands out as the region in which almost all of the models perform best. The higher CSI scores in Lokoja are likely a reflection of the region's narrow confined floodplain, and the relative simplicity of modelling the flood where extent is not sensitive to flood flow magnitude. The increased complexity in modelling flooding in flat extensive floodplains such as the one in Idah is reflected in the lower CSI scores for the region. The overlap of the observed and modelled extents (Figure 3) illustrates the varied success of the GFMs at modelling floodplain inundation in Idah. GLOFRIS, which uses a simple flood volume distribution method for modelling inundation, had the lowest CSI score and showed large regions of underprediction within the Idah floodplain. CaMa-Flood and SSBN performed somewhat better in Idah due to the greater connectivity modelled within the floodplain by their native sub-grid models. CIMA-UNEP's 1D cross-section approach to modelling floodplains proved successful at modelling much of the central floodplain missed by GFMs as the 1D section implicitly connects low areas along the cross-section, although this can also lead to overprediction.

JRC was the best performing model averaged across the three regions. This may stem from the more complex 2D hydrodynamic flood modelling scheme at the core of the model. The only other 2D hydrodynamic model in this study is SSBN's, which performed favourably, but not as well as JRC's model. This suggests that other aspects of the flood model setup also played a role. Both JRC's model and SSBN's model have a schematic sub-grid representation of river channels, however, they are based on different parameters. JRC's sub-grid representation may have been more representative of the three regions in this study, or its input flows were more accurate, resulting in the higher performance scores across the three regions.

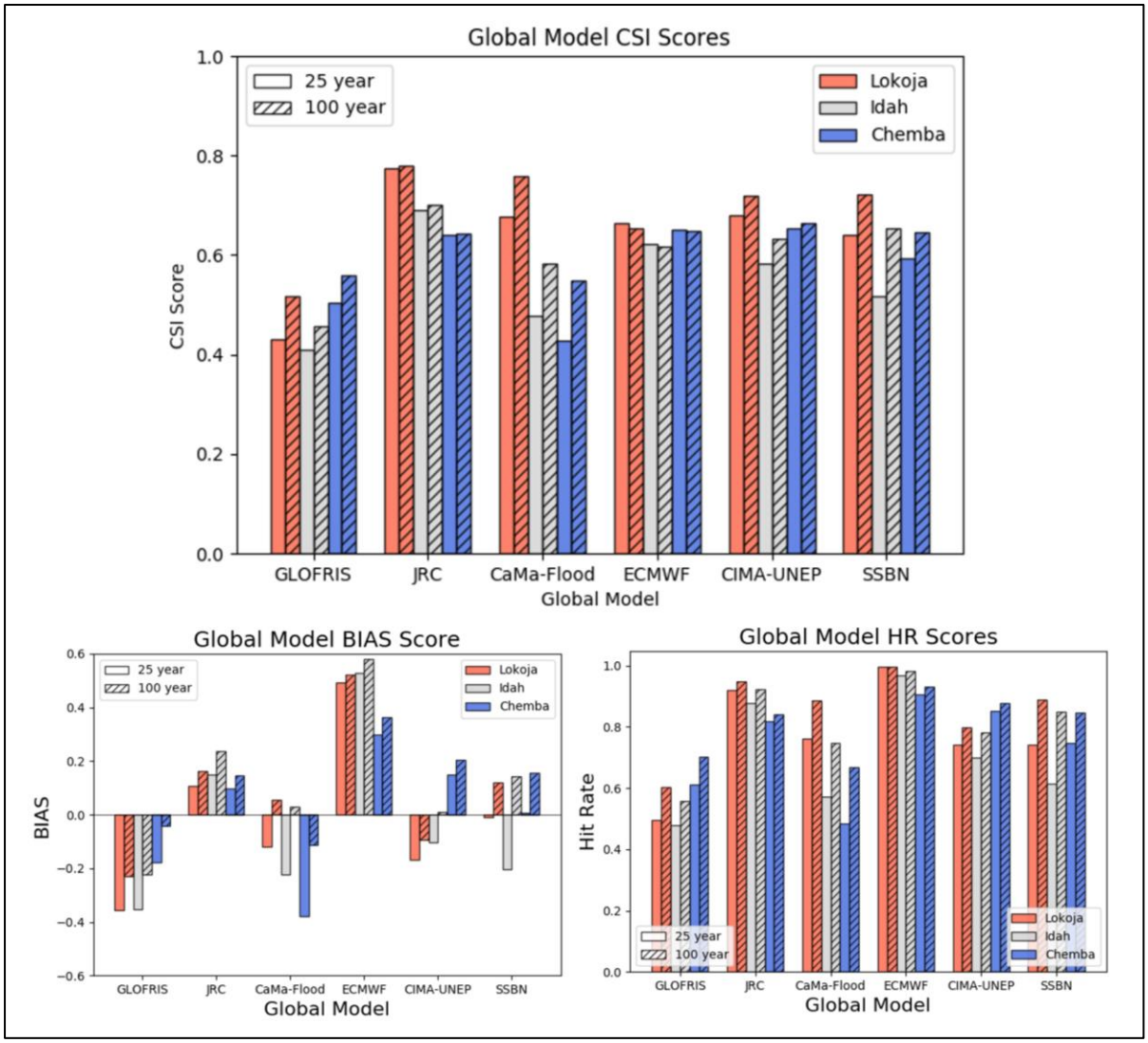

Figure 2. Individual GFM Critical Success Index, Bias, and Hit Rate performance scores when compared against observed events in Lokoja, Idah, and Chemba. Results shown for 25 and 100 year return period GFM output.

The GFM with one of the highest CSI scores in Chemba is ECMWF, whereas the GFM with the lowest CSI score in Chemba is CaMa-Flood. This highlights the importance of input flow in GFM performance: CaMa-Flood and ECMWF share the same core hydrodynamic model, but differ in their flow generation model. The performance of CaMa-Flood also significantly improves as the modelled return period is increased from 25 years to 100 years. This suggests that the input flow was the limiting factor affecting the performance scores of the 25 year return period output. Apart from ECMWF, increasing the return period from 25 years to 100 years generally increased the CSI scores of the GFMs. Increasing the GFM return period resulted in averaged CSI percentage increases of 14% (GLOFRIS), 0.1% (JRC), 5% (CIMA-UNEP), 19% (CaMa-Flood), and 15% (SSBN). These findings show that in these three study areas GLOFRIS, CaMa-Flood and SSBN are sensitive to input flow. JRC continues to perform the best of all the models at either return period when averaged across the three regions. The variation in input flow is reflected in the HR and BIAS scores of the GFMs. Averaged across all three regions, ECMWF captures almost all the flooding, with an HR of 0.96 for a 25 year flood. However, it

is also the GFM that showed the largest bias in either direction: 0.44 for the 25 year return period and 0.49 for the 100 year return period. These results suggest that ECMWF is significantly overestimating input flow at both return periods.

Although we are able to make hypotheses about GFM input flow based on our validation results we have also highlighted the limitations of validating exclusively against observed events. A more comprehensive validation that includes explicit flow comparison would allow conclusions to be drawn

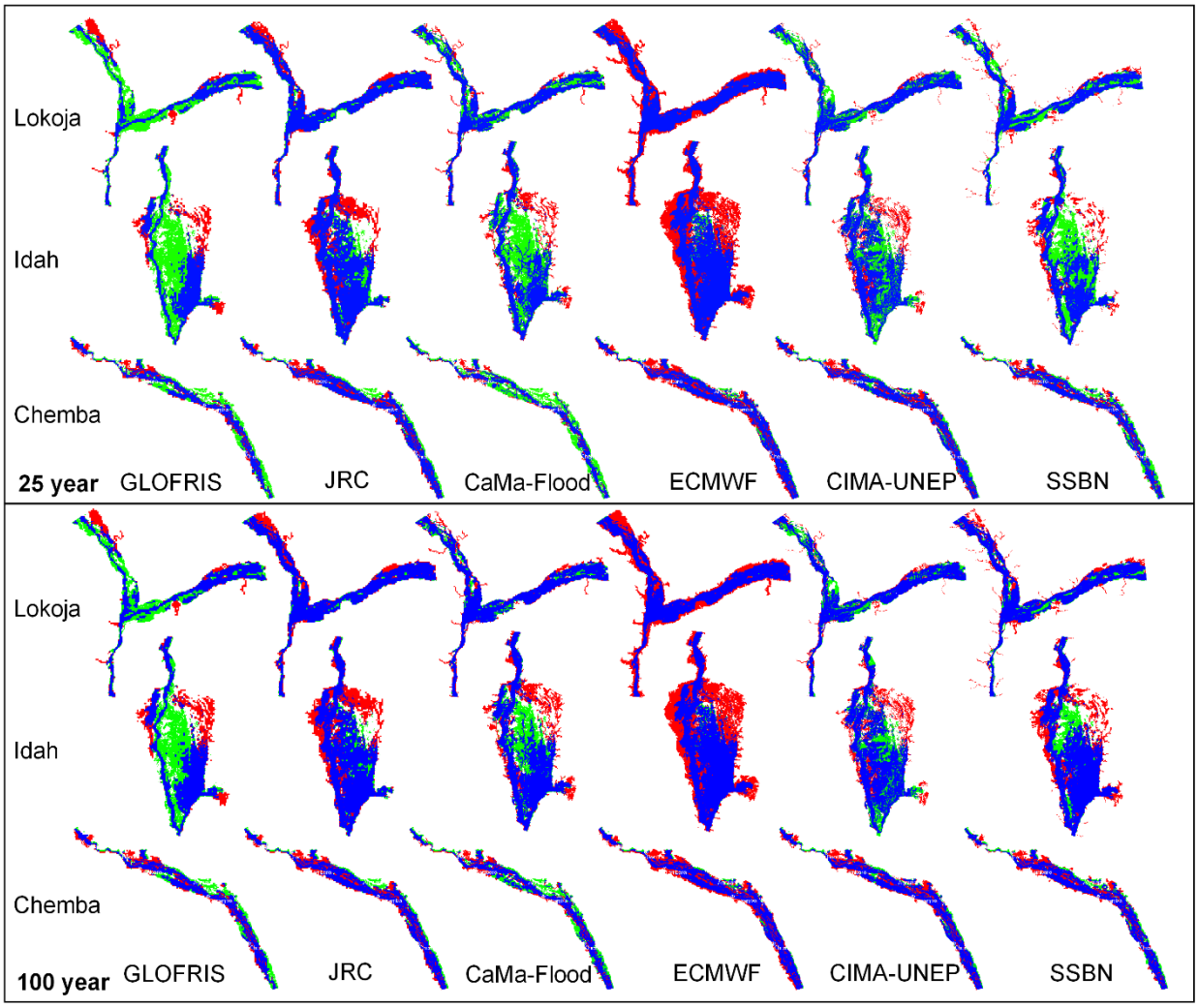

Figure 3. Overlap of individual GFM modelled extent for return period flows of 25 and 100 years and MODIS observed flood extent for Lokoja, Idah, and Chemba.

with more certainty. Additionally, validation of the GFMs across different temperate regions would allow for more conclusions to be drawn about the flow generation techniques of the GFMs. The regions in this study are limited to the tropics; expanding the study to include sub-tropical and temperate regions would give an insight into the effect that geographical zone has on GFM performance.

The improved connectivity offered by higher resolution GFMs is evident in the Idah floodplain (Figure 3). CIMA-UNEP and SSBN, both with outputs of 90 m resolution at the equator, are able to model

some of the smaller channels within the floodplain (either implicitly or explicitly). Despite the improved connectivity representation, there is no discernible correlation between the performance scores and GFM resolution, indicating that the models still need further improvements in capturing river/floodplain connectivity. At present, there is currently no well-developed method to represent channel bifurcation in 1D river models. A better representation of bifurcation would improve the performance of both 1D and 2D sub-grid models in areas of high bifurcation, such as floodplains [42]. The comparative usefulness of GFMs and regional flood models is a point of contention in flood modelling literature [12, 43]. Thomas [44] developed a regional flood risk model for southern Nigeria and validated it against the 2012 floods. The model incorporated local bathymetric and hydrographic data. The regional model's CSI scores when compared with MODIS data of the flood event were 0.73 and 0.53 for Lokoja and Idah respectively. Comparison with the best GFM performance scores show that JRC and CaMa-Flood outperform the regional model in Lokoja with CSIs of 0.78 and 0.75 respectively. The case for the GFMs is even stronger in Idah as five GFMs outperform the regional model: JRC, SSBN, CaMa-Flood, ECMWF and CIMA-UNEP with CSI scores of 0.70, 0.65, 0.58, 0.62 and 0.58 respectively. Comparison of performance scores between the studies should be approached with some caution as the analysis areas in Thomas' [44] study varied slightly compared with the ones used in this study. However, in the cases shown here, the performance of GFMs is comparable to, or better than, the performance of a locally calibrated regional model.

The performance scores of the different levels of model agreement for the 25 year return period aggregated model (Figure 4a) shows that the CSI peaks at ≥2 and ≥3 model agreement. These results correspond with the results of the individual model validation: two or three models consistently

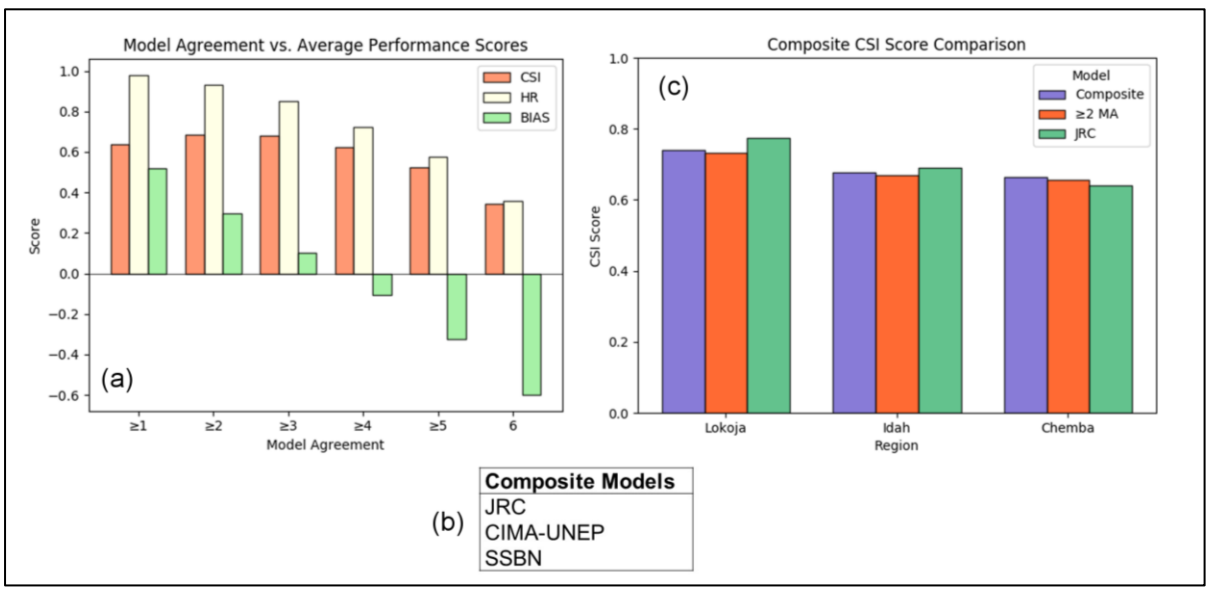

Figure 4. (a) Critical Success Index, Hit Rate, and Bias scores averaged across all three study regions for the 6 different levels of aggregated model agreement for a 25 year return period flood. (b) The three best performing individual GFMs that are included in the composite model as determined by the composite score. (c) Critical Success Index scores for the composite model, the best performing individual GFM (JRC), and the best performing aggregated model agreement level (≥2 Model Agreement) for a 25 year return period flood.

outperform the rest. A hit rate of 0.36 at 6 model agreement shows that all 6 models are correctly capturing at least 36% of the observed flood events. The bias scores trend fairly linearly from overprediction to underprediction as the model agreement level increases. The least bias in either direction occurs at ≥3 and ≥4 model agreement, this is likely due to the fact that the opposite bias of the individual models shown in Figure 2 balanced one another out.

The results of individual and aggregate model validation indicate that the optimum number of models to include in the composite model is three. The individual models included in the composite model, chosen using the composite score, are JRC, CIMA-UNEP, and SSBN (Figure 4b). The validation performance scores of the composite model are compared (Figure 4c) with the best performing models from the individual and aggregate group: JRC and ≥2 model agreement. Results shows that there is little difference between the CSI scores of the composite model, JRC, and ≥2 model agreement. Furthermore, the JRC GFM scores higher than the composite model in Lokoja and Idah. The aim of a composite model is to reduce the uncertainty associated with using a single model. For a composite model to perform better than individual models, the individual models that make up the composite model need to compensate for the uncertainty in the other models either through different input data or different modelling methods. Judging from the results of the analysis, it seems that the combination of individual models did not improve the results as a whole. If anything, they added to the uncertainty in the form of increased overprediction, which resulted in the reduced CSI scores. Although the composite model did not outperform the best individual model, it did score comparably well. Where there are regions in the world where it is not possible to validate flood models, the use of a multiple model composite could reduce the uncertainty associated with using only one model, whilst not significantly reducing the flood hazard prediction accuracy.

It would be imprudent to discuss our validation findings without making some reference to the observational data used and the inherent uncertainty that is associated with flood observation mapping. This study used extents from the DFO archive, which is currently the most extensive global flood database. However, work is being done to develop a global database of historic flood events in Google Earth Engine (GEE) [45]. The DFO flood extents used in this study and the equivalent extents from the new global database were analysed to examine the agreement between the two data sources. The results of the analysis show that there is 12% disagreement in Lokoja, 11% disagreement in Idah, and 63% disagreement in Chemba between the observed flood extents from the two data sources. Disagreement in this sense is defined as the area where only one of the two datasets observed flooding. Figures showing the observational agreement and disagreement are included in the Supplementary Material. This observational disagreement between data sources highlights an underlying problem with flood mapping. Satellite imagery, both optical and radar, faces issues with observational bias. Optical imagery is affected by cloud cover and radar imagery is affected by vegetation. Data sources differ in the methods they use to reduce the effects of such observational bias. As a result, flood maps for the same

event can differ if they are obtained from different sources. Neither source captures all of the flooding; each misses different parts. The task faced by the end user when confronted with the uncertainty associated with two disagreeing datasets is to decide which most closely represents the actual event. Even then, the chosen extent is used under the assumption that it is entirely correct. If these observational uncertainties could be incorporated into flood maps it would allow for a measure of confidence to be calculated relating to the accuracy of the observations and as a result, the accuracy of the validation findings.

Conclusion

This paper outlines the first validation intercomparison between global flood models. It also introduces a vital stage of model review that is beneficial both to model producers and model end-users. Validation against observed events in Nigeria and Mozambique show that there is a significant variation in GFM performance, with average CSI scores ranging from 0.45 to 0.7. Input flow was identified as a crucial factor in modelling a representative flood inundation extent and increasing the return period of the GFMs resulted in significant performance improvements for half of the GFMs. The underlying hydraulic models showed varying success in modelling floodplain inundation. The flow connectivity provided by 2D models was evident across the regions; although the 1D cross-section approach proved advantageous in modelling flooding in the complex morphology of the Idah floodplain. Resolution, although showing some improvement in floodplain connectivity, did not obviously improve model performance.

Comparison of the GFMs with a regional flood model developed for Nigeria showed that some of the GFMs outperformed the regional model. Through validation the three best models were identified and combined into a composite model. The validation of the composite model, showed that it performed similarly, but not better than the best individual GFM.

This study has demonstrated the usefulness of a common GFM validation procedure. The comparisons and conclusions that can be drawn from the common validation data cannot be made using the individual internal GFM validation data that has been available thus far. The focus area of this study has been limited to three regions in Africa. Going further, a more extensive validation procedure that incorporates a comparison of flow, inundated depth, and makes use of a global database of observed extents needs to be developed. In order for this to happen, the observational data used for validation should be shared openly. This would encourage and allow for common validation of GFMs to take place. The development of an open-access database of observed events [45] is an important step in the right direction. Future studies should also incorporate more global flood models such as insurance catastrophe models to encourage the knowledge transfer between research and industry. Incorporating more methods of model output validation and applying them across more regions would allow for a truly global validation comparison study of Global Flood Models.

References

[1] Brakenridge G R Global Active Archive of Large Flood Events. (Dartmouth Flood Observatory: University of Colorado)

[2] Berz G, Kron W, Loster T, Rauch E, Schimetschek J, Schmieder J, Siebert A, Smolka A and Wirtz A 2001 World map of natural hazards - A global view of the distribution and intensity of significant exposures *Nat. Hazards* **23** 443-65

[3] UNISDR C The Human Cost of Weather Related Disasters.

[4] Trigg M A, Birch C E, Neal J, Bates P, Smith A, Sampson C, Yamazaki D, Hirabayashi Y, Pappenberger F, Dutra E, Ward P J, Winsemius H C, Salamon P, Dottori F, Rudari R, Kappes M S, Simpson A, Hadzilacos G and Fewtrell T 2016 Aggregated fluvial flood hazard output for six Global Flood Models for the African Continent. (Research Data Leeds Repository

[5] Kummu M, de Moel H, Ward P J and Varis O 2011 How Close Do We Live to Water? A Global Analysis of Population Distance to Freshwater Bodies *PLoS One* **6** 13

[6] Jongman B, Ward P J and Aerts J 2012 Global exposure to river and coastal flooding: Long term trends and changes *Glob. Environ. Change-Human Policy Dimens.* **22** 823-35

[7] Sherwood S C, Bony S and Dufresne J L 2014 Spread in model climate sensitivity traced to atmospheric convective mixing *Nature* **505** 37-+

[8] Alfieri L, Bisselink B, Dottori F, Naumann G, de Roo A, Salamon P, Wyser K and Feyen L 2017 Global projections of river flood risk in a warmer world *Earth Future* **5** 171-82

[9] Nations U 2015 Transforming our World: The 2030 Agenda for Sustainable Development.

[10] Reduction) U U N I S f D 2015 Sendai Framework for Disaster Risk Reduction 2015-2030.

[11] Wood E F, Roundy J K, Troy T J, van Beek L P H, Bierkens M F P, Blyth E, de Roo A, Doll P, Ek M, Famiglietti J, Gochis D, van de Giesen N, Houser P, Jaffe P R, Kollet S, Lehner B, Lettenmaier D P, Peters-Lidard C, Sivapalan M, Sheffield J, Wade A and Whitehead P 2011 Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water *Water Resour. Res.* **47** 10

[12] Ward P J, Jongman B, Salamon P, Simpson A, Bates P, De Groeve T, Muis S, de Perez E C, Rudari R, Trigg M A and Winsemius H C 2015 Usefulness and limitations of global flood risk models *Nat. Clim. Chang.* **5** 712-5

[13] Sampson C C, Smith A M, Bates P B, Neal J C, Alfieri L and Freer J E 2015 A high-resolution global flood hazard model *Water Resour. Res.* **51** 7358-81

[14] Dottori F, Salamon P, Bianchi A, Alfieri L, Hirpa F A and Feyen L 2016 Development and evaluation of a framework for global flood hazard mapping *Adv. Water Resour.* **94** 87-102

[15] Pappenberger F, Dutra E, Wetterhall F and Cloke H L 2012 Deriving global flood hazard maps of fluvial floods through a physical model cascade *Hydrol. Earth Syst. Sci.* **16** 4143-56

[16] Rudari R, Silvestro F, Campo L, Rebora N, Boni G and Herold C 2015 IMPROVEMENT OF THE GLOBAL FLOOD MODEL FOR THE GAR 2015.

[17] Ward P J, Jongman B, Weiland F S, Bouwman A, van Beek R, Bierkens M F P, Ligtvoet W and Winsemius H C 2013 Assessing flood risk at the global scale: model setup, results, and sensitivity *Environ. Res. Lett.* **8** 10

[18] Winsemius H C, Van Beek L P H, Jongman B, Ward P J and Bouwman A 2013 A framework for global river flood risk assessments *Hydrol. Earth Syst. Sci.* **17** 1871-92

[19] Yamazaki D, Kanae S, Kim H and Oki T 2011 A physically based description of floodplain inundation dynamics in a global river routing model *Water Resour. Res.* **47** 21

[20] Ward P J, Jongman B, Aerts J, Bates P D, Botzen W J W, Loaiza A D, Hallegatte S, Kind J M, Kwadijk J, Scussolini P and Winsemius H C 2017 A global framework for future costs and benefits of river-flood protection in urban areas *Nat. Clim. Chang.* **7** 642-+

[21] Trigg M A, Birch C E, Neal J C, Bates P D, Smith A, Sampson C C, Yamazaki D, Hirabayashi Y, Pappenberger F, Dutra E, Ward P J, Winsemius H C, Salamon P, Dottori F, Rudari R, Kappes M S, Simpson A L, Hadzilacos G and Fewtrell T J 2016 The credibility challenge for global fluvial flood risk analysis *Environ. Res. Lett.* **11** 10

[22] Hoch J M, Neal J C, Baart F, van Beek R, Winsemius H C, Bates P D and Bierkens M F P 2017 GLOFRIM v1.0-A globally applicable computational framework for integrated hydrological-hydrodynamic modelling *Geosci. Model Dev.* **10** 3913-29

[23] Muis S, Verlaan M, Winsemius H C, Aerts J and Ward P J 2016 A global reanalysis of storm surges and extreme sea levels *Nat. Commun.* **7** 11

[24] Ikeuchi H, Hirabayashi Y, Yamazaki D, Muis S, Ward P J, Winsemius H C, Verlaan M and Kanae S 2017 Compound simulation of fluvial floods and storm surges in a global coupled river-coast flood model: Model development and its application to 2007 Cyclone Sidr in Bangladesh *Journal of Advances in Modeling Earth Systems* **9** 1847-62

[25] Davies B R, Beilfuss R D and Thoms M C 2000 Cahora Basin retrospective, 1974-1997: effects of flow regulation on the Lower Zambezi River *Verh. Internat. Verein. Limnol.* **27** 1-9

[26] Rana R 2007 A Review of the Mozambique Floods Response Shelter Working Group. International Federation of Red Cross and Red Crescent Societies) p 43

[27] Nigeria T F G o 2013 Nigeria: Post-Disaster Needs Assessment 2012 Floods. p 154

[28] Nigro J, Slayback D, Policelli F and Brakenridge G R 2014 NASA/DFO MODIS Near Real-Time (NRT) Global Flood Mapping Product Evaluation of Flood and Permanent Water Detection.

[29] CES 2014 Ecofarm Irrigation and Organic Sugarcane Project Mozambique, Botanical Specialist Study. (Grahamstown

[30] News B 2007 Mozambique seeks urgent flood aid. (bbc.co.uk

[31] Reuters 2012 Nigeria floods kill 363 people, displace 2.1 mln - agency. (reuters.com

[32] Wilks D 2006 *Statistical Methods in the Atmospheric Sciences* (United States of America: Elsevier)

[33] Alfieri L, Salamon P, Bianchi A, Neal J, Bates P and Feyen L 2014 Advances in pan-European flood hazard mapping *Hydrol. Process.* **28** 4067-77

[34] Wing O E J, Bates P D, Sampson C C, Smith A M, Johnson K A and Erickson T A 2017 Validation of a 30 m resolution flood hazard model of the conterminous United States *Water Resour. Res.* **53** 7968-86

[35] Stephens E, Schumann G and Bates P 2014 Problems with binary pattern measures for flood model evaluation *Hydrol. Process.* **28** 4928-37

[36] Ehrendorfer M 1997 Predicting the uncertainty of numerical weather forecasts: a review *Meteorol. Z.* **6** 147-83

[37] Leith C E 1974 THEORETICAL SKILL OF MONTE-CARLO FORECASTS *Mon. Weather Rev.* **102** 409-18

[38] Demeritt D, Cloke H, Pappenberger F, Thielen J, Bartholmes J and Ramos M 2007 Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting *Environmental Hazards* **7** 115-27

[39] Schellekens J, Dutra E, Martinez-de la Torre A, Balsamo G, van Dijk A, Weiland F S, Minvielle M, Calvet J C, Decharme B, Eisner S, Fink G, Florke M, Pessenteiner S, van Beek R, Polcher J, Beck H, Orth R, Calton B, Burke S, Dorigo W and Weedon G P 2017 A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset *Earth Syst. Sci. Data* **9** 389-413

[40] Siqueira V A, Collischonn W, Fan F M and Chou S C 2016 Ensemble flood forecasting based on operational forecasts of the regional Eta EPS in the Taquari-Antas basin *RBRH-Rev. Bras. Recur. Hidr.* **21** 16

[41] Gneiting T and Raftery A E 2005 Atmospheric science - Weather forecasting with ensemble methods *Science* **310** 248-9

[42] Mateo C M R, Yamazaki D, Kim H, Champathong A, Vaze J and Oki T 2017 Impacts of spatial resolution and representation of flow connectivity on large-scale simulation of floods *Hydrol. Earth Syst. Sci.* **21** 5143-63

[43] Bank W 2014 Understanding risk in an evolving world (Washington, D.C.: World Bank Group) p 220

[44] Thomas E-w I 2017 Application of Open-access and 3rd Party Geospatial Technology for Integrated Flood Risk Management in Data Sparse Regions of Developing Countries. In: *Faculty of Science and Technology*: Lancaster University)

[45] Tellman B, Sullivan J, Doyle C, Kettner A, Brakenridge G R, Erickson T and Slayback D 2017 A Global Geospatial Database of 5000+ Historic Flood Event Extents. In: *American Geophysical Union,* (New Orleans