https://doi.org/10.5194/hess-2021-153
Preprint. Discussion started: 30 March 2021

Hydrology and Earth System Sciences
Open Access
Discussions

Intercomparison of global reanalysis precipitation for flood risk modelling

Fergus McClean[1], Richard Dawson[1], and Chris Kilsby[1,2]

[1]School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[2]Willis Research Network, 51 Lime St., London, EC3M 7DQ, UK

Corresponding author: Fergus McClean <fergus.mcclean@newcastle.ac.uk>

Abstract

Reanalysis datasets are increasingly used to drive flood models, especially for continental and global analysis, and in areas of data scarcity. However, the consequence of this for risk estimation has not been fully explored. We investigate the impact of using four reanalysis products (ERA-5, CFSR, MERRA-2 and JRA-55) on simulations of historic flood events in Northern England. These results are compared to a benchmark national gauge-based product (CEH-GEAR1hr). All reanalysis products predicted fewer buildings would be inundated by the events than the national dataset. JRA-55 was the worst by a significant margin, underestimating by 40% compared with 14-18% for the other reanalysis products. CFSR estimated building inundation the most accurately, while ERA-5 demonstrated the lowest error in terms of river stage (29.4%) and floodplain depth (28.6%). Accuracy varied geographically and no product performed the best across all basins. Global reanalysis products provide a useful resource for flood modelling where no other data is available, but they should be used with caution. Until a more systematic international strategy for the collection of rainfall data ensures more complete global coverage of validation data, multiple reanalysis products should be used concurrently to capture the range of uncertainties.

Introduction

The primary drivers of pluvial and fluvial flooding are precipitation events. The duration, intensity and spatial extent of these events can all affect the depth and extent of any flooding caused. Therefore, the choice of precipitation data when simulating floods is critical. Inaccurate precipitation will undoubtedly lead to a spurious and potentially misleading understanding of the risk posed by a given event. This effect is further exacerbated when low-quality precipitation data is used to project risk into the future, with planning decisions being made based on the results. Unfortunately, understanding which source of precipitation is most appropriate is challenging. There is also spatial variation in the availability and quality of precipitation data. High-quality data is often collected by national or regional authorities but can be inaccessible or difficult to obtain, therefore continental or global precipitation datasets, such as reanalysis products, are a popular option despite their generally lower resolution and accuracy.

Reanalysis precipitation data has been widely used in large-scale flood risk modelling (Alfieri et al., 2013; Andreadis et al., 2017; Pappenberger, Dutra, Wetterhall, & Cloke, 2012; Schumann et al., 2013; Seyyedi, Anagnostou, Beighley, & McCollum, 2015; Winsemius, Beek, Jongman, Ward, & Bouwman, 2013; Xu, Xu, Chen, & Chen, 2016). The main advantages of reanalysis products are their vast spatiotemporal coverage and ease of access. In areas with a limited number of rain gauges that can provide high-quality observations, reanalysis products are often the best or only source of precipitation inputs for

https://doi.org/10.5194/hess-2021-153
Preprint. Discussion started: 30 March 2021

Hydrology and Earth System Sciences
Discussions

flood simulations. However, there is no guarantee that they are able to accurately represent extreme events and subsequently characterise flood risk. As a range of reanalysis datasets are available, there is also the question of which is more suitable for the application.

The influence of reanalysis data on flood risk estimates has previously been explored in part. Sampson et al. (2014) found that the loss ratio decreased by 8.5 times when using a reanalysis product (ERA-Interim) instead of a satellite rainfall product (CMORPH) in their catastrophe risk model of Dublin. Andreadis et al. (2017) compared flood models driven using an ensemble of parameters from 20CRv2 to a benchmark using observed flow boundary conditions and found that, overall, 20CRv2 only captured 15.7% of the benchmark inundated area. Mahto and Mishra (2019) used ERA-5, ERA-Interim, CFSR, JRA-55 and MERRA-2 to drive VIC and simulate monsoon season runoff in India. CFSR and JRA-55 resulted in a strong positive bias, compared to a national precipitation dataset, while MERRA-2 strongly underestimated runoff. The two ERA products showed a much less prominent positive bias. While their study represents one of the first intercomparisons of different reanalysis precipitation products for runoff modelling, it does not go as far as looking at the consequences for flood impacts. Meanwhile, Chawla and Mujumdar (2020) demonstrate a strong negative bias in flood discharge when using CFSR in the Himalayas. This is indicative of the spatial variability in accuracy inherent in reanalysis datasets, driven largely by assimilation data availability. Winsemius et al. (2013) compared flood impacts from the GLOFRIS model cascade, which uses ERA-Interim, with The OFDA/CRED International Disaster Database (EM-DAT). However, the effect of using a different source of precipitation was not assessed and therefore the impact of using reanalysis data on the cascade is unknown.

This paper extends previous studies by undertaking a systematic intercomparison of how modern reanalysis products compare when used to drive a hydrodynamic flood model. This provides important insights to inform the selection of data for flood modelling in data-sparse regions as well as a more general assessment of how well extreme rainfall events are captured in each product. To provide further context and identify the potential effects on flood risk assessments, the flood model outputs are subsequently used to estimate the number of buildings that would be inundated by each rainfall product.

Methodology

Study Area

To assess the performance of global reanalysis precipitation, more reliable gauge-based data is required as a baseline to validate against. However, the quantity and quality of gauge observations are limited across much of the globe, particularly in sparsely populated and poorer regions. Local gauge data may in fact be of lower accuracy than the large scale products if the rain gauges on the ground are of poor quality or have been influenced by human error. There is no way to check which is more correct by looking at precipitation alone and an independent source of data is required. River flow data has been used for this purpose in the past Beck et al. (2017) and presents a viable option for assessing precipitation performance in the context of flood events. To fulfil the requirements of high-quality local precipitation and river flow observations, an area of northern England, encompassing the Tyne, Tees, Eden, Wear and Lune basins (Figure 1) was selected for this study. The relatively simple flood response of these steep, surface water dominated basins, and the occurrence of recent flood events, means they provide a suitable testbed for investigating the effects of using global reanalysis products for more localised flood risk modelling.

Model Setup

The City Catchment Analysis Tool (CityCAT) (Glenis, Kutija, & Kilsby, 2018), a hydrodynamic surface water flood model, was used to simulate flooding in this study. CityCAT represents spatial rainfall fields falling directly onto uniformly gridded elevation surfaces and propagating according to the shallow water equations. The system is suitable for this study as it is able to directly capture the effects of rainfall on flood depths without requiring any intermediate steps. Model domains within the study area

https://doi.org/10.5194/hess-2021-153
Preprint. Discussion started: 30 March 2021

Hydrology and Earth System Sciences
Open Access
Discussions

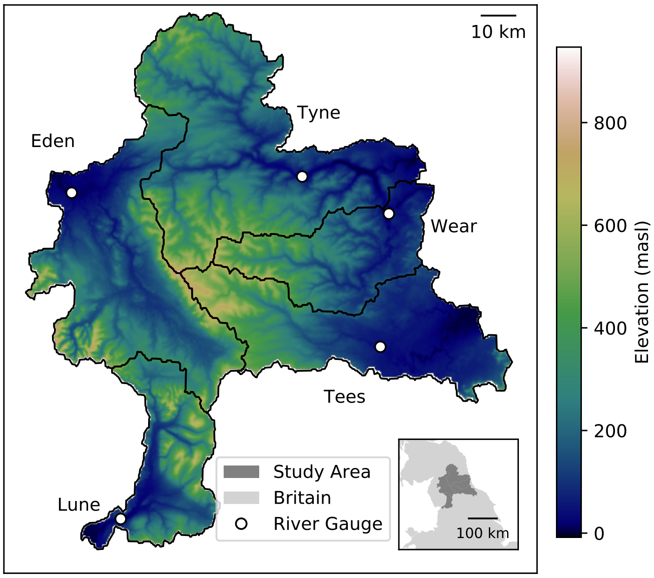

Figure 1: The location and topography of basins within the study area, illustrated using OS Terrain 50.

were delineated using HydroBASINS (Lehner & Grill, 2014). Water depths were output every hour for each grid cell within the domain and then extracted at each gauge location (Figure 1). The Manning's coefficient for all domains was uniformly defined as 0.03 (Chow, 1959) and the land surface was assumed to be impermeable given the extreme nature of the selected events (described below). As this study primarily aims to compare model results with one another, the absolute accuracy of the hydrodynamic model is not the focus and therefore the friction and infiltration configurations are not critical.

Rainfall

Four global reanalysis products (JRA-55, MERRA-2, ERA-5 and CFSR) have been selected and compared against CEH-GEAR1hr, used here as a benchmark. Each rainfall dataset is described below and key characteristics are shown in Table 1. The reanalysis products were selected based on their high spatiotemporal resolution, open availability and suitable duration. Events between the start and end dates of CEH-GEAR1hr (1990-2014) were selected based on the peak stage at the most downstream river gauge within each basin (Table 2). This identified the most extreme rainfall events, independently of the rainfall data itself. Different events were selected for each basin as the largest extremes may have occurred at different times in different areas. Each identified event was only simulated in the basin in which it was observed, to enable river gauge records to be used for validation. Simulations were commenced two weeks before the discharge peaks and ran until one week after. This was to allow model spinup and characterisation of hydrograph recession. The sensitivity to run duration was not explicitly assessed here but the duration was sufficient in all cases to ensure adequate accounting for antecedent rainfall and return to normal flow conditions . The events, according to each dataset, are mapped in Figure 2. CEH-GEAR1hr contained, on average, higher rainfall totals than the reanalysis products and JRA-55 represented only approximately half as much precipitation as other reanalysis products. Each rainfall value was converted into a rate (kg m^{-2} s^{-1}) at the observed times and each point on the grid was converted into an area with a width and height equivalent to the horizontal and vertical resolution of the dataset. These areas were then re-projected into British National Grid as cartesian coordinates

https://doi.org/10.5194/hess-2021-153
Preprint. Discussion started: 30 March 2021

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Dataset	DOI	Resolution	Coverage	Period	Frequency
CEH-GEAR1hr	10.5285/d4ddc781-25f3-423a-bba0-747cc82dc6fa	1 km	Great Britain	1990-2014	Hourly
ERA-5	10.24381/cds.adbb2d47	~30 km	Global	1979-	Hourly
MERRA-2	10.5067/7MCPBJ41Y0K6	~55 km	Global	1980-	Hourly
CFSR	10.5065/D6513W89	~35 km	Global	1979-2011	Hourly
JRA-55	10.5065/D6HH6H41	~60 km	Global	1958-	3 Hourly

Table 1: Precipitation products included in this study. Where the end of the period is not given, the product continues to be updated to the present day at the time of writing.

are required by CityCAT.

Basin	Stage Peak (m)	Peak Time	Start Time	End Time
Wear	4.1	2009-07-18 11:00:00	2009-07-04 11:00:00	2009-07-25 11:00:00
Tyne	6.3	2005-01-08 08:00:00	2004-12-25 08:00:00	2005-01-15 08:00:00
Tees	3.3	1995-01-31 20:15:00	1995-01-17 20:15:00	1995-02-07 20:15:00
Lune	7.1	1995-01-31 21:15:00	1995-01-17 21:15:00	1995-02-07 21:15:00
Eden	7.2	2005-01-08 14:30:00	2004-12-25 14:30:00	2005-01-15 14:30:00

Table 2: Event start and end times for each basin based on the observed stage peak at the most downstream gauge. Start times are two weeks before, and end times two weeks after, the observed stage peak times to allow for model spin-up and inclusion of hydrograph recession.

The Centre for Ecology and Hydrology provide an hourly version of their Gridded Estimates of Areal Rainfall dataset (CEH-GEAR1hr) (Lewis et al., 2019). This hourly product is based on a daily product which interpolates data from rain gauges using natural neighbour interpolation (Tanguy, 2019). CEH-GEAR1hr uses nearest neighbour interpolation to maintain more realistic weather patterns and unmoderated peak values. To ensure consistency between the hourly and daily versions, the daily totals were maintained in the hourly dataset by scaling the interpolated values accordingly (Lewis et al., 2018). This gauge-based dataset was used as a baseline to compare against the reanalysis products identified below.

Japanese Meteorological Agency reanalysis 55 (JRA-55) replaces JRA-25, incorporating higher resolution and better data assimilation, among other improvements (Japan Meteorological Agency, 2013; Kobayashi et al., 2015). Suzuki et al. (2017) were able to effectively simulate continental river discharge using JRA-55, however, they found large biases attributable to precipitation error in some regions.

Modern-Era Retrospective Analysis for Research and Applications 2 (MERRA-2) (Global Modeling and Assimilation Office, 2015) builds upon its predecessor, MERRA (Rienecker et al., 2011), with reduced biases in aspects of the water cycle, among other improvements (Gelaro et al., 2017). MERRA-2 uses observed precipitation products to correct the forecasts and provide better estimates (Reichle et al., 2017). Hua, Zhou, Nicholson, Chen, and Qin (2019) found that MERRA-2 was better at representing rainfall climatology over Central Equatorial Africa than ERA-Interim and JRA-55, among others.

The European Centre for Medium-Range Weather Forecasts Reanalysis 5 product (ERA-5) (ECMWF, 2018) replaces and improves on ERA-Interim (Dee et al., 2011), which stopped being produced in August 2019. It supports an increased spatial and temporal resolution, along with an updated modelling and data assimilation system, which has resulted in better representation of convective rainfall (Mahto & Mishra, 2019). The land surface component is being used to calculate river discharge for the Global Flood Awareness System (Harrigan et al., 2020). Albergel et al. (2018) found that ERA-5 resulted in better estimates of river discharge than ERA-Interim when used to drive a land surface model of the US. It has also been shown to outperform a range of other reanalysis products as part of a hydrological model applied in two Indian basins (Mahto & Mishra, 2019).

The NCEP Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010) replaces the previous NCEP/NCAR reanalysis (Kalnay et al., 1996) and uses a very similar analysis system to MERRA-2

https://doi.org/10.5194/hess-2021-153
Preprint. Discussion started: 30 March 2021

Hydrology and Earth System Sciences
Open Access
Discussions

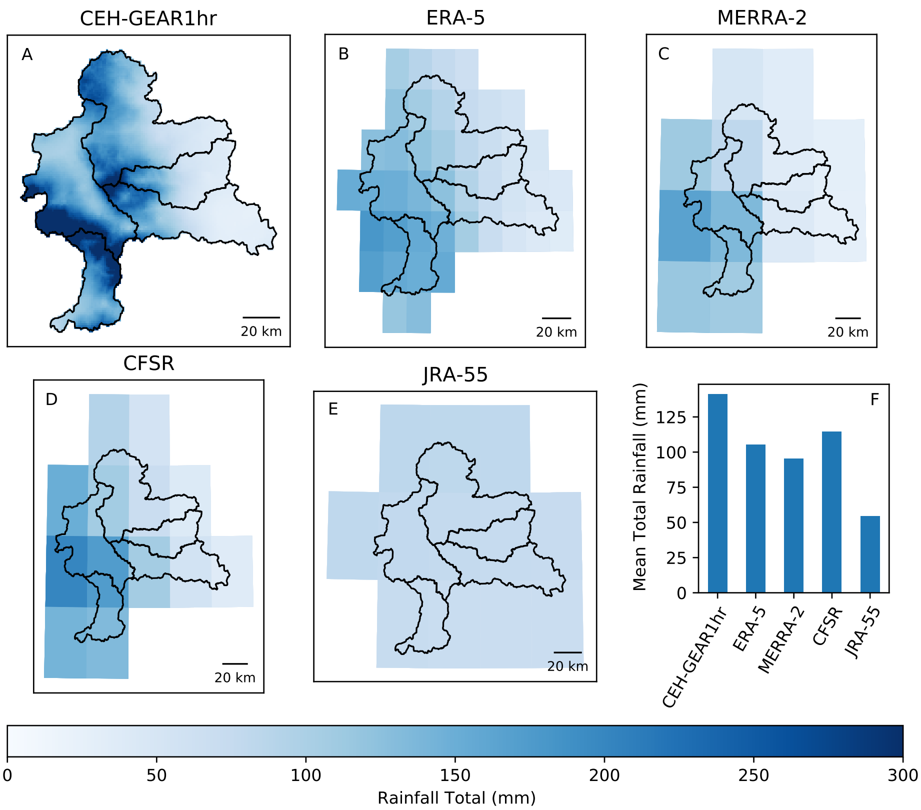

Figure 2: (A)-(E) show rainfall over the study area during the events in Table 2, according to each reanalysis dataset. (F) shows the total rainfall over the domain for all reanalysis datasets.

(Saha et al., 2010). Zhu, Xuan, Liu, and Xu (2016) demonstrated that CFSR was liable to overestimate high streamflow in two Chinese basins using SWAT and highlighted that performance varied between basins.

Digital Elevation Model

The terrain dataset used to represent the domain surface is a nationally and freely available Digital Elevation Model (DEM) product from the Ordnance Survey, known as OS Terrain 50 (Ordnance Survey, 2017). This has been shown to perform best for flood risk modelling in a comparison with other DEMs (McClean, Dawson, & Kilsby, 2020). The product is based on airborne LIDAR and is corrected using a combination of automated and manual processes to create a bare earth surface with raised structures removed. The dataset was clipped to the area of each basin and used directly within the models (Figure 1).

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Exposure

Building outlines from OS VectorMapLocal (VML) (Ordnance Survey, 2020) were used to estimate numbers of buildings inundated by each model. VML only represents individual buildings with a floor area over 20 m^2 and each polygon may represent multiple buildings. Therefore, not all buildings are included in the inundation totals. This is acceptable for this analysis which compares the relative magnitude of flooding, rather than the absolute totals. Buildings from VML were classified as flooded if they intersected any model cell above a typical property threshold of 0.3 m (Environment Agency, 2019).

Results

The performance of each simulation was compared in terms of the magnitude and timing of the hydrograph peak, the flood depth, and the number of buildings inundated (Table 3). ERA-5 outperformed other reanalysis datasets in terms of hydrograph peak error and floodplain depth, however, CFSR produced more similar inundation levels to CEH-GEAR1hr and demonstrated more accurate peak timing. JRA-55 performed significantly worse than other reanalysis products across all of these aggregated measures. The variability of each metric will now be assessed in more detail, including spatial variations in performance.

Rainfall Source	Mean Absolute Peak Error (%)	Mean Absolute Peak Time Error (hrs)	Mean Absolute Inundation Error (%)	Median Floodplain Depth Error (%)
CEH-GEAR1hr	29.4	1.0		
ERA-5	29.4	4.2	16.1	-28.6
MERRA-2	49.2	3.4	18.4	-44.4
JRA-55	70.9	162.6	39.6	-66.7
CFSR	38.0	3.4	14.4	-33.3

Table 3: Summary of performance statistics for each model. Building inundation and floodplain depth errors are relative to CEH-GEAR1hr, therefore the CEG-GEAR1hr model has no values for these measures.

The maximum water depths according to models using each of the rainfall datasets are shown in Figure 3. Overall, the spatial distribution of floodwater is similar, as the same DEM is used in all models. There are noticeably higher depths along main river channels in the CEH-GEAR1hr results. JRA-55 presents less clearly visible channels than the other models, particularly in the Lune and Tees basins. The maps also illustrate that the MERRA-2 model produced lower depths in the Tyne basin than other reanalysis precipitation datasets. Across all basins, ERA-5 and CFSR produced similar distributions of error relative to the CEH-GEAR1hr results. The inter-quartile range of errors in MERRA-2 is narrower but the median error is slightly further below zero than ERA-5 and CFSR. JRA-55 water depths were significantly further below the other reanalysis datasets.

Time series of water depths were extracted from the models at each gauge location and compared with the observed values (Figure 4). In all basins, apart from the Wear, CEH-GEAR1hr was closest to the observed peaks and predicted the highest maximum depth. In the Wear basin, where all models overestimated river stage, CEH-GEAR1hr was actually the least accurate. However, the observed values may be misleading here as flows go out of bank above 3 metres and so peaks are truncated. JRA-55 consistently severely underestimated river stage and only captured peaks in the Eden and Tyne basins. ERA-5 and CFSR display relatively similar performance across all basins. Meanwhile, MERRA-2 underestimated the peaks in the Eden and Tyne. All reanalysis products strongly underestimated the flood peak in the Lune basin.

The total numbers of inundated buildings for each model are shown in Figure 5. In four out of five basins, using CEH-GEAR1hr resulted in the highest number of inundated buildings. ERA-5 inundated the most

https://doi.org/10.5194/hess-2021-153
Preprint. Discussion started: 30 March 2021

Hydrology and Earth System Sciences
Open Access
Discussions

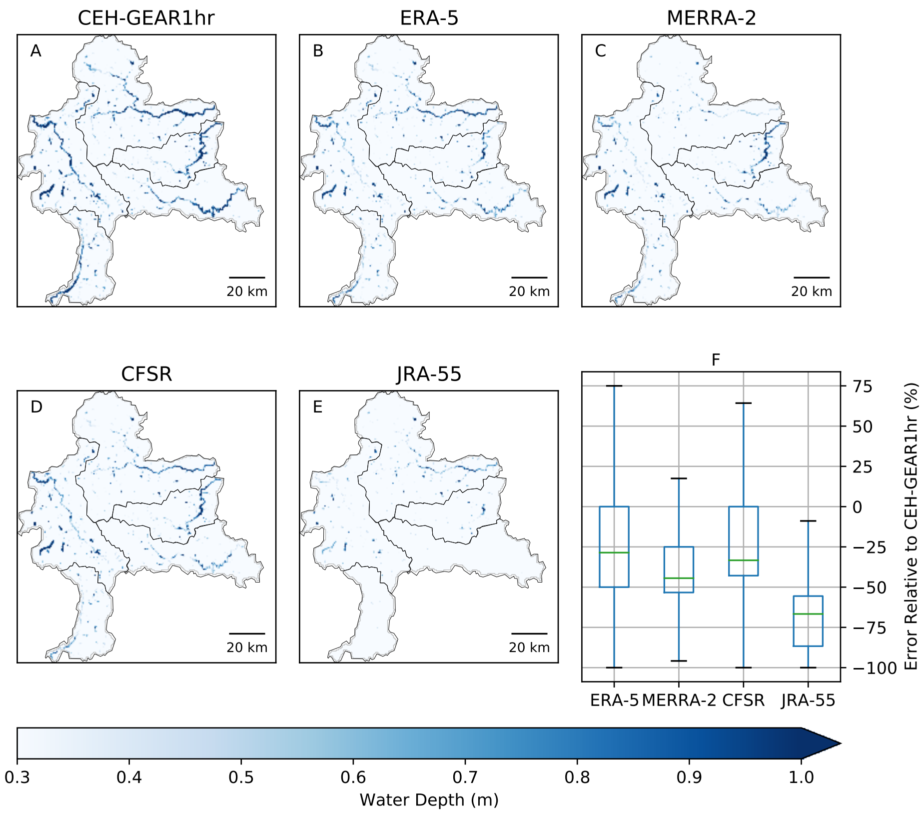

Figure 3: (A)-(E) show maximum water depth throughout the study area from models using each of the rainfall datasets. (F) shows the depth error of the reanalysis datasets relative to CEH-GEAR-1hr across all cells, excluding outliers.

buildings in the Tees basin despite not being consistently higher than the other reanalysis datasets in the other basins. JRA-55 inundated the lowest number of buildings by a large margin in all basins apart from the Tyne, where it exceeded both MERRA-2 and CFSR. CFSR never resulted in either the highest or lowest number of inundated buildings. There is general agreement between the rankings of modelled peak water depth as shown in Figure 4 and the number of inundated buildings. Notable exceptions include the Tees, where the high inundation levels predicted by ERA-5 were not replicated in its depth peak, which was lower than CEH-GEAR1hr by a clear margin.

https://doi.org/10.5194/hess-2021-153
Preprint. Discussion started: 30 March 2021

Hydrology and Earth System Sciences
Discussions

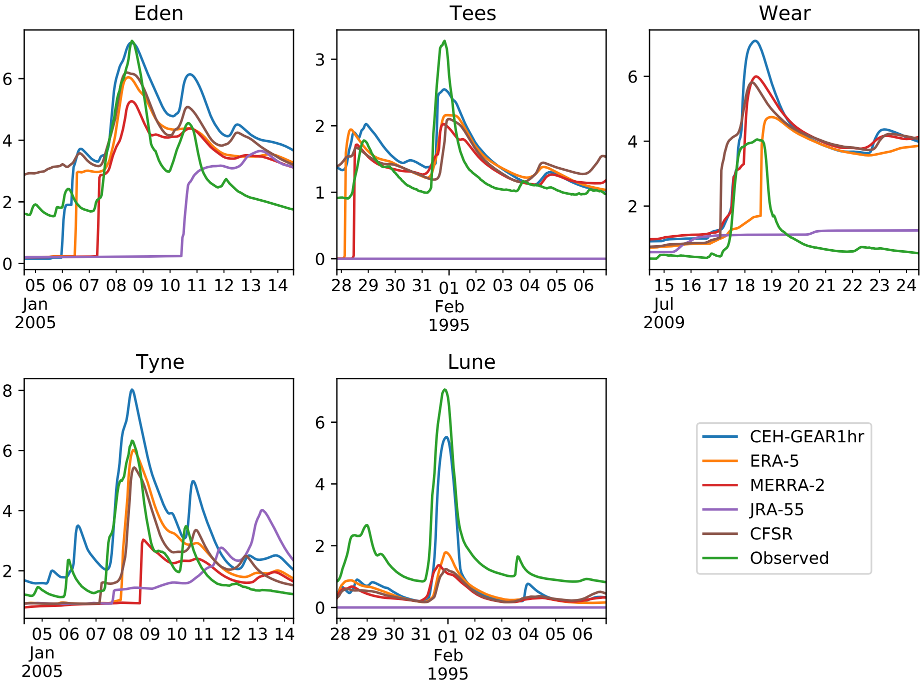

Figure 4: Stage hydrographs comparing water depths (m) from model results and observed values at each gauge.

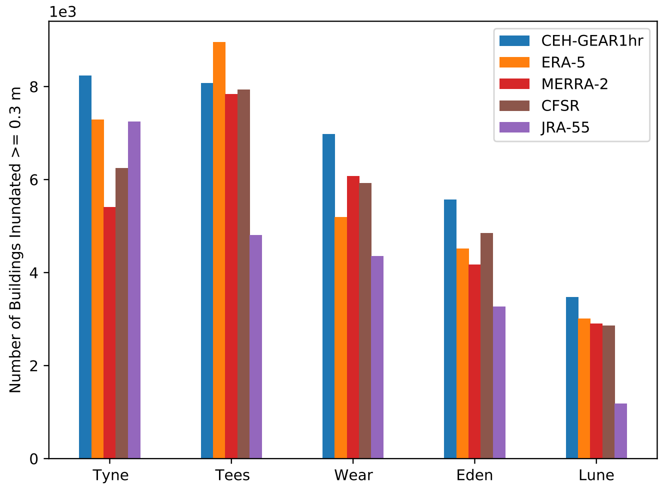

Figure 5: Number of buildings inundated above a threshold (0.3 m) per basin by each model.

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

Discussion

The results presented above demonstrate a persistent bias towards underestimation of flood depths and impacts when using global reanalysis products in place of high-resolution gauge-based rainfall datasets. The negative bias has been shown to exist in depths across the basins studied, at river gauging stations and specifically at the locations of buildings, which correspond to built-up areas exposed to flooding. This finding is in line with Sampson et al. (2014), who show ERA-Interim, an older product, underestimated flood risk. Our results, however, do not indicate such a stark bias, perhaps because the products used here are more modern and advanced than ERA-Interim. This is backed up by Towner et al. (2019) who have demonstrated improved performance of ERA-5 over ERA-Interim using hydrological models of the Amazon. Hirpa et al. (2016) also illustrate that ERA-Interim can underestimate flood risk, with spatial variability, which further reinforces our finding. In contrast, Andreadis et al. (2017) find flood extent to be overestimated (relative to a benchmark simulation) when using the 20CRv2 reanalysis product. However, they did find that outflow discharge was underestimated, which agrees with our results. Their assessment of flood extent did not include flood depths or effects on the inundation of exposed assets, as we have done here, which may explain the observed overestimation to some degree. We also did not replicate the underestimation of streamflow found by Zhu et al. (2016) when using CFSR. Though, it is difficult to draw direct comparisons given the major differences in methodology between studies.

We found that no precipitation product performed better in all models and each product performed differently depending on the basin. This implies that the optimum dataset to use depends on the location of the model. JRA-55 was very poor at capturing extreme rainfall and subsequently hydrograph peak and inundation magnitude in almost all cases. This may be slightly influenced by the lower temporal resolution, but it is unlikely that the small difference in observation frequency would result in such a strong negative effect on model performance. ERA-5 consistently performed better than other reanalysis datasets in terms of capturing the observed hydrograph peak. ERA-5, CFSR and MERRA2 were more evenly matched in terms of floodplain water depth, inundation extent, and impacts. We find no cause to favour any of these three datasets and suggest that all three could be adopted in parallel by reanalysis-based flood models to capture the range of uncertainty.

Links between hydrograph performance and estimated numbers of inundated buildings are present but the relationship is not consistent. For example, in the Tyne basin, CFSR estimates a higher gauge peak than JRA-55 but, at the same time, inundates fewer buildings. Meanwhile, MERRA-2 only has the lowest hydrograph peak in the Tyne, where it estimates the lowest total building inundation compared to other models. CEH-GEAR1hr is also both generally higher in terms of both building inundation and hydrograph peak, but the occasions where this is not the case do not correspond to the same basin. These findings demonstrate that there is generally a positive relationship between peak hydrograph depth and numbers of inundated buildings, but increased river depth does not always lead to greater inundation. Therefore, hydrograph performance is not an entirely reliable metric for assessing the accuracy of flood risk estimated using global reanalysis products.

The underestimation of inundation magnitude caused by using global precipitation data is counter to the overestimation that results from using global DEM data, as demonstrated by McClean et al. (2020). The negative inundation bias caused by using reanalysis precipitation is, however, not as strong as the positive bias from global DEM products. This is because changes in rainfall input have a less significant impact on the spatial distribution of flooding than changes in DEM input. Therefore, it is anticipated that the combined effects of using both global DEM and global reanalysis precipitation would not cancel themselves out and likely to produce a net positive bias.

Undoubtedly the effects shown here are specific to the study area and other locations may present different patterns. Each reanalysis product may behave differently across climatic regions, for example. Areas with highly constrained topography are unlikely to be strongly affected by the choice of precipitation data, in terms of flood extent and numbers of inundated assets. This is because increases in total rainfall volume will not greatly alter flood extent if there are no new available flow pathways. A key limitation to applying our methodology in new locations is the requirement for high-quality gauge-based precipitation datasets and river flow observations to compare against. Despite the caveat of locality, our results do demonstrate the potential for underestimation of flood risk when reanalysis products are involved. This underestimation has been replicated by previous studies in other areas (Sampson et al., 2014) and users of models based on reanalysis data should be aware of this effect.

Conclusions

Using precipitation from global reanalysis datasets results in an underestimation of flood risk by 14-18 % of inundated buildings and 29-44 % of median floodplain depth (Table 3, excluding JRA-55). The effect is location-specific, though, and this study found that no product performed best across all five of the catchments we studied. In some areas, the reanalysis data did result in similar levels of inundation to the national observed precipitation product (CEH-GEAR1hr). This is a positive message for the use of reanalysis data in flood risk modelling generally and future progress in forecast models will undoubtedly reduce this gap even further.

As climatic and land-use changes increase flood hazard, the importance of accurately understanding current and future flood risk is increasing. Reanalysis data has enabled flood risk assessments to be undertaken more widely. However, this analysis shows global or regional reanalysis data should not be considered as a replacement for local, high resolution, observations. Uncertainties in flood risk assessment using reanalysis data need to be properly quantified and communicated to insurers, local and national authorities and communities, to ensure flood risk management decisions are not misinformed.

While reanalysis datasets do show promising and improving results (ERA-5 achieved a mean absolute hydrograph peak error of 29.4 %, equivalent to CEH-GEAR1hr and CFSR only inundated on average 14.4 % fewer buildings than CEH-GEAR1hr), caution should be used when interpreting outputs from any models based on them. We suggest that multiple products, such as ERA-5, CFSR and MERRA-2, should be used where possible to capture the full range of uncertainty. This is because each of these products has been shown to perform better in different areas or when using different performance measures. Based on the comparatively strong negative bias in inundation and flood peak shown here, JRA-55 should not be used in flood risk modelling. However, as highlighted, certain products may perform better in other areas and further research is needed to assess new and existing reanalysis products for flood modelling across a wider range of climatic regions. To enable this, a more systematic international strategy for the collection of rainfall data is needed to ensure more complete global coverage of validation data, building on efforts from Lewis et al. (2019).

Code/Data availability

The reanalysis products can be downloaded using the DOIs in Table 1. All model results and code used to generate figures will be made available by final submission.

Author contribution

FM processed the data and executed the simulations. FM prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors were supported by the Natural Environmental Research Council (NERC) Centre for Doctoral Training for Data, Risk and Environmental Analytical Methods (DREAM), TWENTY 65: Tailored

Water Solutions for Positive Impact and Willis Research Network. The City Catchment Analysis Tool (CityCAT) was developed and provided by Vassilis Glenis. Chris Kilsby acknowledges the support of the Willis Research Network.

References

Albergel, C., Dutra, E., Munier, S., Calvet, J.-C., Munoz-Sabater, J., de Rosnay, P., & Balsamo, G. (2018, jun). ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better? *Hydrology and Earth System Sciences*, *22*(6), 3515–3532. Retrieved from `https://doi.org/10.5194%2Fhess-22-3515-2018` doi: 10.5194/hess-22-3515-2018

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013, mar). GloFAS – global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*(3), 1161–1175. Retrieved from `https://doi.org/10.5194%2Fhess-17-1161-2013` doi: 10.5194/hess-17-1161-2013

Andreadis, K. M., Schumann, G. J.-P., Stampoulis, D., Bates, P. D., Brakenridge, G. R., & Kettner, A. J. (2017, oct). Can Atmospheric Reanalysis Data Sets Be Used to Reproduce Flooding Over Large Scales? *Geophysical Research Letters*, *44*(20), 10,369–10,377. Retrieved from `https://doi.org/10.1002%2F2017gl075502` doi: 10.1002/2017gl075502

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., . . . Wood, E. F. (2017, dec). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217. Retrieved from `https://doi.org/10.5194%2Fhess-21-6201-2017` doi: 10.5194/hess-21-6201-2017

Chawla, I., & Mujumdar, P. (2020, sep). Evaluating rainfall datasets to reconstruct floods in data-sparse Himalayan region. *Journal of Hydrology*, *588*, 125090. Retrieved from `https://doi.org/10.1016%2Fj.jhydrol.2020.125090` doi: 10.1016/j.jhydrol.2020.125090

Chow, V. T. (1959). *Open-channel hydraulics* (Vol. 1). McGraw-Hill New York.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., . . . Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553-597. doi: 10.1002/qj.828

ECMWF. (2018). *ERA5 hourly data on single levels from 1979 to present.* Copernicus Climate Change Service. doi: https://doi.org/10.24381/cds.adbb2d47

Environment Agency. (2019). *What is the risk of flooding from surface water map.* Retrieved from `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment{_}data/file/842485/What-is-the-Risk-of-Flooding-from-Surface-Water-Map.pdf`

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., . . . Zhao, B. (2017, jul). The Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. Retrieved from `https://doi.org/10.1175%2Fjcli-d-16-0758.1` doi: 10.1175/jcli-d-16-0758.1

Glenis, V., Kutija, V., & Kilsby, C. (2018, nov). A fully hydrodynamic urban flood modelling system representing buildings green space and interventions. *Environmental Modelling & Software*, *109*, 272–292. Retrieved from `https://doi.org/10.1016%2Fj.envsoft.2018.07.018` doi: 10.1016/j.envsoft.2018.07.018

Global Modeling and Assimilation Office. (2015). *MERRA-2 tavg1_2d_flx_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Surface Flux Diagnostics V5.12.4.* Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center. Retrieved from `https://doi.org/10.5067/7MCPBJ41Y0K6` doi: https://doi.org/10.5067/7MCPBJ41Y0K6

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., . . . Pappenberger, F. (2020, jan). GloFAS-ERA5 operational global river discharge reanalysis 1979–present. Retrieved from `https://doi.org/10.5194%2Fessd-2019-232` doi: 10.5194/essd-2019-232

Hirpa, F. A., Salamon, P., Alfieri, L., del Pozo, J. T., Zsoter, E., & Pappenberger, F. (2016, apr). The Effect of Reference Climatology on Global Flood Forecasting. *Journal of Hydrometeorology*, *17*(4), 1131–1145. Retrieved from `https://doi.org/10.1175%2Fjhm-d-15-0044.1` doi: 10.1175/jhm-d-15-0044.1

Hua, W., Zhou, L., Nicholson, S. E., Chen, H., & Qin, M. (2019, aug). Correction to: Assessing reanalysis data for understanding rainfall climatology and variability over Central Equatorial Africa. *Climate Dynamics*, *53*(7-8), 5139–5139. Retrieved from https://doi.org/10.1007%2Fs00382-019-04918-7 doi: 10.1007/s00382-019-04918-7

Japan Meteorological Agency. (2013). *JRA-55: Japanese 55-year Reanalysis, Daily 3-Hourly and 6-Hourly Data.* Boulder CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. doi: 10.5065/D6HH6H41

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., … Joseph, D. (1996, mar). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. Retrieved from https://doi.org/10.1175%2F1520-0477%281996%29077%3C0437%3Atnyrp%3E2.0.co%3B2 doi: 10.1175/1520-0477(1996)077⟨0437:tnyrp⟩2.0.co;2

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., … Takahashi, K. (2015). The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5–48. Retrieved from https://doi.org/10.2151%2Fjmsj.2015-001 doi: 10.2151/jmsj.2015-001

Lehner, B., & Grill, G. (2014). *HydroBASINS: Global watershed boundaries and sub-basin delineations derived from HydroSHEDS data at 15 second resolution—Technical documentation version 1. c* (Vol. 27; Tech. Rep. No. 2013). Retrieved from https://www.hydrosheds.org/images/inpages/HydroBASINS{_}TechDoc{_}v1c.pdf

Lewis, E., Fowler, H., Alexander, L., Dunn, R., McClean, F., Barbero, R., … Blenkinsop, S. (2019, jul). GSDR: A Global Sub-Daily Rainfall Dataset. *Journal of Climate*, *32*(15), 4715–4729. Retrieved from https://doi.org/10.1175%2Fjcli-d-18-0143.1 doi: 10.1175/jcli-d-18-0143.1

Lewis, E., Quinn, N., Blenkinsop, S., Fowler, H. J., Freer, J., Tanguy, M., … Woods, R. (2018, sep). A rule based quality control method for hourly rainfall data and a 1km resolution gridded hourly rainfall dataset for Great Britain: CEH-GEAR1hr. *Journal of Hydrology*, *564*, 930–943. Retrieved from https://doi.org/10.1016%2Fj.jhydrol.2018.07.034 doi: 10.1016/j.jhydrol.2018.07.034

Mahto, S. S., & Mishra, V. (2019, aug). Does ERA-5 Outperform Other Reanalysis Products for Hydrologic Applications in India? *Journal of Geophysical Research: Atmospheres*, *124*(16), 9423–9441. Retrieved from https://onlinelibrary.wiley.com/doi/10.1029/2019JD031155 doi: 10.1029/2019JD031155

McClean, F., Dawson, R., & Kilsby, C. (2020, oct). Implications of Using Global Digital Elevation Models for Flood Risk Analysis in Cities. *Water Resources Research*, *56*(10). Retrieved from https://doi.org/10.1029%2F2020wr028241 doi: 10.1029/2020wr028241

Ordnance Survey. (2017). *OS Terrain 50 User Guide v1.3* (Tech. Rep.). Retrieved from https://www.ordnancesurvey.co.uk/documents/os-terrain-50-user-guide.pdf

Ordnance Survey. (2020). OS VectorMap local technical specification. Retrieved from accessedon=Monday April 13 2020{https://www.ordnancesurvey.co.uk/documents/os-vectormap-local-technical-specification.pdf}accessedon={MondayApril132020}

Pappenberger, F., Dutra, E., Wetterhall, F., & Cloke, H. L. (2012, nov). Deriving global flood hazard maps of fluvial floods through a physical model cascade. *Hydrology and Earth System Sciences*, *16*(11), 4143–4156. Retrieved from https://doi.org/10.5194%2Fhess-16-4143-2012 doi: 10.5194/hess-16-4143-2012

Reichle, R. H., Liu, Q., Koster, R. D., Draper, C. S., Mahanama, S. P. P., & Partyka, G. S. (2017, feb). Land Surface Precipitation in MERRA-2. *Journal of Climate*, *30*(5), 1643–1664. Retrieved from https://doi.org/10.1175%2Fjcli-d-16-0570.1 doi: 10.1175/jcli-d-16-0570.1

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., … Woollen, J. (2011, jul). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, *24*(14), 3624–3648. Retrieved from https://doi.org/10.1175%2Fjcli-d-11-00015.1 doi: 10.1175/jcli-d-11-00015.1

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., … Goldberg, M. (2010, aug). The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1058. Retrieved from https://doi.org/10.1175%2F2010bams3001.1 doi: 10.1175/2010bams3001.1

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., … Goldberg, M. (2010). *NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010.* Boulder CO: Research Data Archive at the National Center for Atmospheric

Research, Computational and Information Systems Laboratory. Retrieved from https://doi.org/10.5065/D6513W89

Sampson, C. C., Fewtrell, T. J., O'Loughlin, F., Pappenberger, F., Bates, P. B., Freer, J. E., & Cloke, H. L. (2014, jun). The impact of uncertain precipitation data on insurance loss estimates using a flood catastrophe model. *Hydrology and Earth System Sciences*, *18*(6), 2305–2324. Retrieved from https://doi.org/10.5194%2Fhess-18-2305-2014 doi: 10.5194/hess-18-2305-2014

Schumann, G. J.-P., Neal, J. C., Voisin, N., Andreadis, K. M., Pappenberger, F., Phanthuwongpakdee, N., ... Bates, P. D. (2013, oct). A first large-scale flood inundation forecasting model. *Water Resources Research*, *49*(10), 6248–6257. Retrieved from https://doi.org/10.1002%2Fwrcr.20521 doi: 10.1002/wrcr.20521

Seyyedi, H., Anagnostou, E. N., Beighley, E., & McCollum, J. (2015, oct). Hydrologic evaluation of satellite and reanalysis precipitation datasets over a mid-latitude basin. *Atmospheric Research*, *164-165*, 37–48. Retrieved from https://doi.org/10.1016%2Fj.atmosres.2015.03.019 doi: 10.1016/j.atmosres.2015.03.019

Suzuki, T., Yamazaki, D., Tsujino, H., Komuro, Y., Nakano, H., & Urakawa, S. (2017, dec). A dataset of continental river discharge based on JRA-55 for use in a global ocean circulation model. *Journal of Oceanography*, *74*(4), 421–429. Retrieved from https://doi.org/10.1007%2Fs10872-017-0458-5 doi: 10.1007/s10872-017-0458-5

Tanguy, H.-I. D. V., M.;Dixon. (2019). Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890-2017) [CEH-GEAR]. doi: 10.5285/ee9ab43d-a4fe-4e73-afd5-cd4fc4c82556

Towner, J., Cloke, H. L., Zsoter, E., Flamig, Z., Hoch, J. M., Bazo, J., ... Stephens, E. M. (2019, jul). Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. *Hydrology and Earth System Sciences*, *23*(7), 3057–3080. Retrieved from https://doi.org/10.5194%2Fhess-23-3057-2019 doi: 10.5194/hess-23-3057-2019

Winsemius, H. C., Beek, L. P. H. V., Jongman, B., Ward, P. J., & Bouwman, A. (2013, may). A framework for global river flood risk assessments. *Hydrology and Earth System Sciences*, *17*(5), 1871–1892. Retrieved from https://doi.org/10.5194%2Fhess-17-1871-2013 doi: 10.5194/hess-17-1871-2013

Xu, H., Xu, C.-Y., Chen, S., & Chen, H. (2016, nov). Similarity and difference of global reanalysis datasets (WFD and APHRODITE) in driving lumped and distributed hydrological models in a humid region of China. *Journal of Hydrology*, *542*, 343–356. Retrieved from https://doi.org/10.1016%2Fj.jhydrol.2016.09.011 doi: 10.1016/j.jhydrol.2016.09.011

Zhu, Q., Xuan, W., Liu, L., & Xu, Y.-P. (2016, apr). Evaluation and hydrological application of precipitation estimates derived from PERSIANN-CDR TRMM 3B42V7 and NCEP-CFSR over humid regions in China. *Hydrological Processes*, *30*(17), 3061–3083. Retrieved from https://doi.org/10.1002%2Fhyp.10846 doi: 10.1002/hyp.10846