Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019

Natural Hazards and Earth System Sciences
Discussions

A Comprehensive Evaluation of the National Water Model (NWM) - Height Above Nearest Drainage (HAND) Flood Mapping Methodology

J. Michael Johnson[1]*, Dinuke Munasinghe[2], Damilola Eyelade[1], Sagy Cohen[2]

[1]University of California, Santa Barbara, Santa Barbara, 93106, USA
[2]University of Alabama, Tuscaloosa, AL 35487

Correspondence to: J. Michael Johnson (jmj00@ucsb.edu)

Abstract. Flood maps are needed for emergency response, research, and planning. The Height Above Nearest Drainage (HAND) technique is a low-complexity, terrain-based approach for inundation mapping from elevation, discharge-height relationship, and streamflow data. The recent operational capacities of the NOAA National Water Model (NWM) and pre-processed HAND products from the University of Texas offer an operational framework for real-time and forecast flood mapping across the United States. This paper offers the first detailed evaluation of the coupled NWM-HAND approach by comparing NWM-HAND simulations to a database of remotely sensed flood maps. These comparisons are made both quantitatively and through a detailed evaluation of selected results. The results show that NWM-HAND performs well in the majority of cases with 70% of the calculated flood maps yielding an agreement of 80% or better with the remote sensing observations. Persistent inaccuracies occur in lower-order reaches, regions of low relief, and when the NWM misses substantially either in magnitude or timing.

1. Introduction

Floods are one of the deadliest, recurring natural hazards in the United States. Over the last three decades, floods have killed, on average, 86 people annually. In the last 10 years, that average has increased to 95, and between 2015 and 2017 has increased to more than 100 deaths (Lam, 2018). Moreover, these floods are increasingly occurring inland. In the last decade, 8 of the 10 US states with the largest flooding disasters were landlocked (Lightbody and Tompkins, 2018). Additionally, as climate and land use pressures amplify, floods are expected to increase in frequency and magnitude (Hirabayashi et al., 2013; Yin et al., 2018). While flood damage cannot be eliminated it can be mitigated. This is particularly true when populations are given advanced warning, have faith in the forecast accuracy, and are provided actionable intelligence (Johnson et al., 2018). Moreover, when emergency responders are given maps showing where water is, and where it might be in the future, better choices can be made that systematically save time, energy, and resources (National Research Council, 2009). Combined, these statistics highlight a need to better understand and forecast flood events across the continental USA (CONUS).

Flood maps are commonly used to help stakeholders and the general public better understand and manage flood risk (Henstra et al., 2019; Merz et al., 2007). Most often these are static maps

representing a pre-set risk magnitude (e.g. FEMA 100-year flood zones) (Maidment, 2009). Maps that predict specific events in real-time, are becoming more relevant as computational power increases and forecast domains broaden. While both types of maps require high levels of accuracy, the latter is able to sacrifice some accuracy for gains in forecast domain size and speed of delivery.

Since 2015, efforts towards producing real-time and forecast inundation forecasts for CONUS have resulted in the Height Above Nearest Drainage (HAND) layers for the CONUS (Liu et al., 2018). The methodology for coupling streamflow predictions from the National Water Model (NWM) and HAND was initiated as part of the National Flood Interoperability Experiment (NFIE) (Maidment, 2016) with the aim of better connecting national flood forecasting with local emergency response through a continuous, nationally consistent, and computationally cheap method. The methodology has since been enhanced and added to the National Water Center (NWC) operational US flood forecasting framework.

In this paper we conduct the first extensive evaluation of the NWM-HAND methodology by comparing simulated events against an observational dataset. Through this process we identify areas where the methodology performs well, and where it could be improved. An understanding of these spatial properties is critical to achieving reasonable continental scale inundation forecasts and closing the gap between national flood prediction and local response.

2. Background

To assess the accuracy of the NWM-HAND methodology calculated ## flood maps by coupling the CONUS 10m HAND product and the NWM reanalysis streamflow forecasts. These maps were compared to an observational remote sensing-based flood map repository, the US Flood Inundation Map Repository (USFIMR). Each of these products is described below.

2.1 Height-Above-Nearest-Drainage (HAND) Method

As the name implies, a Height Above Nearest Drainage (HAND) map contains the vertical distance between a location (grid-cell) and its nearest stream reach. Producing a HAND map requires a rasterized terrain dataset and a spatial representation of a region's river network to define local drainages. In the USA these primarily come from the USGS National Elevation and Hydrography Datasets respectively (NED and NHD). Each polyline in NHD is described as a 'reach' and assigned a unique common identifier (COMID). To calculate the 'nearest drainage', the contributing catchment for each reach is rasterized to the spatial resolution of the DEM in use. To illustrate this, a generic reach catchment (reach 101) is shown in Figure 1A. In this example, all cells in the catchment raster inherit the COMID (101) associated with the draining reach (Figure 1B).

To generate a HAND raster, a river network is used to determine in-stream and out-of-stream cells. For all out-of-stream cells, the relief (change in elevation) between that cell and the nearest in-stream cell is calculated (become the HAND value; Figure 1C).

Through recent advances, HAND raster's can be used to derive hydraulic geometries for each NHD reach (Zheng et al., 2018). When paired with an estimated channel length, slope, and roughness, these geometries can be used in Manning's equation to generate a streamflow-stage rating curve.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019

Natural Hazards and Earth System Sciences
Discussions

Channel length and average slope for each reach are attributes of the NHD, while a constant channel roughness of 0.05 has been commonly used for US applications (Johnson and Cole, 2017). Once developed, rating curves can be used to convert streamflow volumes (discharge) to a water levels (stage) (Reilly et al., 1987). When a region includes more than a single reach, a reclassification table storing the water level-COMID relation needs to be created (Figure 1E). Such a table can then be used to reclassify the catchment raster creating a water level raster. Subtracting a HAND raster from a water-level raster yields the cell-level water height in the catchment (Figure 1G). Any value greater than zero is classified as flood water, while values less than zero indicate dry cells (Figure 1H).

In 2017, HAND and catchment raster's, along with synthetic rating curves (SRCs) were generated for CONUS using the 10-meter NED and NHD datasets on the ROGER supercomputing system at the University of Illinois Urbana Champaign (Liu et al., 2018; Zheng et al., 2017) All products are catalogued by HUC6 (the basin level units in the WBD) on the UT Corral Server (https://web.corral.tacc.utexas.edu/nfiedata/).

2.2 National Water Model (NWM)

In August 2016, the NWM version 1.0 was made operational, expanding the Nation's forecasting domain from approximately 9,000 river forecasting locations to 2.7 million locations corresponding to the reaches in the NHD. The model generates hourly streamflow forecasts for the CONUS and high-resolution (1-km) forecasts of soil moisture, runoff, snow water equivalent, and other water balance states. The NWM also serves as a cornerstone of the new NOAA Water Initiative to provide integrated predictive capabilities that promote resilience and mitigation of water risks (Cline and Maidment, n.d.).

The foundation of the NWM is the WRF-Hydro modeling architecture supported by the National Center for Atmospheric Research (NCAR) (D. J. Gochis et al., 2018). The NWM routes water over the NHD, producing streamflow values at the end point of each reach and indexed according to the respective COMID (Figure1A – red dot). The NWM runs in four configurations including analysis and assimilation, short range, medium range, and long range (NOAA, 2016a; 2016b; Salas et al., 2017). The analysis and assimilation output provide an hourly snapshot of current hydrologic conditions out to three hours, while the short-range configuration produces hourly deterministic streamflow and hydrologic state forecasts out to 18 hours. The medium range configuration produces 3-hourly forecasts out to 10 days; and the long-Range configuration generates 30-day, four-member forecast ensembles (NOAA, 2016a). With respect to flooding, the most valuable models for prediction and tracking are the hourly short-range and 6-hour medium-range products while the analysis and assimilation provide the most accurate information for real-time inundation mapping.

NWM products are provided to a number of official National Weather Service forecast outlets including the River Forecast Centers (RFCs), Weather Forecast Offices, the National Centers for Environmental Computing and the NWC (D. Gochis et al., 2016). Operational data is publicly available for a 48-hour rolling window on the NOAA NOMAD server and can be accessed directly or through a number of user-contributed packages and end-points (Johnson, 2018a).

In addition to the operational products, reanalysis studies have been run for NWM versions 1.0 and 1.2. Each providing analysis and assimilation output from January 1993 to December 31,

2016/2017 respectively. Version 1.2 is available through Amazon Web Services (https://registry.opendata.aws/nwm-archive/).

2.3 U.S. Flood Inundation Map Repository (USFIMR)

Following NFIE, academic partners at the University of Alabama developed the USFIMR to provide high-resolution inundation maps for past U.S. flood events. These maps were derived by image classification techniques from a number of satellite sensors (e.g. Landsat, Sentinel-1, 2) with some ground proofing based on secondary sources (e.g. news reports, social media). Such maps are useful for model calibration and flood susceptibility assessment (Cohen et al., 2018, (Munasinghe et al., 2018; Zhang et al., 2018). Additionally, USFIMR provides download links for the surrounding 10-meter DEM (from the National Elevation Dataset (NED)) and a link to the nearest USGS river gage site through an open and free web portal (http://sdml.ua.edu/usfimr).

3. Methods

A number of studies have shown that the HAND methodology can produce accurate inundation maps from a known input streamflow (Afshari et al., 2018; Rennó et al., 2008; Rodda, 2005; Zheng et al., 2018). There has not been, however, an exploration of how errors in the methodology and its individual components may compound and propagate through the coupled NWM-HAND implementation. Possible errors could originate in the regional terrain profiles (e.g. low relief/homogeneous terrain) (Bernhofen et al., 2018), synthetic rating curve (SRC) generation (over-generalized width, slope and roughness), or in the NWM forecasts (magnitude and/or timing). To fill this gap, a catalogue of 30 USFIMR inundation maps (Table 1) were compared to NWM-HAND inundation maps generated using the publicly available HAND products and the retrospective NWM streamflow rates.

3.1 Flood map calculation

For each USFIMR map, the AOI (Johnson, 2018a) and HydroData (Johnson, 2018b) R packages were used to determine the minimum bounding box of the flood to facilitate the subset of the NHD. A list of COMIDs was extracted from the NHD attributes in addition to unique HUC6 units. For each HUC6, the HAND, catchment and rating curve products were downloaded cropped and, when necessary, mosaicked. The timestamp of each USFIMR satellite image was used to query the appropriate NWM output (NetCDF) file and subset it to the list of COMIDs using the NWM R package (Johnson, 2018c). Streamflow values were converted to stage using the COMID-specific rating curves, and a water level raster was generated by reclassifying the HAND catchment raster. The HAND raster was then subtracted from the water level raster yielding a raster of inundation depths. All values less than 0 were set to zero (dry cells) and, since this study is interested in the binary Flood/Not Flooded zones, all cells with value greater than zero were set to one (flooded cells). This process was repeated for each USFIMR flood map and the process is formalized in the Flood Mapping R package (Johnson, 2019) .

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019

Natural Hazards and Earth System Sciences Discussions

3.2 Areal Comparisons

With a complete set of NWM-HAND and USFIRM maps, areal agreement can be determined for each case study. Since the effective area of comparison is limited by the USFIMR extent, the concave hull was used to crop our NWM-HAND rasters. Perennial NHD water bodies were also set to NULL to reduce ambiguities from the comparison. NWM-HAND rasters were reclassified to 2 (flooded) and 0 (non-flooded) while observed rasters remained classified as 1 (flooded) and 0 (non-flooded). This convention allows observed and calculated floods to be added together to reveal areas where there is: no flood present in either observed or calculated maps (0); flood present in observed but not calculated maps (1); flood present in calculated but not observed maps (2); and flood present both in observed and calculated maps (3) (Figure 2).

To quantify overall agreement between these 4 classes, the Advanced Fitness Index (AFI) (adaptation of the 'fitness index' (Bates and De Roo, 2000)) was used as an aerial statistic (Munasinghe et al., 2018; Zhang et al., 2018) and can be calculated as:

$$AFI = \frac{Fpred \cap Fobs + NFpred \cap NFobs}{A} \times 100 \quad (1)$$

Where *Fpred* is the predicted flood; *NFpred* is no predicted flood; *Fobs* is regions of observed flood; *NFobs* is regions of no observed flood; and *A* is the Area of total extent.

AFI values were categorized (based on Bates and De Roo, 2000; Horrit, 2000; Mason et al., 2009), to determine thresholds. They include Poor: 0-40%; Fair: 40-60%; Moderate: 60-80%; Good: 80-90%; Excellent: 90-100%.

4. Results

Seventy (70) percent of flood maps (21 out of the 30) were classified as either 'good' or 'excellent' (Figure 3). This shows that NWM-HAND can be used to quite confidently to generate flood maps using the CONUS-scale HAND product and the NWM predictions. A notable anomaly in our comparisons was the Brazos River flooding near Hampstead TX (FloodID 6). This event featured a combination of two issues that are found to be consistent under some circumstances: poor horizontal water volume control in flat terrain, and poor flood mapping calculation in 1st and 2nd order stream reaches. All floods with 'fair' or 'moderate' agreement suffered from issues including these along with timing/magnitude issues in the NWM forecast, and remote sensing image classification challenges. Below we discuss the underlying factors that led to over and under prediction. While we manually examined all the case studies, we focus here on instances that represent consistent patterns.

4.1 Errors related to the HAND methodology

The primary focus of this paper is identifying how HAND-NWM performs and if there are systematic biases in the way it performs. Since HAND products are static, recyclable products that interface with the temporal NWM predictions, there is a greater priority in ensuring that HAND performs as expected when given accurate streamflow values. So, while there will always be uncertainties in NWM forecasts,

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019


Natural Hazards and Earth System Sciences Discussions

this first discussion is principally concerned with patterns that emerge when HAND is provided with accurate streamflow values. In our evaluations, three main type of cases emerged where HAND struggled. These are in lower order streams, areas of low relief, and where the underlying data is inadequately represented.

4.1.1 Raster resolution systematically suppresses stage in lower order reaches

HAND yielded poor results for 3rd order reaches and underestimates inundation for virtually all 1st and 2nd order streams. Figure 5A (AFI = 69.8%; FloodID 8), shows an example focused on a gaged 2nd order reach. According to the SRC, 300 m^3/s of streamflow is required to generate the observed inundation depth (3.96 m). However, the NWM simulated 50 m^3/s (stage = 1.82m) while the gage recorded approximately 80 m^3/s. This type of error was particularly prominent in flood IDs 1, 4, 6, 8, 25 and 27.

This issue stems from the raster resolution of the HAND dataset. Functionally, the resolution of the raster (currently 10m) defines the minimum bank-width for all in-stream cells (cells identified as 0 in Figure 1C). In a survey of 418 rivers, the mean width of 1st-5th order streams were found to be 1.9m, 2.6m, 7.5m, 27.5m and 72.7m with standard deviations of 1.1, 1.8, 6.3, 42, and 98.1, respectively (Downing, 2012). As such, it appears that reaches with a mean width less than the resolution of the raster are those that systematically under predict stage heights. Physically speaking, this stems from a given flow volume being passed through a channel with an over-prescribed cross-sectional area (Figure 4).

A calculation for the above example shows that, if the cross-sectional base width was scaled to the mean reach value of a 2nd order steam (2.6 m), the required discharge would mirror the observed NWIS conditions at that gage at the time of the event:

10 m / 2.6m = 3.846 (HAND resolution divided by second order width average)
300 m3/s / 3.846 = 78.00 m3/s (HAND SRC required flow scaled by width ratio)

While a more rigorous evaluation of the effectiveness of this fix is outside the scope of this paper, it certainly deserves attention for improving the current HAND SRCs in lower order reaches.

4.1.2 Overprediction in areas of low-relief

While HAND struggles to allocate water volumes vertically on lower order reaches, regions of low relief can struggle to allocate water volumes horizontally. Where the vertical issues lead to systemic underprediction, the horizontal errors lead to over-predicted flood extents. The HAND methodology relies on elevation differences to predict if individual cells will be flooded. In areas of low relief (were the vertical change in neighboring cells is very small), HAND is very sensitive to even small increases in the magnitude of discharge when stage height exceeds bank height.

An example of this effect is shown on the Washita river at Anadarko OK (Figure 5B; AFI = 82.3%). A cross-section of the landscape indicates the terrain is relatively uniform, resulting in HAND spreading the flood water as far as 4 kilometers from the original stream. Compounding this issue is the fact that areas of low relief tend to also have much larger contributing catchments, resulting in large low

relief areas contributing to a single reach. Such a situation results in over-predicted inundated extents. This type of error is most prominent in FloodIDs 3 and 6.

4.1.3 Data representation

Errors within the NHD can also lead to forecasts errors. Such errors include missing 1st and 2nd order streams or inaccuracies in the spatial representation of stream segments. For the example shown in Figure 5C (Flood ID 12), the NHD does not capture the meandering segment of the Canadian River's main channel (AFI = 85.6%). In such a situation, even with accurate NWM forecasts and rating curve relationships, the predicted flood extent will be spatially incorrect. A similar situation can also occur in areas with very low gradients where a river may have a braided channel or wide estuary. Standard DEM-based stream delineation (used in NHD) only allow one channel to be designated for each reach. This results in poor representation of the stream network density or channel width in certain reaches. Therefore, these situations result in an inability to capture the spatial extent of flooding. Prediction Errors due to missing lower order streams can be seen in FloodID 8, while errors due to channel and or coastal feature representation can be seen in FloodIDs 19, 23 and 24.

4.2. NWM streamflow predictions

With an understanding of some systematic biases in HAND products, namely issues with lower order streams, poor water volume control in low relief regions and challenges with misrepresented or simplified stream features, we turn to evaluating how sensitive is the NWM-HAND inundation generation methodology to NWM forecast accuracy. To compare NWM streamflow predictions to observed streamflow records, all USGS National Water Information System (NWIS) stations within each flooded region were identified using the R HydroData package. Hourly (instantaneous) streamflow records for a time period of 20 days before and 20 days after each flood were downloaded from the NWIS servers via the dataRetrieval R package (Hirsch and De Cicco, 2015).

4.2.1 Timing Errors

HAND-based inundation calculation is based on a discharge value in a singular timestep. Each timestep is independent of the last and the method is unable to simulate how water might pond or recede from previous timesteps. Further, the method is unable to pass water between individual reach catchments. That is, no matter how fast, or high, water might pile it cannot extend beyond the specific catchment defined by contributing cells (Figure 1A). In the example shown in Figure 6A (Flood ID 7; AFI =75.7%), the NWM-predicted peak flow occurs at the same time as the observed flows and is within 5% of the observed peak magnitude. However, the predicted recession was much steeper than observed. Even though the image captures the inundation one day after the peak, the slight lag in time and the rapid recession predicted by NWM leads to significantly reduced inundation compared to the observed extent. This example illustrates how timing may be more critical then magnitude for accurate NWM-HAND forecasts. Similar errors can also be seen in FloodID 25.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019

Natural Hazards and Earth System Sciences Discussions

4.2.2 Magnitude Errors

While timing appears to be the principle driver in NWM induced errors, the NWM-HAND methodology both over and under predicts inundation extents due to inaccuracies in NWM-predicted streamflow magnitude. For example, Figure 6B (Flood ID 4) highlights an example where the NWM-HAND under predict inundation (AFI = 69.6%). Here, the NWM peak streamflow occurred about 5 days before the actual peak (presenting a timing issue), and at the time of the actual peak, was approximately half of the observed value (presenting a magnitude mismatch). In this case, even if the peak would have aligned, the quantity of water in the reach was approximately 2/3rds of what it should have been. These errors also occurred in floods 5, 25 and 27.

In other locations, NWM-HAND over predicts inundation due to NWM over predictions. Such an example is shown in Figure 6C (Flood ID 9; AFI = 66.3%) centered on the Brazos River at Richmond TX. A comparison of predicted and the observed hydrographs show that the NWM predicted peak flow of approximately 250% of the observed peak. These errors also occurred in flood 12.

4.3. Remote sensing (RS) Challenges

In all comparison case studies, we treated the USFIMR maps as ground-truth. For the sake of a robust comparison it is worth highlighting some of the issues that arise in the generation of remotely sensed inundation maps and their consequences in these comparisons.

4.3.1 Pixel misclassification

The spectral similarity of dark vegetation and water pixels in optical satellite imagery contributes to AFI uncertainties in both directions. Most often, these errors resulted in isolated patches of dark vegetation being classified as water (e.g. Figure 7A FloodID 2). Other cases pixel misclassification occurred in floods 1, 3, 16, 17, 18 and 23.

Water was also incorrectly classified in a few notable cases. For example, pixels clearly within the main channel of the Mississippi river were classified as unflooded in Figure 7B (Flood ID 13). In this example, the NWM-HAND methodology correctly identified such locations as being flooded (despite over predicting the overall flood extent). These error types were also present in floods 19 and 23.

4.3.2 Image artifacts

Two imagery artifacts also impact our comparison. The first is the well-known scan line error present in LANDSAT 7 images processed after 2003 (Scaramuzza and Barsi, 2005). Despite gap filling, some data gaps still remain. The gaps created by the scan line error (Figure 7C; Flood ID 4; AFI = 67.3%), created situations where both methods missed a flood or HAND-NWM alone predicts a flood, even though flooding could be detected from surrounding pixels. The overall extent covered by scan line gaps did not have a large impact on AFI scores, particularly as this issue affected only a relatively small percentage of the area in one case study.

The second artifact is cloud cover which had a strong effect in one case study (Figure 7D; Flood ID 5). The image used to generate the USFIMR map was taken from a LANDSAT scene with 40% cloud cover, making it difficult to accurately classify some segments of the image. This comparison also contained large regions of freshly melted snow (visible in the LANDSAT image as well as from the road network and farm fields in the classified image). The combination of clouds and melted snow result in a relatively low AFI score of 72% for this case. Cloud related classification errors also affected FloodID 15, but to a lesser degree.

5. Discussion

Based on the results, we can categorize flood comparisons with low AFI scores according to their predominant error type. Table 2 shows these errors for all floods with an AFI below 80%.

In areas were the terrain is suitable, the performance of NWM-HAND is driven by the accuracy of the NWM forecast. The NWM-HAND method seems to be more sensitive to the errors in timing than it is to errors of magnitude. The most important characteristics impeding HAND-NWM flood mapping accuracy are poor rating curves in low Strahler order streams and an inability to properly represent data in coastal and low-lying terrain. The results also show that HAND cannot accurately capture inundation extents in areas of low relief and that a mechanism for better controlling water volume distribution needs to be developed. Perhaps there is an opportunity for using the NWM velocity forecasts to do so.

Lastly, it should be stated that while AFI provides a detailed look at how maps compare across flooded and non-flooded regions, the overall metric is naturally biased toward areas of non-inundation since, even in regions of flooding, the dominant land cover is non-flooded pixels. This was suitable for the specific aim of this study, which is the assessment of the overall levels of agreement between NWM-HAND and observed extents. In cases where users are more interested in pixel-level agreement other metrics such as mean error, mean average error or the original fitness index (Bates and De Roo, 2000) might be more suitable.

6. Conclusion

In this study, we evaluated the HAND-NWM flood inundation mapping methodology using 30 observed flood inundation maps. Using the AFI metric we find that 21 of 30 (70%) flood extent case studies had 80% or more agreement with observed floods. The general level of agreement indicates that the HAND-NWM methodology performs well in its stated intentions of operational flood forecasting. In case studies with moderate and fair agreement, we have found systematic errors due to a mix of issues including HANDs inability to characterize lower order stream geometries, control for water volumes across flat terrain, poor representation underlying data, and inaccuracies in NWM forecasts. In general, NWM-HAND has a bias towards underprediction except in cases where the NWM provides excessive flow rates or in low relief regions where water volume control played a role.

Further, a spatially explicit understanding of where the NWM performs well not only for average streamflow but during extreme conditions is needed, as is a more methodological assessment of what features could be excluded or improved in the NWM-HAND methodology. Currently, accessing

and combining the needed NWM-HAND products is cumbersome, and for regions straddling one or more HUC6 units can be data and processing intensive. While services like the Flood Mapping R package can remove some of this headache, HAND products could be distributed via a web feature (or coverage) service, or even better, a THREDDS OPeNDAP server like those used to distribute climate data and NWM output. Building on this, in order to lower the barriers of entry to operational flood mapping, a GUI driven platform could be easily implemented to aggregate and map future flood extents. On this vein, cartographic research regarding the communication of flood maps should be undertaken, in addition to viable nodes for putting rapidly developed flood information into action. Finally, as noted in the NWM sections, work needs to be done to better understand and communicate the systematic reasons that HAND under predicts flood extents, namely mechanisms for evaluating 3rd and lower order reaches, controlling volume in areas of low relief, and better representing the hydraulic geometries in braided, meandering and parallel channels. That said, the products stemming from the NFIE experiment, and the operational NWM provide a promising means for revolutionizing the state of USA flood mapping.

Acknowledgements: USFIMR is funded through the NOAA National Water Center (NWC) via the UCAR COMET Program (2016/17), the AOI, HydroData, and Flood Mapping R packages were developed through the NOAA National Water Center (NWC) and UCAR COMET Program (2017/18).

References

Afshari, S., Tavakoly, A.A., Rajib, M.A., Zheng, X., Follum, M.L., Omranian, E., Fekete, B.M: Comparison of new generation low-complexity flood inundation mapping tools with a hydrodynamic model. Journal of Hydrology 556, 539–556, 2018.

Bates, P.D., De Roo, A: A simple raster-based model for flood inundation simulation. Journal of Hydrology 236, 54–77, 2000.

Cline, D., Maidment, D.: Community Advisory Committee for Water Prediction (CAC-WP), 2019.

Cohen, S., Brakenridge, G. R., Kettner, A., Bates, B., Nelson, J., McDonald, R., Huang, Y., Munasinghe, D., Zhang, J.: Estimating floodwater depths from flood inundation maps and topography. JAWRA Journal of the American Water Resources Association, 54(4), 847-858, 2018.

Downing, J.A., Cole J.J., Duarte C.M., Middelburg J.J., Melack J.M., Prairie Y.T., Kortelainen P., Striegl R.G., McDowell W.H. & Tranvik L.J.: Global abundance and size distribution of streams and rivers. IW 2, 229–236. doi:10.5268/IW-2.4.502, 2012.

Gochis, D., Dugger, A., McCreight, J., Karsten, L.R., Logan, Yu, W., Pan, L., Yates, D., Zhang, Y., Sampson, K., Cosgrove, B., Salas, F., Clark, E., Graziano, T., Maidment, D., Phan, C., Cui, Z., Liu, Y., Feng, X., Lee, H.: Technical Description of the National Water Model Implementation of WRF-Hydro, CUAHSI Technical Report, 2016.

Gochis, D.J., Dugger, A., Barlage, M., Fitzgerald, K., Karsten, L., McAllister, M., McCreight, J., Mills, J., Rafieeinasab, A., Read, L., Sampson, K., Yates, D., Yu, W.: The NCAR WRF-Hydro Modeling

System Technical Description [WWW Document]. ral.ucar.edu. URL https://ral.ucar.edu/sites/default/files/public/WRF-HydroV5TechnicalDescription.pdf (accessed 8.21.18), 2018.

Henstra, D., Minano, A., Thistlethwaite, J.: Communicating disaster risk? An evaluation of the availability and quality of flood maps. Natural Hazards and Earth System Sciences 19, 313–323, 2019.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S.:Global flood risk under climate change. Nature Publishing Group 3, 816, 2013.

Hirsch, R.M., De Cicco, L.A.: User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data. US Geological Survey, 2015.

Horritt, M. S.: Calibration of a two-dimensional finite element flood flow model using satellite radar imagery. Water Resources Research, 36(11), 3279-3291, 2000.

Johnson, J.M.: AOI - An R Package for find, process, and describe "areas of interest", Github Repository. GitHub. doi:10.5281/zenodo.1318687, 2018a.

Johnson, J.M.: HydroData – An R Package for finding and getting geospatial earth systems data, Github Repository. GitHub. doi:10.5281/zenodo.1401393, 2018b.

Johnson, J.M.: NWM – An R Client for the National Water Model, Github Repository. GitHub. doi:10.5281/zenodo.1401394, 2018c.

Johnson, J.M., Flood Mapping – An R package for flood forecasting via the National Water Model, Github Repository. Github. doi: 10.5281/zenodo.2588009, 2019.

Johnson, J.M., Coll, J.M.: National Water Center Innovators Programs Summer Institute Report 2017. doi:10.4211/technical.20171009, 2017.

Johnson, J.M., Coll, J.M., Ruess, P.J., Hastings, J.T.: Challenges and Opportunities for Creating Intelligent Hazard Alerts: The “FloodHippo” Prototype. Journal of the American Water Resources Association, 2018.

Lam, L.: A Concerning Trend: Flooding Deaths Have Increased in the U.S. the Last Few Years | The Weather Channel, 2018.

Lightbody, L., Tompkins, F: Where It Rains, It Floods, Pew Trust, 2018.

Liu, Y.Y., Maidment, D.R., Tarboton, D.G., Zheng, X., Wang, S.: A CyberGIS Integration and Computation Framework for High-Resolution Continental-Scale Flood Inundation Mapping. Journal of the American Water Resources Association, 2018.

Maidment, D.R.: Conceptual Framework for the National Flood Interoperability Experiment. J Am Water Resour Assoc 53, 245–257. doi:10.1111/1752-1688.12474. 2016.

Maidment, D.R.: FEMA flood map accuracy: Presented at the World Environmental and Water Resources Congress 2009: Great Rivers, pp. 1–10, 2009.

Mason, D. C., Bates, P. D., & Dall'Amico, J. T.: Calibration of uncertain flood inundation models using remotely sensed water levels. Journal of Hydrology, 368(1-4), 224-236, 2009.

Merz, B., Thieken, A.H., Gocht, M.: Flood risk mapping at the local scale: concepts and challenges, in: Flood Risk Management in Europe. Springer, pp. 231–251, 2007.

Munasinghe, D., Cohen, S., Huang, Y.F., Tsang, Y.P., Zhang, J., Fang, Z.: Intercomparison of Satellite Remote Sensing-Based Flood Inundation Mapping Techniques. Journal of the American Water Resources Association 54, 834–846, 2018.

National Research Council: Mapping the zone: Improving flood map accuracy. National Academies Press, 2009.

NOAA: National Water Model [WWW Document]. httpwater.noaa.govdocumentswrn-national-water-model.pdf. URL http://water.noaa.gov/documents/wrn-national-water-model.pdf (accessed 8.21.18a), 2016a.

NOAA: Experimental Water Information Interface Webpage Product Description Document [WWW Document]. water.noaa.gov. URL http://water.noaa.gov/documents/OWP-interface-PDD.pdf (accessed 8.21.18b), 2016b.

Reilly, T.E., Franke, L., Bennett, G.D.: Techniques of Water-Resources Investigations of the United States Geological Survey, 1987.

Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., Waterloo, M.J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. Remote Sensing of Environment 112, 3469–3481, 2008.

Rodda, H.J.: The development and application of a flood risk model for the Czech Republic. Natural hazards 36, 207–220, 2005.

Salas, F.R., Somos-Valenzuela, M.A., Dugger, A., Maidment, D.R., Gochis, D.J., David, C.H., Yu, W., Ding, D., Clark, E.P., Noman, N.: Towards Real-Time Continental Scale Streamflow Simulation in Continuous and Discrete Space. Journal of American Water Resources Association 51, 10078–21. doi:10.1111/1752-1688.12586, 2017.

Scaramuzza, P., Barsi, J.: Landsat 7 scan line corrector-off gap-filled product development: Presented at the Proceeding of Pecora, pp. 23–27, 2005.

Yin, J., Gentine, P., Zhou, S., Sullivan, S.C., Wang, R., Zhang, Y., Guo, S.: Large increase in global storm runoff extremes driven by climate and anthropogenic changes. Nature Communications 9, 4389, 2018.

Zhang, J., Huang, Y.F., Munasinghe, D., Fang, Z., Tsang, Y.P., Cohen, S.: Comparative analysis of inundation mapping approaches for the 2016 flood in the Brazos River, Texas. Journal of the American Water Resources Association 54, 820–833, 2018.

Zheng, X., Maidment, D.R., Tarboton, D.G., Liu, Y.Y., Passalacqua, P.: GeoFlood: Large-Scale Flood Inundation Mapping Based on High-Resolution Terrain Analysis. Water Resources Research, 2018.

Zheng, X., Tarboton, D.G., Maidment, D.R., Liu, Y.Y., Passalacqua, P.: River channel geometry and rating curve estimation using height above the nearest drainage. Journal of the American Water Resources Association, 2017.

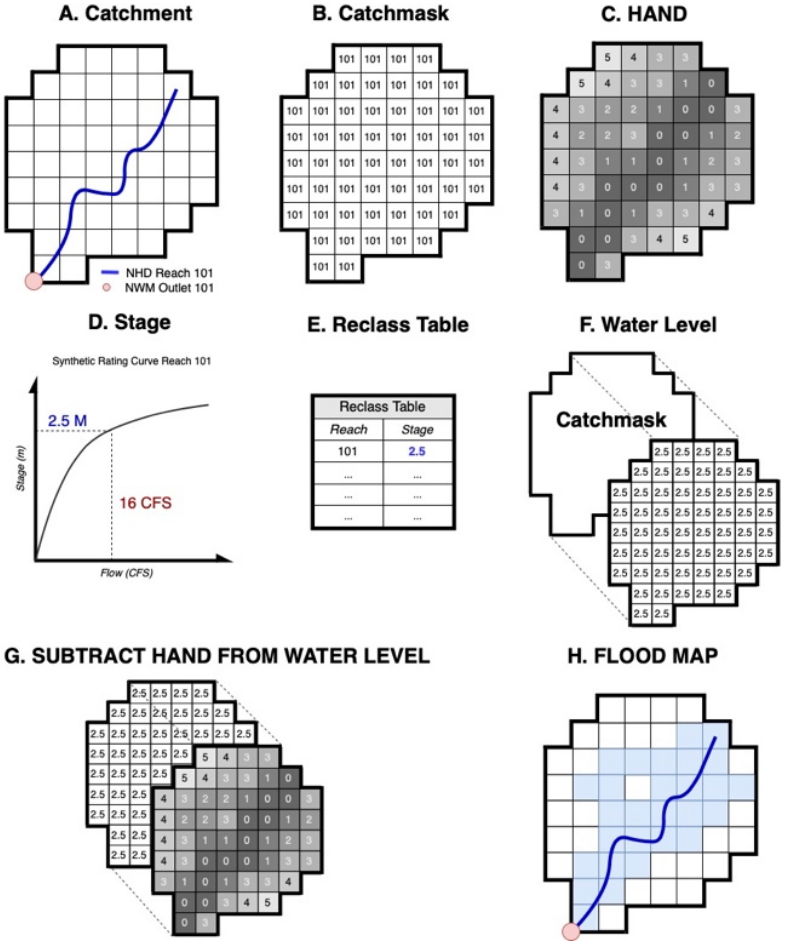

Fig 1: The HAND methodological workflow: ***(A) The contributing catchment to a defined outlet is rasterized to the resolution of the supporting DEM (B) This raster is reclassified to match the identifying code of the outlet. When using the NHD, the COMID can serve this purpose. (C) In and out-of-stream cells are identified using the flowline vector. The relief between all out-of-stream cells from their nearest in-stream-cell is calculated to define a HAND raster. (D) From a given flow rate, a rating curve can be used to convert flow to stage. (E) A reclassification table can be built relating the reach COMID to the current stage. (F) The table can then be used to reclassify the catchmask, (B), into a water-level. (G) Subtracting C from F yields a water-level above surface raster. All values less than 0 can be set to 0 and the remaining show the estimated flood height at each cell (H).***

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019

Natural Hazards and Earth System Sciences
Discussions

Table 1. Flood events used for comparison

Index	River name (of flood)	Flood Date	Satellite Sensor used for Observation	State	Flood area (km^2)
1	Choctawhatchee River	4-Jan-16	Landsat 8 OLI	FL	1493
2	Holmes Creek	4-Jan-16	Landsat 8 OLI	FL	129
3	Pea River	4-Jan-16	Landsat 8 OLI	AL	407
4	Pascagoula River	17-Mar-11	Landsat 7 ETM+	MS	1692
5	Red River	18-Apr-97	Landsat 5 TM	ND	7777
6	Brazos River	28-May-16	Landsat 8 OLI	TX	74
7	Spring Creek	28-May-16	Landsat 8 OLI	TX	160
8	Willow Creek	28-May-16	Landsat 8 OLI	TX	48
9	Washita River	29-May-15	Sentinel-1 VH Polarization	OK	623
10	Brazos River	30-May-16	Sentinel-1 HV Polarization	TX	951
11	San Jacinto River	30-May-16	Sentinel-1 HV Polarization	TX	212
12	Canadian River	17-Jun-15	Sentinel-1 VH Polarization	OK	739
13	Mississippi River	16-Aug-16	Earth-Observing 1 ALI	LA	3029
14	Trempealeau River	26-Sep-10	Landsat 5 TM	WI	276
15	Wisconsin River	26-Sep-10	Landsat 5 TM	WI	8350
16	Cedar River	26-Sep-16	Landsat 8 OLI	IA	4425
17	Iowa River	26-Sep-16	Landsat 8 OLI	IA	3541
18	Maquoketa River	26-Sep-16	Landsat 8 OLI	IA	2249
19	Mississippi River	26-Sep-16	Landsat 8 OLI	IA	1937
20	Wapsipinicon River	26-Sep-16	Landsat 8 OLI	IA	4148
21	Minnesota River	1-Oct-10	Landsat 5 TM	MN	7876
22	Redwood River	1-Oct-10	Landsat 5 TM	MN	1414
23	Mississippi River	3-Oct-10	Landsat 5 TM	MN	6957
24	Ashley River	13-Oct-16	Landsat 8 OLI	SC	1804
25	Black River	13-Oct-16	Landsat 8 OLI	SC	431
26	Cooper River	13-Oct-16	Landsat 8 OLI	SC	2096
27	Lumber River	13-Oct-16	Landsat 8 OLI	SC, NC	1116
28	Neuse River	15-Oct-16	Landsat 8 OLI	NC	1054
29	Tar River	15-Oct-16	Landsat 8 OLI	NC	1702
30	Rio Grande River	14-Jul-10	Landsat 7 ETM+	TX	6836

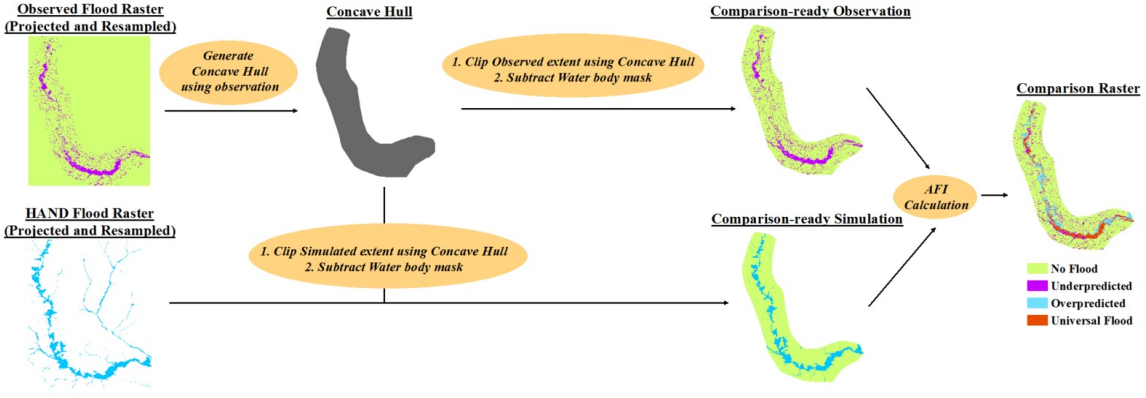

Figure 2: Comparison methodology between NWM-HAND and USFIMR floods.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019

Natural Hazards and Earth System Sciences
Discussions

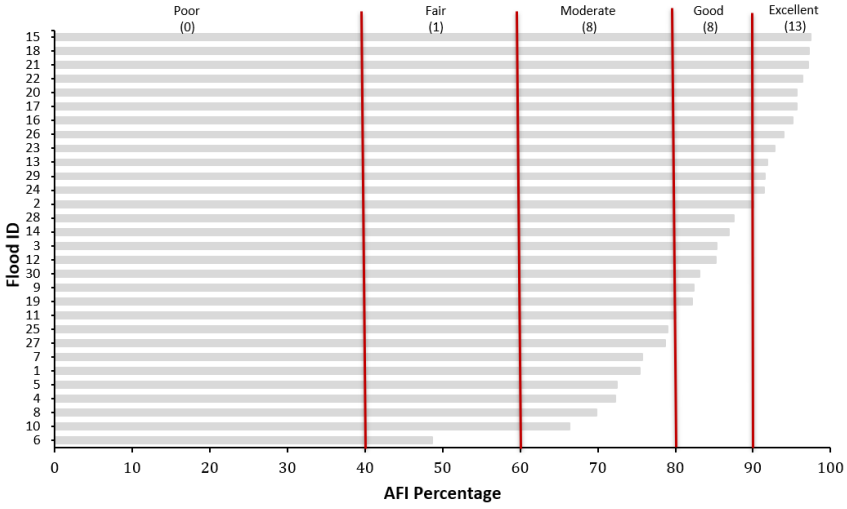

Figure 3: The Advanced Fitness Index (AFI) and rating for the flood events comparisons

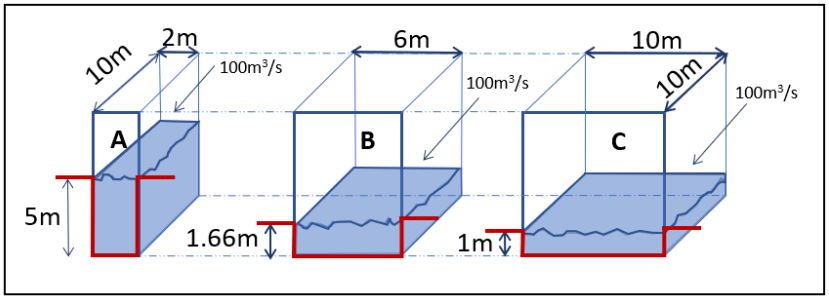

Figure 4: For narrow streams, stage is dependent on resolution of the HAND raster: (A) 100m³/s of flow results in a stage of 5m when cell area is 20m² (B) 100m³/s of flow results in a stage of 1.66m when cell area is 60m² (C) 100m³/s of flow results in a stage of 1m when cell area is 100m².

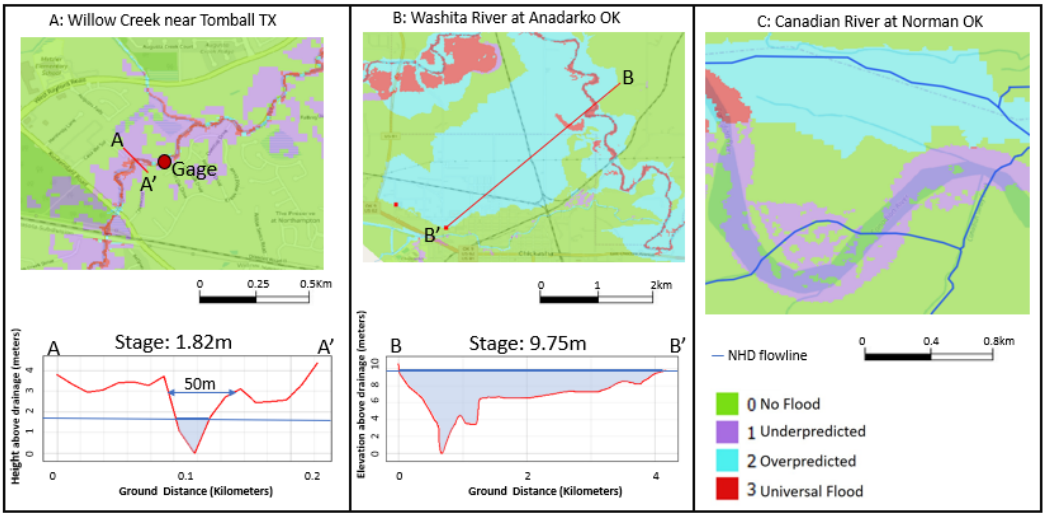

Fig 5: Prediction errors due to HAND methodology (A) Inaccurate HAND stage-discharge relationship causes underprediction in a 2nd order stream; (B) In flat terrain, HAND poor volume control causes unrealistic flooding extents; (C) Actual flood extent missed because main channel is not correctly located in NHD data and not detectable from the HAND raster.

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019

Natural Hazards and Earth System Sciences
Discussions

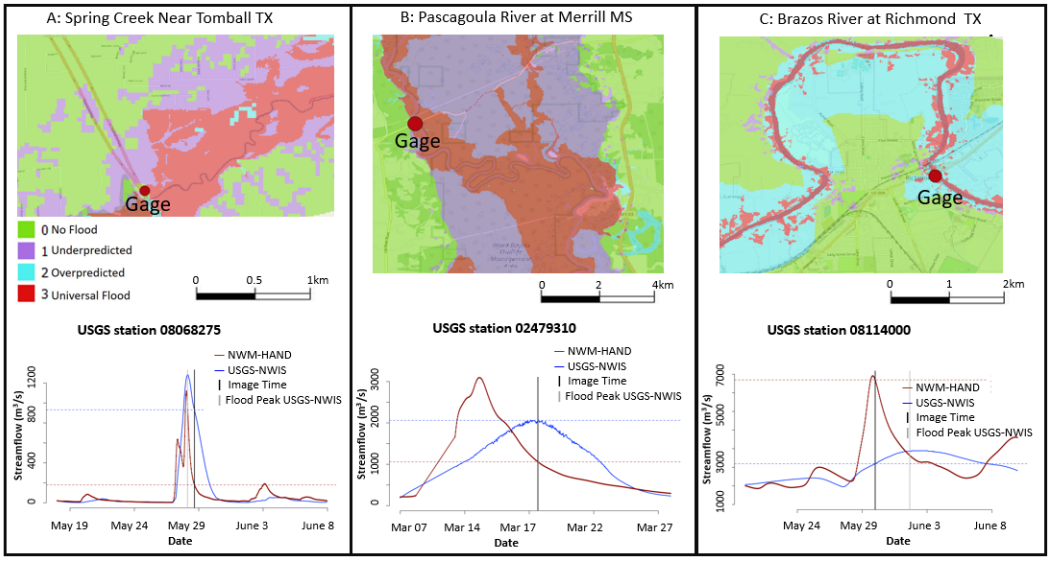

Fig 6: Prediction errors due to NWM flow forecast: (A) HAND flood underprediction due to NWM timing of recession. (B) HAND flood underprediction due to NWM low flow forecast at actual flood peak. (C) HAND flood overprediction due to MWM high discharge forecast.

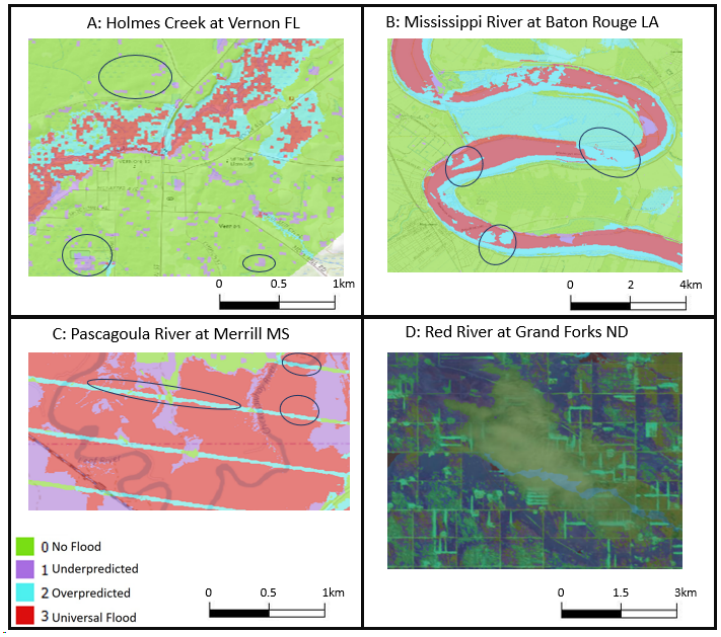

Fig 7: Errors due to classification issues and image artefacts: (A) Dark vegetation misclassified as water due to spectral similarity. (B) Misclassified channel segments due to spectral similarity. (C) Scan line errors in input RS image create gaps in flooded extent. (D) clouds and melted snow create ambiguities in flooded extent recognition.

Table 2: Low AFI values explained by predominate error type

Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2019-82
Manuscript under review for journal Nat. Hazards Earth Syst. Sci.
Discussion started: 2 April 2019

Natural Hazards and Earth System Sciences
Discussions

Flood	Date	AFI	Reason	Topic
Brazos River near Hampstead Texas (6)	5/28/2016	48.2	Low Relief / Stream Order	Terrain
Lumber River Lumberton North Carolina (27)	10/13/2016	61.7	Underprediction / Stream Order	NWM
Brazos River near Richmond Texas (10)	5/30/2016	66.2	Overprediction / Stream Order	NWM
Pascagoula River at Merrill Missouri (4)	01/04/2016	67.3	Underprediction / Stream Order	NWM
Spring-creek near Tomball (7)	05/28/2016	75.7	Timing error	NWM
Red River at Grand Forks North Dakota	*04/18/1997*	*72.5*	*Clouds and Snow*	*RS*
Choctawhatchee River near Bellwood Alabama (1)	01/24/2016	68.7	Data Representation (coastal features) / Stream Order	HAND
Willow Creek near Tomball Texas (8)	5/28/2016 - 5/30/2016	69.8	Stream Order	HAND
Black River at Kingstree South Carolina (25)	10/13/2016	78.9	Stream Order	HAND