

1. Overview & Background

Predicting the impact of floods on transportation infrastructure is crucial for emergency response and preparedness. Specifically, improvements in predictive accuracy significantly reduces capital investments needed to protect critical roads [1]. In the past few years, multiple researchers investigated flood risk and the resulting vulnerabilities in the transportation network [2, 3, 4, 5, 6]. The primary objective of such studies was to identify critical transportation links by quantifying the likelihood and impact of floods.

Data drive methods for quantifying flood risk exploit recent advances in wireless sensor networks and flood information systems [7, 8, 9]. Specifically, these methods generate statistical predictive models that exploit existing data on flood condition and watershed characteristics to quantify road flood probability [2, 5, 10]. Other data driven studies include identifying statistical relationships between the functional performance of the roadway and the flood intensity [4].

Another class of methods for estimating road network flood risk rely on terrain models generated using LiDAR data along with hydrologic models describing flood dynamics [3, 6, 11]. One possible approach is to first extract roadway centerlines using LiDAR data, and then determine the flood extent for a given water level using a digital terrain model [11]. Such methods highlight the importance of topographic attributes that could be integrated with additional landscape features and hydrologic models in geographic information systems [10, 11]. In comparison with statistical data driven methods, approaches that use terrain data and hydrologic models may provide increased accuracy of flood risk. However, this increased accuracy is at the expense of computational cost and predictive complexity [10].

Recently, the National Water Model (NWM) was developed to provide large scale flow forecasts on about 2.7 million stream and river reaches [12]. The NWM core is a weather research and forecasting hydrologic model (WRF-Hydro) that is set to perform Muskingum-Cunge channel routing through stream reaches obtained from the national hydrography dataset (NHDplus). This framework uses a nudging scheme to assimilate streamflow observations from United States geological survey (USGS) streamgages [13].

Thus, instead of relying on computationally expensive hydrologic models, the NWM provides streamflow forecasts that are freely available for download. However, streamflow forecasts do not give a direct indication of water level and corresponding road network flood risk. This led researchers to explore methods for estimating rating curves at a large-scale by relying on channel geometry using the height above nearest drainage method (HAND) [14]. In turn, the HAND method can also be used to determine inundation for any particular stage height computed using the aforementioned rating curves. This coupling between HAND and NWM forecasts enables near real-time inundation forecasts for the continental United States [15, 16].

2. Objective

The primary objective of this project is to analyze the applicability of HAND-NWM inundation forecasts for determining road network flood damage. Specifically, we aim to compare inundation maps from the HAND-NWM framework with observed flood incidents. Then, we will provide a framework for probabilistic predictions that can be used to (1) quantify the uncertainty associated with HAND-NWM inundation maps and (2) integrate road network incident observations to improve road network disruption predictions at unobserved locations.

3. GIS Data Processing

In this section, we will discuss methods for extracting data generated by the national flood interoperability experiment (NFIE) and the NWM. In addition, we will discuss data processing methods that are needed to determine if a road is inundated or not using HAND-NWM inundation maps. A central theme in the following discussion is the impact of data resolution on the applicability of road network HAND-NWM inundation mapping. Specifically, we show that the relatively coarse resolution of watershed information and digital elevation models leads to difficulties in processing the road network. To address these limitations, we develop python scripts that process the road network data, incident information, and HAND-NWM inundation maps. The outcome of the developed data analysis modules is a dictionary with road ID, inundation level at lowest point predicted by HAND-NWM inundation maps, location of lowest point in projected coordinates, and whether an incident was observed in the field or not. This dictionary is saved as a pckl file and uploaded to Hydroshare along with the python code. The Hydroshare resource link is as follows: <http://www.hydroshare.org/resource/45122a8fc2ea4bc984d370e5b543d3d9> [17].

3.1 Study Area & Road Network Data

The study area and flood event considered correspond to the Harvey hurricane that passed through Houston in 2017. Figure 1 shows the area of interest. To analyze road network inundation in the Houston area, we resort to Harvey Hydroshare resource for road network data [18]. In particular, under the “Harvey Basemap Data Collections” collection resource, there is a “Texas Basemap - Transportation Map Data” composite resource that contains shapefiles for the road network maintained by the Texas Department of Transportation (TxDOT). For the road network flood incidents, a shape file with the incident data was uploaded to [17]. Note that we restrict the analysis to incidents observed on 08-28-2017. By restricting to this date, we can easily extract NWM forecasts that corresponds to incidents observed throughout that day. As for the hydrology data, these were downloaded from the NFIE database for regional 12. This database is available as a Hydroshare resource at <http://www.hydroshare.org/resource/1d78964652034876b1c190647b21a77d> [19]. The data was extracted to the region of interest using the extract by mask feature with the subwatershed boundary surrounding the Houston area as the mask. Note that all feature classes were projected to the NAD 1983 UTM Zone 14N coordinate system using a transverse Mercator projection.

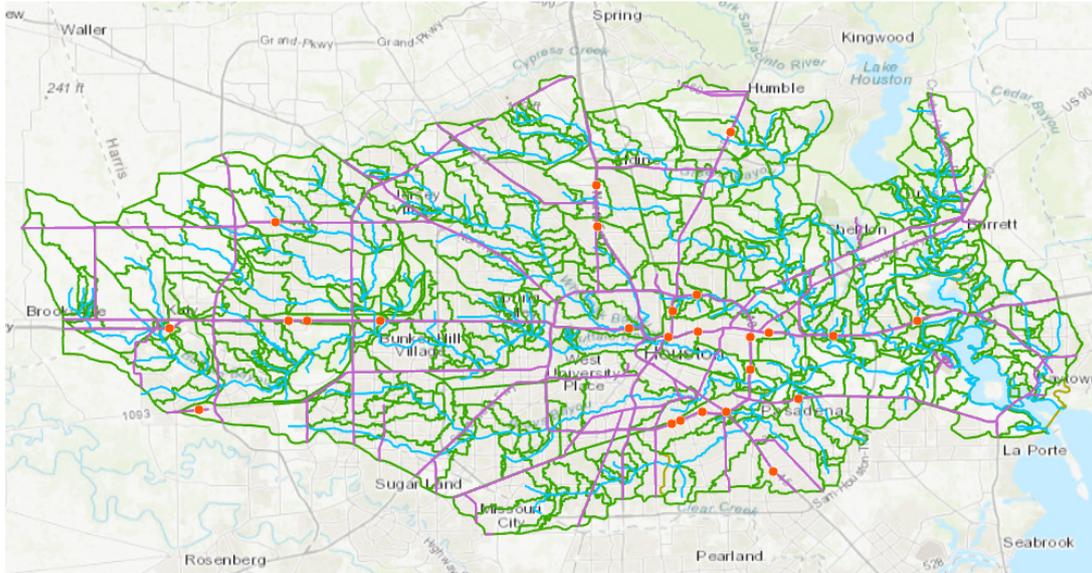


Figure 1: Study area where flood incidents observed by TxDOT are shown as red points

3.2 Processing Road Network Data: Resolution Effects

Consider a road segment defined as the stretch between two intersections. To determine if a particular road is inundated using HAND-NWM inundation maps, we need to first identify the catchment in which the road segment lies. However, in certain cases road segments cross several catchments. In these situations, the road should be divided into smaller sub-segments where each sub-segment belongs to a specific catchment. To divide the road network layer into the aforementioned sub-segments, we intersect the road polyline feature class with the catchments feature class. Then, we split the road segment at the resulting intersection points using the ‘split line at a point’ geo-processing tool.

However, due to the catchment resolution, this leads to a large number of road sub-segments. In particular, by observing Figure 2, we can see that the catchment boundary is often along the road segment itself. Due to the non-smooth boundary of the catchments (as a result of the resolution of data used to build the catchments dataset), the catchment boundary intersects the road segment at multiple points. In turn, this results in a large number of road sub-segments.



Figure 2: Catchment boundary along a road results in dividing the road into a large number of small sub-segments (one specific sub-segment is circled)

Managing a large number of tiny road sub-segments is computationally expensive. Moreover, analyzing such short sub-segments is not necessary when considering network level road damage in the subwatershed. Therefore, we delete these road sub-segments that are at irregular catchment boundaries. To do so, we create a road network layer by spatial joining the splitted road layer with catchments using a search radius of 20m. In that layer (after spatial join), every road segment at the boundary will be associated with catchments on both sides and will have duplicate entries (one to many spatial join). Then, we write python code to process the resulting layer and delete duplicates. This code is available within *process-roads.py* at [17]. The road segments lost to the cleaning process are shown in black in Figure 3. While this leads to a loss of information on certain parts of the network, the number of road segments and sub-segments was reduced from 2120 to 407.

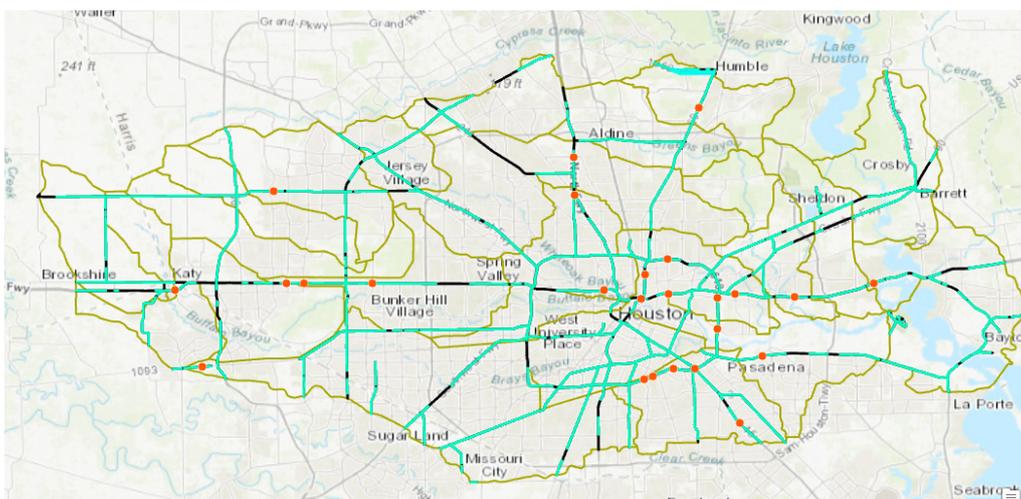


Figure 3: Roads lost due to cleaning segments at catchment boundaries are shown in black

3.3 Height Above Nearest Drainage

In the HAND-NWM inundation mapping, we need to compute the HAND value along road segments. The HAND raster was downloaded from the NFIE database at the Texas Advance Computing Center (TACC). The HUC-6 area that includes Houston is 120401. Then, the HAND raster was projected to the UTM Zone 14 coordinate system (cubic convolution). The resulting raster is shown in Figure 4. We note that the HAND raster downloaded from NFIE has some negative values and does not fully cover the Houston area. Specifically, Figure 5 shows a portion of Houston that did not have any HAND values (even after downloading the adjacent HAND raster for HUC-6 #120701).

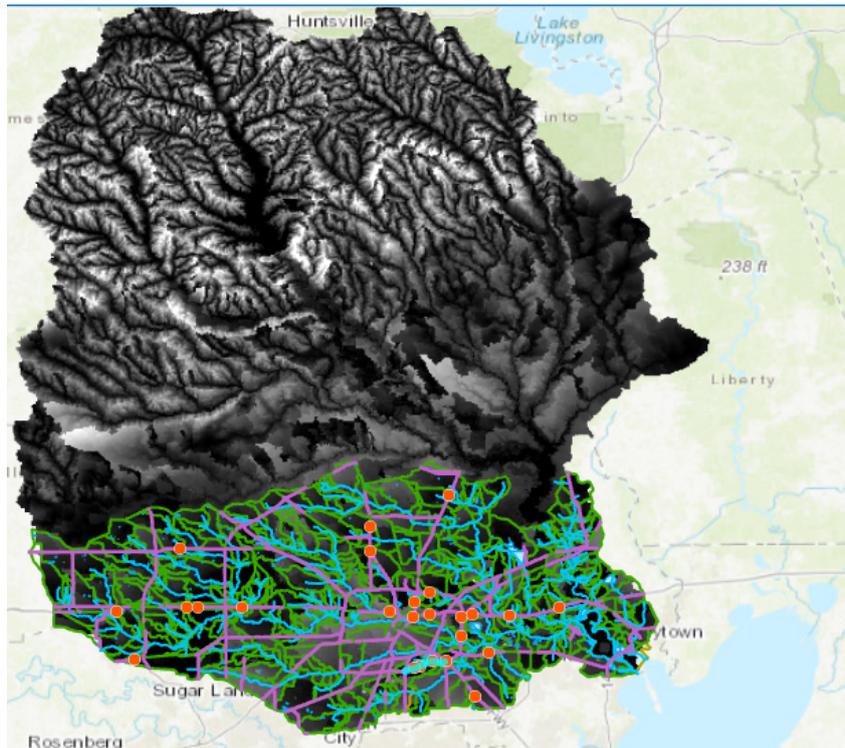


Figure 4: HAND raster for the HUC-6 region # 120401

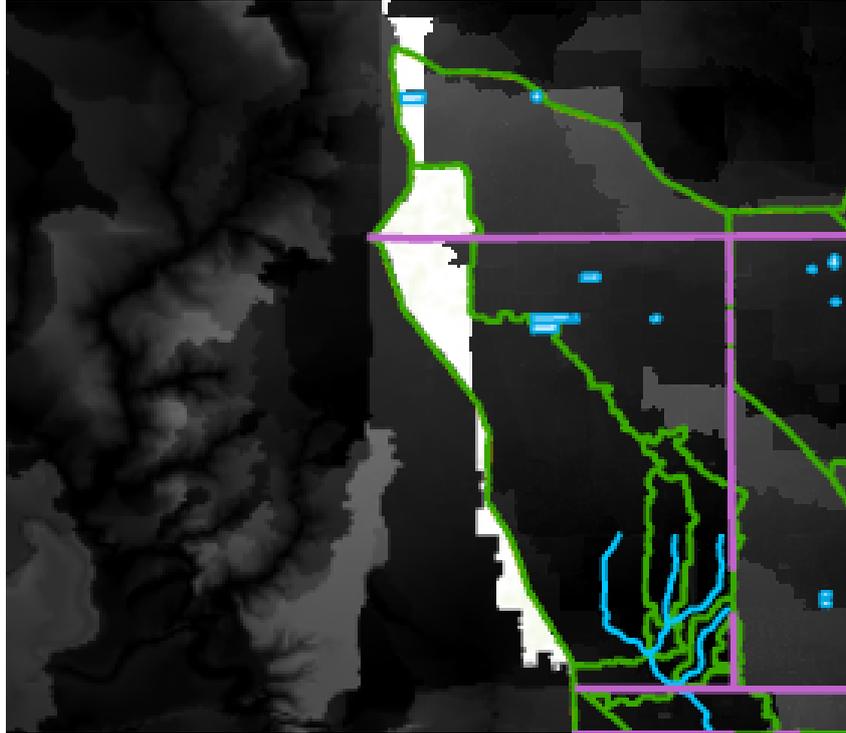


Figure 5: Missing HAND values in the Houston area using HAND rasters downloaded from the NFIE database

To extract the HAND values to the road segments, we first convert the polyline feature into a raster, and then convert the resulting raster to a point layer. The resulting raster and overlaid point layer are shown in Figure 6. The point layer is also shown in Figure 7. Then, using the ‘extract values to points’ geo-processing tool to combine the HAND raster and the point layer associated with the road segments, we generate another point layer that has the corresponding road segment (sub-segment) ID and HAND value for every point.

However, to determine if a road segment is damaged by a flood, we only need to find the point along the road segment with minimum HAND value. Specifically, if the road is inundated, then the minimum HAND point along the road will have the largest inundation. To find the minimum HAND point for every road segment, we write python functions in a module referred to as *computeRoadHand.py* at [17]. In addition, we create a layer out of the minimum HAND points and use the ‘add XY coordinates’ geo-processing tool to find the coordinates of every minimum HAND point.



Figure 6: Raster and point layers associated with the road segment polyline feature class

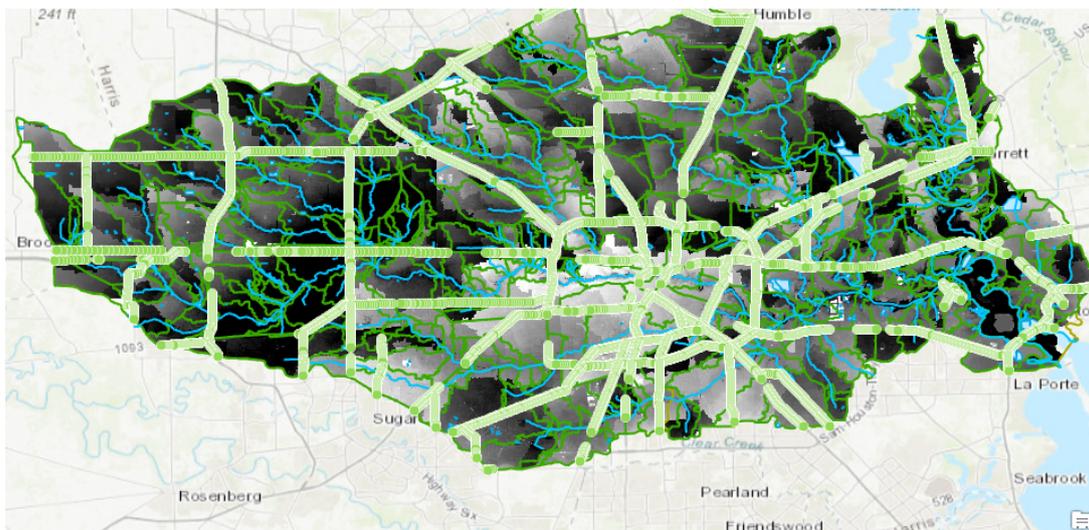


Figure 7: Point layer and HAND raster after extraction to subwatershed area

3.3 National Water Model Data & NFIE Rating Curves

After extracting the HAND data at the lowest points along road segments, we need to obtain streamflow forecasts and translate the forecasts into stage height to determine the corresponding inundation. The NWM forecast data are in the order of terabytes. Thus, it is difficult to manipulate this data by downloading the entire data set and searching for time series within a specific time horizon and over a particular region. However, on Hydroshare, there are resources for effectively retrieving the NWM forecasts using Thredds and OpenDAP technologies. Specifically, we modify Castronova's code that is available at [20] to extract data for the Houston area. The uploaded *getNWMdata.py* module contains code for extracting NWM short range forecasts within the Houston area on 08-28-17 [17]. Then, the NWM discharge value associated with any catchment

is chosen to be the maximum reported value within the 18 hour short range forecast. It is worthy of note that the NWM streamflow forecasts were not available for certain catchments shown in Figure 8. While this is not convenient for automated processing of road segment inundation levels, it is clear that the roads in this region will be inundated due to their proximity to the Galveston Bay.

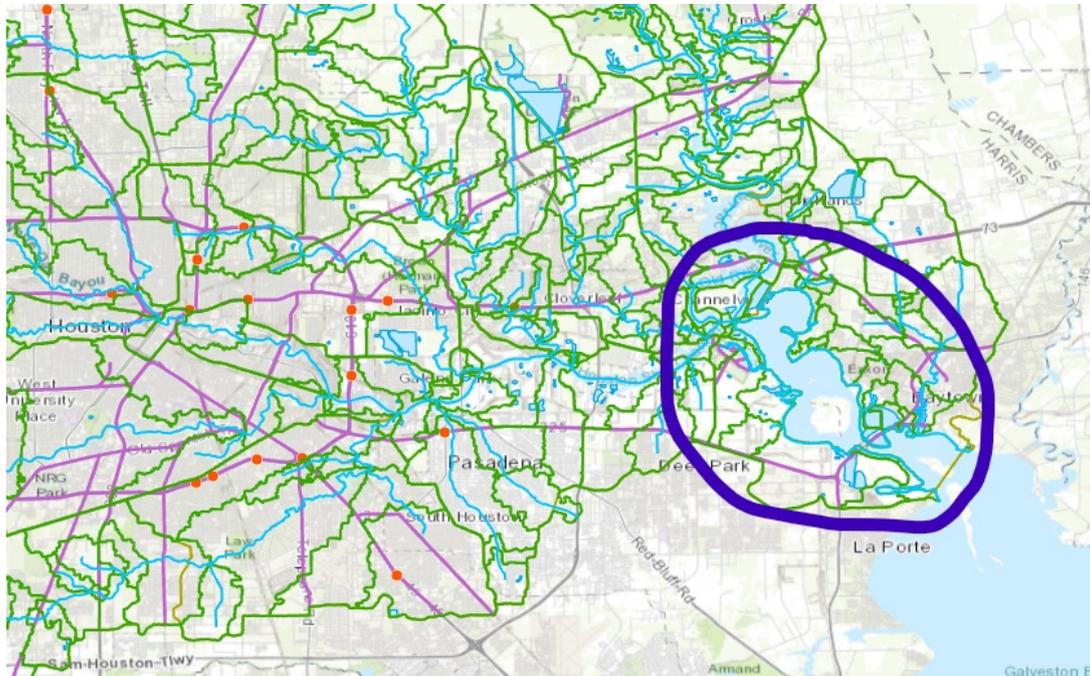


Figure 8: Region with missing NWM streamflow forecasts

To obtain rating curves we rely on data generated from the NFIE experiment to map discharge for different levels of stage heights for every catchment within the continental United States. Specifically, within the HUC-6 #120401 NFIE database, there is a csv file referred to as ‘hydroprop-fulltable-120401.csv’ that has rating curve data points that are generated using the procedure in [14] based on river channel geometry. For every catchment, the csv file has 82 discharge-stage data points. To extend this rating curve for any discharge-stage pair, we write code in *discharge2stage.py* that does piecewise linear interpolation between the NFIE data points. This code is also uploaded to [17].

4. Comparison of Inundation Maps with Observed Incidents

After generating NWM streamflow forecasts for 08-28-17 and translating the forecasts to stage height using NFIE rating curves, we can compare the stage heights within every catchment to the minimum HAND value on road segments within the catchment to determine if the road segments are damaged by the flood. In particular, the inundation level at any road segment would be the minimum HAND value subtracted from the stage height.

For the 26 flood incidents that caused road obstructions identified by the Texas department of transportation, 5 incident locations had an inundation level of zero computed

through the HAND-NWM inundation maps. Thus, for the available incident data set, the HAND-NWM inundation map had an accuracy of about %80. This inability to detect inundation at all field observed incident locations is due to the assumptions and data resolution limitations of the HAND-NWM inundation mapping.



Figure 9: TxDOT observed incidents on 08-28-17 that are not detected by the corresponding HAND-NWM inundation mapping

5. Spatial Probabilistic Model that Incorporates Observed Incident Information

In this section, we outline a model for generating probabilistic predictions using the HAND-NWM inundation maps and the observed road network incidents. The probabilistic nature of the model allows us to quantify the uncertainty due to the inaccuracy of HAND-NWM inundation maps. Furthermore, the model conditions on observed incidents to improve the probabilistic predictions. Specifically, we propose the use of a Gaussian process (GP) to generalize a linear regression spatially and generate probabilistic predictions. In particular, the GP is a non-parametric model that specifies a distribution over a random function f as shown in Equation 1. In this equation, m specifies a mean function and k is a symmetric positive-definite kernel. The Gaussian process can be interpreted as follows: for every argument x_i of the random function f there is an associated random variable $f(x_i)$ such that the joint distribution for a finite set of random variables $\{f(x_1), f(x_2), \dots, f(x_N)\}$ is multivariate Gaussian with a mean vector specified by m and a covariance matrix specified by k [21].

$$f \sim \mathcal{GP}(m, k) \quad (1)$$

In the case of road network inundation, the mean function m can reflect the inundation level computed using HAND-NWM inundation maps. The covariance matrix represents the relationship between observations at different locations. For example, an exponential covariance kernel defined in Equation 2 implies that as the road segments are further apart the corresponding correlation between water depth measurements decreases [22]. In this equation, x refers to the location of the minimum HAND point along a road

segment.

$$k(x_i, x_j) = e^{-\phi(x_i - x_j)^2} \quad (2)$$

To incorporate road network observations, we can decompose the multivariate Gaussian distribution associated with the random variables as shown in Equation 3 where the vector \mathbf{f}_y represents the random variables at the monitored locations and the vector \mathbf{f}_x represents the random variable at unmonitored locations. μ and Σ reflect the corresponding mean and covariance matrix, respectively.

$$\begin{bmatrix} \mathbf{f}_x \\ \mathbf{f}_y \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{x,y} \\ \Sigma'_{x,y} & \Sigma_y \end{bmatrix} \right) \quad (3)$$

Then, using properties of the multivariate Gaussian distribution, the posterior distribution at locations that are not monitored for a given set of observations $\mathbf{f}_x = f_x$ at monitored locations can be given as shown in Equation 4. Therefore, using the GP and HAND-NWM inundation maps (to specify the mean function) we are able to predict the posterior distribution at unobserved road segments. This posterior distribution determines the expected water depth and variance at an unobserved location given the observed measurements. In turn, the variance can be also used to quantify the uncertainty on predictions.

$$\mathbf{f}_x | \mathbf{f}_y \sim \mathcal{N}(\mu_x + \Sigma'_{x,y} \Sigma_y^{-1} (f_y - \mu_y), \Sigma_x - \Sigma'_{x,y} \Sigma_y^{-1} \Sigma_{x,y}) \quad (4)$$

6. Conclusion

In this report, we examined the accuracy of HAND-NWM inundation maps in determining road network flood damage. Specifically, we compare HAND-NWM inundation maps from NWM stream flow forecasts on 08-28-17 against observed road network flood incidents on that same day. We show that the inundation maps report a positive inundation value at only 80% of observed road incident locations. This reflects the limitations and uncertainties that are associated with HAND-NWM inundation maps. To address these limitations, we propose a probabilistic kernel regression scheme based on Gaussian processes.

In addition, we identify certain data limitations that impact road network inundation mapping. Specifically, the resolution at which catchments are delineated results in irregularities when the catchment boundary is along a road segment. In addition, the HAND rasters downloaded from NFIE do not fully cover the area of interest and include negative values. Similarly, the NWM streamflow forecast data was not found for a small number of catchments.

Python modules for data processing and extraction are uploaded to the Hydroshare resource [17].

References

- [1] A. Asadabadi and E. Miller-Hooks, “Assessing strategies for protecting transportation infrastructure from an uncertain climate future,” *Transportation Research Part A: Policy and Practice*, vol. 105, 2017.
- [2] A. Michielsen, Z. Kalantari, S. W. Lyon, and E. Liljegren, “Predicting and communicating flood risk of transport infrastructure based on watershed characteristics,” *Journal of Environmental Management*, vol. 182, 2016.
- [3] N. H. Nielsen, M. R. A. Larsen, and S. F. Rasmussen, “Development of a screening method to assess flood risk on Danish national roads and highway systems,” *Water Science & Technology*, vol. 63, no. 12, 2011.
- [4] M. Pregnolato, A. Ford, S. M. Wilkinson, and R. J. Dawson, “The impact of flooding on road transport: A depth-disruption function,” *Transportation Research Part D: Transport and Environment*, vol. 55, 2017.
- [5] Z. Kalantari, M. Cavalli, C. Cantone, S. Crema, and G. Destouni, “Flood probability quantification for road infrastructure: Data-driven spatial-statistical approach and case study applications,” *Science of The Total Environment*, vol. 581-582, 2017.
- [6] A. Pedrozo-Acuña, G. Moreno, P. Mejía-Estrada, P. Paredes-Victoria, J. A. Breña-Naranjo, and C. Meza, “Integrated approach to determine highway flooding and critical points of drainage,” *Transportation Research Part D: Transport and Environment*, vol. 50, 2017.
- [7] I. Demir and W. F. Krajewski, “Towards an integrated Flood Information System: Centralized data access, analysis, and visualization,” *Environmental Modelling & Software*, vol. 50, 2013.
- [8] M. Bartos, B. Wong, and B. Kerkez, “Open storm: A complete framework for sensing and control of urban watersheds,” *Environmental Science: Water Research & Technology*, vol. 4, no. 3, 2018.
- [9] D. R. Maidment, “Open Water Data in Space and Time,” *Journal of the American Water Resources Association*, vol. 52, no. 4, 2016.
- [10] Z. Kalantari, A. Nickman, S. W. Lyon, B. Olofsson, and L. Folkesson, “A method for mapping flood hazard along roads,” *Journal of Environmental Management*, vol. 133, 2014.
- [11] H. Cai, W. Rasdorf, and C. Tilley, “Approach to Determine Extent and Depth of Highway Flooding,” *Journal of Infrastructure Systems*, vol. 13, no. 2, 2007.
- [12] “NOAA launches America’s first national water forecast model,” 2016.
- [13] D. J. Gochis, M. Barlage, A. Dugger, K. FitzGerald, L. Karsten, M. McAllister, J. McCreight, J. Mills, A. RafieeiNasab, L. Read, K. Sampson, D. Yates, and W. Yu, “The WRF-Hydro modeling system technical description, (Version 5.0).,” NCAR Technical Note, 2018.

- [14] X. Zheng, D. G. Tarboton, D. R. Maidment, Y. Y. Liu, and P. Passalacqua, “River Channel Geometry and Rating Curve Estimation Using Height above the Nearest Drainage,” *JAWRA Journal of the American Water Resources Association*, 2018.
- [15] D. R. Maidment, “Conceptual Framework for the National Flood Interoperability Experiment,” *JAWRA Journal of the American Water Resources Association*, vol. 53, no. 2, 2017.
- [16] Y. Y. Liu, D. R. Maidment, D. G. Tarboton, X. Zheng, and S. Wang, “A Cyber-GIS Integration and Computation Framework for High-Resolution Continental-Scale Flood Inundation Mapping,” *JAWRA Journal of the American Water Resources Association*, 2018.
- [17] C. Yahia, “A probabilistic model for predicting road network disruption by integrating large-scale flood forecasts, topographic characteristics, and local sensor data, HydroShare.” <http://www.hydroshare.org/resource/45122a8fc2ea4bc984d370e5b543d3d9>, 2018.
- [18] D. Arctur, E. Boghici, D. Tarboton, D. Maidment, J. Bales, R. Idaszak, M. Seul, and A. M. Castronova, “Hurricane Harvey 2017 Collection, HydroShare.” <https://doi.org/10.4211/hs.2836494ee75e43a9bfb647b37260e461>, 2018.
- [19] C. Fagan, “NFIE-Geo Texas-Gulf Region, HydrShare.” <http://www.hydroshare.org/resource/1d78964652034876b1c190647b21a77d>, 2015.
- [20] A. Castronova, “Hurricane Harvey NWM Subsetting Exercise, HydroShare.” <http://www.hydroshare.org/resource/3db192783bcb4599bab36d43fc3413db>, 2018.
- [21] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [22] P. Z. G. Qian, H. Wu, and C. F. J. Wu, “Gaussian process models for computer experiments with qualitative and quantitative factors,” *Technometrics*, vol. 50, no. 3, pp. 383–396, 2008.