NASA DEVELOP National Program
Maryland – Goddard

Spring 2020

Ellicott City Disasters II
Enhancing a Statistical Flood Risk Model to Continue Improving Early Warning Systems and Public Safety in Ellicott City, Maryland

DEVELOP Technical Report

Final Draft – April 2nd, 2020

Alina Schulz (Project Lead)
Scott Cunningham
Jonathan Donesky
Matthew Pruett

Dr. John Bolten, NASA Goddard Space Flight Center (Science Advisor)
Dr. Sujay Kumar, NASA Goddard Space Flight Center (Science Advisor)

Previous Contributors:
Terra Edenhart-Pepe
Julio Peredo
Caroline Resor
Callum Wayman

1. Abstract

As flooding events in the United States grow in frequency and intensity, the use of technological advancements and applied science are increasingly necessary for effective flood monitoring and warning systems. The NASA DEVELOP Ellicott City Disasters II project investigated the use of machine learning for applications in flood risk detection to support the improvement of early warning systems. To strengthen the efforts of the Howard County Office of Emergency Management (OEM) in building a more robust flood monitoring system, the project improved the original statistical flood risk model, FLuME (Flood Learning Model Environment), programmed by the first DEVELOP term. The enhancements incorporated an additional six years of precipitation and soil moisture data from the North American Land Data Assimilation System (NLDAS), modeled using Aqua Advanced Microwave Scanning Radiometer for EOS and Tropical Rainfall Measuring Mission (TRMM) Microwave Imager. These Earth observations were supplemented by stream gauge data from the OEM and the US Geological Survey. The resultant flood risk model FLASH (Flood Learning Environment and Severity Assessment Hub) was trained to evaluate input variables and predict stage height in Ellicott City in real time. The addition of an advanced deep learning framework known as long short-term memory improved the model's ability to capture relationships between variables. To assess the effectiveness of the new model, FLASH produced a model efficiency metric of 0.99, a significant improvement over the 0.85 value produced by the previous model. The project assisted the OEM in pursuing the integration of open data and NASA Earth observations into a threat matrix capable of informing near real-time decision making.

Keywords
remote sensing, flash flooding, machine learning, neural network, long short-term memory (LSTM), emergency management

2. Introduction

2.1 Background Information

Flooding is one of the most destructive natural disasters in the United States, costing billions of dollars in damages annually and endangering thousands of lives (Thakali, Bhandari, Kandissounon, Kalra, & Ahmad, 2017). Because flood frequency and severity are expected to increase due to climate variability, it is of great socio-economic importance to develop mechanisms to reduce vulnerability and improve resilience (Ganguly, Nahar, & Hossain, 2018; Hirabayashi et al., 2013). Flash floods, in particular, pose a unique threat to infrastructure and livelihoods due to their rapid onset, and increasing the warning time before a flood event can significantly reduce negative outcomes (Alipour, Ahmadalipour, Abbaszadeh & Moradkhani, 2020).

The Mid-Atlantic region of the United States is a frequent victim of flooding (Smith & Smith, 2015). Ellicott City, Maryland, shown in *Figure 1*, is particularly at risk because it lies in a steep river valley composed of a layer of low-drainage soil on top of low-permeability granite bedrock (United States Department of Agriculture Natural Resources Conservation Service [USDA NRCS], 2008). A contiguous United States-wide analysis that quantified relative flash flood severity on a scale of zero to one (where "flashiness" is the difference between the peak discharge and action stage discharge normalized by the flooding rise time and basin area) gave Ellicott City a maximum score of 1.0 (Saharia et al., 2017). The historic downtown is specifically situated at the junction of several tributaries and the Patapsco River, making it uniquely vulnerable to flash flooding. According to the National Weather Service, the city has been struck with two 1,000-year rainfall events in recent years: one in July 2016 and another in May 2018 (Halverson, 2018). The unusual frequency of flooding in the region is likely due to a confluence of factors, including extreme weather and urbanization, with the latter leading to an increase in impervious surface area (Gori, Blessing, Juan, Brody, & Bedient, 2019; Hirabayashi et al., 2013). The city's drainage systems are designed to manage 100-year floods, but the recent 1,000-year rainfall events have proven too violent for current infrastructure.

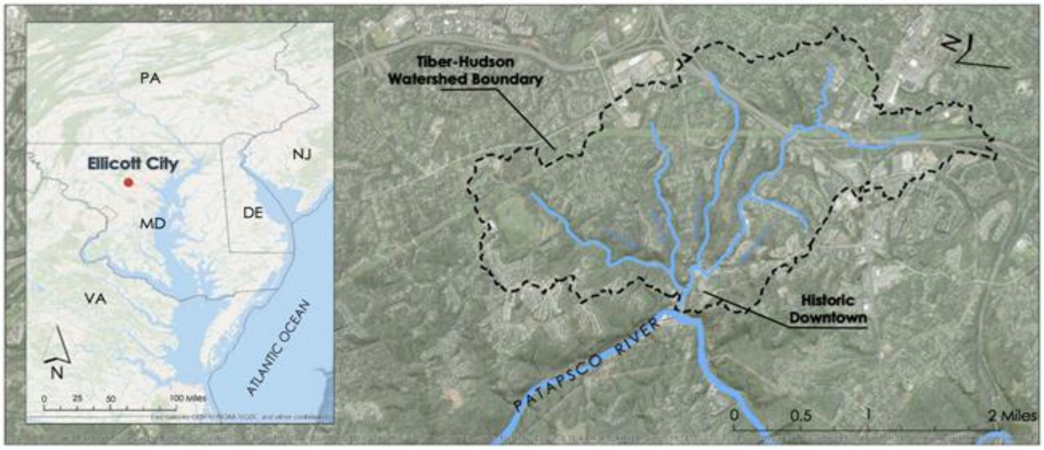

Figure 1. This map displays the project study area of Ellicott City in relation to the Patapsco River and the Tiber Hudson watershed.

Using machine learning applications for flood and risk predictions is a field of active research. Neural networks, a subset of machine learning models, are frequently used to analyze floods (Mosavi, Ozturk, & Chau, 2018). These models attempt to iteratively explore the importance of various input parameters on flood severity. They require training data to perform their analyses, which can be provided from ground-based and remote sensing data (see Appendix for a list of relevant studies using these methods). One particularly well-performing recurrent neural network used for modeling complex, non-linear runoff situations is known as the Long Short-Term Memory (LSTM) network (Xiang, Yan, & Demir, 2020). The advantage of the LSTM is its ability to learn long-term dependencies between variables provided as inputs and outputs of the network. This type of memory is particularly useful for modeling relationships between variables that do not have linear relationships in a time series, such as precipitation rates and soil moisture (Kratzert, Klotz, Brenner, Schulz, & Herrnegger, 2018).

The first term of this project created a statistical flood risk model called the Flood Learning Model Environment (FLuME) and confirmed that this machine learning tool provides a feasible way to integrate NASA Earth observations into public safety and emergency management protocols. FLuME's outputs helped identify upstream discharge as the strongest link in predicting downstream discharge. While additional inputs such as soil moisture and precipitation had the potential to bolster model performance, the model had to be optimized with a memory component to better incorporate these indicators of flood severity and to extend its predictive capability.

This project expanded the FLuME flood risk model by incorporating the deep learning LSTM architecture to enhance the predictive capability of the model. The data acquisition years were expanded to 2011 to 2020 to incorporate information from a 2011 flood event. The upgraded model, the Flood Learning Environment and Severity Hub (FLASH), identifies conditions necessary for potentially damaging stage levels. The results will be leveraged for a third term project to ingest socio-economic data and the flood severity model for a customized operational decision support system capable of disseminating warnings in Ellicott City and surrounding areas.

2.2 Project Partners & Objectives

This project was conducted in collaboration with the Howard County Government Office of Emergency Management (OEM), the Howard County Government Stormwater Management Division, and the National Oceanic and Atmospheric Administration (NOAA) - National Weather Service (NWS) Baltimore-Washington Weather Forecast Office. OEM is responsible for managing and overseeing emergency preparedness throughout Howard County and was interested in the potential of applying remote sensing to

decision-making processes. This project focused on enhancing a statistical model to predict flood severity in Ellicott City. The team approached this through machine learning, combining NASA Earth observations with a variety of ground data to train the model to recognize characteristics of a severe flooding event prior to occurrence. The primary benefits of the machine learning approach for the end users include the ease of replicability with near-real-time inputs and the model's predictive capability despite limited data inputs.
In accordance with the first project objective, the team finalized the statistical flood severity model in FLASH using Python to integrate the LSTM network. Another main objective focused on filling data gaps identified during the first term. This involved updating input datasets to include precipitation, soil moisture, discharge and stage data from 2011 to 2020 and writing new data acquisition scripts. Lastly, the team pursued the analysis of each input variable's value to the model's performance and predictive ability.

3. Methodology

3.1 Data Acquisition
The team used a variety of both NASA meteorological data and *in situ* data to train and test the machine learning model. NASA meteorological data provided soil moisture and precipitation while discharge and stage on the Patapsco were provided via *in situ* gauges operated by the Howard County OEM and the United States Geological Survey (USGS).

3.1.1 North American Land Data Assimilation System
The North American Land Data Assimilation System Phase 2 (NLDAS-2) combines multiple sources of observations, such as precipitation gauge data, radar measurements, and observations from NASA satellites like AQUA Advanced Microwave Scanning Radiometer for EOS (AMSR-E) and Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), to produce climatological field measurements at or near the Earth's surface (Table 1). The precipitation field is derived from NOAA Climate Prediction Center (CPC) gauge-only precipitation measurements and disaggregated to an hourly time scale using radar-based estimates or modeling techniques. The non-precipitation fields are derived from the analysis fields of the National Centers for Environmental Prediction (NCEP), spatially interpolated to a finer resolution and temporally disaggregated to hourly frequency.

Table 1
This table lists the NASA resources from which input data was acquired for training FLASH.

Source	Level	Parameters	Digital Object Identifier	Available Years
NLDAS-2 Primary Forcing Dataset	4	Precipitation, Convective Available Potential Energy, Surface Atmospheric Pressure	10.5067/6J5LHHOHZHN4	1979 to 2020
NLDAS-2 VIC Land Surface Model	4	Soil Moisture	10.5067/ELBDAPAKNGJ9	1979 to 2020

Both NLDAS-2 models utilize data from NASA satellites to derive their values.

The team extracted and used the following hourly fields from the NLDAS-2 Primary Forcing dataset: total precipitation, convective available potential energy, and surface atmospheric pressure. Google Earth Engine was used to find the average value of each field over the extent of the Tiber Hudson watershed and extract the values. Soil moisture measurements were acquired from the NLDAS-2 Variable Infiltration Capacity (VIC) model dataset. The VIC model is a grid-based hydrologic model that derives soil moisture in the top 100 centimeters of the soil. Hourly measurements of soil moisture were extracted for the Tiber Hudson watershed from the NASA Goddard Earth Science Data and Information Services Center (GES DISC). All data used to develop the model were collected for the period between January 2011 and January 2020.

3.1.2 On-site Gauges

The Howard County OEM and the USGS operate several *in situ* gauges in and around Ellicott City that provide data of high temporal resolution that proved vital to our research (Table 2). Howard County operates three gauges within the bounds of the study area: one measuring stage height and discharge of the Patapsco River; one measuring the discharge of the Tiber Run, a stream that runs through the center of the watershed; and a rain gauge within the city. These data were obtained using the Howard County Web Emergency Operations Center (WebEOC), or OneRain, portal. The USGS operates two stream gauges on the Patapsco River, one upstream from the city at Hollofield, and one downstream at Catonsville. These values were obtained from the National USGS Water Data Portal. *Figure 2* shows the locations of all of the gauges.

Table 2
The in-situ datasets and parameters used to train the model and evaluate its performance.

Parameter	Provider	Source	Temporal Resolution	Use
Stage Height, Discharge	USGS	USGS Surface Water Database	5 min	Model Training & Evaluation
Stage Height, Discharge, Precipitation	Howard County OEM	OneRain Portal, Howard County OEM	15 min	Model Training & Evaluation

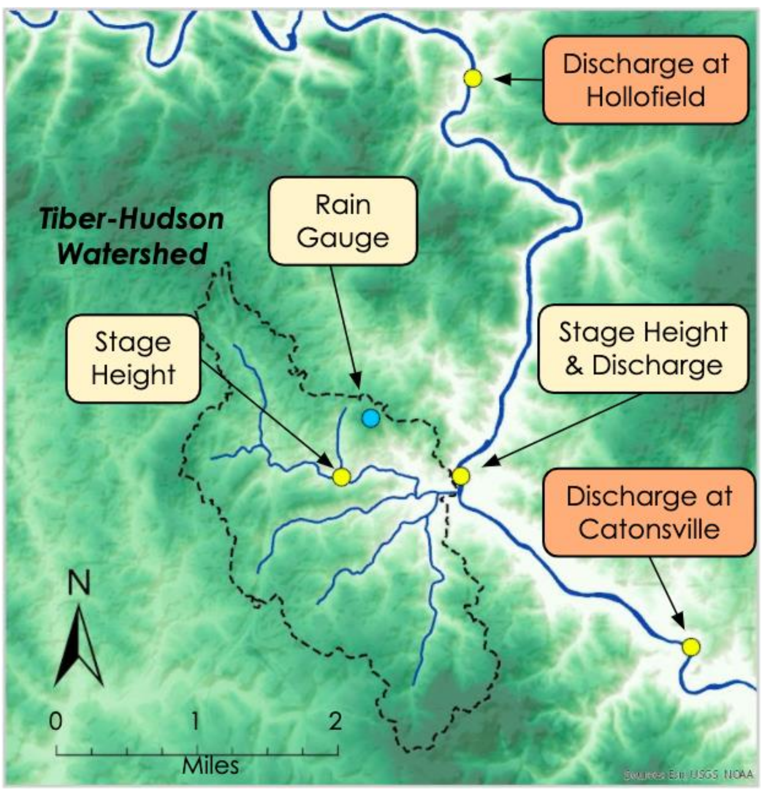

Figure 2. A map showing the boundary of the Tiber-Hudson watershed in relation to the Patapsco River and the USGS stream gauges used in this study.

3.2 Data Processing

Several Python scripts developed for FLuME were replicated or modified to process new data for FLASH. One script was used to extract precipitation data from NLDAS files and compile these values in a two-column comma-separated file, with the columns representing the timestamp and precipitation amount. This method was repeated for the NLDAS soil moisture product. A second script was used to collect precipitation, soil moisture, and discharge data and merge columns based on shared timestamps. NLDAS data are hourly, while discharge data are occasionally sub-hourly; in these cases, the script averaged discharge values for the previous hour. Last, the script checked for and filled missing values.

Taking in this merged dataset, FLASH allows the users to designate a point in the timeline at which to split the data into training and testing subsets. The model learns the relationship between inputs and outputs in the training subset and attempts to predict outputs given inputs in its testing data subset. This team set the split such that the years 2016 and 2017 were used for training and 2018 and 2019 for testing. This portioning was chosen so that each subset would contain one of the two major floods, which occurred in 2016 and 2018.

Data were initially collected for an extended timeline of observations from January 2011 onward. However, the additional seven years of data caused a significant increase in training time and therefore was shortened to four years for the final run (2016 to 2020). Nevertheless, more data inputs are essential for regularizing the model to forecast floods accurately under a broader range of conditions. With that in mind, additional Python scripts were developed to process data from similarly small, flashy watersheds in the Midwest and Northeast US.

FLASH performs two final preprocessing steps within the model. First, it reshapes the linear time series data into a sliding window of arrays of inputs and outputs. Second, it reshapes those arrays into three-dimensional tensors that go into the PyTorch-supported LSTM layers of the neural network.

In the sliding window, the inputs consist of an array of values representing each input variable at each timestep between the present time and a user-specified timestep in the past. In *Figure 3*, this is represented by all the data contained within the bounded box labeled Input Data. The output paired to each window of inputs is a single value at a user-specified timestep in the future, represented in *Figure 3* by the point labeled Predicted Value. The width of the window can be adjusted to use more or fewer past timesteps as inputs, and the forecast point can be moved to predict nearer or further into the future. This team tested predictions at 15-minute intervals ranging between 15 minutes to 3 hours ahead.

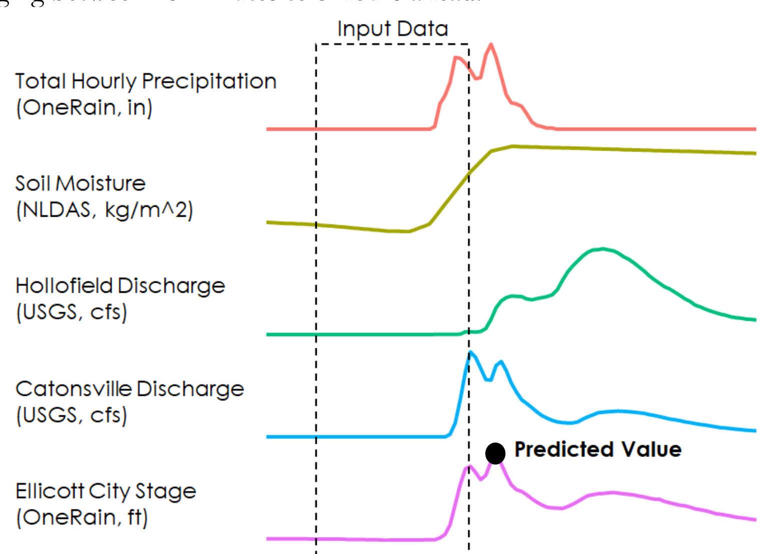

Figure 3. Time series of input variables with example box showing input data and example predicted data point.

3.3 Data Analysis
The LSTM model developed during this term is a type of recurrent neural network (RNN) commonly used in machine learning but is a novel approach for application within the field of hydrology. The LSTM improves upon a typical RNN by creating a feedback connection where the output of one LSTM cell is included as an input to the next cell (*Figure 4*).

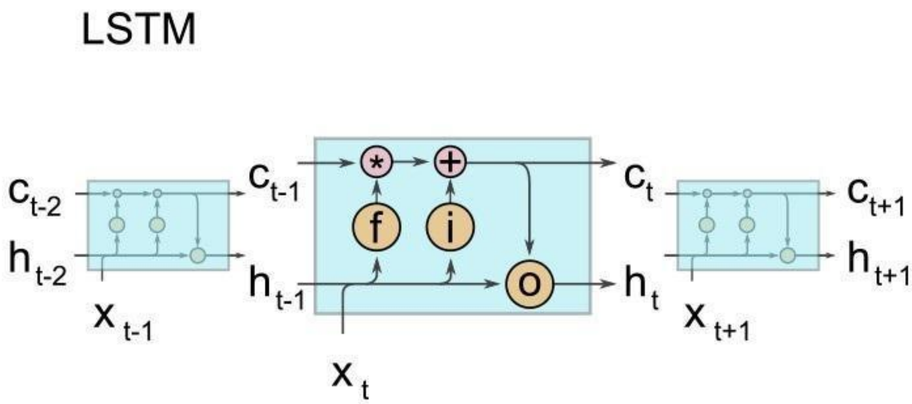

Figure 4. Illustration of the internals of a LSTM cell, where ***f*** stands for the forget gate, ***i*** for the input gate, and ***o*** for the output gate. $\boldsymbol{c_t}$ denotes the cell state at time step *t* and $\boldsymbol{h_t}$ is the hidden state. Figure taken from Kratzert et al, 2018.

To evaluate the effectiveness of the model, the team used the Nash-Sutcliffe efficiency coefficient (NSE), a commonly used metric for evaluating hydrological model performance (Equation 1). NSE is defined by:

$$\mathrm{NSE}=1-\frac{\sum_{t=1}^{T}\left(Q_m^t-Q_o^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t-\bar{Q}_o\right)^2} \tag{1}$$

where Q_o is the mean of observed discharges, and Q_m is modeled discharge. Q_{ot} is observed discharge at time t. Model predictors were evaluated by iteratively removing each predictor and training the model and evaluating the model over 1000 epochs (iterations) with the same model parameters to get a sample of NSE's for the model without each predictor. The NSE allowed the model's performance and its input parameters to be evaluated in addition to making it comparable to other models, such as FLuME.

In the development of FLASH, we tested two common packages for developing machine learning models in Python, TensorFlow, and PyTorch. The TensorFlow model exhibits overfitting but may be a viable alternative if overfitting issues are resolved. The results that are shown in this report utilize the model developed in PyTorch.

4. Results & Discussion

4.1 Model Performance
The FLASH model was trained on data from 2016 to 2017 and tested on data from 2018 to 2019 (*Figure 5*). Over the testing period, the model displayed a maximum NSE of 0.99, an improvement over FLuME, which had an NSE of 0.85 over the same period.

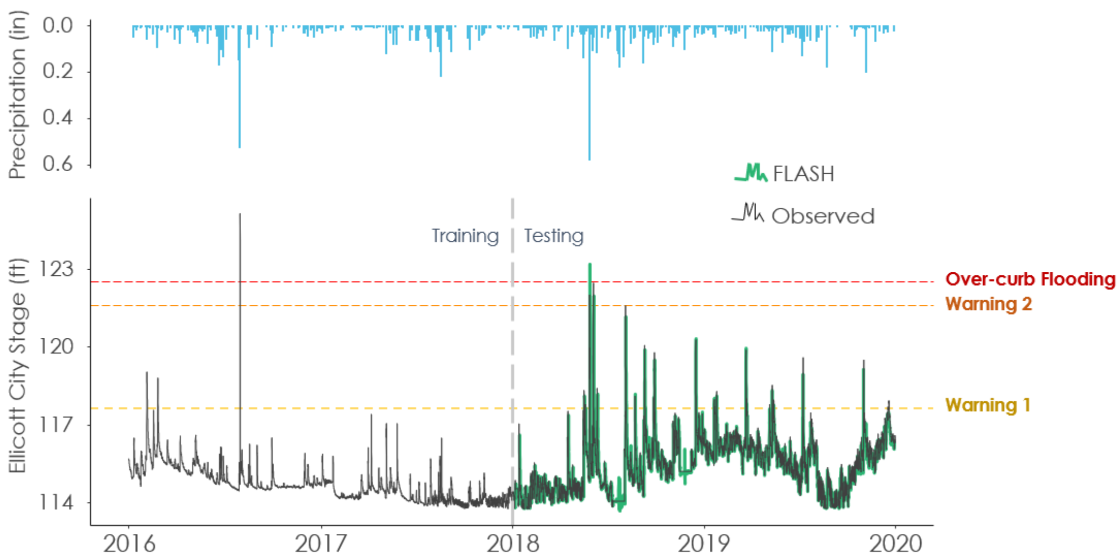

Figure 5. Time series of modeled and observed discharge at Catonsville with the gray dashed vertical line indicating the split between training and testing data.

The switch from a feed-forward network to an LSTM network allowed FLASH to predict the flood stage on the Patapsco in Ellicott City. The 2018 flood was the landmark feature from the testing set that we used to evaluate the model (*Figure 6*). The modeled stage shows the average of 10 model runs predicting 15 minutes ahead bounded by a 95 percent confidence interval. The model misses the first peak but closely models the second peak.

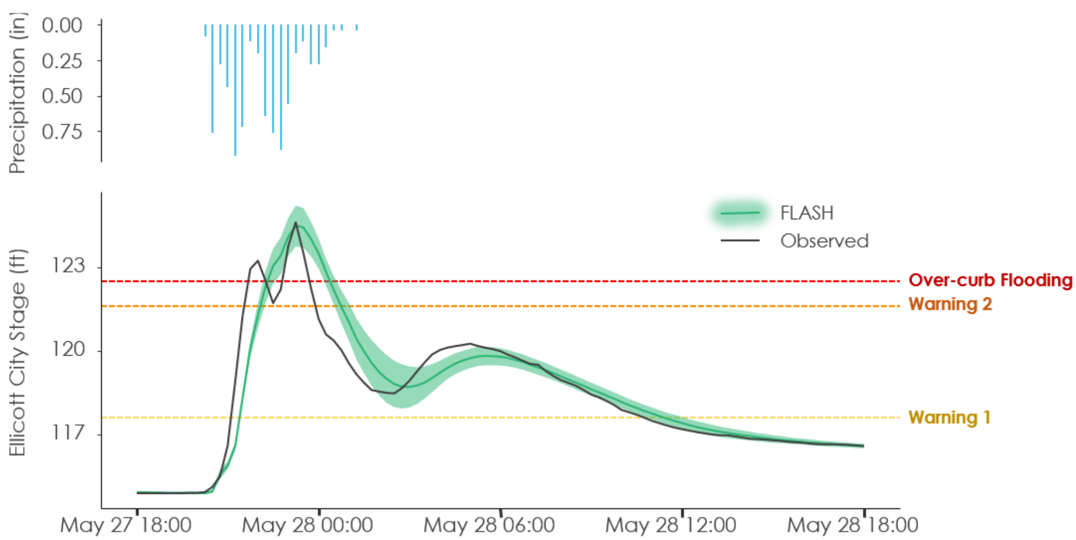

Figure 6. Time series of the 2018 Ellicott City flood with modeled and observed stage height.

4.2 Comparison to FLuME

Because FLASH differs fundamentally in its operation to FLuME, direct comparison is difficult but we can control some parameters including input data and temporal resolution to make some comparisons. FLuME is a feed-forward network and uses data at a discrete time step to model a value at that same time step; for example, the original model used hourly soil moisture, precipitation, and upstream discharge at Hollofield to model discharge downstream at Catonsville. This differs from the operation of FLASH described above, so to make comparisons we set up FLASH using the same input dataset used with FLuME. To model discharge one hour into the future, 24 hours of input data prior to that prediction was used. With the model configured in this fashion, FLASH was able to achieve an average NSE of 0.95 while FLuME had an average NSE of 0.85. It should be noted that although the same input dataset was used with FLASH, the advantages of an LSTM allowed us to use the previous record of discharge to predict the next value increasing overall model efficiency. In addition, the dataset used to maintain consistency contained time zone errors which led to precipitation and soil moisture lagging the discharge data by 5 hours. This offset may have led to less than ideal model performance in both models.

4.3 Predictive Capability

The model is currently set up to operate as a many-to-one model, meaning that it takes an input sequence of variable length and outputs a singular value. This allows the model to utilize 24 hours of data to predict 15 minutes or an hour into the future but does not allow for multiple predictions. A many-to-many or sequence-to-sequence model may take in 24 hours of input data and output a predictive value every 15 minutes until the next hour (i.e. 4 total values). While a sequence-to-sequence model is preferable, in the model's current form we can vary the prediction timestep to gain an understanding of how the model's performance might degrade as we predict farther out. To assess this function, the model was evaluated using 24 hours of input data to predict 15, 30, and 45 minutes out (*Figure 7*).

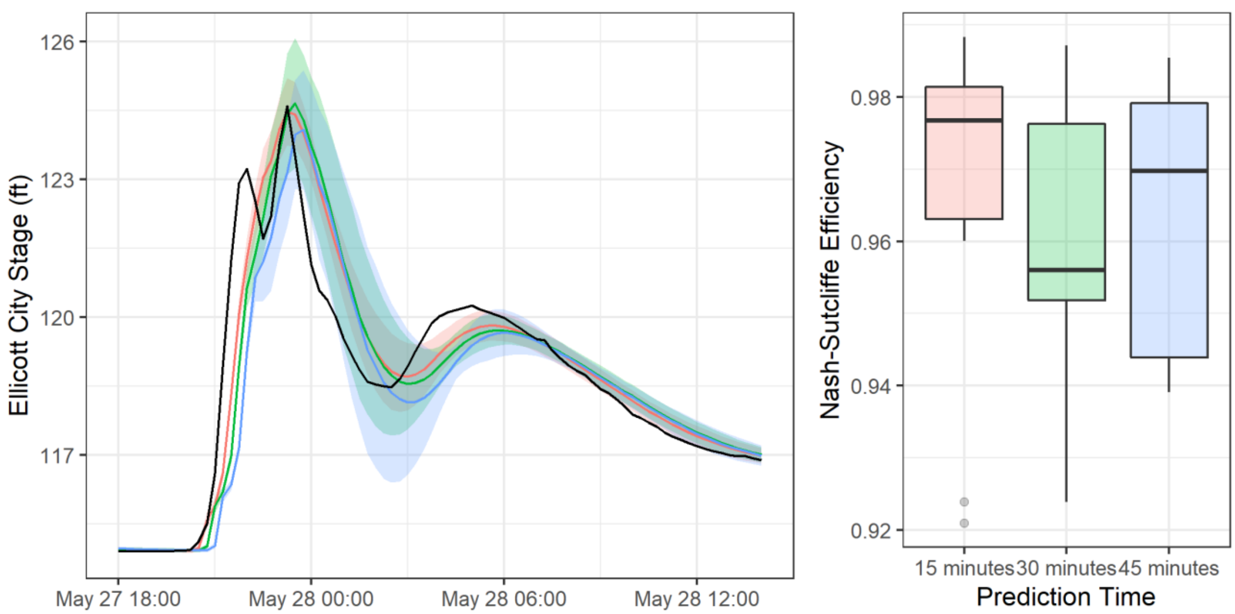

Figure 7. left: Time series of the 2018 Ellicott City flood with 3 lines for predicting 15, 30, and 45 minutes ahead (red, green, and blue, respectively). *right:* Boxplot of Nash-Sutcliffe values for the associated prediction times.

As expected, the model performance degrades as we predict farther out. This can be seen as the Nash-Sutcliffe efficiency (NSE) value drops from a mean of 0.966 to 0.960. As we go farther out the mean NSE increases slightly to 0.963 but so does the standard deviation. Additional model runs may be needed to determine the degradation in model performance between 30 and 45 minutes out. We can see similar performance degradation in the time series plot shown on the left in the figure above. When the flood starts, all three models lag behind the initial rise in stage. This lag is amplified for predictions farther out. At the first peak when the observed stage is 123 feet, the 15-minute model predicts 120 feet, the 30 minute 119 feet and 45 minute 117 feet.

We also performed a simple sensitivity analysis to identify which driver variable provided the most value to the model. The sensitivity analysis of each contributing input factor indicated that some predictors such as "number of hours since the last rain event" are not useful in this model and can be abandoned. It also showed that both gauged and modeled precipitation from NLDAS were equally valid as inputs. The best predictor for stage was lagged stage. This is unsurprising as stage height is not expected to vary wildly over 15 minutes and should be self-correlated. This finding is also a strength of FLASH, by integrating this predictor the model performance is greatly increased.

4.4 Future Work

While FLASH is an improvement on the previous term's model, and the introduction of a predictive component greatly enhances the model's capability, it continues to underestimate peak discharges during large events which are exactly the type of events we are looking to capture. This behavior is typical for hydrological models because of the relative infrequency of high discharge events (Kratzert, 2018). The model shown here was only able to train on one event (2016 flood) to predict the flood in 2018. Using precipitation forecasts from the National Weather Service as an input to train the model could dramatically increase the model's stage forecasting ability.

Nonetheless, the team's contributions as the second term helped highlight the direction in which the final term of the project could develop. One potential next step toward model refinement would include training and testing the model using the expanded time range of data inputs (2011 to 2020). Another approach to developing the model could include using the scripts developed during this term to gather data from the ten other "flashy" watersheds identified to have similar characteristics to that of the Ellicott City watershed. This would help make the model capable of performing under a broader range of conditions. Future work could integrate the model into a graphical user interface for added ease of use. Additionally, considering socio-economic factors and the three flood severity stages designated by the OEM, work toward assimilating FLASH into the actionable risk scorecard would bolster existing decision-making frameworks.

5. Conclusions

Modeling rainfall-runoff behavior in urban watersheds involves complex non-linear relationships with diverse temporal and spatial dynamics. The team approached this challenge from the machine learning angle, implementing a deep learning technique known as Long Short-Term Memory (LSTM) with known success in hydrological modeling scenarios. The LSTM network in the FLASH model programmed during this term incorporated a forecasting feature capable of using the past 24 hours of real-time inputs to predict stage height on the Patapsco River at Ellicott City. Adjusting the temporal frequency of the model's input data from hourly to 15-minute intervals allowed the model to adjust the prediction window to 15 minutes, 30 minutes, 45 minutes, or any 15-minute interval from the current time step. Enhancements to data acquisition and model architecture for FLASH improved model performance by 14%, as the NSE increased to 0.99 from FLuME's 0.85 performance metric. Nonetheless, the team identified opportunities for additional model improvement, such as programming a sequence-to-sequence structure to strengthen the LSTM network's existing predictive capabilities. Community concerns related to public safety during flash flood events have been addressed using this model since it enables officials to recognize the characteristics of a flood event with greater response times. Partners can benefit from the accuracy, flexibility, and near real-time predicting capacity found in the flood severity model built into FLASH. Once integrated into the OEM's threat response matrix, this end product will enhance public safety by refining early warning systems with stage height predictions in Ellicott City.

6. Acknowledgments

The team would like to thank everyone that contributed to the progress made over the course of this term: our Science Advisors at Goddard Space Flight Center, Dr. John Bolten, Dr. Sujay Kumar, and Perry Oddo. We would like to thank our DEVELOP center lead, Darcy Gray for her continued support, the Ellicott City Disasters I team for the progress made during the first term, and Brian Cleary of the Howard County

Stormwater Management Division for his insight during our visit to Ellicott City. Lastly, we appreciate the input of our partners at the Howard County Government and the NOAA- NWS Washington-Baltimore Forecast Office.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C.

7. Glossary

1,000-year rainfall event – hydrologic event that has a 0.1 percent chance of happening in a given year based on its extreme magnitude
Earth observations – satellites and sensors that collect information about the Earth's physical, chemical, and biological systems over space and time
Flashy Watershed – experiences a rapid increase in flow shortly after the onset of a precipitation event, and an equally rapid return to base conditions shortly after the end of the precipitation event
Google Earth Engine API – cloud-based platform for geospatial analysis
Hidden layer – a layer between model input and output layers where the model has generated an equation that assigns weights for each input to approximate the output
Long Short-Term Memory (LSTM) – an artificial recurring neural network used in machine learning that allows for memory retention in a learning environment
Machine learning – a branch of artificial intelligence in which a computer system employs algorithms and statistical models to effectively extract meaning inferred from patterns related to a specific task or inquiry; this process is typically iterative and involves a learning component by which inputs are acted upon and outputs are fed back through statistical models to improve outputs
Nash-Sutcliffe Efficiency (NSE) coefficient – common metric used to evaluate predictive power of hydrological models, expressed as a value between negative infinity and 1.0 where 1.0 corresponds to a perfect match between modeled and observed values
Neural network – computing systems inspired by animal brains that "learn" by completing example tasks using artificial neurons called "nodes".
Sequence-to-sequence structure – deep learning method that pairs with an LSTM to enhance forecasting by considering long term dependencies and multiple outputs
Stage – level (measured by height, often in feet) of water in a channel
Statistical model – mathematical model that embodies a set of assumptions concerning data generation and can be used to make predictions
Stream discharge – the rate at which water passes a specified point, measured in volume per second
Tensor – A multi-dimensional array which supports the machine learning training process
United States Geological Survey (USGS) – federal agency responsible for collecting data from and maintaining the river gauges used in this study

8. References

Alipour, A., Ahmadalipour, A., Abbaszadeh, P., & Moradkhani, H. (2020). Leveraging machine learning for predicting flash flood damage in the Southeast US. *Environmental Research Letters, 15*. https://doi.org/10.1088/1748-9326/ab6edd

Ganguly, K., Nahar, N., & Hossain, B. (2018). A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh. *International Journal of Disaster Risk Reduction, 34,* 283-294. https://doi.org/10.1016/j.ijdrr.2018.12.002

Gori, A., Blessing, R., Juan, A., Brody, S., & Bedient, P. (2019). Characterizing urbanization impacts on floodplain through integrated land use, hydrologic, and hydraulic modeling. *Journal of Hydrology*, *568*, 82-95. https://doi.org/10.1016/j.jhydrol.2018.10.053

Halverson, J. B. (2018). "Flood City, USA": The sorrowful tale of Ellicott City, Maryland. *Weatherwise*, *71*(5), 48. Retrieved from https://search.proquest.com/docview/2159933219?accountid=28155

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., … Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816-821. https://doi.org/10.1038/nclimate1911

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall--runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences, 22,* 6005-6022. https://doi.org/10.5194/hess-22-6005-2018

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 40. https://doi.org/10.3390/w10111536

NASA/GSFC, Greenbelt, MD, USA, NASA Goddard Earth Sciences Data and Information Services Center (GES DISC). DOI: https://doi.org/10.1029/2010EO340001

Saharia, M., Kirstetter, P.-E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2016). Mapping flash flood severity in the United States. *Journal of Hydrometeorology*, *18*(2), 397–411.

Smith, B. K., & Smith, J. A. (2015). The flashiest watersheds in the contiguous United States. *Journal of Hydrometeorology, 16*, 2365-2381. https://doi.org/10.1175/JHM-D-14-0217.1

Thakali, R., Bhandari, R., Kandissounon, G. A., Kalra, A., & Ahmad, S. (2017). Flood risk assessment using the updated FEMA floodplain standard in Ellicott City, Maryland, United States. *Proceedings of the World Environmental and Water Resources Congress 2017*, (pp. 280-291). Sacramento, California: American Society of Civil Engineers. Retrieved from https://digitalscholarship.unlv.edu/fac_articles/449

United States Department of Agriculture Natural Resources Conservation Service (USDA NRCS) (2008). *Soil Survey of Howard County, Maryland*. Retrieved from https://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx

Vrije Universiteit Amsterdam (Richard de Jeu) and NASA GSFC (Manfred Owe) (2011), AMSR-E/Aqua surface soil moisture (LPRM) L3 1 day 25 km x 25 km ascending V002, Edited by Goddard Earth Sciences Data and Information Services Center (GES DISC) (Bill Teng), Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Accessed, March 30, 2020. doi:10.5067/X3K5V3NNLYAV

Vrije Universiteit Amsterdam (Richard de Jeu) and NASA GSFC (Manfred Owe) (2012), TMI/TRMM surface soil moisture (LPRM) L3 1 day 25 km x 25 km nighttime V001, Edited by Goddard Earth Sciences Data and Information Services Center (GES DISC) (Bill Teng), Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Accessed March 30, 2020. doi:10.5067/GWHRZEL8SA21

Xiang, Z., Yan, J., Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Research Research, 56*(1), 910-922. https://doi.org/10.1029/2019WR025326

9. Appendix

List of Relevant Machine Learning Studies

Kwak, Y.-J., Pelich, R., Park, J., & Takeuchi, W. (2018). Improved flood mapping based on the fusion of multiple satellite data sources and in-situ data. *The Institute of Electrical and Electronics Engineers, Inc. (IEEE) Conference Proceedings*, pp. 3521–3523. DOI: 10.1109/IGARSS.2018.8517336

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 40. https://doi.org/10.3390/w10111536

Petković, V., & Kummerow, C. D. (2015). Performance of the GPM passive microwave retrieval in the Balkan Flood Event of 2014. *Journal of Hydrometeorology*, *16*(6), 2501-2518. https://doi.org/10.1175/jhm-d-15-0018.1

Suriya, S., Mugral, B., & Nelliyat, P. (2012). Flood damage assessment of an urban area in Chennai, India, part I: Methodology. *Natural Hazards, 62*(2)*,* 149-167. https://doi.org/10.1007/s11069-011-9985-3

Wan, L., Liu, M., Wang, F., Zhang, T., & You, H. J. (2019). Automatic extraction of flood inundation areas from SAR images: A case study of Jilin, China during the 2017 flood disaster. *International Journal of Remote Sensing*, *40*(13), 5050-5077. https://doi.org/10.1080/01431161.2019.1577999

Zaji, A. H., Bonakdari, H., & Gharabaghi, B. (2019). Applying upstream satellite signals and a 2-D error minimization algorithm to advance early warning and management of flood water levels and river discharge. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(2), 902-910. https://doi.org/10.1109/tgrs.2018.2862640