

AUTOMATED MAPPING OF FLOOD EVENTS IN THE MISSISSIPPI RIVER BASIN UTILIZING NASA EARTH OBSERVATIONS

Mercedes Bartkovich, Project Coordination Fellow

Helen Blue Baldwin-Zook, Assistant Center Lead

Dashiell Cruz, Team Member

Nicholas McVey, Team Member

Chris Ploetz, Team Member

Olivia Callaway, Team Member

NASA DEVELOP Program – SSAI

320 Sparkman Dr.

Huntsville, AL 35805

bartkovichm@gmail.com

helenbluebaldwin@gmail.com

cruz.dashiell@gmail.com

nam0014@uah.edu

ploetcj@gmail.com

ogc0003@uah.edu

ABSTRACT

The Mississippi River Basin is the fourth largest drainage basin in the world, and is susceptible to multi-level flood events caused by heavy precipitation, snow melt, and changes in water table levels. Conducting flood analysis during periods of disaster is a challenging endeavor for NASA's Short-term Prediction Research and Transition Center (SPoRT), Federal Emergency Management Agency (FEMA), and the U.S. Geological Survey's Hazards Data Distribution Systems (USGS HDDS) due to heavily-involved research and lack of manpower. During this project, an automated script was generated that performs high-level flood analysis to relieve the workload for end-users. The script incorporated Landsat 8 Operational Land Imager (OLI) tiles and utilized computer-learning techniques to generate accurate water extent maps. The script referenced the Moderate Resolution Imaging Spectroradiometer (MODIS) land-water mask to isolate areas of flood induced waters. These areas were overlaid onto the National Land Cover Database's (NLCD) land cover data, the Oak Ridge National Laboratory's LandScan data, and Homeland Infrastructure Foundation-Level Data (HIFLD) to determine the classification of areas impacted and the population density affected by flooding. The automated algorithm was initially tested on the September 2016 flood event that occurred in Upper Mississippi River Basin, and was then further tested on multiple flood events within the Mississippi River Basin. This script allows end users to create their own flood probability and impact maps for disaster mitigation and recovery efforts.

KEYWORDS: Mississippi River Basin, disaster relief, Landsat 8 OLI, flood impact, automated

INTRODUCTION

The Mississippi River Basin is the world's fourth largest drainage basin, draining 41% of the contiguous U.S. to the Gulf of Mexico (U.S. Army Corps of Engineers, n.d.). The millions of residents within the Mississippi River Basin area, which includes portions of 31 states, undergo social and economic damage each time a flood event occurs within the Basin (Kusky, 2013). Since 2010, flooding has incurred more than \$34 billion in damages within the basin (Gerencer, 2015) and has continuously been an issue for farmers within the Midwest region (Garber, 2008). In a recent flood event in September 2016, over 10,000 people were evacuated from flood areas in Iowa and Wisconsin (Blau, 2016), and an estimated tens of thousands of acres of crops were lost as a result of flooded farmland (Eller, 2016; Towne, 2016). This project addressed the needs of communities in the basin by providing more precise data and maps to disaster response and relief organizations.

The objective of this project was to create a Python script that provided the user the ability to conduct flood analysis of any given area and date across the United States. This project partnered with FEMA, the USGS Hazards Data Distribution Systems (HDDS), and NASA's Short-term Prediction Research and Transition Center (SPoRT).

FEMA is the primary disaster relief organization in the United States with a mission to “support...citizens and first responders to ensure that as a nation we work together to build, sustain, and improve...capability to prepare for, protect against, respond to, recover from and mitigate all hazards” (Federal Emergency Management Agency, 2016). The USGS HDDS provides a multitude of data on various natural hazards and ecosystems to scientists worldwide, and NASA SPoRT is a NASA project designed to distribute unique observations and data acquisitions to enhance understanding of various weather events.

The project partners expressed a need for an automated, easy to operate script for flood analysis due to shrinking budgets and lack of manpower to conduct such extensive studies. The USGS HDDS and FEMA expressed an interest in the analysis of flood that occurred in Cedar Rapids, Iowa during September 2016. The automated script was initially tested on the September 2016 flood, and was then used to analyze multiple flood events throughout the entire Mississippi River Basin (Figure 1).

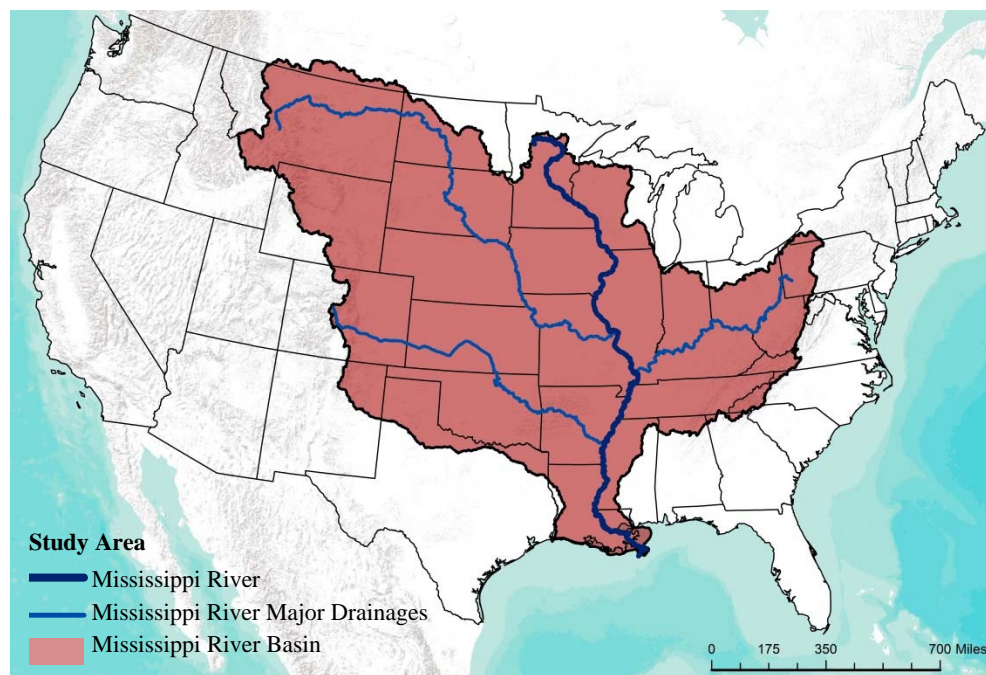


Figure 1. Project study area – the Mississippi River Basin

METHODOLOGY

Data Acquisition

The Python script was tested on the Cedar Rapids, Iowa flood event that reached its peak on September 23, 2016. This project primarily utilized Landsat 8 OLI multispectral imagery of flood events at a 30m spatial resolution. Landsat 8 OLI captured data soon after the flood peak, on September 26, 2016, providing extremely clear images of the flood extent. We acquired these images manually using USGS’s Earth Explorer and Global Visualization Viewer and saved to the workspace defined by the user prior to flood analysis. In an earlier study, we determined that McFeeters (1996) Normalized Difference Water Index (NDWI) identified flood water more accurately than other vegetation indices. To produce this NDWI, bands that represent reflected green light and near infrared radiation are necessary; for Landsat 8 OLI, this includes band 3, which records visible green light from 0.53 μm to 0.59 μm wavelengths, and band 5, which records near infrared radiation from 0.85 μm to 0.88 μm wavelengths (USGSa, 2016). We also used band 9, known as the cloud mask, in our analysis to remove unusable data.

At the start of each flood analysis, the script set local variables, retrieved relevant imagery from the workspace with the Python glob.glob function, and imported Arcpy, including 3D Analyst and Spatial Analyst extensions (Figure 2). Local variables included a water mask, as well as agriculture, population, and infrastructure data. The Global Water Mask was obtained from the University of Maryland and NASA’s Global Land Cover Facility

website, and provided a complete map of surface water at 250m spatial resolution. National Land Cover Database 2011 data were obtained from the Multi-Resolution Land Characteristics consortium website. The 2011 NLCD dataset categorizes different types of land cover into 16 classes at a 30m spatial resolution and provided valuable information regarding flood impact on agriculture. Homeland Infrastructure Foundation-Level Data were acquired from the Homeland Infrastructure Foundation-Level Data website. The HIFLD were used to map flood impact on infrastructure. LandScan 2014 data were downloaded from the Oak Ridge National Laboratory and were used to analyze flood impact on local populations.

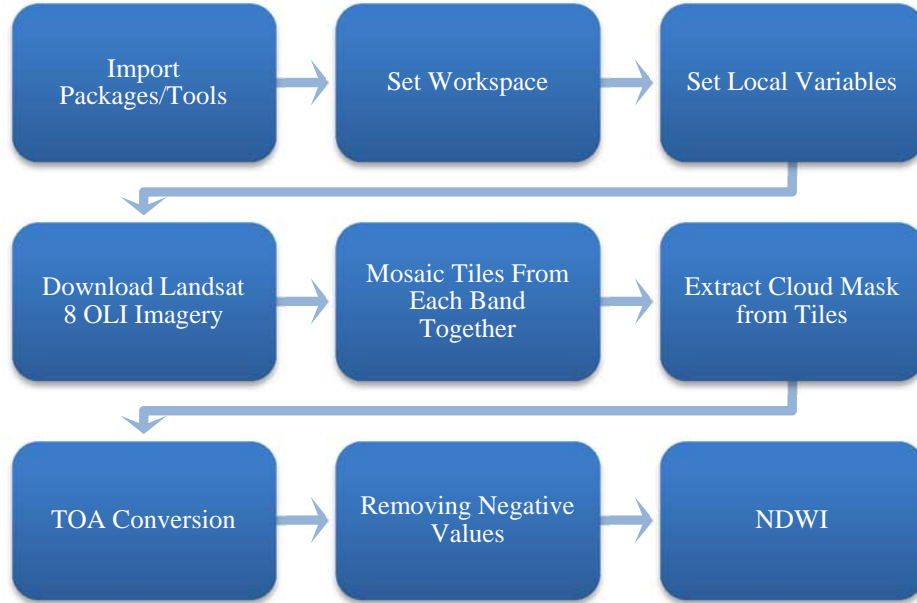


Figure 2. Algorithm used to create the data acquisition and processing sections of the Flood Probability Python script

Data Processing

After data acquisition, manually downloaded Landsat images were processed resulting in an NDWI (Figure 2). In order to conduct analysis on a flooded area larger than a single Landsat 8 OLI tile at one time, the individual tiles were mosaicked into a single tile according to band type using a “for” function loop and “Mosaic to New Raster” tool from the Arcpy package. Once mosaicked, the cloud mask (band 9) was extracted from the mosaicked bands 3 and 5. To ensure an accurate value for reflectance during processing, bands 3 and 5 were converted from Digital Numbers (DN) to top of atmosphere (TOA) reflectance. While surface reflectance would have produced even more accurate results, the turn-around time required for a Landsat surface reflectance product is three to five days (USGS, 2016a). We decided that even a short lag would reduce the benefits of the script to the project partners and proceeded with TOA reflectance.

Digital Numbers are the actual amount of radiance measured by the satellite, whereas reflectance is a property of the object being observed and represents the ratio of light reaching and being reflected from the target. TOA refers to the value of this reflectance at the top of the atmosphere, as opposed to at the Earth’s surface. This conversion was necessary as DN does not account for changes in the radiance measured. The TOA reflectance is calculated from the DN values of the radiance image using Equation 1:

$$\rho\lambda' = M\rho Q_{cal} + A\rho \quad (1)$$

where $\rho\lambda'$ is TOA planetary reflectance, $M\rho$ is band-specific multiplicative rescaling factor from the metadata, Q_{cal} is quantized and calibrated DN values, and $A\rho$ is band-specific additive rescaling factor from the metadata (USGS, 2016b).

Once the TOA conversion was complete, invalid image areas were eliminated by using the Raster Calculator through the Arcpy module to remove negative values. This step ensured the proper execution of a NDWI. The NDWI was calculated using Equation 2:

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (2)$$

where Green is visible green light and NIR is near-infrared radiation. NDWI helps distinguish water from non-water by using the green and NIR bands to produce an image where negative values are typically non-water features and positive values are typically open water areas (USGS, 2016c). NDWI outputs tend to appear cleaner with less noise, and are not as influenced by seasonality (e.g. growing seasons) as other spectral signatures (Gao, 1996).

Data Analysis

During the data analysis stage, flood-induced water and impacted agriculture, population, and infrastructure were identified in the processed images (Figure 3). The script distinguished land and water within the study area by applying a maximum likelihood classification to the NDWI using an NDWI signature file that we developed. This signature file was a collection of samples of what water and land should look like in an NDWI output. These served as a baseline to distinguish water from land. These samples were then incorporated into the computer learning classification process.

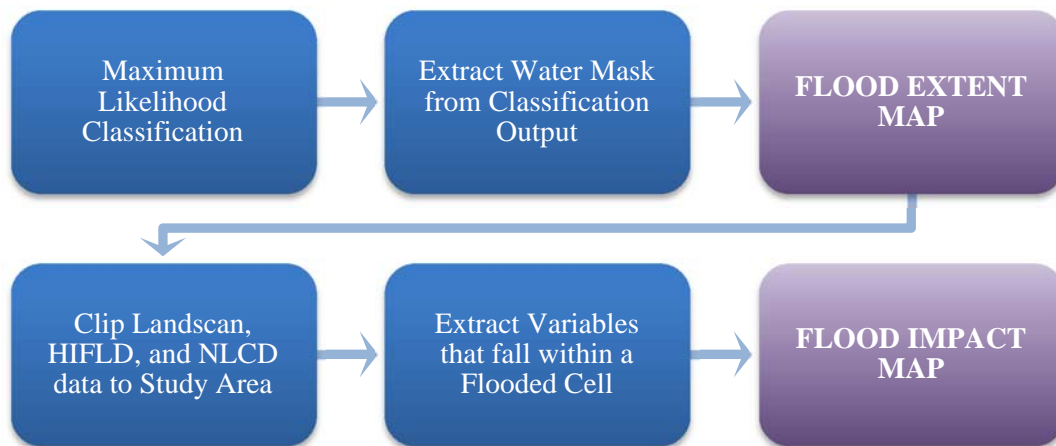


Figure 3. Algorithm used to create the data analysis section of the Flood Probability Python script

Once the classification was complete, the water mask was used to remove non-flood waters from the NDWI signature, resulting in only flood induced water (Figure 4). After the Flood Extent Map was created, the LandScan, Homeland Infrastructure Foundation-Level Data (HIFLD), and NLCD 2011 data were clipped to the study area, and data that intersected with the flood waters were extracted to create a Flood Impact Map (Figure 5). This impact map provided information about the effect of the flood on population, infrastructure, and agriculture – in the case of the September 2016 flood event only population and agriculture were impacted.

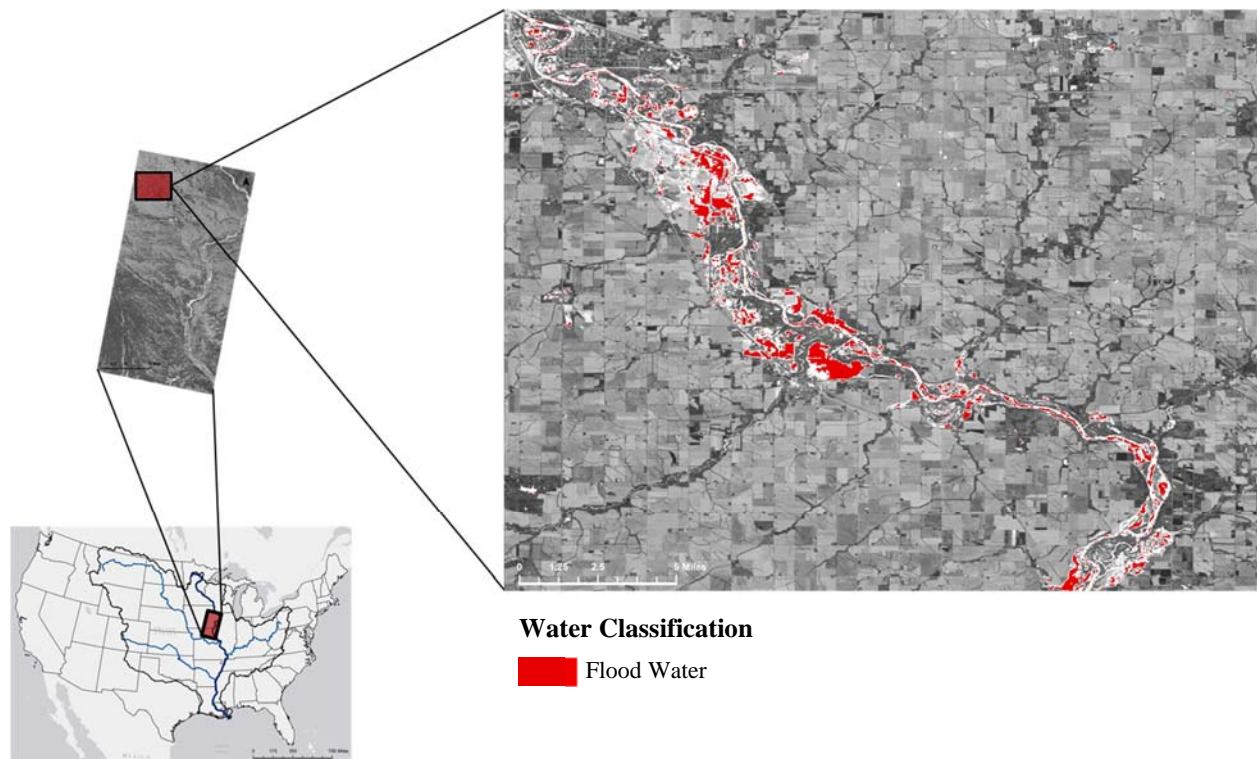


Figure 4. Flood Extent Map of the northwest corner of the Iowa flood event, identifying areas of flood induced water

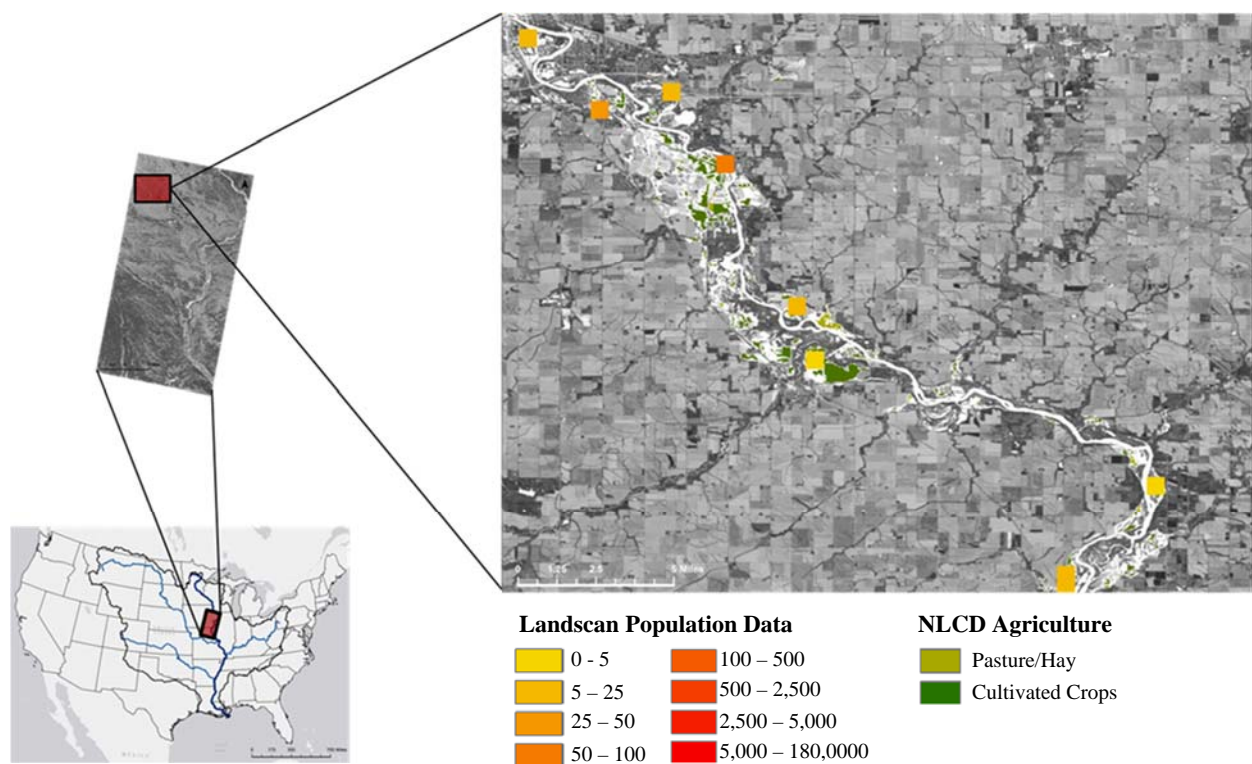


Figure 5. Flood Impact Map of the northwest corner of the Iowa flood event, identifying impacted areas of population, agriculture, and infrastructure

The automated algorithm was further tested on multiple flood events that occurred between September 2013 and March 2016 throughout the Mississippi River Basin. At FEMA's request, data was specifically extracted from an Iowa flood event that took place in September 2016.

RESULTS & DISCUSSION

The script was used to analyze 11 different flood events that occurred between August 2013 and March 2017, as seen in Table 1 and Figure 6, to test its ability to identify flood water and flood impact on a variety of landscapes and physiographic regions. A flood extent map and flood impact map were created for each of these flood events to analyze the flood impact on the region. In addition, these flood events fell within five of the six physiographic regions within the Mississippi River Basin, providing confidence in the script's and the signature file's ability to analyze flood events throughout the basin.

Table 1. List of flood events that were analyzed using the automated Python script

Location	Flood Date	Image Date
Reno Co., Kansas	8/4/2013	8/6/2013
Boulder, Colorado	9/15/2013	9/17/2013
Claremore, Oklahoma	5/23/2015	5/26/2015
St. Louis, Missouri	12/28/2015	12/31/2015
Mississippi/Louisiana	3/11/2016	3/13/2016
Fremont Co., Wyoming	5/11/2016	5/13/2016
St. Paul, Minnesota	6/23/2016	7/10/2016
Fayette Co., West Virginia	6/23/2016	6/30/2016
Indiana/Ohio	8/28/2016	8/30/2016
Cedar Rapids, Iowa	9/26/2016	9/27/2016
Waynesburg, Pennsylvania	3/1/2017	3/4/2017

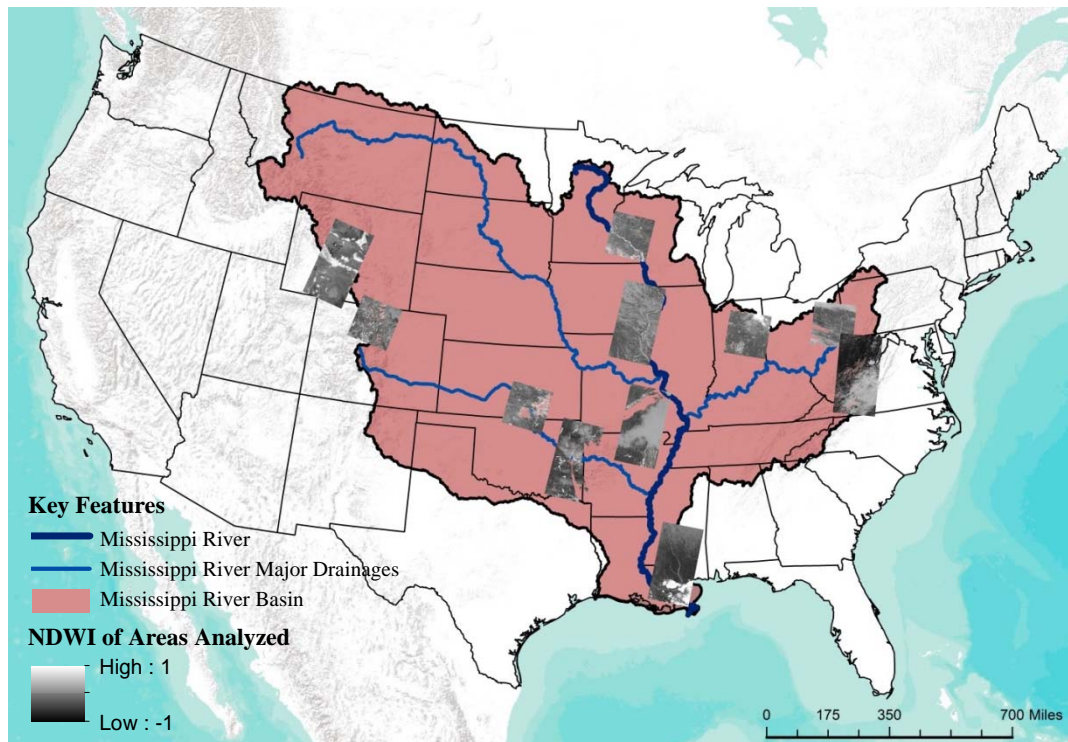


Figure 6. Map of flood events that were analyzed using the automated Python script

The 11 flood events impacted nearly 20,000 people, over 74,000 acres of agriculture (i.e. 16%), and five large infrastructures. The flood that occurred in Mississippi and Louisiana during March 2016 had the greatest numerical population impact of the 11 events, affecting almost 8,000 people, while the Missouri flood event had the largest agricultural impact, effecting over 55,000 acres of agriculture in the area (Table 2). Abnormally, according to our analysis, the flood event in Minnesota had minimal impact on the population or agriculture. This lack of results could to be due to the delay in available Landsat 8 OLI imagery, which was not available until 14 days after the peak of the flood event. On the contrary, news articles that covered the Minnesota flood stated that there was at least \$32 million in damages and 53% of farmland was flooded (Davies, 2014).

Table 2. Impact Analysis of each flood event using the Flood Probability Python script

Flood Event	Flood Impact			
	No. People	Acres of Ag.	Percent of Ag.	Infrastructure
Reno Co., KS	63	321.14	0.09%	0
Boulder, CO	6128	1,741.57	3.77%	Grade School and Nursing Home/Assisted Living Facility
Claremore, OK	219	6,719.43	1.73%	0
St. Louis, MO	648	55,258.14	3.18%	0
MS/LA	7857	4,427.21	5.24%	College/University
Fremont Co., WY	225	29.36	0.02%	0
St. Paul, MN	0	8.22	0.00%	0
Fayette Co., WV	52	19.35	0.01%	Child Care Facility
IN/OH	925	424.77	0.07%	0
Cedar Rapids, IA	953	4737	0.77%	0
Waynesburg, PA	2447	630.93	0.57%	Child Care Facility

CONCLUSIONS

Flooding can cause catastrophic damages to a community, and a disaster relief organization's ability to identify flooded areas in a timely manner is crucial. In the past, organizations such as FEMA and USGS HDDS have relied on flood detection models to plan relief efforts. These methods have a tendency to overcompensate and can lead relief efforts to areas that are not as highly impact. Now, with satellite imagery and the ability to automate the process, flood impacts can be detected in much greater detail throughout the United States. This project automated a flood probability algorithm that was used to analyze 11 flood events from August 2013 to March 2017, including events that occurred in five of the six physiographic regions within the Mississippi River Basin. For each event, the script produced a Flood Extent Map and Flood Impact Map allowing us to quantify impacts on agriculture, population, and infrastructure. The script, Flood Extent Maps and Flood Impact Maps created during the term will provide our project partners with valuable information about flood events that can be used to inform their decision-making processes.

Future Work

The project partners expressed an interest in an automated imagery downloading process being supported within our code. This automated process would require the user to input the path and row of the tile of interest, and then the code would call and download the tile using the `dnppy` package. Acknowledging that data acquisition is one of the most challenging components to data analysis, the self-contained process would eliminate the need for an external site to be accessed that can confuse or complicate the analysis. The team developed a working version of this code, but it had a low success rate and incorrectly operated approximately 50% of the time. After searching for a solution, we have concluded that a 64-bit background processing on the 10.4 version of ArcMap would be sufficient to add this feature; however, this is an untested solution.

There was also consideration in incorporating an input section that would allow the user to select the geographic region that he or she was interested in. This selection would incorporate premade signature files specific for each region, ensuring the accuracy of all analysis being conducted regardless of location. In addition, incorporating more robust infrastructure data would improve the usefulness of the product. Such data would enable our code to identify smaller infrastructure and residential houses impacted by flooding, oppose to only identifying major infrastructure like hospitals and schools.

Furthermore, finding better methods to mask clouds and shadows, as well as mapping prevailing water in the United States, would improve the overall accuracy of the product. The cloud masking was conducted through

utilization of the cloud bands provided by the USGS from the Landsat 8 OLI data download and a water mask generated from MODIS. The utilization of these two features were satisfactory for a basic analysis but to move into more detailed analysis there are some issues with each that need to be addressed. The 250m mask poses a problem when areas of prevailing water are too small to be detected by MODIS. This leads to our product erroneously classifying these waters as flood-induced instead of as prevailing water bodies. The incorporation of a 30m water mask, such as the NLCD open water classification, would be beneficial to eliminating false detection. For clouds, the team simply deleted all data covered by the clouds making analysis impossible on cloudy days.

Lastly, using a data from a satellite that has a higher spatial and temporal resolution would be beneficial. For example, WorldView-4 data has both a high spatial resolution (31cm) and is available much more frequently on a 3-day return interval. The incorporation of this data, or the more assessable high resolution Sentinel-2 MSI data, with the team's process for flood mapping could provide even more accurate flood analysis data to the project partners. This would also afford more opportunities to avoid cloud covered days and get the best available picture of the flood event.

ACKNOWLEDGEMENTS

We would like to thank the mentors and partners who dedicated their time and assistance to this project, including Dr. Jeffrey Luvall, Dr. Robert Griffin, Dr. Andrew Molthan, Leigh Sinclair, Maggi Klug, Brenda Jones and Glen Russell. We would also like to thank all previous contributors to this project including, Nicholas McVey, Helen Baldwin-Zook, Dashiell Cruz, Olivia Callaway, and Chris Ploetz. Without any of them, this project would not have been possible.

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 and cooperative agreement NNX14AB60A.

REFERENCES

- Blau, M. (2016). Iowa flooding: Thousands evacuate Cedar Rapids. Retrieved from <http://www.cnn.com/2016/09/27/us/iowa-wisconsin-flooding/>
- Eller, D. (2016). Iowa flooding sucker-punches harvest. Retrieved from <http://www.desmoinesregister.com/story/money/agriculture/2016/09/27/flooding-sucker-punches-iowa-harvest/91135314/>
- Davies, R. (2014). Cost of June 2014 Minnesota Floods. Retrieved from <http://floodlist.com/america/usa/cost-minnesota-floods-june-2014>
- Federal Emergency Management Agency (2016). About the Agency. Retrieved from <https://www.fema.gov/about-agency/>
- Garber, K. (2008). Midwest Floods Ruin Crops. Retrieved from <http://www.usnews.com/news/national/articles/2008/06/18/midwest-floods-ruin-crops/>
- Gerencer, T. (2015). U.S. Floods Cost \$34 Billion. Retrieved from <http://moneynation.com/u-s-floods-cost-34-billion/>
- Gao, B. (1996). NDWI-A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote Sensing of Environment*, 58:257-266.
- Kusky, T.M. (2013). A Brief History of Flooding and Flood Control Measures Along the Mississippi River Basin, in Boulter, S., Palutikof, J., Karoly, D.J., and Guitart, D. (Eds.), *Natural Disasters and Adaptation to Climate Change* (31-41). Cambridge: Cambridge University Press.
- McFeeters, S.K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17:1425-1432.
- Towne, J. (2016). Harvest on Hold as Farmers Wait for Fields to Dry. Retrieved from <http://whotv.com/2016/09/27/harvest-on-hold-as-farmers-wait-for-fields-to-dry/>
- US Army Corps of Engineers. (n.d.). The Mississippi Drainage Basin. Retrieved from <http://www.mvn.usace.army.mil/Missions/Mississippi-River-Flood-Control/Mississippi-River-Tributaries/Mississippi-Drainage-Basin/>
- U.S. Geological Survey. (2014). Provisional Landsat OLI Surface Reflectance. U.S. Geological Survey Earth Resources Observation and Science Center. <https://doi.org/10.5066/F78S4MZJ>

USGS (2016a). Provisional Landsat 8 Surface Reflectance Data Available. Retrieved from <https://landsat.usgs.gov/provisional-landsat-8-surface-reflectance-data-available>

USGS (2016b). Using the USGS Landsat 8 Product. Retrieved from <https://landsat.usgs.gov/using-usgs-landsat-8-product/>

USGS (2016c). Normalized Difference Water Index. Retrieved from https://deltas.usgs.gov/fm/data/data_ndwi.aspx