Lake Ontario Disasters

Employing NASA Earth Observations in the Greater Toronto Area to Improve Flood Preparedness for Coastal Communities

**https://lh3.googleusercontent.com/Gc_uPlteRsS_xrm0MwVT3mw588Sfg7z4Y6PSc7mz1HrjV36TfC-6oiN-wkSubOLzDsDzaIOdBkuYF0dkTud3v4OX7z5B5NgseTjADACaVNW1PYbc31edPJU7sbzu6XdweJdBhOG- Technical Report**

Final Draft – April 5th, 2019

Tyler Hennessee (Project Lead)

Isaac Gamoran

Meryl Kruskopf

Elizabeth Swanson

Dr. David Hondula, Arizona State University (Science Advisor)

Dr. Cedric Fichot, Boston University (Science Advisor)

Other Contributors:

Christie Fleming

Oyut Amarjargal

Eric Deutsch

Celeste Gambino

**1. Abstract**

From late January through the beginning of May 2017, an extraordinary amount of precipitation fell in the Lake Ontario watershed. By late April, large swaths of the Greater Toronto Area (GTA), including numerous lakefront properties, beaches, and public recreation facilities, had become inundated with water. Partnering with the City of Toronto, the City of Mississauga, the Great Lakes and St. Lawrence Cities Initiative (GLSLCI), the Credit Valley Conservation (CVC), and the Toronto and Region Conservation Authority (TRCA), this project provided tools for local and regional organizations to improve their response to flood events. These tools used Earth observations from six satellites to produce two end products. Sentinel-1 C-Band Synthetic-Aperture Radar (C-SAR) and Sentinel-2 Multispectral Instrument (MSI) provided data for flood extent maps using a classification algorithm. Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG), Soil Moisture Active Passive (SMAP) L-band Radiometer, Terra Moderate Resolution Imaging Spectrometer (MODIS), and Shuttle Radar Topography Mission (SRTM) contributed precipitation, soil moisture, snow cover, and elevation data used in a Google Earth Engine (GEE) data visualization tool. The results of this project demonstrated the feasibility of GEE as a platform for providing municipalities in the Toronto metropolitan area with tools to understand and visualize flood behavior and associated patterns in hydroclimatic variables. Thus, these municipalities are better prepared to protect the most vulnerable flood-prone areas in the GTA.

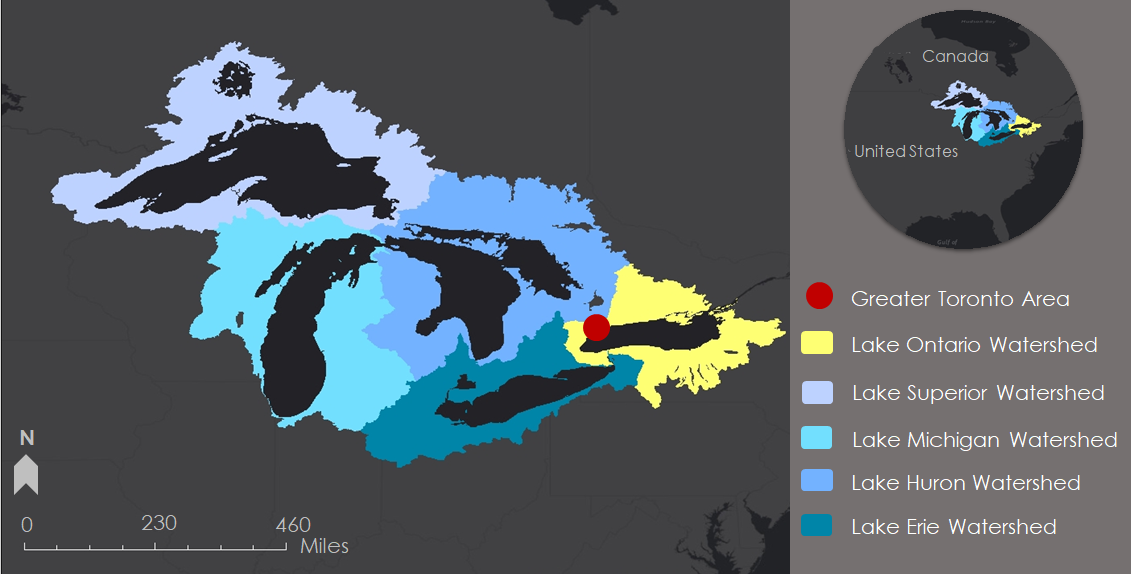
**Keywords**

flooding, Google Earth Engine (GEE), GPM IMERG, MODIS, SMAP, remote sensing, Sentinel-1 C-SAR

**2. Introduction**

* 1. ***Background Information***

Flooding was the most frequent natural disaster event in Canada from 1900 to 2013, and it sits atop the list of increasingly damaging climate hazards (Struzik, 2015). The Great Lakes Basin (*Figure 1*), located in eastern North America, is highly susceptible to such flooding. This basin contains the five Great Lakes (Superior, Michigan, Huron, Erie, and Ontario) as well as the watershed of each lake. As the easternmost watershed and farthest downstream, the Lake Ontario watershed plays a vital role in the basin. All water in the basin must flow through the Lake Ontario watershed as it moves towards the Saint Lawrence River. Furthermore, this watershed is home to 25 percent of Canada’s total population (Great Lakes Guide, 2018). A number of cities, including the partner cities of Toronto and Mississauga, are located within this watershed.



*Figure 1.* This image shows the study area of the Great Lakes Basin. The Greater Toronto Area is represented with a red dot, and the Lake Ontario Watershed is highlighted in yellow.

As Canada’s largest urban area with a population of over 6.4 million people as of 2016, the Greater Toronto Area (GTA) sits precariously close to Lake Ontario and is vulnerable to intense flooding (Toronto Population 2018, 2018). As a result of flooding, Toronto communities have experienced population displacement, monetary loss from capital damage and displaced workers, and impacts on both physical and mental health (Henstra & Thistlethwaite, 2017). Flooding events have also initiated costly response efforts by city municipalities across the Toronto area. Heightened awareness of flooding is bringing attention to outdated design codes and standards as well as decaying infrastructure in the GTA (Nirupama, Armenakis, & Montpetit, 2014).

In the spring of 2017, a combination of factors led to severe flooding in the GTA. The region experienced an unusually wet winter from January to March, record precipitation in both April and May, historically warm spring temperatures, and highly variable ice conditions. All of these factors caused water levels to rise sharply in Lake Ontario and forced inundation across the GTA by late May (Bechard et al., 2018). Portions of the city remained inundated for weeks, and Toronto Island Park was closed for 88 days due to the flooding. A similar event in the spring of 2013 led to record flooding throughout the GTA that cut off power to 300,000 residents and forced the municipal government to provide over $65 million for response and recovery efforts (Henstra & Thistlethwaite, 2017). Both events illustrated the continued challenges of flooding facing city municipalities and the need for reliable flood mapping and preparedness across the region.

To produce products for flood preparedness, this project referenced and combined scientific methods from a variety of research. Past studies for land classification used cloud-free optical satellite imagery from both Landsat 8 Operational Land Imager (OLI) and Sentinel-2 Multispectral Instrument (MSI) to differentiate water from land (Nandi, Srivastava, & Shah, 2017; Xu, 2006). When cloud-free imagery was unavailable, other studies used Sentinel-1 C-Band Synthetic Aperture Radar (C-SAR) to differentiate flat from rough surfaces. This project combined these methods to compromise the limitations of each (Kyriou & Nikolakopoulos, 2015). Additionally, this study referenced the user interface capability of a Google Earth Engine (GEE) script developed by the DEVELOP fall 2018 Osa Peninsula Water Resources III project to implement a similar user interface in the Flooding in the Lake Ontario Watershed (FLOW) Tool.

***2.2 Project Partners & Objectives***

For this project, the team partnered with the City of Toronto, the City of Mississauga, the Great Lakes and St. Lawrence Cities Initiative (GLSLCI), the Credit Valley Conservation (CVC), and the Toronto and Region Conservation Authority (TRCA). The City of Toronto uses the expertise and reports from local water conservation authorities and drives research initiatives in hazard identification to help prepare for an emergency. During the flooding of 2017, city employees manually measured water levels in the field to monitor flood levels. The City of Mississauga uses static maps and GIS to make decisions regarding storm drainage but does not have a data-driven approach to disaster planning and management. The GLSLCI uses research and expertise on the policy side of planning to inform disaster management. The CVC and the TRCA monitor current watershed conditions and weather forecasts, predict river and creek conditions, and communicate findings to the public, municipalities, and media. For this project, the team set out to: (1) create a flood extent map for the GTA to show where flooding has occurred during previous floods; (2) analyze trends in precipitation, snow cover, and soil moisture; and (3) provide the partners with data to understand historic flooding events.

**3. Methodology**

***3.1 Data Acquisition***

Through GEE, this project obtained and processed remotely sensed data from a number of different satellites (Table 1). For the FLOW Tool, the study acquired all available imagery from four sensors: Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG), Soil Moisture Active Passive L-band radiometer (SMAP), Terra Moderate Resolution Imaging Spectroradiometer (Terra MODIS), and Shuttle Radar Topography Mission (SRTM). For the land classification maps, the study acquired data for the summer of 2017 and May of 2017 from two sensors: Sentinel-1 C-SAR and Sentinel-2 MSI.

Table 1

*Satellite datasets utilized for flood mapping during the study period*

|  |  |  |
| --- | --- | --- |
| **Platform & Sensor** | **Data Product** | **Dates Available** |
| **Sentinel-1 C-SAR** | Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar Ground Range Detected, log scaling | October 3, 2014 to February 18, 2019 |
| **Sentinel–2 MSI** | Sentinel-2 Multispectral Instrument Level 1C | June 23, 2015 to February 2019 |
| **GPM IMERG** | GPM: Global Precipitation Measurement (GPM) v5 | March 12, 2014 to February 16, 2019 |
| **SRTM** | SRTM Digital Elevation Data 30 m | February 11, 2000 to February 22nd, 2000 |
| **SMAP L-band radiometer** | NASA-USDA SMAP Global Soil Moisture Data | April 15, 2015 to February 10, 2019 |
| **Terra MODIS** | MOD10A1.006 Terra Snow Cover Daily Global 500 m | March 5, 2000 to February 10, 2019 |

***3.2 Data Processing Land Classification***

The team processed data from Sentinel-1 C-SAR and Sentinel-2 MSI before combining them into a single image to be used for classification. The user chosen Area of Interest (AOI), time period, and instrument mode provided the filtering parameters for the Sentinel-1 C-SAR data. The instrument mode was Interferometric Wide Swath. The team defined functions to remove streaks from images by masking outlying values identified visually from an image histogram and to convert image bands (Vertical Vertical (VV) and Vertical Horizontal (VH)) back to power values from decibels so that the ratio VV/VH could be calculated and added to the image. The team then applied these functions to every image in the collection. Lastly, the collection was reduced to the median image, a smoothing kernel was applied with a radius of 10 meters, and the image was clipped to the AOI.

Sentinel-2 MSI data were cloud masked and normalized difference bands were calculated and added. The cloud masking algorithm used was taken directly from the GEE application programming interface and utilized the QA60 band. The QA60 band holds bit-wise information indicating the presence of clouds and cirrus. The data were first filtered by time period and AOI. The data were then filtered by CLOUDY\_PIXEL\_PERCENTAGE less than 20 percent and the cloud mask was applied. The resulting image was mosaicked and clipped to the AOI. Next, normalized difference bands were calculated to improve the model’s accuracy based on previous research showing their ability to differentiate water from land. The Normalized Difference Water Index (NDWI) has been shown to differentiate between wet and dry land (Acharya & Yang, 2015). The NDWI is a ratio using green and near-infrared (NIR) frequencies that enhances water while decreasing the influence of vegetation (Equation 1).

(1)

In addition, studies have shown that modifying the NDWI can better highlight standing water and decrease noise from built areas and vegetation (Nandi et al., 2017; Xu, 2006). While these studies both show increased success of enhancing standing water identification through changing the NIR band, they use different bands. Xu uses mid-infrared (MIR) while Nandi et al. use shortwave infrared (SWIR). The Modified NDWI (MNDWI) using SWIR was added to the image since Sentinel-2 data do not contain a MIR band (Equation 2). The MNDWI is an improvement over NDWI as water absorbs more MIR light than NIR light (Kyriou & Nikolakopoulos, 2015). Lastly, the Normalized Difference Vegetation Index (NDVI) (Equation 3) was calculated to provide another parameter for the random forest classification to use. Functions were defined to calculate NDVI, NDWI, and MNDWI then applied to the image. The Sentinel-1 image bands were added to the Sentinel-2 image, and the resulting image was made a global variable so that it could be accessed outside the script.

(2)

(3)

The cloud masking algorithm used was not rigorous enough to provide any optical data during flooding periods, so two separate scripts were created to make images for flooding and non-flooding time periods. The script for the non-flooding time period is Sentinel1\_2\_Image and contains the exacts steps as listed above. The script Sentinel1\_Flooded\_Image contains the steps for processing the Sentinel-1 data but does not use any Sentinel-2 data.

***3.3 FLOW Google Earth Engine Tool***

The FLOW GEE Tool provided the partners with an easy way to access and download precipitation, snow cover, and soil moisture data during the focus period of April 2015 to February 2019. The tool gathered precipitation data from GPM IMERG, soil moisture data from SMAP, and snow cover data from Terra MODIS. In the user interface, the partners can visualize the variables as monthly medians by pixels as well as the monthly deviation from the normal. Finally, the tool gathered elevation data from SRTM to provide a reference layer in the tool.

The team provided GPM IMERG precipitation data for each individual month between April 1, 2015, and February 28, 2019. To do so, the team first filtered the data by date (month and year), by the study area region (using filter bounds), and by the band “precipitationCal”. This band contains precipitation data in millimeters calibrated with *in situ* data to improve accuracy. After filtering the data, the team used the sum function in GEE to add up all the individual pixel values for one month together and obtain the total precipitation for the month by pixel. Finally, the team clipped the data to the study area.

The team also provided GPM IMERG precipitation data for monthly normals calculated using all the available data during the study period. For GPM IMERG, the team acquired about 5 years of data between April 1, 2014, and February 28, 2019. The group created individual data layers for each month during this time period by filtering the data by date (month and year), by the study area region (using filter bounds), and by the band “precipitationCal”. Then, the team summed the data for the entire month and clipped them to the study area. For each month, the team then created an image collection of all images for the month of each year. For example, the team calculated the January normal. First, the team created an image collection of median values for each January from 2015 to 2019. From there, the team took the median of all five Januarys to find the median value during the focus period.

The team provided SMAP soil moisture data by month for each individual month between April 1, 2015, and February 28, 2019. To do so, the team first filtered the data by date (month and year) and by the band “ssm”. This band contains surface soil moisture data from 0 to 28 mm. Pixels along the lakeshore displayed inaccurate values, so the team created a function to mask these out. After removing these values, the team calculated the median values by pixel for each month and clipped the resulting images for each month to the study area.

The team also provided SMAP soil moisture data for monthly normals calculated using all the available data during the study period. For SMAP, the team acquired about 4 years of data between April 1, 2015, to February 28, 2019. The team created individual data layers for each month during this time period by filtering the data by month within the years above and by the band “ssm”. Then, the team masked out the inaccurate shoreline values, calculated the median values for the entire month, and clipped the data to the study area. For each month, the team then created an image collection of all images for the month of each year. For example, the team calculated the January normal. First, the team created an image collection of median values for each January from 2016 to 2019. From there, the team took the median of all four Januarys to find the median value during the focus period.

The team provided Terra MODIS snow cover data by month for each individual month between April 1, 2015, and February 28, 2019. To do so, the team first filtered the data by date (month and year), by the study area region (using filter bounds), and by the band “NDSI\_Snow\_Cover”. This band contains the Normalized Difference Snow Index (NDSI) (Equation 4) data that indicates the likelihood of a pixel having snow cover. After filtering the data, the team calculated the median values by pixel for each month and clipped the resulting images for each month to the study area.

(4)

The team also provided Terra MODIS snow cover data for monthly normals calculated using all the available data. For Terra MODIS, the team acquired 19 years of data between March 1, 2000, and February 28, 2019. The group created individual data layers for each month during this time period by filtering the data by month within the above years, by the study area region (using filter bounds), and by the band “NDSI\_Snow\_Cover”. Then, the team calculated the median values for the entire month and clipped the data to the study area. For each month, the team then created an image collection of all images for the month of each year. For example, the team calculated the January normal. First, the team created an image collection of median values for each January from 2001 to 2019. From there, the team took the median of all nineteen Januarys to find the median value during the focus period.

***3.4 Data Analysis of Land Classification***

The team used image bands, prepared as described previously, as the variable inputs into a random forest classification algorithm to classify pixels into surface features: buildings, paved, soil, vegetation, and water and a binary presence classification. There were three processes that went into the algorithm. These processes included preparing the image bands (variables), variable selection, and training data.

Variable selection was informed by two sources of information: 1) R to produce a correlation and importance matrix, and 2) knowledge gained from literature demonstrating the ability of the variables to differentiate water from land such as NDWI, and MNDWI (Acharya & Yang, 2015; Nandi et al., 2017; Xu, 2006). The R packages dplyr, corrplot, and rfUtilities were used to create the correlation matrix. First, the variables were ranked by importance using rf.modelsel. Next, the correlation coefficient matrix was calculated using the corrplot package. This script was run twice, designating the independent variable as ‘presence’ and then ‘surface’. Using presence and surface as the independent variables, two matrices were produced and are provided in Appendix Tables B1 and B2.

Two sets of variables were chosen for experimentation: Set 1 was based on the ‘presence’ matrix and includes [SWIR2, MNDWI, VV, VH] while Set 2 was based on the ‘surface’ matrix and includes [blue, NDWI, VV, VH]. The variables for each set were chosen based on the variable with the highest importance, a variable that has been shown to differentiate between land and water (MNDWI or NDWI), a correlation lower than 0.7, and including optical and SAR data (VV, VH).

Training data were created using two different methods. The first set of training data was created by generating 300 random points in GEE over the GTA and exporting them to a comma-separated values (CSV) file. Next the CSV file was opened in Google sheets, presence and surface columns were added, and each point was manually assigned by a team member by Googling the coordinates and zooming in to designate the appropriate class: Building (0), Paved (1), Soil (2), Vegetation (3), or Water (4). The excel sheet was then imported into ArcMap, converted to a shapefile, and uploaded to GEE as a feature collection Earth Engine (EE) asset. With the goal of improving the accuracy of the classification, the second set of training data was created in a smaller area of interest. This area merged municipal boundaries of the cities of Mississauga and Toronto with a 500 m buffer to include the coastline, referred to as MT. This was done to limit the land types that occurred in the given region. First, 200 random points were generated over the MT region. Next, classes were created by clicking a random point using the interactive GEE map interface and selecting ‘FeatureCollection’ under ‘Import as’. The point was given one of five names: Soil, Building, Paved, Vegetation or Water. Each point was assigned both the properties of ‘surface’ (0,1,2,3,4) and presence (0,1). In this dataset, soil was an empty class. After the first dataset showed very few soil points, soil was moved to 0 since ArcMap assigns empty pixels the value of 0 but properties must begin with 0 and increase in even integer intervals. In addition, the team added 9 water points to skew the model to recognize water.

Once the training data and variable selection were complete, the random forest algorithm could be run. Two separate scripts (Sentinel1\_2\_RF\_Classification\_Non\_Flood and Sentinel1\_RF\_Classification\_Flood) were created for running the random forest classification for flood and non-flood time periods using the prepared images. Both scripts followed the same steps. First, the prepared image was imported and the AOI was defined and added to the map. Second, the classification type was defined (‘surface’ or ‘presence’) and the bands were renamed for ease of understanding (e.g. ‘B2’ became ‘blue’). Third, the training data were imported, and a new image was created only containing the selected variables. Fourth, the training data were sampled on the new image so that the samples contained both the variables and the designated classifications. Fifth, the data was split 70/30 into training and testing points. Sixth, the classifier was trained and run on the image. Lastly, the image was visualized and displayed, the classifier was applied to the test points, and the confusion matrices were calculated for an accuracy assessment. In addition, the classified image maps of either land classification or presence-absence were exported to Google Drive.

***3.5 Data Analysis of FLOW Tool***

The team compared the monthly data during the focus period for precipitation, soil moisture, and snow cover to the normal values for each month. To do so, the team subtracted each individual month from the calculated normal for that month to show the deviation from normal. For example, the team subtracted the January 2018 layer from the January median layer to produce the January 2018 deviation from normal layer.

The team also included charts using the deviation from normal data during the focus period. The team based the calculations for these charts on two study areas, the Great Lakes Basin and the Lake Ontario Watershed. The team reduced the monthly median values by pixel across the study area into one mean value for each month. From there, the team subtracted the mean values by month during the focus period from the monthly normal mean values. For example, the team calculated the mean value for the median pixel values from the January 2018 layer for both the Great Lakes Basin and the Lake Ontario Watershed and then subtracted these values from the normal values for January. Finally, the team displayed the deviation values on a chart.

**4. Results & Discussion**

***4.1 Land Classification Results***

The model was run over every combination of the study region (GTA or MT), variable set (1 - SWIR2, MNDWI, VV, VH, and 2 - blue, NDWI, VV, VH), and classification type (surface or presence). Tables of accuracies and confusion matrices can be found in Appendix Tables C1, C2, and D1 through D6. Overall, the model produced higher accuracies when run over the GTA area than the MT region. This most likely has to do with the quality of the training data created for each of these regions, which are summarized in Table 2. The GTA, while bigger, contained more points and was made with more attention to detail. The MT data were created as a quick attempt to increase accuracy by decreasing the region size to an area more relevant to the study. During the creation of the MT dataset, the team decided to assign ‘building’ to points located on the edge of roofs when previously (GTA) they would have been assigned ‘paved’. This turned out to be an inaccurate assignment and caused the model to confuse buildings with paved areas and vegetation. The GTA contains a large amount of agriculture, resulting in 66.67 percent of training data to be vegetation. The MT is much more urban and only 36.84 percent of the training data was vegetation. This caused the model to overpredict vegetation in the GTA.  In addition, neither region contained a high percentage of water, with 9.0 percent in the GTA and 7.18 percent in the MT. This may have made it difficult for the model to identify smaller water bodies.

Table 2

*Land Type Distribution of Training Data*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Region** | **Total # of Training Data Points** | **Building**  **%** | **Paved**  **%** | **Soil**  **%** | **Vegetation %** | **Water**  **%** |
| **Greater Toronto Area (GTA)** | 300 | 8.67 | 12.0 | 3.67 | 66.67 | 9.0 |
| **Mississauga and Toronto boundaries (MT)** | 209 | 30.14 | 25.84 | 0.48 | 36.84 | 7.18 |

The images below demonstrate how creating training data over different regions can highly affect accuracy. Notice how water appears in the middle of the city and the airport on the island also appears to be inundated in the first image bounded by the MT (*Figure 2*). In the second image, vegetation points appear in the water as the model is skewed by all the vegetation in the GTA. However, paved areas such as downtown and the airport are mostly shown as paved (*Figure 3*).



*Figure 2*. Non-flooded image of the Toronto Islands, MT, Variable Set 1 with 52.9% Accuracy (from Sentinel-1 and Sentinel-2).



*Figure 3*. Non-Flooded image of the Toronto Islands, GTA, Variable Set 1 with 77.1% Accuracy (from Sentinel-1 and Sentinel-2).

The table below provides information on the testing points that were correctly classified for the binary classification (Table 3). As opposed to the surface classification, the presence or binary classification consistently produced accuracies near 100% (Appendix Tables C1 and C2). Given the low number of test points in water, seen as 7 in the table, compared to those on land, the seemingly high accuracy of the presence classification does not reflect the ability of the model to differentiate water from land for smaller water bodies. Overall, in both presence and surface classifications, neither variable set 1 (blue, NDWI, VV, VH) nor variable set 2 (swir2, MNDWI, VV, VH) provided consistently higher accuracy results.

Table 3

*Confusion Matrix for Presence in the GTA using variable set 2*

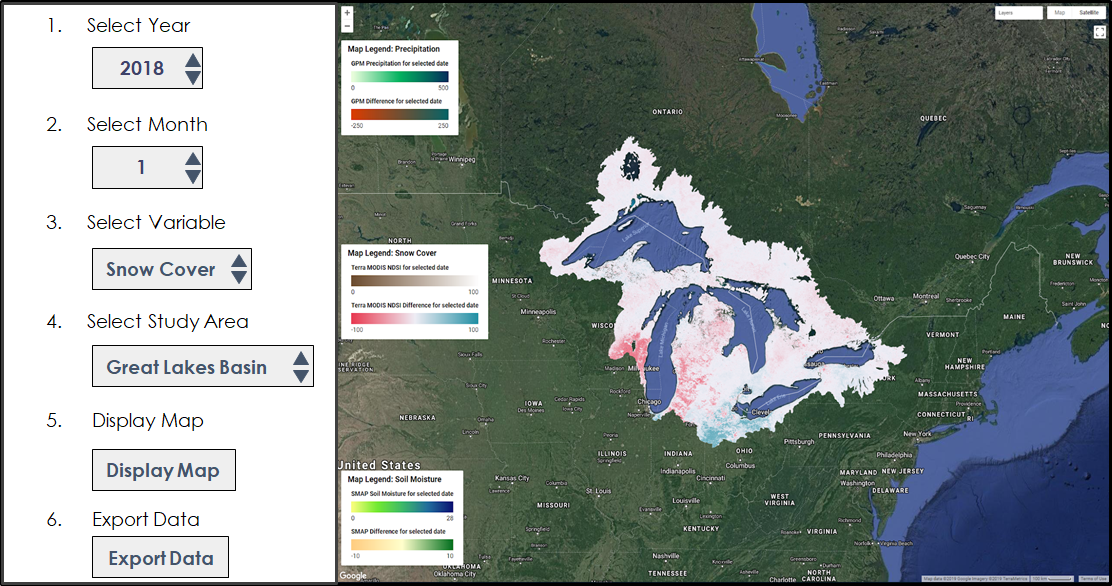
|  |  |  |
| --- | --- | --- |
| **Presence/GTA/Variable Set 2**  **Non-Flood**  **Accuracy 97.6 %** | **Land** | **Water** |
| **Land** | 78 | 0 |
| **Water** | 3 | 7 |

***4.2 Limitations, Errors, and Uncertainties for Land Classification***

The limitations, errors, and uncertainties for the land classification process are primarily associated with the training data. The number of points used, the size of the region, the variation in land cover, and the classification decisions made by the team all have unknown effects on the model. In addition, the team did not comprehend exactly how the random forest model functioned so it could be manipulated beyond providing training data. Even when a ‘seed’ was designated, the model continued to produce different results for the same data inputs. This indicated that there was a random process that was not being controlled. The team did not have the time to investigate this further, but this element contributed significant uncertainty.

***4.3 Explanation of FLOW Tool Outputs***

The resulting output from the FLOW Tool varies based on the user input. The team programmed the tool with a user interface to allow partners to use a computer mouse to select a year from 2015 to 2019, a month, a variable, and an area of study instead of typing these values into the code (*Figure 4*). The tool generates two maps for each selection: a map containing the values for the selected variable for that month and a map showing the deviation from normal for the selected variable for that month.



*Figure 4.*  Example image displaying snow cover differences from normal for January 2018 and an enlarged version of the user interface selection options (from Terra MODIS).

In addition to the user interface and maps, the team also produced three charts to show how the difference from normal values varied by month for both the Great Lakes Basin and the Lake Ontario Watershed. By comparing the partner-provided flooding months for this time period (April 2017 to July 2017, February 2018, and August 2018) to the outputs, the reader can clearly discern the major flooding date of May 2017 in the precipitation and soil moisture charts as shown in *Appendix Figures A1* and *A2*.

During May 2017, the Lake Ontario Watershed received 209 millimeters, or 8.23 inches, more precipitation than normal. In the soil moisture chart, values increased between May and July 2017, reflecting the impact of flooding on soil moisture. Given the hydrophobic nature of soil, the increase in water could have caused increased runoff and exacerbated flooding during the event. Snow cover values did not differ significantly, indicating that both the Great Lakes Basin and the Lake Ontario Watershed received less snow than usual during the winter preceding May 2017. These findings are shown in *Appendix Figure A3* and Table A1.

***4.4 Limitations, Errors, and Uncertainties for the FLOW Tool***

The FLOW Tool is not predictive and will not determine when or where a flood may occur. The tool displays the behavior of some of the variables that contribute to flooding over the study months. Additionally, the team calculated normals by month based on the available data, which included four years for SMAP, five years for GPM, and twenty years for MODIS. These normal values are not indicative of actual climatological normals and may be inaccurate as a result of this. Finally, while the extreme flooding event in May 2017 was clearly visible in the data, the other minor flooding events provided by the partners were not easy to discern.

***4.5 Future Work***

The existing code for land classification was originally created to highlight flooding, so further work could achieve that goal. This could be done by subtracting a non-flood image from a flood image. The difference between two NDWI images has been shown to produce high levels of accuracy in identifying flooded versus non-flooded areas (Ogashawara, Curtarelli, & Ferriera, 2013). Further development of the existing products could identify flooded areas by using the Normalized Difference Flood Index (NDFI) (*Equation 5*), which also uses pre- and post-flood images (Cian, Marconcini, & Ceccato, 2018).

(5)

In addition to highlighting flooding, future work should continue to refine the model and eliminate some of the uncertainties discussed. This requires continued experimentation with variable selection, training data creation, and debugging and understanding the random processes in the GEE random forest model. Lastly, improving the methodology overall would require a more vigorous cloud masking algorithm for Sentinel-2 data so that optical data could be used in flood conditions and researching methods of increasing the accuracy of models only using SAR data.

For the FLOW Tool, future work could include incorporating additional datasets to observe additional variables related to flooding. Possible datasets could include lake water levels, stream gauges, and water flows through the Great Lakes Basin to determine how existing water moving through Lake Ontario relates to flooding. Additionally, future studies could evaluate the FLOW Tool results through inferential statistics to improve the accuracy of the tool and understand how each variable contributes to flooding in the region.

**5. Conclusions**

***5.1 Conclusions of Land Classification***

Overall, a prominent factor in the accuracy of land classification is the quality and classes of training data. Quality refers to the number of points and how carefully the researcher identifies and assigns classes. Google Earth Engine can produce land classification images using Sentinel-1 and Sentinel-2 data that could be used to address other issues, but the resolution and accuracy using this method are not able to provide useful flood extent mapping in urban areas.

***5.2 Conclusions of FLOW TOOL***

The data showed an increase in both precipitation and soil moisture during the flooded time period, which pointed to the use of these variables in understanding the nature of floods. Precipitation values reached maximum amounts in May 2017, the same month in which flooding impacted the GTA. Soil moisture values reached maximum amounts in the months immediately following the event, from June to August 2017, which pointed to this variable being the most useful as a variable after the flood. Furthermore, the team noted soil moisture values well below average in the summer of 2016, which could have contributed to flooding in the spring of 2017. Interaction with the data and difference values is facilitated by a user interface that gives the partners easy access to desired data during the time frame of April 2015 to February 2019.

**6. Acknowledgments**

The Lake Ontario Disasters team would like to thank Erika Higa, our Center Lead, Megs Seeley, our Assistant Center Lead, and David Hondula, our Science Advisor. Additionally, we would like to thank the 2019 spring Niagara Falls Disasters team for their help on our project. We would also like to thank Arizona State University for hosting our node. Finally, we’d like to thank our project partners at the GLSLCI, the City of Mississauga, the City of Toronto Office of Emergency Management, the CVC, and the TRCA.

This material contains modified Copernicus Sentinel data (2014-2019), processed by ESA.

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**

**C-SAR** – C-Band Synthetic Aperture Radar

**CVC** – Credit Valley Conservation

**FLOW** – Flooding in the Lake Ontario Watershed

**EO** – Earth observations; satellites and sensors that collect information about the Earth’s physical, chemical, and biological systems over space and time

**GEE** – Google Earth Engine; a cloud-based geospatial processing platform

**GLSLCI** – Great Lakes and St. Lawrence Cities Initiative

**GPM** – Global Precipitation Measurement

**GPM IMERG** – Global Precipitation Measurement Integrated Multi-Satellite Retrievals for GPM

**GTA** – Greater Toronto Area

**MNDWI** –Modified Normalized Difference Water Index

**MODIS** – Moderate Resolution Imaging Spectroradiometer

**MSI** – Multispectral Instrument

**NDWI** – Normalized Difference Water Index

**NDFI** – Normalized Difference Flood Index

**TRCA** – Toronto and Region Conservation Authority

**SMAP** – Soil Moisture Active Passive

**SRTM** – Shuttle Radar Topography Mission

**SWIR** – Short-Wave Infrared

**8. References**

Acharya, T. D., & Yang, I. (2015). Exploring Landsat 8. *International Journal of IT, Engineering and Applied Sciences Research (IJIEASR)*, *4*(4), 4-10. Retrieved from https://www.researchgate.net/publication/311901147\_Exploring\_Landsat\_8

Bechard, G., Durrett, S., Aubry-Morin, J., Brown, T., Clavet, P., Campany, R., … Sciremammano, F. (2018). *Observed Conditions and Regulated Outflows in 2017* (Report prepared for the International Lake Ontario St. Lawrence River Board). Retrieved from http://ijc.org/files/tinymce/uploaded/ISLRBC/ILOSLRB\_SummaryReport.pdf

Cian, F., Marconcini, M., & Ceccato, P. (2018). Normalized Difference Flood Index for rapid flood mapping:

Taking advantage of EO big data. *Remote Sensing of Environment*, *209*, 712-730. [https://doi.org//10.1016/j.rse.2018.03.006](https://doi.org/10.1016/j.rse.2018.03.006)

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., … Alsdorf, D.E. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, *45*(2), RG2004. https://doi.org/10.1029/2005RG000183.

Great Lakes Guide. (2018). *Lake Ontario*. Retrieved 1 April 2019 from https://greatlakes.guide/watersheds/ontario

Hall, D. K., Salomonson, V. V., & Riggs, G. A. (2016). MODIS/Terra Snow Cover Daily L3 Global 500m Grid, Version 6. Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed February 2019. https://doi.org/10.5067/MODIS/MOD10A1.006

Henstra, D., & Tistlethwaite, J. (2017). Climate change, floods, and municipal risk sharing in Canada. *IMFG Papers on Municipal Finance and Governance, 30*, 1-43. Retrieved from <https://tspace.library.utoronto.ca/bitstream/1807/81204/1/imfgpaper30_henstra_thistlethwaite_Feb_23_2017.pdf>

Huffman, G. (2017). GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree, V05. Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed 2 April 2019. https://doi.org/10.5067/GPM/IMERG/3B-HH/05

Kyriou, A., & Nikolakopoulos, K. (2015). Flood mapping from Sentinel-1 and Landsat-8 data: A case study

from River Evros, Greece. *Earth Resources and Environmental Remote Sensing/GIS Applications VI*, *9644*. <https://doi.org/10.1117/12.2194449>

Nandi, I., Srivastava, P. K., & Shah, K. (2017). Floodplain mapping through support vector machine and optical/infrared images from Landsat 8 OLI/TIRS sensors: Case study from Varanasi. *Water Resources Management,* *31*(4), 1157-1171. <https://doi.org/10.1007/s11269-017-1568-y>

Nirupama, N., Armenakis, C., & Montpetit, M. (2014). Is flooding in Toronto a concern?. *Natural Hazards,* *72*(2), 1259-1264. <https://doi.org/10.1007/s11069-014-1054-2>

Ogashawara, I., Curtarelli, M. P., & Ferreira, C. M. (2013). The use of optical remote sensing for mapping flooded areas. *International Journal of Engineering Research and Applications*, *3*(5), 1956-1960. Retrieved from <https://pdfs.semanticscholar.org/5456/583ca4bb93a4b7533b894917041e98e0ed46.pdf>

O'Neill, P. E., Chan, S., Njoku, E. G., Jackson, T., & Bindlish, R. (2018). SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 5. Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center, accessed 2 April 2019. https://doi.org/10.5067/ZX7YX2Y2LHEB.

Struzik, E. (2015, November 16). Climate change: Floods in Canada are going to get uglier. *Huffington Post*. Retrieved from <https://www.huffingtonpost.ca/2015/11/16/alberta-flood-2013-climate-change_n_8572420.html>

Toronto Population 2018. (2018). Retrieved 6 February 2019 from

https://canadapopulation.org/toronto-population/

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025-3033. <https://doi.org/10.1080/01431160600589179>

**9. Appendices**

**Appendix A.** FLOW Tool results

*Figure A1.* Chart showing the deviation from normal for precipitation values from April 2015 to February 2019 for both the Great Lakes Basin and the Lake Ontario Watershed (from GPM IMERG).

*Figure A2.* Chart showing the deviation from normal for surface soil moisture values from April 2015 to February 2019 for both the Great Lakes Basin and the Lake Ontario Watershed (from SMAP).

*Figure A3.* Chart showing the deviation from normal for NDSI snow cover values from April 2015 to February 2019 for both the Great Lakes Basin and the Lake Ontario Watershed (from Terra MODIS).

Table A1

*Median values for each month and data set across the Great Lakes Basin and the Lake Ontario Watershed*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Date | SMAP Great Lakes Basin | SMAP Lake Ontario Watershed | Terra MODIS Great Lakes Basin | Terra MODIS Lake Ontario Watershed | GPM IMERG Great Lakes Basin | GPM IMERG Lake Ontario Watershed |
| April 2015 | -1.213 | -3.189 | 2.935636 | 0.820728 | 21.224 | 10.105 |
| May 2015 | -1.311 | -2.685 | 0.039356 | 0.014116 | 19.513 | -11.747 |
| June 2015 | 1.214 | 1.746 | 0.035715 | 0.04137 | 32.396 | 115 |
| July 2015 | 0.697 | 2.041 | 0.043956 | 0.021815 | -20.064 | -33.818 |
| August 2015 | -0.267 | 0.581 | 0.030681 | 0.0101 | -15.465 | -5.581 |
| September 2015 | -0.87 | 0.253 | 0.017532 | 0.015854 | -8.105 | 12.477 |
| October 2015 | -3.439 | -0.208 | 0.024518 | 0.013935 | -40.641 | 8.735 |
| November 2015 | -0.991 | -1.643 | -1.24102 | -0.157 | 35.366 | -40.796 |
| December 2015 | 0.568 | -1.366 | -19.067 | -21.5944 | 132.507 | 45.795 |
| January 2016 | 1.352 | 0.31 | -3.78521 | -5.54264 | -22.909 | -59.786 |
| February 2016 | 0.493 | 0.578 | -16.6966 | -37.142 | 2.596 | 48.722 |
| March 2016 | 0.546 | 0.311 | -10.6448 | -16.4 | 60.319 | 18.156 |
| April 2016 | -0.008 | -0.675 | -1.68135 | -0.233 | -68.855 | -101.481 |
| May 2016 | -1.745 | -3.734 | -0.04055 | 0.01537 | -31.212 | -24.284 |
| June 2016 | -2.47 | -6.128 | 0.001826 | 0.0066 | -15.732 | -65.574 |
| July 2016 | -0.794 | -4.794 | 0.009843 | 0.0133 | 26.503 | -46.491 |
| August 2016 | 0.805 | -1.421 | 0.009879 | 0.0078 | 23.515 | 34.294 |
| September 2016 | 1.102 | -0.747 | 0.019517 | 0.0016 | 25.906 | 4.058 |
| October 2016 | 0.938 | 0.06 | 0.012225 | -0.012 | -9.592 | 68.136 |
| November 2016 | -1.596 | -1.291 | -2.36959 | -0.05 | -3.681 | -21.596 |
| December 2016 | 0.144 | 0.564 | 10.61343 | 13.67 | 9.454 | 18.312 |
| January 2017 | 0.852 | 0.838 | -10.196 | -11.19 | 62.84 | 60.688 |
| February 2017 | 1.587 | 0.287 | -16.8033 | -15.87 | 33.718 | 1.321 |
| March 2017 | 0.277 | 0.039 | -8.08746 | -15.87 | 31.133 | 40.337 |
| April 2017 | 0.597 | 0.776 | -2.37755 | -0.202 | 30.006 | 31.59 |
| May 2017 | 1.875 | 3.778 | 0.090665 | 0.0396 | 52.85 | 209.023 |
| June 2017 | 1.765 | 3.954 | 0.019134 | 0.0056 | 69.895 | 32.5 |
| July 2017 | 3.27 | 5.864 | 0.017521 | 0.0199 | 22.52 | 52.983 |
| August 2017 | 1.879 | 2.131 | 0.024299 | 0.0252 | -0.596 | -17.734 |
| September 2017 | -0.438 | 0.907 | 0.032416 | 0.0377 | -35.265 | -12.309 |
| October 2017 | -0.367 | 0.069 | 0.007683 | 0.0211 | 42.575 | 106.166 |
| November 2017 | 0.775 | 1.036 | 7.617814 | 0.5087 | 1.762 | 33.003 |
| December 2017 | -0.289 | -0.363 | -0.77462 | 6.9091 | -8.226 | -16.945 |
| January 2018 | -1.408 | -0.999 | -3.42945 | 0.5454 | -0.921 | 20.494 |
| February 2018 | -0.915 | -0.508 | -15.6813 | -19.4 | 24.017 | -13.101 |
| March 2018 | -3.278 | -1.938 | -0.73373 | -19.4 | -32.525 | -6.774 |
| April 2018 | -1.248 | 1.163 | 14.12707 | 3.1108 | -72.037 | 3.79 |
| May 2018 | 0.601 | 1.992 | 0.137151 | -0.004 | 2.48 | -25.804 |
| June 2018 | -0.746 | -1.081 | 0.026898 | 0.0216 | -39.064 | -65.178 |
| July 2018 | -2.015 | -2.471 | 0.011586 | 0.0033 | -6.485 | 11.574 |
| August 2018 | -1.8 | -1.108 | 0.038801 | 0.02139 | 35.098 | -8.692 |
| September 2018 | 0.169 | -0.441 | 0.035432 | 0.0477 | -10.019 | 16.456 |
| October 2018 | 2.315 | -0.085 | 0.964032 | 0.0721 | 28.254 | -47.019 |
| November 2018 | 0.771 | 0.589 | 17.50714 | 17.468 | -15.172 | 52.788 |
| December 2018 | -1.148 | 0.462 | -4.25757 | -0.141 | -33.416 | 34.103 |
| January 2019 | -1.545 | -0.552 | -9.4013 | -10.19 | 64.519 | 164.277 |
| February 2019 | -1.112 | -0.801 | -1.47437 | -4.159 | 31.398 | 62.534 |

**Appendix B.** Importance and correlation matrices produced in R

Table B1

*Importance and correlation matrix using ‘presence’ as the independent variable*

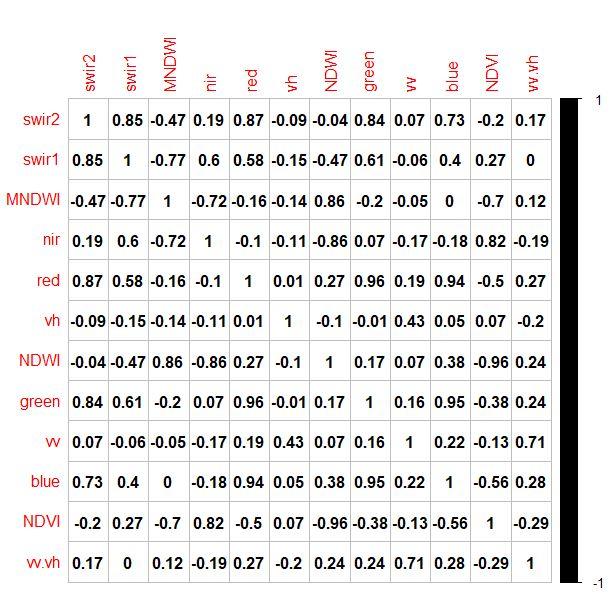
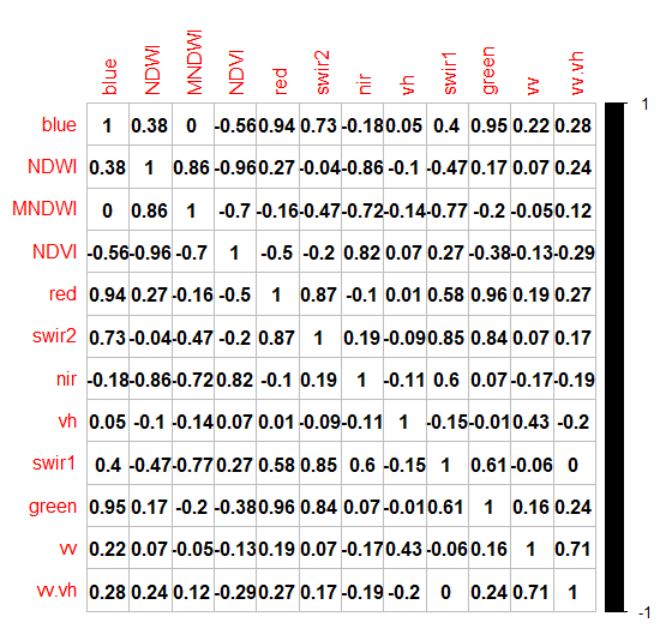


Table B2

*Importance and correlation matrix using ‘surface’ as the independent variable*



**Appendix C.** Tables of accuracies

Table C1

*Non-flood accuracies (06 -01-2017 to 09-30-2017)*

|  |  |  |  |
| --- | --- | --- | --- |
| **Class Type** | **Variable Set** | **Expected Accuracy** | **Area** |
| surface | 2 | 0.7692 | GTA\_1km |
| surface | 1 | 0.7356 | GTA\_1km |
| surface | 2 | 0.5273 | MT\_500 |
| surface | 1 | 0.6545 | MT\_500 |
| presence | 2 | 0.977 | GTA\_1km |
| presence | 1 | 1.00 | GTA\_1km |
| presence | 2 | 0.9818 | MT\_500 |
| presence | 1 | 1.00 | MT\_500 |

Table C2

*Flood accuracies (05-01-2017 to 05-31-2017); flood images could only contain VV, VH, and VV/VH bands, so those were the input variables for all classification trials*

|  |  |  |
| --- | --- | --- |
| **Class Type** | **Accuracy** | **Area** |
| surface | 0.5147 | GTA\_1km |
| surface | 0.4727 | MT\_500m |
| presence | 0.9706 | GTA\_1km |
| presence | 0.9818 | MT\_500m |

**Appendix D.** Land classification confusion matrices

Variable Set 1 (for presence): SWIR2, MNDWI, VV, VH

Variable Set 2 (for surface): blue, NDWI, VV, VH

Table D1

*Flood surface confusion matrix GTA*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Overall Expected Accuracy 67.7 %** | **Actual Values**  **Flood (GTA)** | | | | | |
| **Predicted Values**  **Flood (GTA)** |  | **Building** | **Paved** | **Soil** | **Vegetation** | **Water** |
| **Building** | **0** | **0** | **0** | **6** | **0** |
| **Paved** | **3** | **1** | **0** | **6** | **0** |
| **Soil** | **0** | **1** | **0** | **4** | **0** |
| **Vegetation** | **5** | **2** | **0** | **56** | **0** |
| **Water** | **0** | **0** | **0** | **1** | **6** |

Table D2

*Flood surface confusion matrix MT*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Overall Expected Accuracy**  **46.8 %** | **Actual Values**  **Flood (MT)** | | | | |
| **Predicted Values**  **Flood (MT)** |  | **Building** | **Paved** | **Vegetation** | **Water** |
| **Building** | **10** | **1** | **5** | **0** |
| **Paved** | **4** | **3** | **5** | **0** |
| **Vegetation** | **11** | **6** | **10** | **0** |
| **Water** | **0** | **0** | **1** | **6** |

Table D3

*Non-flood surface confusion matrix GTA Variable Set 2*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Overall Expected Accuracy**  **78.5 %** | **Actual Values**  **Non-Flood - Variable Set 2 - GTA** | | | | | |
| **Predicted Values Non-Flood**  **(GTA)** |  | **Building** | **Paved** | **Soil** | **Vegetation** | **Water** |
| **Building** | **3** | **0** | **0** | **2** | **0** |
| **Paved** | **5** | **5** | **0** | **2** | **0** |
| **Soil** | **0** | **1** | **0** | **4** | **0** |
| **Vegetation** | **2** | **1** | **1** | **61** | **0** |
| **Water** | **0** | **0** | **0** | **0** | **4** |

Table D4

*Non-flood surface confusion matrix MT Variable Set 2*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Overall Expected Accuracy**  **55.2 %** | **Actual Values**  **Non-Flood - Variable Set 2 - MT** | | | | |
| **Predicted Values Non-Flood (MT)** |  | **Building** | **Paved** | **Vegetation** | **Water** |
| **Building** | **8** | **2** | **1** | **0** |
| **Paved** | **11** | **8** | **4** | **0** |
| **Vegetation** | **7** | **1** | **12** | **0** |
| **Water** | **0** | **0** | **0** | **4** |

Table D5

*Non-flood surface confusion matrix GTA Variable Set 1*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Overall Expected Accuracy**  **81.4 %** | **Actual Values**  **Non-Flood - Variable Set 1 - GTA** | | | | | |
| **Predicted Values Non-Flood**  **GTA** |  | **Building** | **Paved** | **Soil** | **Vegetation** | **Water** |
| **Building** | **4** | **0** | **0** | **6** | **0** |
| **Paved** | **1** | **2** | **0** | **7** | **0** |
| **Soil** | **0** | **0** | **0** | **2** | **0** |
| **Vegetation** | **1** | **1** | **1** | **56** | **0** |
| **Water** | **0** | **0** | **0** | **3** | **8** |

Table D6

*Non-flood surface confusion matrix MT Variable Set 1*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Overall Expected Accuracy**  **52.9 %** | **Actual Values**  **Non-Flood - Variable Set 1 - MT** | | | | |
| **Predicted Values Non-Flood MT** |  | **Building** | **Paved** | **Vegetation** | **Water** |
| **Building** | **8** | **5** | **3** | **0** |
| **Paved** | **7** | **6** | **6** | **0** |
| **Vegetation** | **5** | **4** | **22** | **0** |
| **Water** | **0** | **0** | **0** | **4** |