

Article

## Real Time Estimation of the Calgary Floods Using Limited Remote Sensing Data

Emily Schnebele <sup>1,\*</sup>, Guido Cervone <sup>2</sup>, Shamanth Kumar <sup>3</sup> and Nigel Waters <sup>4</sup>

<sup>1</sup> Department of Geography and GeoInformation Science, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA

<sup>2</sup> Department of Geography and Institute for CyberScience, The Pennsylvania State University, 201 Old Main, University Park, PA 16802, USA; E-Mail: cervone@ucar.edu

<sup>3</sup> School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, 699 S. Mill Avenue, Tempe, AZ 85281, USA; E-Mail: Shamanth.Kumar@asu.edu

<sup>4</sup> Center for Excellence in GIS, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA; E-Mail: nwaters@gmu.edu

\* Author to whom correspondence should be addressed; E-Mail: eschnebe@gmu.edu; Tel.: +1-703-993-1210.

Received: 16 December 2013; in revised form: 28 January 2014 / Accepted: 8 February 2014 /

Published: 18 February 2014

---

**Abstract:** Every year, flood disasters are responsible for widespread destruction and loss of human life. Remote sensing data are capable of providing valuable, synoptic coverage of flood events but are not always available because of satellite revisit limitations, obstructions from cloud cover or vegetation canopy, or expense. In addition, knowledge of road accessibility is imperative during all phases of a flood event. In June 2013, the City of Calgary experienced sudden and extensive flooding but lacked comprehensive remote sensing coverage. Using this event as a case study, this work illustrates how data from non-authoritative sources are used to augment traditional data and methods to estimate flood extent and identify affected roads during a flood disaster. The application of these data, which may have varying resolutions and uncertainties, provide an estimation of flood extent when traditional data and methods are lacking or incomplete. When flooding occurs over multiple days, it is possible to construct an estimate of the advancement and recession of the flood event. Non-authoritative sources also provide flood information at the micro-level, which can be difficult to capture from remote sensing data; however, the distribution and quantity of data collected from these sources will affect the quality of the flood estimations.

**Keywords:** flood assessment; volunteered geographical data; data fusion

---

## 1. Introduction

Flood disasters are a global problem capable of causing widespread destruction, loss of human lives, and extensive damage to property and the environment [1]. Flooding is not limited to a particular region or continent and varies in scale from creek and river flooding to tsunami or hurricane driven coastal flooding. Flood events in 2011, which include disasters resulting from the Japanese tsunami and flooding in Thailand, affected an estimated 130 million people and caused approximately \$70 billion spent on flood recovery [2].

The ability to produce accurate and timely flood assessments before, during, and after an event is a critical safety tool for flood disaster management. Furthermore, knowledge of road conditions and accessibility is crucial for emergency managers, first responders, and residents. Over the past two decades, remote sensing has become the standard technique for flood identification and management because of its ability to offer synoptic coverage [3]. For example, [4] illustrate how MODIS (Moderate Resolution Imaging Spectroradiometer) data can be applied for near real-time flood monitoring. The large swath width of the MODIS sensor allows for a short revisit period (1–2 days), and can be ideal for large continental scale flood assessments, but data are relatively coarse with a 250 m resolution. While [5] developed an algorithm for the almost automatic fuzzy classification of flooding from SAR (synthetic aperture radar) collected from the COSMO-SkyMed platform.

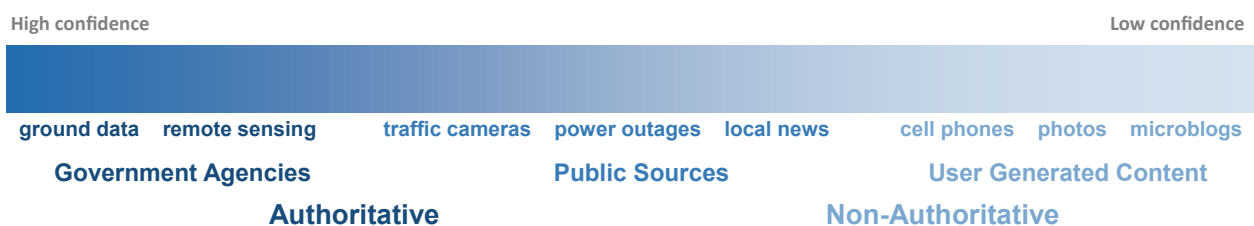
Unfortunately, satellite remote sensing for large scale flood disasters may be insufficient as a function of revisit time or obstructed due to clouds or vegetation. Aerial platforms, both manned and unmanned, are particularly suited for monitoring after major catastrophic events because they can fly below the clouds, and thus acquire data in a targeted and timely fashion, but may be cost prohibitive. Thus, it can be difficult to generate a complete portrayal of an event. The integration of additional data, such as multiple imagery, digital elevation models (DEM), and ground data (river/rain gauges) is often used to augment flood assessments or to combat inadequate or incomplete data. For example, [6] combine SAR with Landsat TM imagery and a DEM to derive potential areas of inundation. Reference [7] illustrate how fusing near real-time low spatial resolution SAR image and a Shuttle Radar Topography Mission (SRTM) DEM can produce results similar to hydraulic modeling. Reference [8] propose the integration of Landsat TM data with a DEM and river gauge data to predict inundation areas under forest and cloud canopy. While [9] used TerraSAR-X data with LiDAR to identify urban flooding.

The utilization of data from multiple sources can help provide a more complete description of a phenomena. For example, data fusion is often employed with remote sensing data to combine information of varying spatial, temporal, and spectral resolutions as well as to reduce uncertainties associated from using a single source [10]. The fused data then provides new or better information than what would be available from a single source [11]. The incorporation of multiple data sources or methods for improved performance or increased accuracy is not limited to the field of remote sensing. Boosting, a common machine learning technique, has been shown to be an effective method for generating accurate

prediction rules by combining rough, or less than accurate, algorithms together [12]. While the individual algorithms may be singularly weak, their combination can result in a strong learner.

This research extends this concept of employing multiple data sources for improved identification or performance by utilizing data from non-authoritative sources, in addition to traditional sources and methods, for flood assessment. Non-authoritative data describes any data which are not collected and distributed by traditional, authoritative sources such as government emergency management agencies or trained professionals. There is a spectrum of non-authoritative sources and the credibility, or level of confidence, in the data will vary by source characteristics (Figure 1). These sources can range from those considered to be somewhat “authoritative” such as power outage information garnered from local power companies or flooded roads collected from traffic cameras to what is clearly non-authoritative, such as texts posted anonymously on social media. Even data which lean toward the authoritative can be categorized as non-authoritative because of the lack of a traditional scientific approach to their collection, processing, or sampling. For example, Volunteered Geographic Information (VGI), a specific type of user generated content, is voluntarily contributed data which contains temporal and spatial information [13]. Because of the massive amount of real-time, on-the-ground data generated and distributed daily, the utilization of VGI during disasters is a new and growing research agenda. For example, VGI has been shown to document the progression of a disaster as well as promote situational awareness during an event [14–16]. Non-authoritative data are not limited solely to VGI and may include data which were originally intended for other purposes, but can also be harnessed to provide information during disasters. For example, traffic cameras provide reliable, geolocated information regarding traffic and road conditions in many cities worldwide, but have yet to be used for flood extent estimation during disasters.

**Figure 1.** Spectrum of confidence associated with authoritative and non-authoritative data sources.



The integration of non-authoritative data with traditional, well established data and methods is a novel approach to flood extent mapping and road assessment. The application of non-authoritative data provides an opportunity to include real-time, on-the-ground information when traditional data sources may be incomplete or lacking. The level of confidence in the data sources, ranging from authoritative to varying degrees of non-authoritative, impart levels of confidence in the resulting flood extent estimations. Recently, methods for flood estimation are including confidence or uncertainty in their results. For example, [17] used coarse resolution ENVISAT Advanced Synthetic Aperture Radar (ASAR) and high resolution ERS-2 SAR data to create various “uncertain flood extent” maps or “possibility of inundation” maps for calibrating hydraulic models. Reference [18] fused remote sensing flood extent maps to create an uncertain flood map representing varying degrees of flood possibility based on the aggregation of

multiple deterministic assessments. Reference [19] highlight uncertainty in satellite remote sensing data for flood prediction, especially in areas near the confluence of rivers, to create a probability of flooding versus a binary interpretation.

In June of 2013, the combination of excessive precipitation and saturated ground caused unprecedented flooding in the Canadian province of Alberta. The Bow River, the most regulated river in Alberta with 40% of its total annual flow altered by eight major upstream reservoirs, flows through the heart of downtown Calgary [20]. Intersecting the Bow River in downtown Calgary is the Elbow River which flows south to the Glenmore Dam, a reservoir for drinking water. The June flooding of the Bow River in 2013 is the largest flood event since 1932 with a peak discharge of 1470 m<sup>3</sup>/s, almost fifteen times the mean daily discharge of 106 m<sup>3</sup>/s [21]. The City of Calgary, in particular, experienced sudden and extensive flooding causing the evacuation of approximately 100,000 people [22]. The damage and recovery costs for public buildings and infrastructure in the City of Calgary are estimated to cost over \$400 million [23]. Because of extensive cloud cover and revisit limitations, remote sensing data of the Calgary flooding in June 2013 are extremely limited. A case study of this flood event provides an opportunity to illustrate a proof of concept where the incorporation of freely available, non-authoritative data of various resolutions and accuracies are integrated with traditional data to provide an estimate of flood extent when remote sensing data are sparse or lacking. Furthermore, this research presents an estimation of the advancement and recession of the flood event over time using these sources.

## 2. Data and Methods

### 2.1. Overview

The proposed methodology is based on the fusion of different layers generated from various data sources. This integration of multiple layers, which may have varying resolutions, sparse data, or different levels of uncertainty can provide information when they would not do so if used in isolation. The result is a flood extent map which is generated from the integration of these data layers. Layers are created using available remote sensing data, DEM, or ground information as traditionally used for flood assessment. The novelty of this methodology is the use of non-authoritative data to create additional data layers which are used to augment traditional data sources when they may be lacking or incomplete.

The creation of a fused data product for flood extent estimation utilizes a three step approach:

1. Generate layers;
2. Merge layers;
3. Create flood estimation map.

### 2.2. Data Sources and Layer Generation

Water identification and flood extent mapping can be accomplished using a variety of methodologies. The goal of this step is to generate multiple data layers which identify areas where water is detected. The task is method-independent and can be accomplished using any method best suited for a particular combination of data and location.

For this work, individual layers were created from eight different sources of available data for each day from 21–26 June 2013. Data sources, availability, and quantity will often vary by event and may also vary during an event. This was the case for the Calgary flooding where the quantity and variety of available data fluctuated over the course the flood event. Table 1 summarizes the data sources and their availability used in this study.

**Table 1.** Data sources and availability.

<b>Data</b>	<b>6.21</b>	<b>6.22</b>	<b>6.23</b>	<b>6.24</b>	<b>6.25</b>	<b>6.26</b>
River Gauge	X	X	X	X	X	X
Street Closures	X					X
Tweets	X	X	X	X	X	X
Photos	X					
RGB photo		X				
SAR		X				
Traffic Cameras	X					

### 2.2.1. Traditional Data

Traditional methods of flood classification are employed for four data sources:

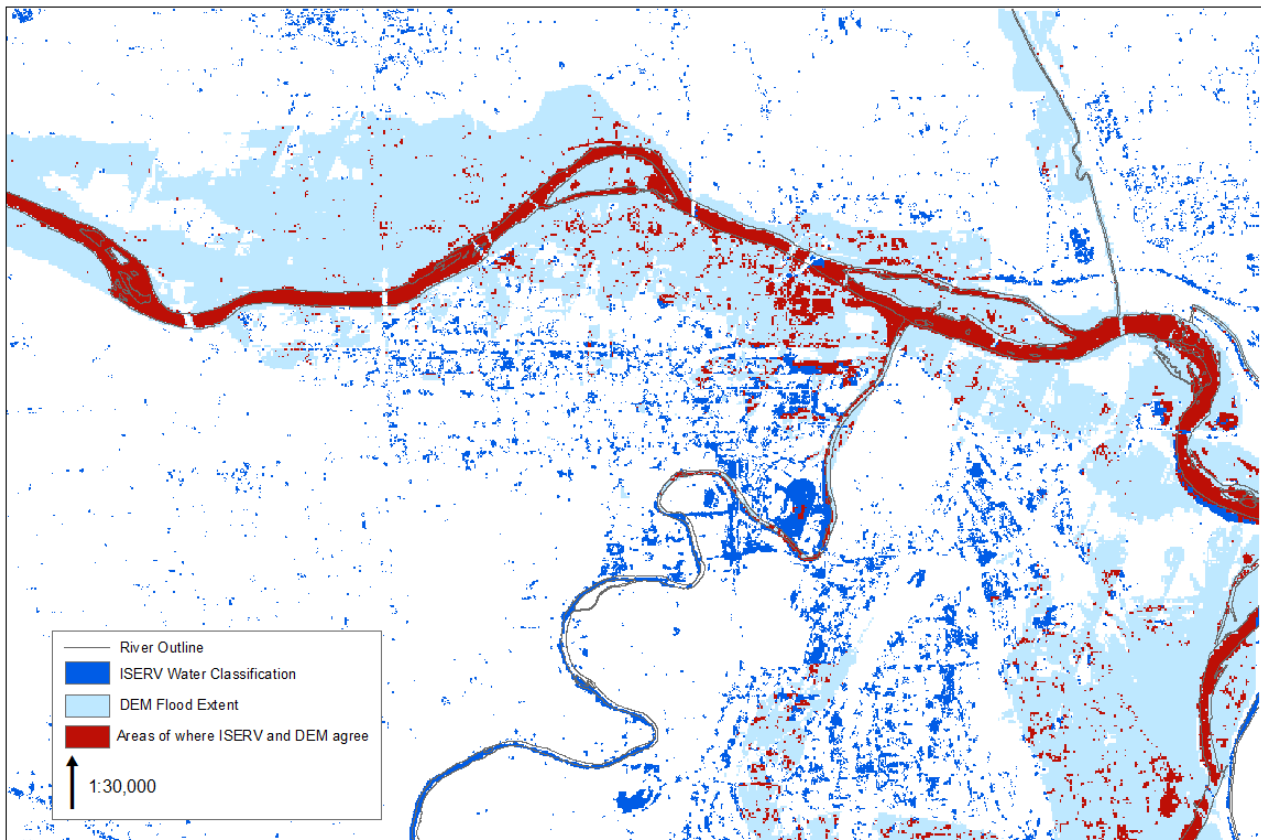
(1) A RGB composite photograph of Calgary with a resolution of 3.2 m was captured by the International Space Station’s Environmental Research and Visualization System (ISERV) on 22 June 2013, one day after the flood peaked in the downtown area. The image did not contain projection information so it was manually georectified and then projected to a UTM coordinate system in ArcGIS 10. A supervised maximum likelihood classification was performed to identify water in the scene.

Although the image was captured at almost the peak of the flood event, the classification of water in RGB composite photos is not optimal because of the difficulty distinguishing between water and land in the visible spectrum. In the ISERV image of Calgary, it is difficult to differentiate between the flood water, which appears as very brown, from roads and concrete in the urban areas. Although not ideal and containing noise, it is possible to identify large areas of water, for example, the main channels of the Bow and Elbow Rivers and flooding in the downtown area. There also is a large mismatch between the classification of the ISERV image and the flooding estimate using the paired DEM and river gauge for 22 June (Figure 2). This is likely the result of an over estimation of flooding using the DEM method and underestimation using the ISERV data. The areas where the two estimates intersect are regions where there is the highest confidence in the flood estimation (indicated in red in Figure 2). It is difficult to assign a value to either method to determine which approach may be more accurate. This challenge underlines the theme of this research, namely the application of multiple sources for improved performance.

(2) Synthetic Aperture Radar (SAR) imagery of Calgary and the surrounding High River area were collected by RADARSAT-2 on 22 June 2013. MDA Geospatial Services processed the SAR data as path-image SGF and then converted to calibrated backscatter (sigma) which was orthorectified using elevation SRTM data. The change detection was derived by thresholding a SAR difference image. The

thresholds were exported as bitmaps and later converted to polygons. The downtown Calgary area was selected from the available shapefile data and projected to a UTM coordinate system.

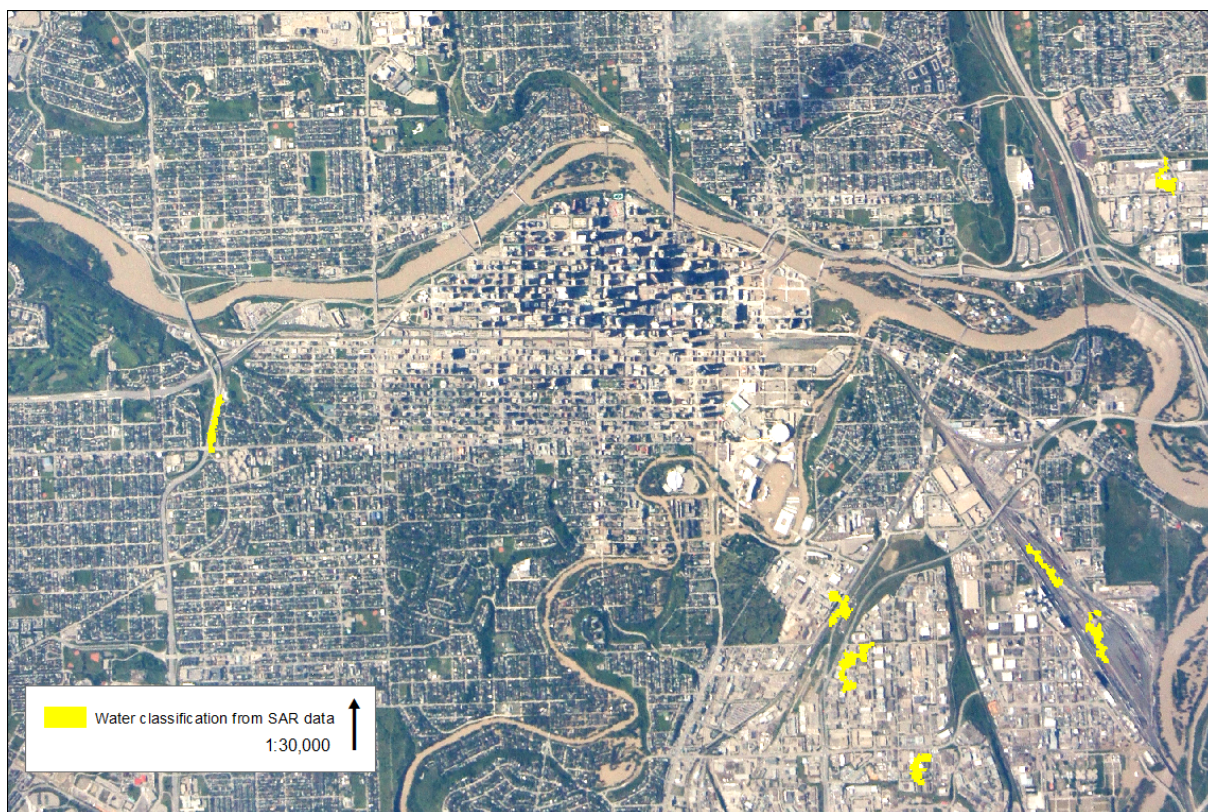
**Figure 2.** Supervised classification of water in the International Space Station's Environmental Research and Visualization System (ISERV) image and flood extent estimation using digital elevation model (DEM) paired with height of Bow River on 22 June 2013.



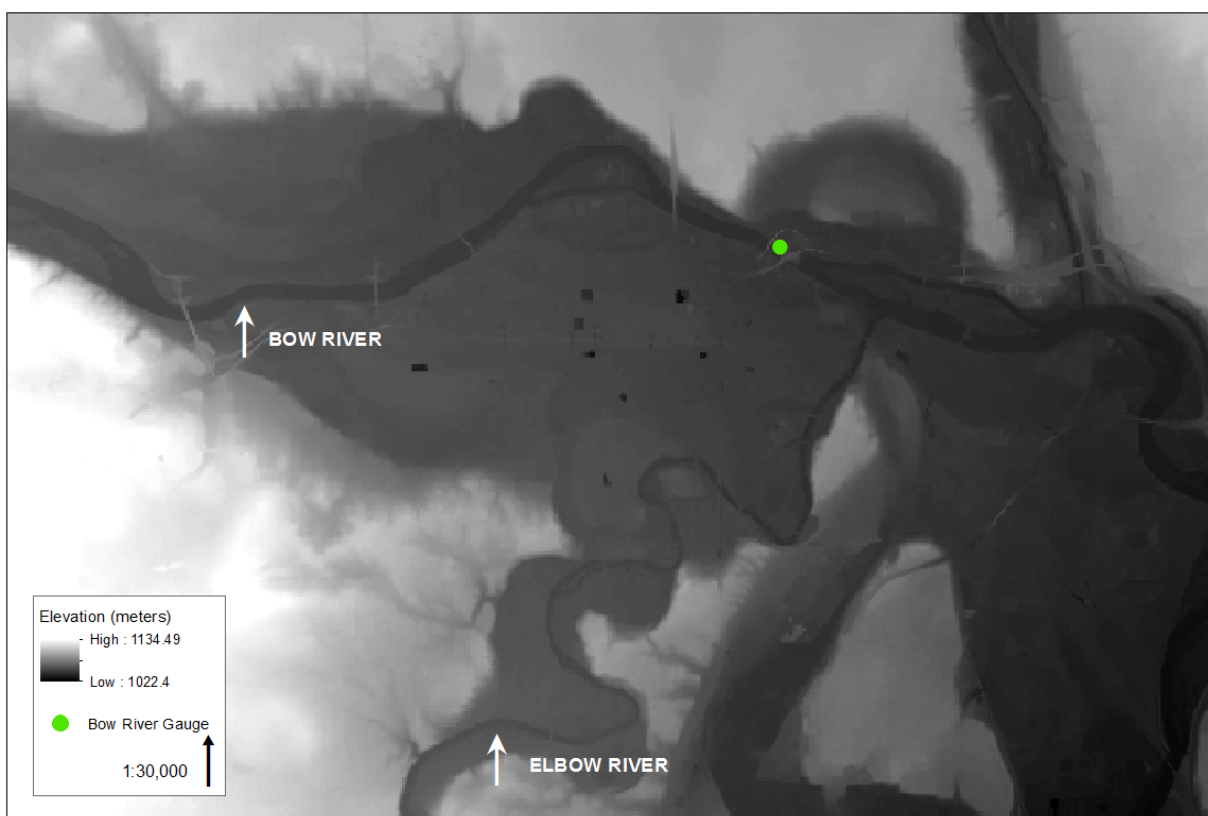
Because the scenes were originally planned for a separate purpose they were obtained using a wide beam, covering an area of 150 km<sup>2</sup> with a 30 m resolution. Consequently, the ground resolution was lower than would be optimally employed when tasking specifically for flood analysis. In addition, the lack of RADAR return off the water mixed with an oversaturated return from buildings made it difficult to accurately identify flood extent in the urban downtown area. As a result, the SAR data layer significantly underestimates the flood extent when compared to photos which document the presence of water in large areas of downtown Calgary (Figure 3). Regardless, the SAR data is included in this research because any information documenting the presence of water contributes and further strengthens the flood extent estimation as a whole.

(3) An AltaLIS LiDAR DEM with a 30 cm vertical and 50 cm horizontal accuracy was provided by the University of Calgary. The DEM was converted from an ESRI Arc Grid ASCII format into a GeoTiff layer with UTM coordinates in ArcGIS 10 (Figure 4).

**Figure 3.** Water classification from Synthetic Aperature Radar (SAR) data plotted over the ISERV photo from 22 June 2013.

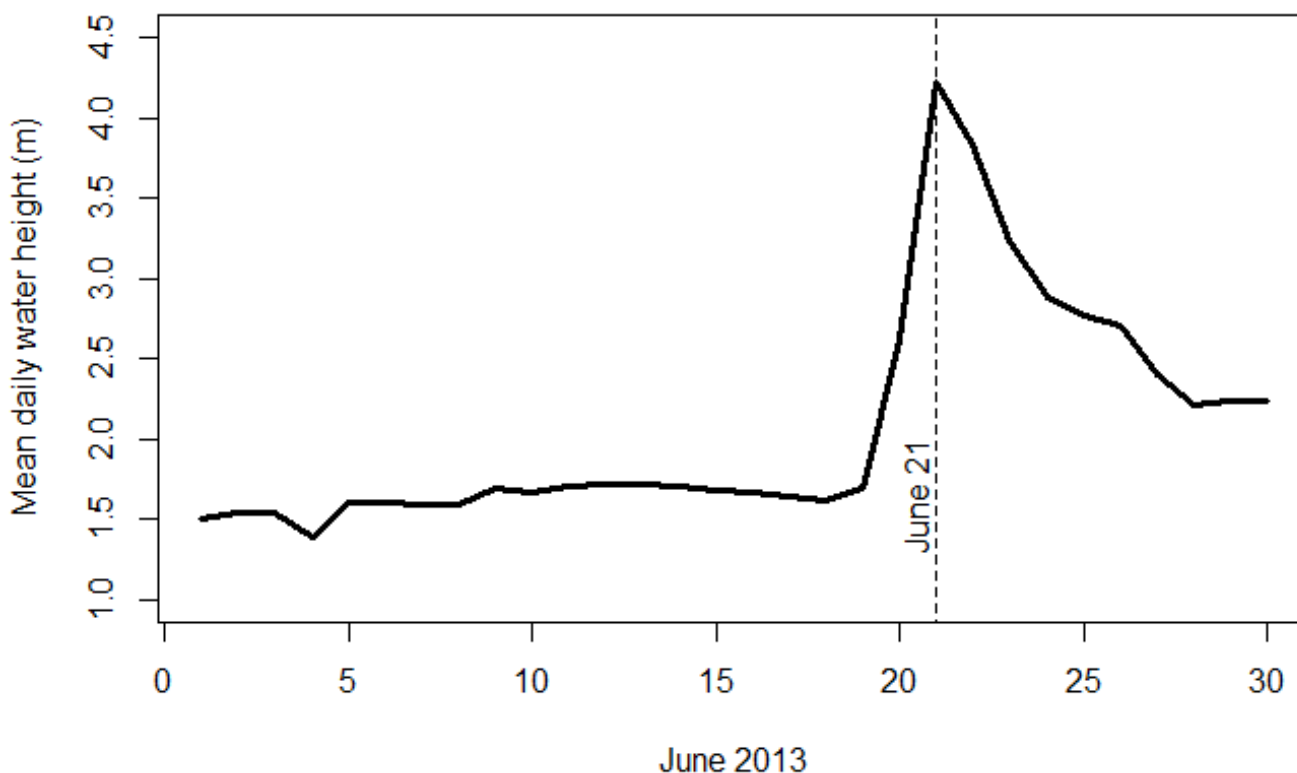


**Figure 4.** Digital Elevation Model for Calgary.



(4) Water height data for the Bow River were downloaded from the Environment Canada website. The data are provided by the Water Survey of Canada, the national authority for water levels and flow data in Canada. The water height data used for this study were collected in downtown Calgary from the Bow River gauge, station 05BH004, located at longitude:  $-114.051389\text{W}$ , latitude:  $51.05\text{N}$ . Mean daily water height for June 2013 were calculated from the quarter-hourly observations (Figure 5). Water height (or river stage) were converted to river height (elevation relative to sea level) by adding 1038.03 m to convert the data to the Geodetic Survey of Canada datum.

**Figure 5.** Mean daily water height for June 2013 on the Bow River in downtown Calgary.



Estimates of daily flood extent from 21–26 June 2013 were generated by pairing the DEM with the mean daily river height. Pixels in the DEM with a height less than or equal to the mean river height for each date are set as water pixels. This method is used to rapidly estimate a rough approximation of flood extent. However, the location and topography of Calgary, essentially a basin at the confluence of the Bow and Elbow Rivers, did not lend itself to a straightforward application. The elevation of the river gauge in downtown Calgary is approximately 1039 m while the elevation of the Bow River at its most western point in this study area is 1052 m and at the eastern edge is 1029 m. Consequently, when using water height data from the Bow River at Calgary station (located approximately in the center of the domain), the western and eastern reaches of the Bow River under or over flooded, respectively, when subtracting elevation from river height. A normalized DEM was created by incrementally decreasing the elevation west of the gauge as well as increasing east of the gauge. The new DEM was calibrated using the mean water height from 2012. Therefore, the daily flood extent estimations used in this work were generated using the normalized DEM created for this purpose.



### 2.2.2. Non-Authoritative Data

Four sources of non-authoritative data, which consists of point and line data identifying the presence of water, are employed:

(1) Volunteered Geographic Information (VGI) in the form of geolocated photos ( $n = 39$ ) which documented flooding within the study domain (51.064456N to 51.013595N latitude and  $-114.136188W$  to  $-114.003663W$  longitude) were downloaded using the Google search engine.

(2) Arizona State University's TweetTracker provided Twitter data for this project [24]. Geolocated tweets ( $n = 63$ ) generated in the study domain during 21–26 June 2013 containing the word “flood” were utilized.

(3) The City of Calgary maintains 72 traffic cameras which provide real-time traffic conditions for major roads around the city. The images collected by the cameras were manually inspected on the beyond.ca website on 26 June 2013. At that time all of the cameras were offline with time stamps for 8:30 am, 21 June 2013. A few cameras ( $n = 7$ ) provided information regarding the state of the roads (clear/flooded) on the morning of 21 June, while the majority did not have imagery available.

(4) A list of Calgary road and bridge closures on 21 June 2013 ( $n = 36$ ) were collected from an on-line news source. Using a road network of Calgary downloaded from the OpenStreetMap website, the data were digitized in ArcGIS 10 to recreate road closures for 21 June [25]. Road closures for 26 June 2013 were downloaded from a Google Crisis map accessed from The City of Calgary website. The data were downloaded into ArcGIS 10 and converted from a KML format to a GeoTiff layer.

Data layers are created from each non-authoritative source for each day data is available from 21–26 June 2013. The layers are generated by first plotting and georeferencing flooded areas which are identified in photos or traffic cameras or inferred from Twitter or road closures. These areas begin as point or line features and are assigned a value of 1, with the remaining non-flooded areas assigned values of 0. A kernel density smoothing operation is then applied to each layer. The kernel smoothing was accomplished using ArcGIS 10 which employs a quadratic kernel function as described by [26]. Let  $(x_1, x_2, \dots, x_n)$  be samples drawn from a distribution with an unknown density  $f$ , the goal is to estimate the shape of this function. The general kernel density estimator is

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (1)$$

where  $K$  is the kernel function and  $h$  is the bandwidth or smoothing parameter.

The density smoothing is employed with the point and line data to spatially extend their representation in preparation for Step 2. This is a necessary step because point data can become insignificant when combined or merged with data from other sources, such as flood extent estimated from a DEM and river gauge data.

In the process of performing the smoothing operation, the bandwidth was varied by data source. Bandwidth is a parameter which determines the amount of smoothing. As the bandwidth is increased, the smoothness will increase, yielding progressively less detail. If the bandwidth is too narrow, the representation can be too irregular, making interpretation difficult. By increasing the bandwidths during the smoothing operation the density values were not significantly changed, but by incorporating a large number of surrounding cells, it is possible to create a more generalized grid. An increase in smoothing is

more important for some data types than others. For example, the bandwidth used for the road data was smaller compared to the bandwidth, or amount of smoothing, utilized for the tweet or photo data. The choice to increase or decrease bandwidth was based on the assumption that when using road closures as an indication of flooding this would be more localized information than flood information from photographs or tweets. Following the kernel smoothing, the layers are converted to raster format to facilitate the layer merging in Step 2.

### 2.3. Layer Merge

Following the generation of individual data layers, a weighted sum overlay application is utilized to merge them together. The use of a weighted sum overlay approach allows for two processes to be accomplished in one step: (i) weights are assigned to each data layer based on data characteristics (ii) multiple data layers are integrated into a single comprehensive layer per time interval. The presence of water ( $W_i$ ) at cell  $i$  is given by

$$W_i = \sum_{i=1}^n w_i x_i \quad (2)$$

where weight  $w$  is a user selected weighting scaler chosen for each data layer and  $x$  is the value of a cell  $i$ . The weight describes the importance of a particular observation, or the confidence associated with a data source. Source confidence is a function of multiple variables: confidence in the producer (anonymous vs. authoritative or trusted), accuracy of geolocation (manually geolocated, automatic, or fixed (*i.e.*, traffic cameras)), trust in the method of water identification (machine learning algorithm vs. processing of text). While all data contain some level of uncertainty, data collected from social media, in particular Twitter, this uncertainty can be particularly high because of producer anonymity and questions related to the reliability of filtering methods and geolocation accuracies. However, Twitter can still provide relevant and timely information, with its uncertainty moderated by using it in conjunction with other sources considered more reliable (*i.e.*, traffic cameras). To account for this higher uncertainty, or low level of confidence, Tweets (microblogs) are assigned the lowest weight of all data sources with the remaining sources weighted linearly following the scale in Figure 1. Following the addition of a weight to each layer, the layers are summed together yielding a comprehensive merged data layer for each day from 21–26 June 2013.

### 2.4. Flood Estimation Map

A flood estimation map is then generated for each day using the comprehensive merged layer. This may be accomplished using a variety of mathematical, statistical, or machine learning approaches. For this article, kriging is used to create a geostatistical representation from the merged layer. The geostatistical technique of kriging creates an interpolated surface from the spatial arrangement and variance of the nearby measured values [27]. Kriging allows for spatial correlation between values (*i.e.*, locations/severity of flooding) to be considered and is often used with Earth science data [28–30]. Kriging utilizes the distance between points, similar to an inverse weighted distance method, but also considers the spatial arrangement of the nearby measured values. A variogram is created to estimate

spatial autocorrelation between observed values  $Z(x_i)$  at points  $x_1, \dots, x_n$ . The variogram determines a weight  $w_i$  at each point  $x_i$ , and the value at a new position  $x_0$  is interpolated as:

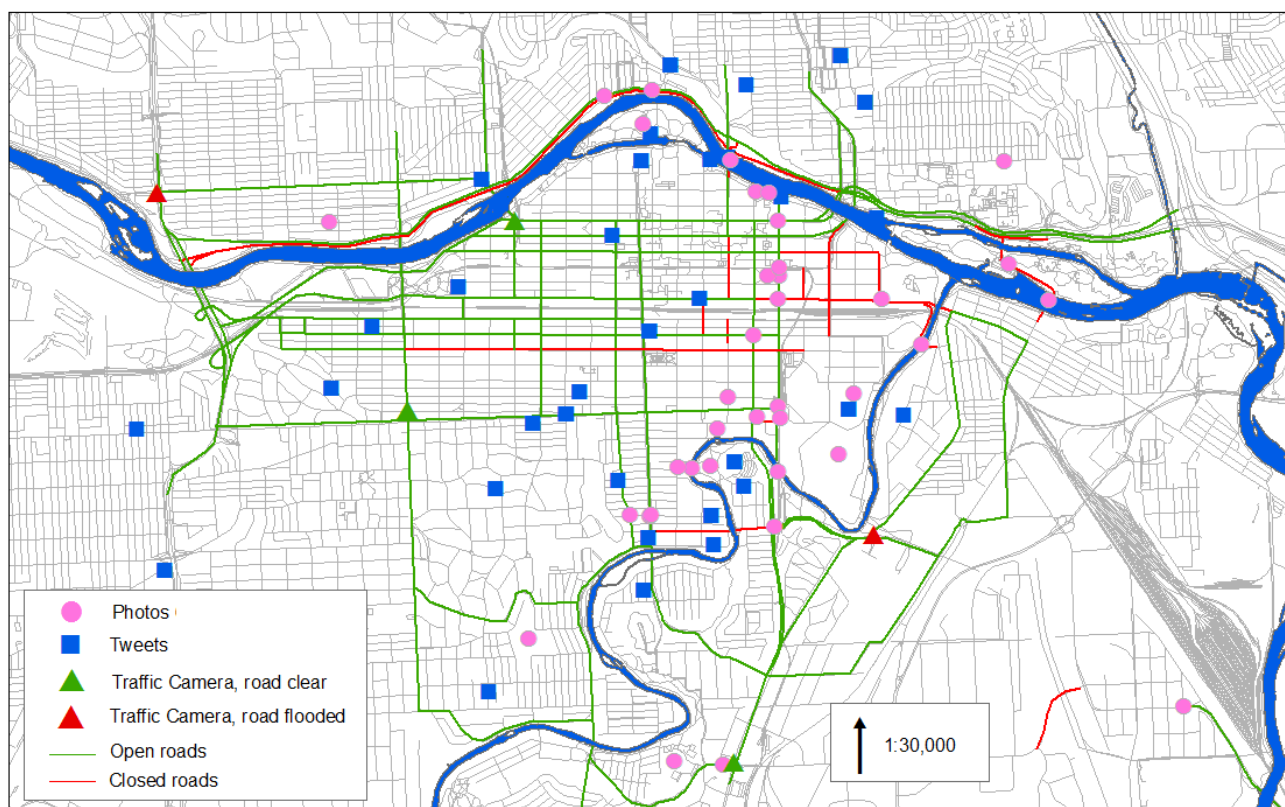
$$\hat{Z}(x_0) = \sum_{i=1}^n w_i Z(x_i) \quad (3)$$

The geostatistical interpolation yields a flood extent product for each time interval based on the fusion of traditional and non-authoritative data.

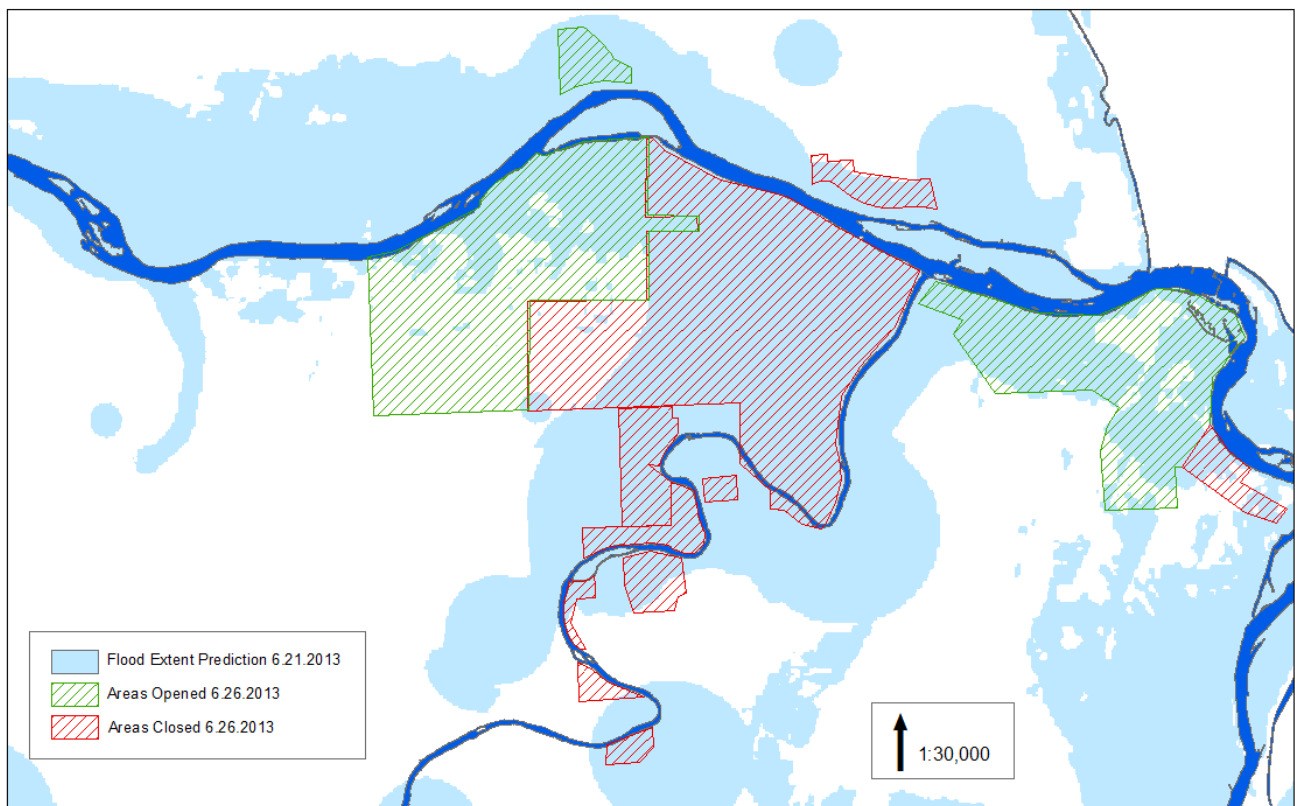
### 3. Results

Using the methodology described in Section 2.3 the data layers are merged together for each date, 21–26 June 2013, yielding 6 daily layers for geostatistical interpolation. The layer weights are assigned linearly based on confidence in the data source following the scale in Figure 1. Specifically, the Tweets were assigned a weight of 1, photos a 2, road closures (local news) and traffic cameras a 3, and remote sensing and DEM data a 4. Because this research extended over a 6 day period, there were more data available on some days compared to others (Table 1). This did not affect the actual methodology for layering, as the layers are weighted and summed together in one step regardless of the number of layers used. The locations of the non-authoritative data were generally well distributed across the domain (Figure 6). Although the volume of non-authoritative data varied from day to day with some days only having a sparse amount, it has been shown that even a small amount of properly located VGI data can help improve flood assessment [31].

**Figure 6.** Distribution of non-authoritative data.



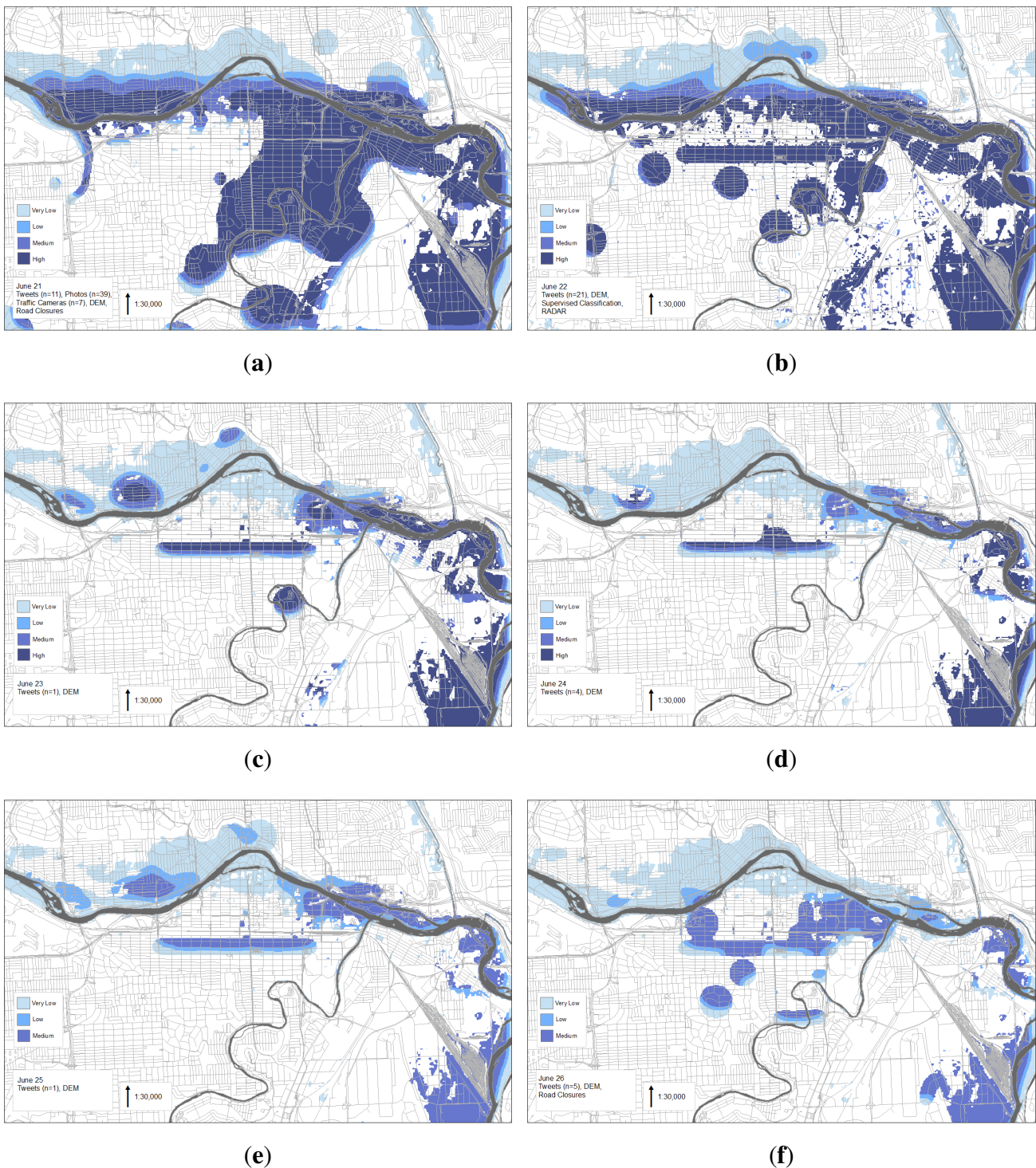
**Figure 7.** Flood extent estimated for 21 June as compared to areas which had been previously closed (and opened as of 26 June) and areas still closed on 26 June.



Following the merging of layers, flood estimation maps are generated as discussed in Section 2.4. Figure 7 is a comparison of the maximum flood extent which was estimated on 21 June (Figure 8a) and areas indicated as closed from a Google Crisis map available for this flood event. The maps agree well in some areas and not in others. Some of the areas of over estimation are likely due to the DEM utilized which had been manually normalized to account for changes in elevation in the scene. In addition, non-authoritative data may be providing flood information not captured in the Google Crisis map, specifically flooding which have receded or localized neighborhood flooding at a “micro-level”. Figure 8 illustrates daily estimated flood extent for 21–26 June 2013. The flood time series maps are presented as an estimate of flood extent over the 6 day period as well as the level of confidence in the estimations. The daily series demonstrates a progression from severely flooded, 21 June, through a flood recession. The quantity of data available each day does appear to affect the map results. For example, only two days had road closure data, 21 June and 26 June. Because of the quantity and variety of the data for 21 June, the road closure layer is well assimilated with the rest of the data (Figure 8a). This is not the case for 26 June, where a much smaller amount of data were available. This results in the road closure layer being evident as indicated by the horizontal flood estimated in the center of the image (Figure 8f). An assumption was made that the road categorized as flooded on 26 June was likely flooded on previous days as well, but because of a lack of road data for 22–25 June, it was not included in the original analysis. Therefore, a decision was made to include the closed road on 26 June into the data sets for previous days. This results in the horizontal flooded area in (Figure 8b–e). The sparseness of

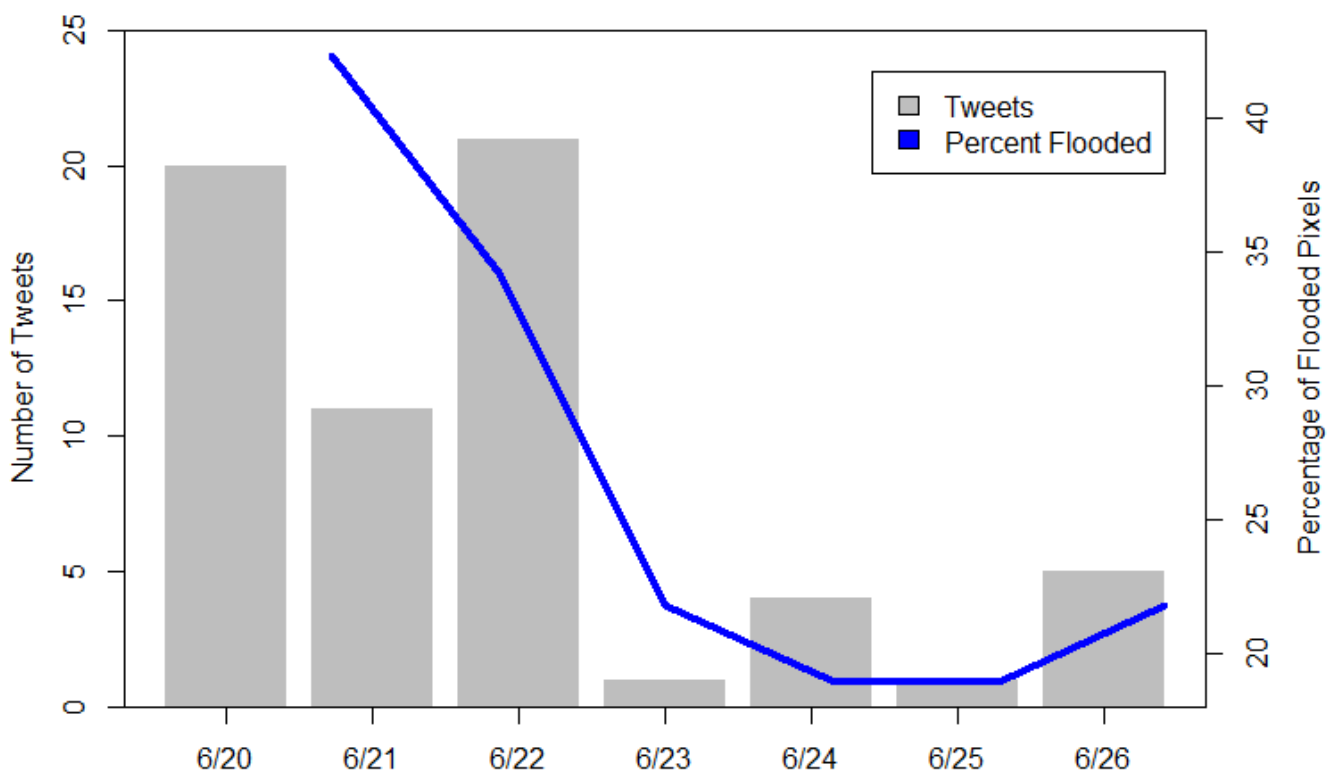
data is also evident by the circular areas of flooding. These are the result of individual tweets which are located too far from the the majority of data in the scene to be properly integrated (Figure 8b,c,f). By comparing these flood maps to the one created for 21 June (Figure 8a) it is clear that a smoother and richer estimation can be accomplished as data volume and diversity increases.

**Figure 8.** Flood extent estimation and road assessment. The categories (very low, low, medium, high) represent the confidence in a pixel being affected by flooding.(a) 21 June; (b) 22 June; (c) 23 June; (d) 24 June; (e) 25 June; (f) 26 June.



The overall tweet volume corresponds well to the progression of the flood event (Figure 9). The maximum number of tweets are posted during the peak of the flood and then reduce as the flood recedes. It is unclear why there are small increases in the number of tweets during the later days of the flood event. These tweets may be related to flood recovery with information regarding power outages, drinking water, or closures/openings of public facilities. Figure 9 also illustrates the area of the flood as a function of time. By using the flood extent estimations created with this methodology, flood area is represented as the percentage of pixels classified as flooded each day in (Figure 8a–f). Flood area does increase slightly the last day of the study. This is likely the result of a corresponding increase in tweets for the same day and not an actual increase in flood area.

**Figure 9.** Progression of tweet volume and flooded area over time.



The estimation of flood extent can be further processed by applying an additional kernel smoothing operation. This may be necessary for layers with lower data quantities. For this research, a smoother flood extent was desired. The flood maps were exported from ArcGIS as GeoTiff files and then smoothed using **R** statistical software. The same kernel density estimator as in Equation (1) was applied. The specific methodology used for kernel smoothing and its **R** implementation is described by [32].

#### 4. Discussion and Conclusions

The June 2013 flooding in Calgary is a good example of how remote sensing data, although a reliable and well tested data source, are not always available or perhaps cannot provide a complete description of a flood event. As a case study, this work illustrates how the utilization and integration of multiple data sources offers an opportunity to include real-time, on-the-ground information. Further, the identification of affected roads can be accomplished by pairing a road network layer with the flood extent estimation.

Roads which are located within the areas classified as flooded are identified as regions which are in need of additional evaluation and are possibly compromised or impassable (Figure 8a–f). Roads can be further prioritized as a function of distance from the flood source (*i.e.*, river or coastline) or distance from the flood boundary. This would aid in prioritizing site inspections and determining optimal routes for first responders and residents. In addition, pairing non-authoritative data with road closures collected from news and web sources provides enhanced temporal resolution of compromised roads during the progression of the event.

The addition of weights allows for variations in source characteristics and uncertainties to be considered. In this analysis weight was assigned based on confidence in the source, for example, observations published by local news are assumed to have more credibility than points volunteered anonymously. However, other metrics can be used to assign weight. For example, the volume of the data can be used to assign higher weight to data with dense spatial coverage and numerous observations. The timing of the data could also be used as a metric for quality. As shown, tweet volume decreases during the progression of the event, with perhaps non-local producers dropping out as interest fades. This may possibly yield a more valuable data set of tweets, those just generated by local producers, which could be inferred to be of higher quality and thus garner a higher weight. However, it is not possible to set the weights as an absolute because each flood event is unique and there will be differences in data sources, availability, and quantity.

Currently, authoritative flood maps from this event have not been made available to the general public, making the validation of the flood extent estimations in Figure 8 difficult. However, even when available, the issue of correctness or accuracy in official estimates should be addressed. Non-authoritative data often provide timelier information than that provided through authoritative sources. In addition, these data can be used for the identification of flooding at a micro-level, which is often difficult to capture using authoritative sources or traditional methods. Although not considered ground truth, non-authoritative data does provide information in areas where there might otherwise be none. However, the flood estimations are controlled by the distribution and quantity of the data. For example, landmark areas are more likely to receive public attention and have flooding documented than other less notable areas. Therefore researchers should be aware of, and recognize, the potential for skewness in the spatial distribution of the available data, and thus the information garnered from it. Moreover, a lack of ground data can be simply an indication of no flooding or can be the result of differences in the characteristics of places within the domain. The importance of data quantity is evident in (Figure 8) where a decrease in quantity and variability of data during the progression of the event creates a less consistent flooded surface, with single tweets standing out in isolation on days when the quantity of data is low. However, the fusion of data from multiple sources yields a more robust flood assessment providing an increased level of confidence in estimations where multiple sources coincide.

While the analysis presented in this work was performed after the flood event, this methodology can be extended for use during emergencies to provide near real-time assessments. The use of automated methods for the ingestion, filtering, and geolocating of all sources of non-authoritative data would decrease processing time as well as provide a larger volume of data which could also enhance results. In addition, the time required to collect and receive remote sensing data is moving toward real-time availability. For example, unmanned aerial vehicles (UAVs) were deployed during the Colorado floods

in September 2013 with images processed and available to the Boulder Emergency Operations Center within an hour [33]. Recent research is also utilizing social media to identify areas affected by natural disasters for the tasking of satellite imagery [34]. Although in this work specific data sources were used, this methodology can be applied with any data available for a particular event.

## Acknowledgments

The authors would like to thank the two anonymous reviewers for their insightful comments on earlier versions of this article.

Work performed under this project has been partially supported by US DOTs Research and Innovative Technology Administration (RITA) award # RITARS-12-H-GMU (GMU #202717). **DISCLAIMER:** The views, opinions, findings and conclusions reflected in this presentation are the responsibility of the authors only and do not represent the official policy or position of the USDOT/RITA, or any State or other entity.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Jha, A.; Bloch, R.; Lamond, J. *Cities and Flooding: A Guide to Integrated Urban Flood Risk Management for the 21st Century*; World Bank Publications: Washington, DC, USA, 2012.
2. EMDAT, 2013. EM-DAT: The OFDA/CRED International Disaster Database, Université Catholique de Louvain, Brussels, Belgium. Available online: <http://www.emdat.be/database> (accessed on 1 August 2013).
3. Smith, L. Satellite remote sensing of river inundation area, stage, and discharge: A review. *Hydrol. Processes* **1997**, *11*, 1427–1439.
4. Brakenridge, R.; Anderson, E. MODIS-based flood detection, mapping and measurement: The potential for operational hydrological applications. In *Transboundary Floods: Reducing Risks through Flood Management*; Springer: Dordrecht, The Netherlands, 2006; pp. 1–12.
5. Pulvirenti, L.; Pierdicca, N.; Chini, M.; Guerriero, L. An algorithm for operational flood mapping from Synthetic Aperture Radar (SAR) data based on the fuzzy logic. *Nat. Hazard Earth Syst. Sci.* **2011**, *11*, 529–540.
6. Townsend, P.; Walsh, S. Modeling floodplain inundation using an integrated GIS with RADAR and optical remote sensing. *Geomorphology* **1998**, *21*, 295–312.
7. Schumann, G.; di Baldassarre, G.; Alsdorf, D.; Bates, P. Near real-time flood wave approximation on large rivers from space: Application to the River Po, Italy. *Water Resour. Res.* **2010**, *46*, doi:10.1029/2008WR007672.
8. Wang, Y.; Colby, J.; Mulcahy, K. An efficient method for mapping flood extent in a coastal floodplain using Landsat TM and DEM data. *Int. J. Remote Sens.* **2002**, *23*, 3681–3696.



9. Mason, D.; Speck, R.; Devereux, B.; Schumann, G.; Neal, J.; Bates, P. Flood detection in urban areas using TerraSAR-X. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 882–894.
10. Zhang, J. Multi-source remote sensing data fusion: Status and trends. *Int. J. Image Data Fusion* **2010**, *1*, 5–24.
11. Pohl, C.; van Genderen, J. Review article multisensor image fusion in remote sensing: Concepts, methods and applications. *Int. J. Remote Sens.* **1998**, *19*, 823–854.
12. Freund, Y.; Schapire, R.; Abe, N. A short introduction to boosting. *J. Jpn. Soc. Artif. Intell.* **1999**, *14*, 1–14.
13. Goodchild, M. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.
14. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake shakes Twitter users: Real-time event detection by social sensors. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; ACM: New York, NY, USA, 2010; pp. 851–860.
15. Vieweg, S.; Hughes, A.; Starbird, K.; Palen, L. Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. In Proceedings of the 28th International Conference on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; ACM: New York, NY, USA, 2010; pp. 1079–1088.
16. Acar, A.; Muraki, Y. Twitter for crisis communication: Lessons learned from Japan’s tsunami disaster. *Int. J. Web Based Commun.* **2011**, *7*, 392–402.
17. Di Baldassarre, G.; Schumann, G.; Bates, P.D. A technique for the calibration of hydraulic models using uncertain satellite observations of flood extent. *J. Hydrol.* **2009**, *367*, 276–282.
18. Schumann, G.; di Baldassarre, G.; Bates, P.D. The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2801–2807.
19. Stephens, E.; Bates, P.; Freer, J.; Mason, D. The impact of uncertainty in satellite data on the assessment of flood inundation models. *J. Hydrol.* **2012**, *414*, 162–173.
20. BRBC. Bow River Basin Council: Dams and Reservoirs, 2013. Available online: <http://wsow.brbc.ab.ca> (accessed on 10 December 2013).
21. Water Survey of Canada, 2013. Available online: <http://www.ec.gc.ca/rhc-wsc> (accessed on 10 December 2013).
22. Upton, J. Calgary Floods Trigger an Oil Spill and a Mass Evacuation, 2013. Available online: <http://grist.org/news/calgary-floods-trigger-an-oil-spill-and-a-mass-evacuation/> (accessed on 25 June 2013).
23. Fletcher, R. Calgary Flood Costs Now Total \$460 Million: A Report, 2013. Available online: <http://metronews.ca/news/calgary/783593/calgary-flood-costs-now-total-460-million-report/> (accessed on 2 September 2013).
24. Kumar, S.; Barbier, G.; Abbasi, M.A.; Liu, H. TweetTracker: An analysis tool for humanitarian and disaster relief. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, (ICWSM), Barcelona, Spain, 17–21 July 2011.
25. OpenStreetMap, 2013. OpenStreetMap. Available online: <http://www.openstreetmap.org/> (accessed on 25 June 2013).

26. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC Press: Boca Raton, FL, USA, 1986; Volume 26.
27. Stein, M.L. *Interpolation of Spatial Data: Some Theory for Kriging*; Springer Verlag: New York, NY, USA, 1999.
28. Oliver, M.A.; Webster, R. Kriging: A method of interpolation for geographical information systems. *Int. J. Geogr. Inf. Syst.* **1990**, *4*, 313–332.
29. Olea, R.A.; Olea, R.A. *Geostatistics for Engineers and Earth Scientists*; Kluwer Academic Publishers: Dordrecht, The Netherlands 1999.
30. Waters, N. Representing surfaces in the natural environment: Implications for research and geographical education. In *Representing, Modeling and Visualizing the Natural Environment: Innovations in GIS 13*; Mount, N., Harvey, G., Aplin, P., Priestnall, G., Eds.; CRC Press: Boca Raton, FL, USA, 2008; pp. 21–39.
31. Schnebele, E.; Cervone, G. Improving remote sensing flood assessment using volunteered geographical data. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 669–677.
32. Wand, M.; Jones, M. *Kernel Smoothing*; Chapman & Hall: New York, NY, USA 1995; Volume 60.
33. FALCON, 2013. Falcon UAV Supports Colorado Flooding until Grounded by FEMA. Available online: <http://www.falcon-uav.com/falcon-uav-news/2013/9/14/-falcon-uav-supports-colorado-flooding-until-grounded-by-fem.html> (accessed on 14 September 2013).
34. Waters, N.; Cervone, G. Using Social Networks and Commercial Remote Sensing to Assess Impacts of Natural Events on Transportation Infrastructure. Available online: <http://trid.trb.org/view/2012/P/1243850> (accessed on 25 June 2013).

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).