

Article

Mapping Seasonal Inundation Frequency (1985–2016) along the St-John River, New Brunswick, Canada using the Landsat Archive

Ian Olthof

Canada Centre for Mapping and Earth Observation (CCMEO), Natural Resources Canada, 560 Rochester, Ottawa, ON K1A 0E4, Canada; ian.olthof@canada.ca; Tel.: +1-613-759-6275

Academic Editors: Magaly Koch and Prasad S. Thenkabail

Received: 14 December 2016; Accepted: 4 February 2017; Published: 10 February 2017

Abstract: Extreme flood events in recent years in Canada have highlighted the need for historical information to better manage future flood risk. In this paper, a methodology to generate flood maps from Landsat to determine historical inundation frequency is presented for a region along the St-John River, New Brunswick, Canada that experiences annual springtime flooding from snowmelt and river ice. 1985–2016 Landsat data from the USGS archive were classified by combining See5 decision trees to map spectrally variable water due to spring ice and sediment, and image thresholding to map inundated floodplains. Multiple scenes representing each year were overlaid to produce seasonal time-series of spring (March–May) and summer (June–August) maximum annual water extents. Comparisons of annual surface water maps were conducted separately for each season against historical hydrometric water depth as a measure of relative springtime flood severity, and 1 m water masks from digital orthophotos were used to perform a formal accuracy assessment of summer water. Due to Landsat's 16-day revisit time, peak flood depth was poorly related to flood extent; however, spring depth measured during Landsat acquisitions was significantly related to extent ($\tau = 0.6$, p -value < 0.001). Further, summer maps validated against 30 m water fractions scaled from 1 m water masks were over 97% accurate. Limitations with respect to the assessment of flood extent from depth, timing differences between peak flood depth and extent due to Landsat revisit time and cloud cover, and suggestions to overcome limitations through multi-sensor integration including radar are discussed.

Keywords: Landsat; flood; inundation

1. Introduction

Climate change is expected to increase the amount and intensity of precipitation received over high-latitude regions [1,2], producing greater flood risk in the future. Extreme flood events have caused property damage costing in the millions to billions of dollars in Canada in recent years. In 2013 alone, flooding in southern Alberta caused damage in excess of \$6 billion while in Toronto, heavy rainfall produced urban flooding that led to property damages approaching \$1 billion [3]. Although these events represent extremes, in other parts of Canada springtime flooding caused by snowmelt in combination with river ice jamming and precipitation occurs annually. The St-John River in Canada's province of New Brunswick has experienced 14 springtime flood events in the past decade, with an average cost of more than \$9 million in damages [4]. In order to minimize impacts, better information on the area and frequency of historical flooding is required to manage risks associated with future events.

Flood risk may be obtained from statistical modelling with climate data as a main input [5,6]. A more direct way of assessing flood risk is by using historical maps of flood extent, which exist for

some regions and events in Canada but are far from comprehensive. Flood extent may be estimated by combining inundation depth with a digital elevation model (DEM) [7–9]. However, sufficiently accurate and detailed DEMs are not available for most floodplains in Canada, while those that exist may not accurately represent man-made or other barriers to water flow. Global hydraulic modelling using a conditioned 90 m DEM based on the Shuttle Radar Topographic Mission (SRTM) and regionalized flood frequency analysis based on climate and catchment characteristics produced flood risk maps that captured between two-thirds and three-quarters of the flood risk area in benchmark flood hazard datasets [10]. Remote sensing provides an alternate and more straightforward means of obtaining accurate historical flood extents through direct mapping of surface water.

In order to accurately map flood risk with historical remote sensing data, annual peak flood conditions must be imaged to capture the full extent of the floodplain that is inundated on an annual or semi-annual basis. Capturing peak flood conditions depends on the timing of the flood event in relation to the timing of the satellite overpass, with more frequent imaging providing a better chance of capturing the maximum annual flood extent. In Canada, routine flood mapping is currently performed using Synthetic Aperture Radar (SAR) from Canada's RadarSat 2 satellite with plans to migrate to RadarSat constellation data when operational in 2018. RadarSat 2 has a repeat cycle of 24 days, but is side looking up to 29.8 degrees off nadir [11], effectively reducing the revisit time to a minimum of 2–3 days depending on acquisition mode and latitude. In practice, surface water monitoring with RadarSat 2 has been performed bi-weekly in Canada's prairie pothole region in [12] to every three weeks in Canada's Athabasca region by [13].

Landsat satellites provide optical data with identical coverage and geometry every 16 days, while two satellites were in orbit simultaneously over much of the Landsat program period, effectively doubling the number of acquisitions. The number of observations may be further increased if the area of interest falls within the overlap region of two adjacent orbits with more overlap occurring at higher latitudes. However, the usability of Landsat data is affected by cloud cover while radar is able to penetrate cloud due to its all-weather imaging capabilities. Landsat's consistent repeat coverage provides the ability to minimize the effect of cloud by selecting and combining clear-sky areas from multiple scenes through compositing, which produces a variable number of observations by location. In RadarSat, imagery can be tasked with different modes and geometry to increase the number of scenes for monitoring. Unlike Landsat, tasking produces variable scene extents, resulting in different locations being imaged from scene to scene and a variable number of observations by location.

Decreasing the revisit time through multi-sensor integration enhances the likelihood of imaging peak flood extent. This paper evaluates flood extent mapping from Landsat for future integration with other sensors including radar. A validated methodology for springtime flood extent mapping that deals with problems of water turbidity and ice is required to increase the number of observations during the flood season for near real-time operations. This methodology is also applied to historical Landsat 5, 7, and 8 data acquired from the freely available USGS archive to generate a time-series of flood maps that are used to locate permanent water and periodically flooded zones for mapping areas vulnerable to flood risk. Groups including the Australian Government have mapped surface water dynamics from regional to continental scales using a similar approach to generate historical flood information for flood risk assessment and ecosystem monitoring [14–16]. Notable differences in this paper are due to the Canadian context, where flooding occurs most often during the spring snowmelt season.

In this paper, two separate Landsat stacks were built from 1985 to 2016 Landsat imagery for the spring and summer seasons. The summer stack was used to map permanent and ephemeral water caused by precipitation and drought, while the spring stack provided similar information due to snowmelt, river ice jamming, and precipitation. Combining information from the two image stacks offers a more complete picture of flood extents, frequencies, and their causes. Approaches for verifying spring and summer flood extents extracted from Landsat are described based on best available reference data. One approach compares hydrometric data collected at flood gauge stations

with spring flood extents, while the other uses high-resolution water masks generated from 1 m digital orthophotos to validate the water extraction methodology during summer.

2. Materials and Methods

2.1. Study Area

The St-John River flows from Maine to the Bay of Fundy and forms the Canada-US border along parts of its length (Figure 1). The river is approximately 673 km long, with most of it located in the province of New Brunswick. Its widest point is nearly 3 km across and is located just north of the City of Saint-John near its mouth into the Bay of Fundy. While annual springtime flooding occurs along much of its length, it is most severe along a 55 km stretch from Fredericton to Gagetown where ice jams form in an area of expansive floodplains with low relief. Hydrometric flood gauge data from Environment Canada show that maximum depth was generally reached between mid to late April from 1985 to 2016 at the Fredericton station.

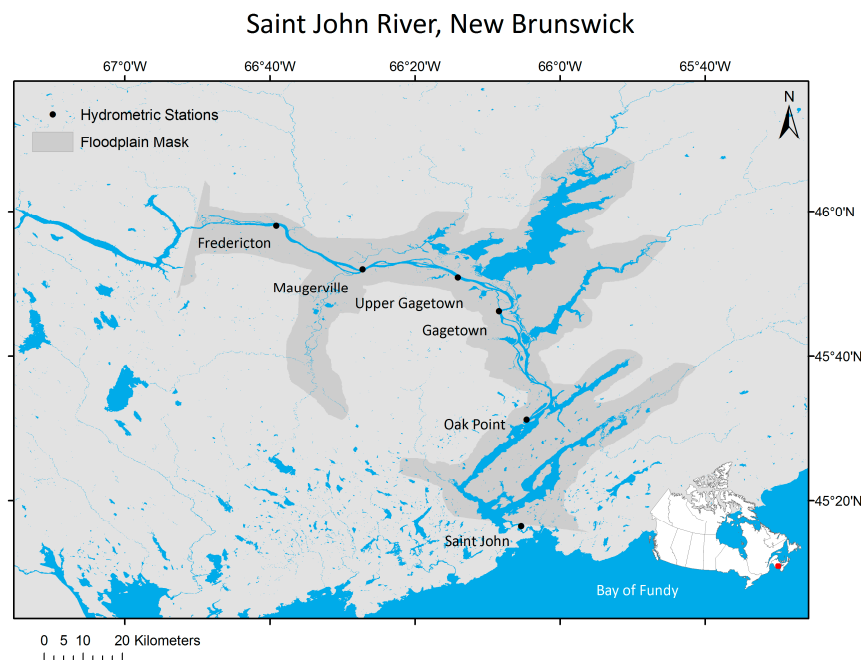


Figure 1. Study area along the St-John River including the floodplain mask and six hydrometric stations.

2.2. Data

2.2.1. Landsat

Landsat data were downloaded from the USGS through EarthExplorer™ representing years from 1985 to 2016 and WRS Path/Row 10/28. Scenes from Landsat 5 TM, 7 ETM+ and eight OLI satellites were 30 m resolution terrain corrected Level 1T with geodetic accuracies generally less than one-quarter of a pixel. Two image stacks corresponding to the spring melt season containing 72 scenes acquired between the months of March and May, and 130 summer scenes between the months of June and August were built separately. Scenes were visually screened to include only those that had cloud and haze-free data. This resulted in an average of nearly 2.5 scenes per year for the springtime stack with a range of 1–6 scenes per year excluding three years where no suitable springtime images were found. The summer stack had slightly more than four scenes per year with a range of 1–11 scenes and no years missing.

Landsat data were converted to Top-of-Atmosphere reflectance using coefficients in the header file following [17]. The fmask algorithm was used to identify areas of cloud and cloud shadow [18],

which were treated as no-data along with no-data strips in Landsat 7 due to scanline correction (SLC-off) failure post-2003.

2.2.2. Orthophotos

Digitally scanned natural colour orthophoto images acquired on unspecified anniversary dates between the summers of 1996 and 2002 were downloaded from the Province of New Brunswick's geographic information data catalogue (<http://www.snb.ca/geonb1/e/index-E.asp>). Twenty seven orthophotos each covering an area approximately $8 \text{ km} \times 6 \text{ km}$ at 1 m spatial resolution along the St-John River from Fredericton to Gagetown (including Grand Lake) were processed to generate high resolution water masks for validating Landsat-based water extraction. Images were downloaded in MrSID format and converted to GeoTIFF for processing. Lakes were extracted using Adobe Photoshop CS5 raster seeding that automatically infills homogeneous features based on colour similarity and spatial adjacency. A combination of automated and manual processing and editing in Photoshop produced highly accurate water masks based on visual assessment.

From the fine resolution water masks, water fraction images were generated showing the percentage of 1 m water pixels located in each 30 m pixel aligned with the Landsat grid (Figure 2). This was performed by locating the center of each Landsat pixel in the 1 m resolution lake mask image and calculating the number of water pixels within the surrounding 31×31 pixel window as a percentage of the window size (961 pixels).

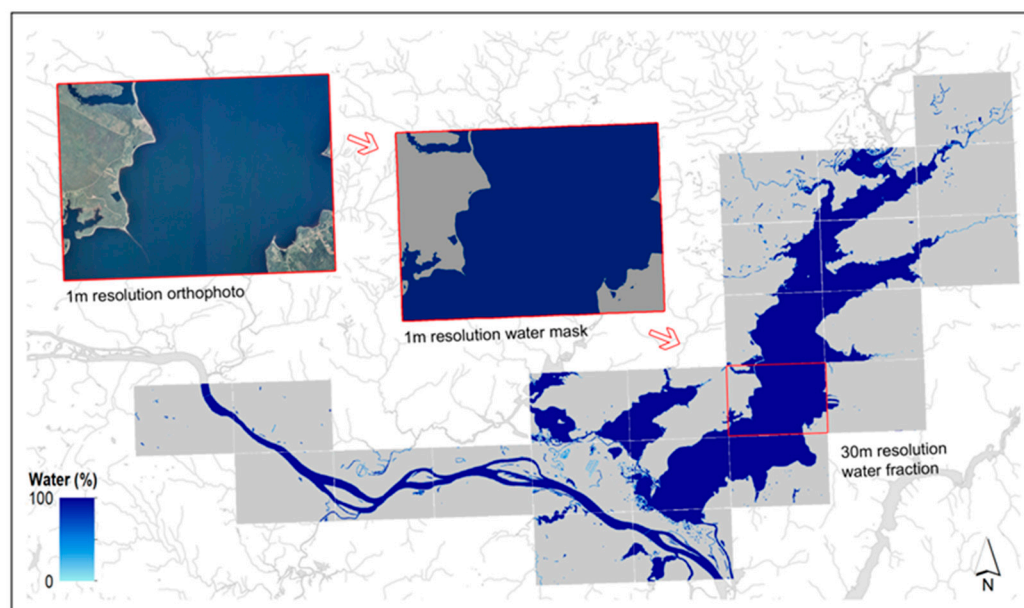


Figure 2. Reference 30 m water fractions generated from 1 m orthophotos covering nearly 1300 km^2 in the floodplain region of the St-John River from Fredericton to Gagetown.

2.2.3. Hydrometric Data

Hydrometric data represent the best available measure of flood severity coincident with Landsat flood extents. Archived historical hydrometric data were downloaded from Environment Canada's Water Survey portal (<https://www.ec.gc.ca/rhc-wsc/>) for comparison with Landsat. Data consisted of daily water depth from six stations located along the St-John River from Fredericton to the river mouth at the Bay of Fundy near the City of St-John. Daily data were complete from 1985 to 2014 for four of six stations except for Gagetown and Upper Gagetown, for which archived data only existed from 1994 to 2014. Maximum annual flood depth and date were obtained from daily data at each station location. To explore the limitations of using Landsat to map peak flood extents, timing, and depth differences

between annual hydrometric peak flood depth and hydrometric flood depth, Landsat acquisition dates were checked. Peak annual flood depths and dates were compared to hydrometric flood depths measured on dates when corresponding annual Landsat scenes were acquired to calculate the average depth difference and number of days between peak flood and each Landsat acquisition. Because annual Landsat images were combined to obtain a maximum annual flood extent, peak annual flood depths and dates were also compared to the maximum water depth measured on Landsat acquisition dates in corresponding years.

2.2.4. CanVec Data

CanVec are digital reference data over Canada representing the best sources of geographical information available. The hydrography layer representing lakes and rivers at 1:50,000 scale was gridded at the same 30 m resolution and extent of Landsat scenes in the image stacks. These data were used to extract classification training data consisting of reflectance values of water and land from individual Landsat scenes in the stack.

2.3. Water Classification

Surface water can have high spectral variability caused by suspended sediment, shallow water and ice cover during flooding in early springtime. This variability makes water classification challenging due to numerous spectral signatures representing water that can vary from scene-to-scene depending on the timing of Landsat. Land is also spectrally variable due to multiple cover types present in a Landsat scene that change because of springtime vegetation phenology. To address these challenges, a sample of scene-specific reflectance values representing both land and water needed to be input into a classifier that can handle data that may not be normally distributed. See5 decision trees [19] were chosen to classify land and water because they meet all the above criteria and have been shown to perform as well or better than other classifiers [20,21]. The See5 decision tree algorithm has been applied extensively to remote sensing classification problems [22,23], most notably by the USGS to generate the Landsat-based NLCD land cover product [24].

Historical water extents were mapped for each Landsat scene from 1985 to 2016 using CanVec permanent waterbodies to sample scene-specific water and land reflectance for input into See5. The assumption with this approach is that differences due to changes in water extent between CanVec and seasonal water are small compared to the whole scene that forms the training sample, and are seen as error in the training data. Because these errors form a small percentage of the overall training data, decision tree classifiers that are known to be robust to training error [25] will tend to predict correctly in their presence. Ephemeral water was not input into the classifier since it may not be present in an individual scene. Rather, ephemeral water that is present will be correctly classified if it is spectrally similar to permanent water somewhere in the scene. See5 was implemented in R Statistics [26] using default parameters of 10 boosting iterations, a confidence factor of 0.25, and no winnowing.

Trials revealed shortcomings of the classification approach mainly in mapping shallow water that were addressed using a complementary method involving thresholding. The Landsat shortwave infrared (SWIR—band 5: 1.55–1.75 μm in TM and ETM+, band 6: 1.56–1.66 μm in OLI) has been shown to separate water from land in summer better when used alone than in combination with other bands [27–29]. Shortwave infrared radiation is strongly absorbed by water and is minimally affected by atmospheric aerosols, suspended sediment, or plant material. A typical SWIR frequency histogram of land and water is bimodal [27], with the darkest mode representing water and a second mode land (Figure 3). A scene-specific minimum value in the mixed pixel region between water and land modes was determined to be a suitable threshold value below which pixels were assigned to water [29].

Final water extents were obtained on a scene basis by combining See5 and threshold water maps where water occurred in one or the other. Maximum annual water extents were subsequently obtained by overlaying all scene maps in a given year and assigning water to the annual flood map anywhere water was present in one of the scenes. Annual flood maps were stacked separately for each season

and inundation frequency was calculated on a pixel basis for each stack. Inundation frequency was determined as the percentage of valid observations unobstructed by cloud or cloud shadow that was classified as water.

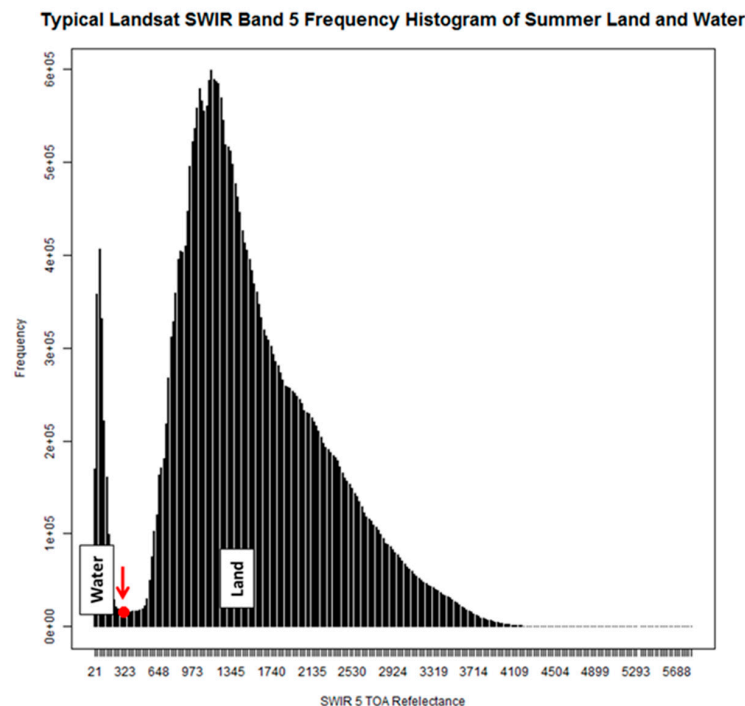


Figure 3. Landsat SWIR bimodal histogram of land and water in summer, with the threshold value determined as the minimum value between the two modes shown in red.

2.4. Validation

The quality of surface water maps was assessed separately for summer and spring flood images using the best available reference data. High-resolution water masks generated from digital orthoimages were used to assess summer surface water maps, while spring flood extents were compared to hydrometric flood depth using scatterplots and rank-based correlation (Kendall's Tau) to evaluate the level of agreement between the two measures of relative flood severity. The summer validation employed the standard accuracy assessment method of reporting the per-pixel agreement between map and reference data using an error matrix and derived statistics. This approach represents a validation of the flood extraction methodology since each water extent map in the time-series cannot be validated independently and because orthoimages were acquired in summers between 1996 and 2002 with no specific anniversary date or year attached to each image. Therefore, a single set of reference images acquired at unknown times between 1996 and 2002 were used to assess each annual summer water extent map from 1996 to 2002.

Because water masks generated from orthophotos were scaled to 30 m fractions for comparison with Landsat, a less than 50% water fraction threshold was chosen to assign pixels to land while pixels with greater than or equal to 50% were water. Accuracy was calculated separately for each 10% water fraction bin from 0% to 100% to determine if bias existed in the Landsat water extents compared to reference water fractions. For comparison with hydrometric data, validating flood extents directly from depth gauge data is not possible without an accurate DEM. Canadian Digital Elevation Data (CDED) DEM is in integer format with vertical errors attributable to horizontal errors in the source data of up to 100 m at the 1:50 k scale [30]. While a direct one-to-one comparison between flood depth and extent is not possible due to an inaccurate DEM, both provide relative measures of flood severity and therefore their ranks should be similar as measured by tau correlation.

The area of flood extent was extracted beneath a mask covering the St-John River and surrounding floodplain. Because data are missing for certain years beneath this mask due to cloud and shadow, flood extent may represent different areas between years. Therefore, flood extent was normalized by calculating the ratio of clear-sky area occupied by flood water relative to clear-sky area occupied by permanent water where inundation occurred in all summers from 1985 to 2014 (Equation (1)).

$$FE_N = ((FE_{cs} + PW_{cs})/PW_{cs}) \times PW \quad (1)$$

where FE_N is the normalized flood extent calculated from FE_{cs} as the flood extent beneath the clear-sky area, PW_{cs} : the permanent water area beneath the clear-sky area and PW : the total permanent water area beneath the study area mask (Figure 4). Absolute flood extent was also determined separately for a large common clear-sky area in 23 of 29 years representing 46% of the study area.

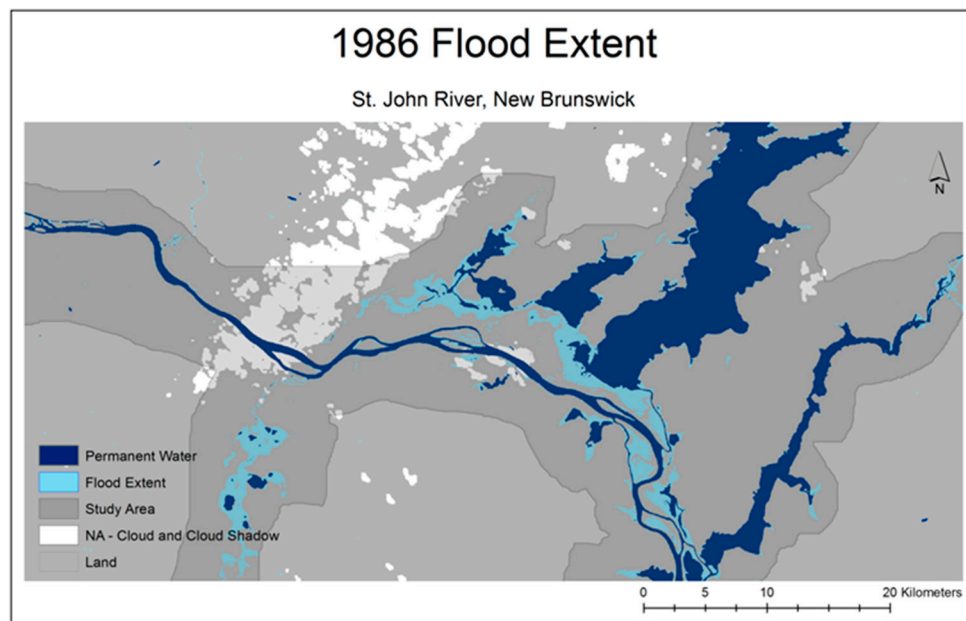


Figure 4. Inputs for calculating the 1986 normalized flood extent beneath the St-John River floodplain mask.

3. Results

1985–2016 annual maps showing the extent of springtime open water due to flooding are shown in Figure 5 with no data areas due to cloud, cloud shadow, and Landsat SLC-off strips in white. While not shown, maps of summer open water extents were generated for the same years with less area covered by no data due to a greater number of summer scenes.

3.1. Assessment—Spring Water Extents vs. Flood Depth

Fredericton was chosen of the six hydrometric stations for comparison between flood extent and depth because it represented a complete water level record from 1985 to 2014. Water depth was highly correlated among all stations with the timing of peak depth having a median occurrence three days later at the St-John station compared to Fredericton located 120 km upstream. The 1985–2014 average daily flood depths and their daily changes were plotted as a function of day of year (DOY) to examine the rate at which flood waters rise and recede (Figure 6). Flood waters rise at an average rate of ~0.1 m per day from DOY 60–110, or end of February to mid-April. Maximum flood depth occurs immediately after at around DOY 112. Waters recede from peak depth at a similar rate of 0.1 m per day until around DOY 170 occurring mid-June.

Annual Springtime Flood Maps

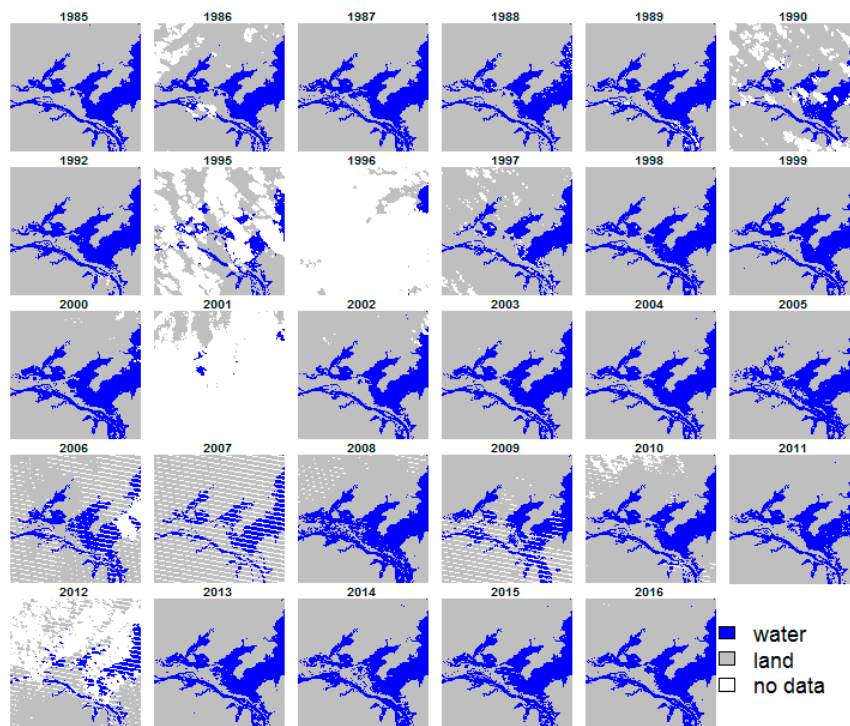


Figure 5. Annual time-series of springtime flood extent maps from Landsat zoomed in on the Area of Interest.

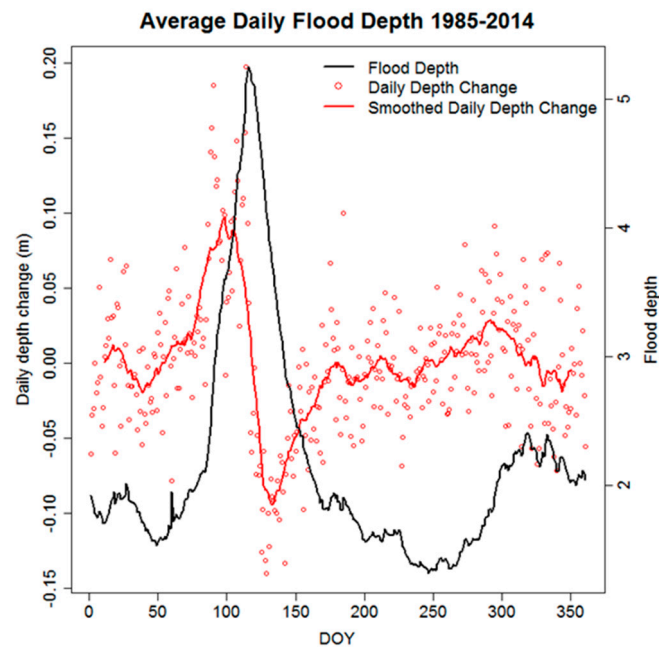


Figure 6. Average 1985–2014 daily flood depth and depth change by Day of Year (DOY) at Fredericton hydrometric station.

The average duration between annual Landsat acquisitions and corresponding annual peak flood at Fredericton was slightly more than 18 days, with a difference of nearly 2.2 m water depth.

When comparing peak flood with the maximum flood depth measured on days when Landsat was acquired in each year, the difference was 12 days and 1.4 m water depth. Thus, the correspondence between peak flood depth and flood extent from Landsat cannot be expected to be perfect due to the timing difference between annual Landsat acquisitions and annual peak depth.

While not shown, the correspondence between peak flood depth and flood extent measured by Landsat is low because of the time lag between the two caused by the 16-day Landsat satellite revisit time and rapidly rising spring flood water levels. However, the correlation between flood depth on Landsat acquisition dates and corresponding flood extent was significant, suggesting the method used to extract flood extent from Landsat produced measurements of relative flood severity that agreed well with hydrometric water depth. Figure 7 shows this relation excluding years 1996, 2015, and 2016 due to missing hydrometric data, with flood depth representing the maximum depth on annual Landsat acquisition dates versus annual normalized flood extent (FE_N) producing a highly significant correlation ($\tau = 0.6$, $N = 26$, $p\text{-value} < 0.001$). Similarly, the correlation between absolute flood extent beneath the common clear-sky area in 23 of 29 years and maximum depth on Landsat acquisition dates was also highly significant ($\tau = 0.66$, $N = 23$, $p\text{-value} < 0.001$).

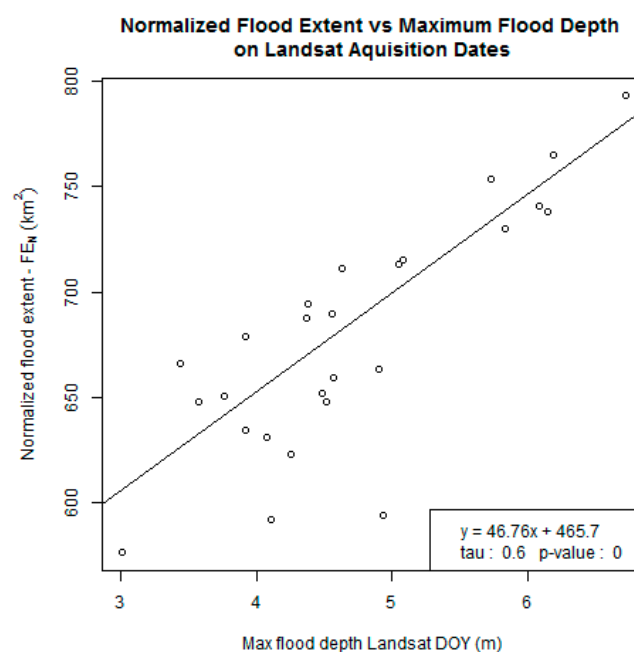


Figure 7. Normalized flood extents (FE_N) compared to the maximum flood depth measured at the Fredericton hydrometric station on corresponding annual Landsat acquisition dates.

3.2. Assessment—Summer Water Extents vs. Ortho Water Fractions

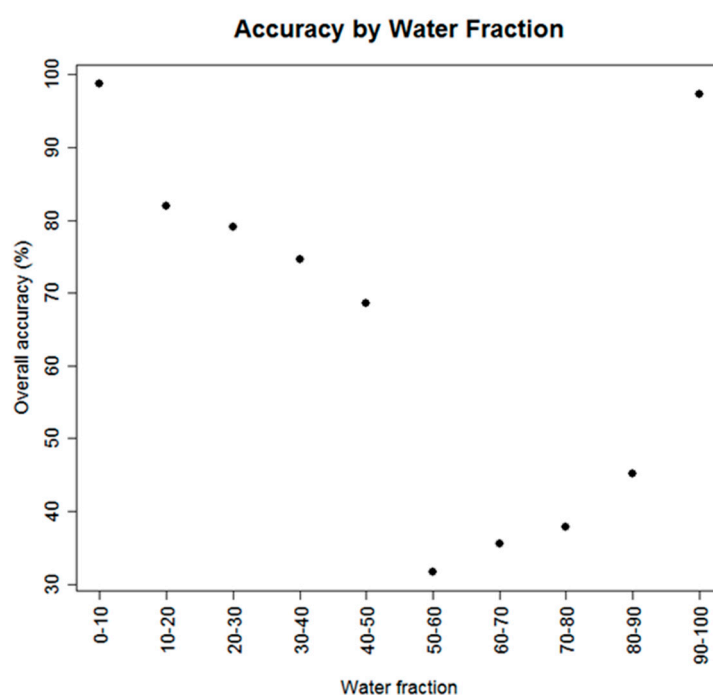
The error matrix and derived statistics comparing the 1999 summer Landsat water extent classification with reference data from orthophoto water fractions is shown in Table 1. The 1999 classification was chosen since it represents the median year of the orthophoto acquisitions, and all years had overall accuracies that were within less than 1% of 1999 accuracy. The assessment was performed on more than one million 30 m pixels.

The overall accuracy was 97.38% with a kappa statistic that accounts for agreement due to random chance of 93.02%, which is highly significant and indicates almost perfect agreement. An inspection of the number of pixels by land and water class in the reference data compared to the classification reveals a slight underestimation of the amount of water in the classification compared to reference data (271,046 water pixels in the classification vs. 272,370 in the reference). Classification accuracy was subsequently examined by binned reference water fractions to determine the source of bias.

Table 1. Error matrix comparing the 1999 summer water extent map from Landsat to reference water from 1 m orthophotos scaled to 30 m resolution.

		Reference Data			
		Water	Land	Total Classification	User's Accuracy (%)
Classification data	Water	257,470	13,576	271,046	94.99
	Land	14,900	802,548	817,448	98.18
	Total reference	272,370	816,124	1,088,494	
	Producer's accuracy (%)	94.53	98.34		
	Overall accuracy (%)	97.38			
	Kappa (%)	93.02			

Figure 8 shows overall the accuracy of 30 m pixels for each 10% water fraction bin from orthophotos. Overall accuracy for pure land (0% water fraction) and pure water (100% water fraction) pixels is in the range of over 97% to nearly 99%. As orthophoto water fraction increases from 0% and pixels become more mixed, overall accuracy decreases to just below 70% for pixels in the 40%–50% water fraction range. Beyond a water fraction of 50%, overall accuracy suddenly drops to just over 30% before increasing gradually to almost 45% in the 80%–90% water fraction range.

**Figure 8.** 1999 summer Landsat water classification accuracy by 30 m orthophoto water fraction.

3.3. Inundation Frequency

Figure 9 shows 1985–2016 inundation frequency maps generated from spring and summer Landsat water extents. Spatial patterns reveal a gradient of decreasing inundation frequency from permanent water to permanent land in both spring and summer, especially in the large floodplain area between St-John River and lakes to the north. The total area beneath the floodplain mask that is subject to periodic summer flooding is 693 km², increasing to 906 km² in spring. Thus, 213 km² of land within the floodplain area is vulnerable to periodic flooding in springtime only. The area of permanent water also varies by season, with 487 km² permanently inundated in summer increasing to 558 km² in spring.

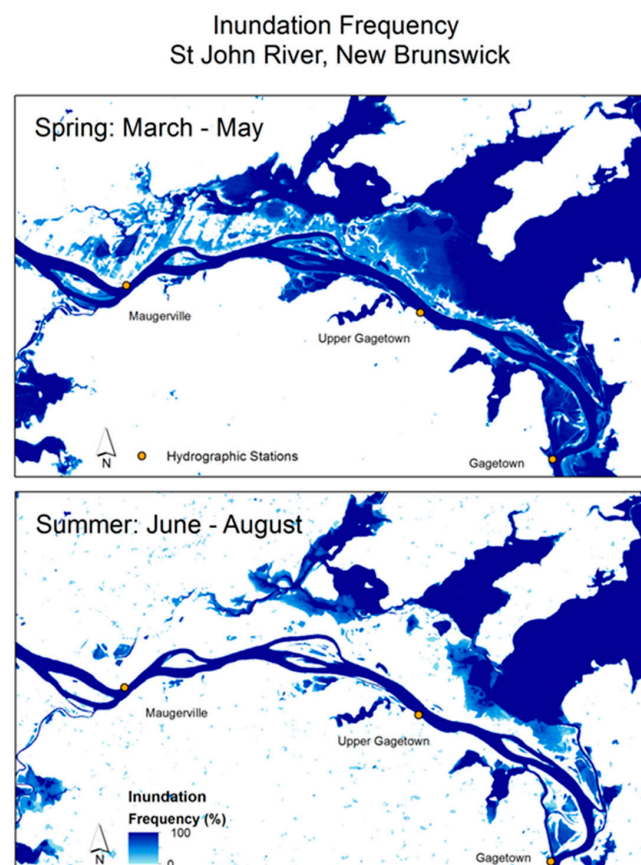


Figure 9. Inundation frequency maps showing the percentage of 1985–2016 annual observations that were flooded along the St-John River in spring (**top**) and summer (**bottom**).

4. Discussion

Validating flood extent is often challenging due to a lack of coincident measures of flood severity other than flood depth. Water depth can be converted to flood extent using a model provided that a current and accurate DEM are available. For most floodplains in Canada however, DEMs do not exist at the required level of accuracy and furthermore, when validating annual flood events spanning multiple years the depth-extent relation changes due to changes in floodplain topography caused by erosion and deposition. Therefore, periodic updating of DEMs is needed in this approach while the required level of elevation accuracy may only be achieved with a LIDAR survey, which is costly. Newer methods of generating point clouds from Structure-from-Motion (SfM) using optical remote sensing surveys from either Unmanned Aerial Vehicles (UAVs) or manned aircraft are promising and more cost-effective, but have rarely been implemented at the scale of most floodplains. Recently, [31] performed a comparison between DEMs derived with aerial and satellite photogrammetry, DEMs generated from LIDAR, Interferometric Synthetic Aperture Radar (InSAR), and SfM over a coastal floodplain in Alaska. Results showed that the LIDAR DEM was most accurate, having a Root Mean Square Error (RMSE) of 8 cm, while SfM was second with a RMSE of 19 cm.

Mapping springtime flood extent by combining See5 classification and SWIR thresholding produced relative measures of flood severity that were significantly related to flood depth from hydrometric stations. Both classification methods proved to be complementary with each method performing better in different circumstances. Flood depth at the time of Landsat acquisition was significantly related to flood extent from See5 alone ($\tau = 0.52$, p -value < 0.001) as well as thresholding ($\tau = 0.58$, p -value < 0.001), while the combined methodology produced a relation that was more significant than either individual method ($\tau = 0.60$, p -value < 0.001). Absolute flood extent

beneath the common clear-sky area in 23 of 29 years was even more strongly related to flood depth ($\tau = 0.66$, p -value < 0.001). Although these relations were highly significant, they were not perfect due to timing differences between hydrometric measurements and flood extents that can change due to water current, natural changes in floodplain topography between years, and human intervention seeking to minimize flood damage.

See5 classification performed well in most cases, except during the earliest part of the springtime season when lakes and rivers were covered with thick ice in the southern part of the study area, and land was covered with snow and ice in the north. In these instances, a high percentage of land covered by snow and ice led the classifier to map snow and ice covered water as land, leading to a high water omission error. In many cases during spring ice breakup however, ice on water was sufficiently transparent to produce distinct reflectance from ice on land (Figure 10). SWIR thresholding failed altogether on 7 of 72 springtime scenes due to unimodal SWIR frequency histograms with significant overlap between land and water and no minimum value in between. Instances where SWIR thresholding proved better included those where springtime water extended significantly beyond the CanVec water body extents. In these cases, and especially where flooded areas were inundated with water that was spectrally distinct from deeper permanent water, See5 predicted land since these areas represented land according to the reference CanVec water mask.

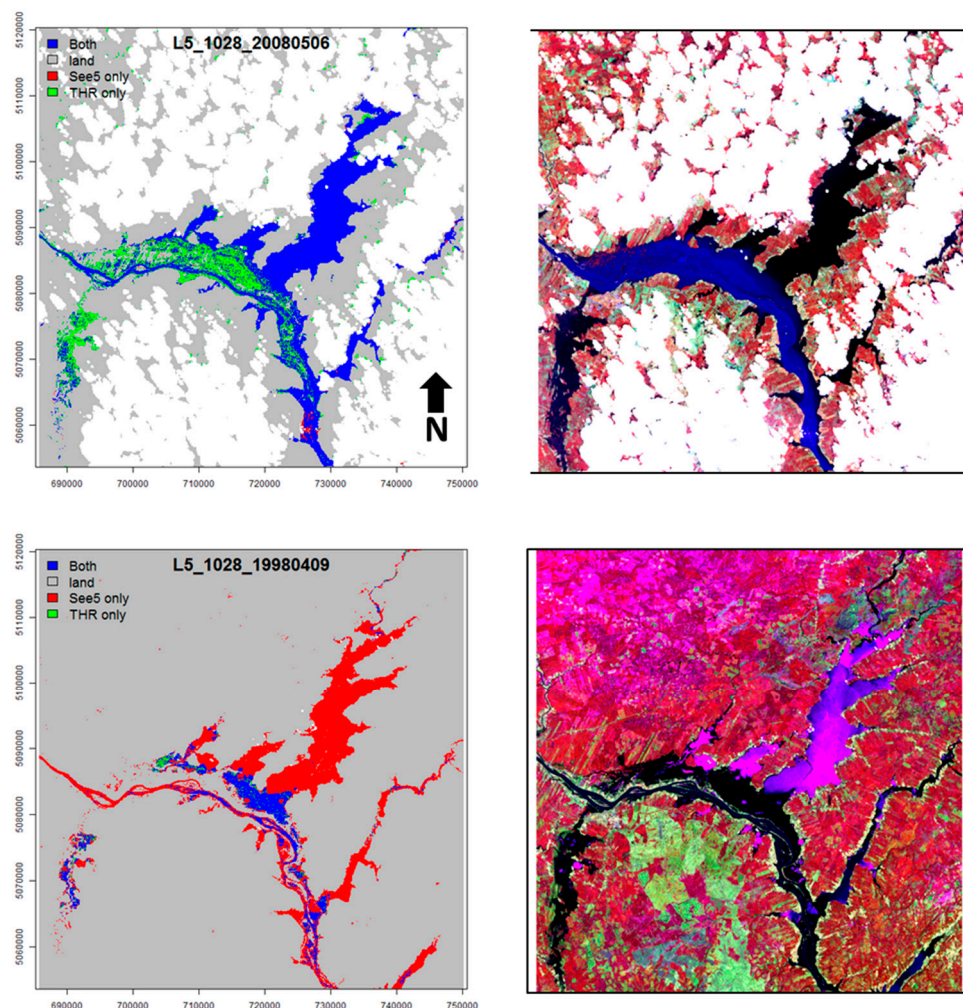


Figure 10. Examples where SWIR thresholding infills water in areas with high sediment load (**top**) with corresponding Landsat bands 4, 5, 3 displayed as R, G, B displayed to the right, and where See5 infills water on lake ice (**bottom**) seen as pink in Landsat.

When exploring the accuracy of mapping summertime water at different reference water fractions from orthophotos, accuracy dropped by almost 40% for pixels with more than 50% water fraction before increasing again to over 97% for pure reference water pixels. The error matrix also revealed a slight bias in reference data, which contained more water pixels than the classification for coincident pixels used to evaluate classification accuracy. It is suspected that some of the mixed pixels in the reference data may have depicted a higher percentage of water than what truly exists due to trees and shadow along shorelines. On southern shorelines, tree shadow was cast northward over the shoreline and onto the water. In many cases where the shoreline was obscured by shadow, it was difficult to visually distinguish between land and water. Further, onshore canopy shadow was sometimes merged with water and shoreline shadow when using raster seeding, since all appeared dark and were spatially adjacent. Manual editing was performed in many areas to try to reduce this bias; however, uncertainty in the precise location of shorelines in these cases may have caused interpreters to miss shoreline editing and maintain a slight bias in reference water. This bias may have been less in Landsat data due to the inclusion of all reflectance bands in a classifier that optimized the separation between land and water.

Orthophotos used to generate water masks may not have been acquired on the same anniversary dates as Landsat, but instead represented the best source of available reference data. According to Figure 6, water levels remain fairly constant between mid-June to early September and therefore within-season water level changes should not have produced large differences between reference and mapped water extents. In case of a significant rainfall event preceding either orthophoto or Landsat acquisition dates, the water depth difference between the two would be higher, producing poorer correspondence and therefore lower accuracy. Because orthophoto acquisitions may have spanned multiple years, pairwise t-tests between water levels for the months of June to August from 1996 to 2002 were conducted and revealed no statistically significant differences between any two years at $p\text{-value} < 0.05$. Regardless, accuracies of 97% or higher for all years between 1996 and 2002 likely represented a realistic assessment of water extent and in case of a large precipitation event immediately prior to orthophoto or Landsat acquisitions, an underestimation of true accuracy.

A limitation of flood extent mapping from Landsat is the 16-day revisit time of individual satellites and the inability of optical sensors to see through cloud, further reducing the amount of usable data. Analysis of seasonal 2004–2016 global cloud fraction from MODIS [32] suggests that the St-John River area is covered by cloud approximately two-thirds of the time in spring, increasing to over 85% of the time in summer. In Australia where inundation frequency has also been mapped with Landsat using all available data from 1987 to 2014, cloud cover is about half. The number of clear observations ranged from 10 in cloudy and mountainous areas to 1200 in the clearest regions of Australia with high overlap between adjacent satellite overpasses [16]. Our analysis had a maximum of 72 observations for spring flooding and 130 for summer based on the number of scenes in each stack. Cloud cover combined with Landsat revisit time produced a mean lag of 12 days between peak flood needed to accurately map the extent of inundation frequency across the entire floodplain, and corresponding annual Landsat acquisition dates. At an average rate of 0.1 m per day change in water depth during the flood season, this amounts to 1.2–1.4 m depth difference, which translates to an underestimation in flooded area of approximately 50 km² according to Figure 7.

1984–2015 global surface water dynamics have recently been mapped from Landsat in [33], including surface water presence, annual and seasonal timing, change, and persistence. This product was made public to enable better global water budget estimation and an examination of regional changes due to climate and land use. A comparison between this product and independently validated surface water dynamic products over select flood-prone regions across Canada is currently underway to determine if the global product is sufficiently accurate at capturing maximum flood extents to guide land use management decisions in Canada.

While the inundation frequency product presented here is representative of a random sample of historical water extents generated from available clear-sky Landsat data, more observations are needed to capture peak flood extent and map inundation frequency over the entire floodplain. Historical RadarSat 1 and 2 data is limited only by what exists in the archive and not by cloud due to its all-weather imaging capabilities. Further, radar has been shown to produce reliable water extents [34]. Reported accuracy for RadarSat 1 water extraction was 83.67% in [35] using amplitude only, increasing to 96.42% after including terrain information and noise removal. Others have reported overall accuracies of 92% for RadarSat 2 Fine Quad beam mode in mapping water bodies larger than 1 hectare [11]. For Landsat TM, [25] achieved a 96.9% overall accuracy delineating water bodies in Landsat 5 using density slicing of the mid-infrared band 5, which was not significantly improved upon using multispectral classification. These accuracies are similar to results obtained here for summer Landsat imagery and are slightly higher than radar, though a direct comparison cannot be made based on statistics alone without considering sampling. As seen in Figure 8, a higher percentage of mixed pixels in the assessment produces a lower overall map accuracy.

Other optical data available to infill time-series include sensors such as MODIS, VIIRS, and Sentinel-3 that acquire data daily at coarse spatial resolution (250 m +). While their spatial resolution precludes precise mapping of flood extents, these data may be useful to estimate flood magnitude and infill temporal gaps in Landsat time-series data. Other medium resolution optical data from Sentinel-2, SPOT, and future Landsat missions may be used as a virtual constellation to map future flood events with increased revisit time [36]. For assessing past inundation frequency, few alternatives exist to Landsat covering the same period from the mid-1980s to present and none are freely available. The SPOT series of satellites began in 1986 and continues to this day, but their specifications have changed from SPOT 1 to SPOT 7, making time-series analysis more challenging.

Future work plans include incorporating other optical data as well as historical radar imagery in the time-series from 1995 to present with RadarSat 1 and 2. Incorporating other data sources in the time-series first requires an assessment of how comparable and consistent other datasets are for water extraction. Detecting water beneath vegetation is another challenge that needs to be addressed, since a significant portion of flooded areas including floodplains are vegetated with plants as large as trees. Initial work suggests change detection between pre-flood and post-flood imagery offers the greatest potential for detecting flooded vegetation.

5. Conclusions

This paper presents an approach for mapping historical flood extents for inundation frequency based on Landsat data that is unique to Canada's climate and flood regimes. A combination of spring and summer inundation frequencies depicts permanent and ephemeral water extents caused by different seasonal processes including snowmelt, ice jamming, and precipitation. Canada's CanVec base hydrography layer was used to derive image-specific training data to classify springtime water that is spectrally variable due to ice and sediment, and land due to land cover and phenology. Water classification alone was determined to have shortcomings that were addressed by combining classified water extents with those derived through simple SWIR image thresholding. Combined surface water extents were determined to be reliable, as the relative magnitude of coincident Landsat water extent and springtime hydrometric water depth were significantly related, while summer extents had a high overall agreement with reference water scaled from digital orthophotos. While validation proved the reliability of the water extraction methodology, limitations on the use of Landsat with its 16-day revisit time and lack of cloud penetration were noted when attempting to map peak flood conditions. Historical water mapping is limited by available data and multi-sensor integration of current and future optical and radar data is needed to increase the revisit time to better enable mapping of peak flood conditions. Plans for future work also include mapping inundation frequency in other flood-prone areas in Canada.

Acknowledgments: Thanks to Aeron Vaillancourt for preparing figures, and to Robert Fraser and Brian Brisco for internal review. Funding was provided by the Floodplain Mapping Project at Natural Resources Canada.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Pachauri, R.K., Meyer, L.A., Eds.; IPCC: Geneva, Switzerland, 2014.
2. Das, S.; Millington, N.; Simonovic, S.P. Distribution choice for the assessment of design rainfall for the city of London (Ontario, Canada) under climate change. *Can. J. Civ. Eng.* **2013**, *40*, 121–129. [[CrossRef](#)]
3. Environment and Climate Change Canada, Canada's Top Ten Weather Stories 2013. Available online: <http://ec.gc.ca/meteo-weather/default.asp?lang=En&n=5BA5EAFc-1&offset=2&toc=hide> (accessed on 14 September 2016).
4. Government of New Brunswick Flood history Database. Available online: <http://www.elgegl.gnb.ca/0001/en/Flood/Search> (accessed on 12 September 2016).
5. Anselmo, V.; Galeati, G.; Palmieri, S.; Rossi, U.; Todini, E. Flood risk assessment using an integrated hydrological and hydraulic modelling approach: A case study. *J. Hydrol.* **1996**, *175*, 533–554. [[CrossRef](#)]
6. Wheater, H.S.; Chandler, R.E.; Onof, C.J.; Isham, V.S.; Bellone, E.; Yang, C.; Lekkas, D.; Lourmas, G.; Segond, M.-L. Spatial-temporal rainfall modelling for flood risk estimation. *Stoch. Environ. Res. Risk Assess.* **2005**, *19*, 403–416. [[CrossRef](#)]
7. Madsen, H.; Jakobsen, F. Cyclone induced storm surge and flood forecasting in the northern Bay of Bengal. *Coast. Eng.* **2004**, *51*, 277–296. [[CrossRef](#)]
8. Bates, P.D.; Dawson, R.J.; Hall, J.W.; Horritt, M.S.; Nicholls, R.J.; Wicks, J.; Hassan, M.A.A.M. Simplified two-dimensional numerical modelling of coastal flooding and example applications. *Coast. Eng.* **2005**, *52*, 793–810. [[CrossRef](#)]
9. Hunter, N.M.; Bates, P.D.; Horritt, M.S.; Wilson, M.D. Simple spatially-distributed models for predicting flood inundation: A review. *Geomorphology* **2007**, *90*, 208–225. [[CrossRef](#)]
10. Sampson, C.C.; Smith, A.M.; Bates, P.D.; Neal, J.C.; Alfieri, L.; Freer, J.E. A high-resolution global flood hazard model. *Water Resour. Res.* **2015**, *51*, 7358–7381. [[CrossRef](#)] [[PubMed](#)]
11. Morena, L.C.; James, K.V.; Beck, J. An introduction to the RADARSAT-2 mission. *Can. J. Remote Sens.* **2004**, *30*, 221–234. [[CrossRef](#)]
12. Bolanos, S.; Stiff, D.; Brisco, B.; Pietroniro, A. Operational surface water detection and monitoring using RADARSAT 2. *Remote Sens.* **2016**, *8*, 285. [[CrossRef](#)]
13. White, L.; Brisco, B.; Dabboor, M.; Schmitt, A.; Pratt, A. A Collection of SAR Methodologies for Monitoring Wetlands. *Remote Sens.* **2015**, *7*, 7615–7645. [[CrossRef](#)]
14. Qi, S.; Brown, D.G.; Tian, Q.; Jiang, L.; Zhao, T.; Bergen, K.M. Inundation Extent and Flood Frequency Mapping Using LANDSAT Imagery and Digital Elevation Models. *GISci. Remote Sens.* **2009**, *46*, 101–127. [[CrossRef](#)]
15. Thomas, R.F.; Kingsford, R.T.; Lu, Y.; Hunter, S.J. Landsat mapping of annual inundation (1979–2006) of the Macquarie Marshes in semi-arid Australia. *Int. J. Remote Sens.* **2011**, *32*, 4545–4569. [[CrossRef](#)]
16. Mueller, N.; Lewis, A.; Roberts, D.; Ring, S.; Melrose, R.; Sixsmith, J.; Lymburner, L.; McIntyre, A.; Tan, P.; Curnow, S.; Ip, A. Water observations from space: Mapping surface water from 25 years of Landsat imagery across Australia. *Remote Sens. Environ.* **2016**, *174*, 341–352. [[CrossRef](#)]
17. Chander, G.; Markham, B.L.; Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **2009**, *113*, 893–903. [[CrossRef](#)]
18. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sens. Environ.* **2015**, *159*, 269–277. [[CrossRef](#)]
19. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1993.
20. Friedl, M.A.; Brodley, C.E. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* **1997**, *61*, 399–409. [[CrossRef](#)]

21. Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* **2003**, *86*, 554–565. [[CrossRef](#)]
22. Olthof, I.; Fraser, R.H. Detecting Landscape Changes in High Latitude Environments Using Landsat Trend Analysis: 2. Classification. *Remote Sens.* **2014**, *6*, 11558–11578. [[CrossRef](#)]
23. Fraser, R.H.; Olthof, I.; Lantz, T.C.; Schmitt, C. UAV photogrammetry for mapping vegetation in the low-Arctic. *Arct. Sci.* **2016**, *2*, 79–102. [[CrossRef](#)]
24. Fry, J.; Xian, G.; Jin, S.; Dewitz, J.; Homer, C.; Yang, L.; Barnes, C.; Herold, N.; Wickham, J. Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote Sens.* **2011**, *77*, 858–863.
25. Machine Learning. Available online: http://www.goodreads.com/work/best_book/206219-machine-learning (accessed on 12 January 2017).
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.
27. Frazier, P.S.; Page, K.J. Water body detection and delineation with Landsat TM data. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 1461–1467.
28. Roach, J.K.; Griffith, B.; Verbyla, D. Comparison of three methods for long-term monitoring of boreal lake area using Landsat TM and ETM+ imagery. *Can. J. Remote Sens.* **2012**, *38*, 427–440.
29. Olthof, I.; Fraser, R.H.; Schmitt, C. Landsat-based mapping of thermokarst lake dynamics on the Tuktoyaktuk Coastal Plain, Northwest Territories, Canada since 1985. *Remote Sens. Environ.* **2015**, *168*, 194–204. [[CrossRef](#)]
30. Centre for Topographic Information Customer Support Group. CDED Canadian Digital Elevation Data, Standards and Specifications. Geomatics Canada, Natural Resources Canada. Available online: http://www.pancroma.com/downloads/NRCAN_CDED_specs.pdf (accessed on 10 November 2016).
31. Kinsman, N.; Gibbs, A.; Nolan, M. Evaluation of vector coastline features extracted from “Structure from Motion”—Derived elevation data. *World Sci.* **2015**. [[CrossRef](#)]
32. Ackerman, S.; Strabala, K.; Menzel, P.; Frey, R.; Moeller, C.; Gumley, L. *Discriminating Clear-Sky from Cloud with MODIS—Algorithm Theoretical Basis Document (MOD35)*; ATBD Reference Number: ATBD-MOD-06; NASA: Greenbelt, MD, USA, 2010.
33. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422. [[CrossRef](#)] [[PubMed](#)]
34. Brisco, B.; Short, N.; van der Sanden, J.; Landry, R.; Raymond, D. A semi-automated tool for surface water mapping with RADARSAT-1. *Can. J. Remote Sens.* **2009**, *35*, 336–344. [[CrossRef](#)]
35. Hong, S.; Jang, H.; Kim, N.; Sohn, H.-G. Water area extraction using RADARSAT SAR imagery combined with Landsat imagery and terrain information. *Sensors* **2015**, *15*, 6652–6667. [[CrossRef](#)] [[PubMed](#)]
36. Wulder, M.A.; Hilker, T.; White, J.C.; Coops, N.C.; Masek, J.G.; Pflugmacher, D.; Crevier, Y. Virtual constellations for global terrestrial monitoring. *Remote Sens. Environ.* **2015**, *170*, 62–76. [[CrossRef](#)]

