DEVELOPMENT OF AN INTEGRATED HIGH RESOLUTION FLOOD PRODUCT WITH MULTI-SOURCE DATA

by

Sanmei Li
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Earth System and Geoinformation Sciences

Committee:
_________________________________ Dr. Donglian Sun Dissertation Director
_________________________________ Dr. Anthony Stefanidis, Committee Member
Dr. Ruixing Yang, Committee Member
_________________________________ Dr. Long Chiu, Committee Member
_________________________________ Dr. Sheryl Luzzadder Beach, Department Chairperson
_________________________________ Dr. Richard Diecchio, Interim Associate Dean for Student and Academic Affairs, College of Science
_________________________________ Dr. Peggy Agouris, Interim Dean, College of Science

Date: ____________________________ Fall Semester 2013
George Mason University
Fairfax, VA
College of Science

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Sanmei Li
Master of Atmospheric Physics and Atmospheric Environment
Peking University, China, 2008
Bachelor of Geo-Information Science
Nanjing University, China, 2001

Director: Donglian Sun, Professor
Department of Geography and Geoinformation Science

Fall Semester 2013
George Mason University
Fairfax, VA
DEDICATION

This is dedicated to my husband Xu Liu, and my daughter Shuhan Liu.
I would like to thank my advisor and many supporters who have made this happen. My advisor, Prof. Donglian Sun not only provides me a lot of constructive instructions, but also always gives me full trust and confidence to finish this research. I would like to thank Dr. Mitchell D. Goldberg. Without his support, this research would have never made a start. My committee members Prof. Anthony Stefanidis, Prof. Ruixing Yang, and Prof. Long Chiu also give me a lot of help on many details in this research. I would also like to thank many staff and faculty in Department of Geography and GeoInformation Science. Prof. Sheryl Luzzadder Beach, Ms. Teri Fede and Ms. Deborah J Hutton are always so nice and patient to answer all the academic questions and provide us a peaceful and friendly atmosphere in which to work and study. Finally, the greatest gratitude is given to my husband Xu Liu and my sweetheart Shuhan Liu. Without their support and understanding, I would have never got the chance to start and complete my PhD study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter/Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE OF CONTENTS</td>
<td></td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>V</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>X</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>XVI</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Flood detection with optical satellite data</td>
<td>3</td>
</tr>
<tr>
<td>1.1.2 Flood detection with microwave-based satellite data</td>
<td>4</td>
</tr>
<tr>
<td>1.1.3 Flood detection with hydrology models</td>
<td>6</td>
</tr>
<tr>
<td>1.1.4 Flood detection with multi-source data</td>
<td>7</td>
</tr>
<tr>
<td>1.2 Literature review</td>
<td>8</td>
</tr>
<tr>
<td>1.2.1 Algorithms for flood detection with optical satellite data</td>
<td>8</td>
</tr>
<tr>
<td>1.2.2 Algorithms with microwave data</td>
<td>12</td>
</tr>
<tr>
<td>1.2.3 Algorithms with multi-source data</td>
<td>15</td>
</tr>
<tr>
<td>1.3 Research objective and content of this dissertation</td>
<td>16</td>
</tr>
<tr>
<td>1.3.1 Research objective</td>
<td>16</td>
</tr>
<tr>
<td>1.3.2 Study area</td>
<td>18</td>
</tr>
<tr>
<td>1.3.3 Data preparation</td>
<td>19</td>
</tr>
<tr>
<td>1.3.4 Dissertation content</td>
<td>20</td>
</tr>
<tr>
<td>CHAPTER 2 WATER DETECTION WITH OPTICAL SATELLITE DATA</td>
<td>21</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>22</td>
</tr>
<tr>
<td>2.2 Methodology</td>
<td>23</td>
</tr>
<tr>
<td>2.2.1 Basic spectral properties of water surface</td>
<td>23</td>
</tr>
<tr>
<td>2.2.2 Decision-tree approach</td>
<td>24</td>
</tr>
<tr>
<td>2.2.2.1 Simple description of decision tree algorithm</td>
<td>24</td>
</tr>
<tr>
<td>2.2.2.2 Decision tree samples</td>
<td>26</td>
</tr>
<tr>
<td>2.2.3 DT application in water detection</td>
<td>34</td>
</tr>
<tr>
<td>2.2.3.1 Tree comparison</td>
<td>34</td>
</tr>
<tr>
<td>2.2.3.2 Decision tree for automatic water detection</td>
<td>36</td>
</tr>
<tr>
<td>2.3 RESULTS</td>
<td>40</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.3.1 BACKGROUND</td>
<td>184</td>
</tr>
<tr>
<td>7.3.2 FLOOD DETECTION IN NEW YORK CITY WITH VIIRS DATA</td>
<td>185</td>
</tr>
<tr>
<td>7.4 DISCUSSION AND SUMMARY</td>
<td>189</td>
</tr>
<tr>
<td>CHAPTER 8 SUMMARY AND FUTURE WORK</td>
<td>191</td>
</tr>
<tr>
<td>8.1 DISSERTATION SUMMARY</td>
<td>191</td>
</tr>
<tr>
<td>8.2 MAIN INNOVATION OF THIS DISSERTATION</td>
<td>192</td>
</tr>
<tr>
<td>8.2 FUTURE WORK</td>
<td>194</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>196</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1 List of deadliest floods and death toll</td>
<td>2</td>
</tr>
<tr>
<td>Table 2 Accuracy by class of the sample J48graft decision tree</td>
<td>33</td>
</tr>
<tr>
<td>Table 3 TP Rate of Classified Instances in different types of Decision Tree Algorithms (Bare land, water, vegetation, cloud shade)</td>
<td>35</td>
</tr>
<tr>
<td>Table 4 TP Rate of Water Classified Instances with DT Algorithms</td>
<td>35</td>
</tr>
<tr>
<td>Table 5 Analysis on water fraction of TM and simulated MODIS using different methods</td>
<td>89</td>
</tr>
<tr>
<td>Table 6 Comparison of lake areas calculated from MODIS by using DNNS and histogram methods with the SWIR channel</td>
<td>94</td>
</tr>
<tr>
<td>Table 7 Quantitative statistics on validation cases</td>
<td>150</td>
</tr>
<tr>
<td>Table 8 Analysis on the unmatched and matched water pixels with land cover</td>
<td>151</td>
</tr>
<tr>
<td>Table 9 Water surface elevation comparison between MODIS and ground observations (Unit: meter)</td>
<td>156</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 1</td>
<td>Pie-chart of natural disaster statistic between 2000 and 2010</td>
</tr>
<tr>
<td>Figure 2</td>
<td>A sketch image of study area</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Plot of reflectance for different land types from VIS to SWIR band range</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Part of the sample J48Graft Decision Tree for water detection</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Series of scatter points images of samples from NOAA-17/AVHRR</td>
</tr>
<tr>
<td>Figure 6</td>
<td>MODIS false-color composite image (Left) and water map (Right) in Mississippi river on May 04, 2011</td>
</tr>
<tr>
<td>Figure 7</td>
<td>MODIS false-color composite image (Left) and water map (Right) in Mississippi river on March 01, 2011</td>
</tr>
<tr>
<td>Figure 8</td>
<td>MODIS false-color composite image (Left) and water map (Right) in Mississippi river on April 10, 2011</td>
</tr>
<tr>
<td>Figure 9</td>
<td>MODIS false-color composite image (Left) and water map (Right) in Mississippi river on May 06, 2011</td>
</tr>
<tr>
<td>Figure 10</td>
<td>MODIS false-color composite image (Left) and water map (Right) in Mississippi river on May 20, 2011</td>
</tr>
<tr>
<td>Figure 11</td>
<td>MODIS false-color composite image (Left) and water map (Right) in Mississippi river on May 29, 2011</td>
</tr>
<tr>
<td>Figure 12</td>
<td>VIIRS false-color composite image (Top) and the corresponding water map (Bottom) in Alaska on May 26, 2013</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Graph of parallax and shadow</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Graph of parallax and shadow geometry over ideal plane</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Graph of parallax and shadow geometry over spherical surface</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Cloud and cloud shadow inter-determination with multiple points</td>
</tr>
<tr>
<td>Figure 17</td>
<td>MSG/SEVIRI false color image series (Left) and the corresponding water maps (Right) in the morning from 07:30 to 10:00 (UTC) on the 68th Julian day in 2010. Top: 07:30, Middle: 08:30, Bottom: 10:00</td>
</tr>
<tr>
<td>Figure 18</td>
<td>MSG/SEVIRI false color image series (Left) and the corresponding water maps (Right) at noon from 11:00 to 13:00 (UTC) on the 68th Julian day in 2010. Top: 11:00, Middle: 12:00, Bottom: 13:00</td>
</tr>
<tr>
<td>Figure 19</td>
<td>MSG/SEVIRI false color image series (Left) and the corresponding water maps (Right) in the afternoon from 14:00 to 15:00 (UTC) on the 68th Julian day in 2010. Top: 14:00, Bottom: 15:00</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Sample SEVIRI false color images and water maps with complicated situations on the 68th Julian day in 2010: (a), (b), and (c), (d): Over water and</td>
</tr>
</tbody>
</table>
shadow mixing area, 2010068 09:00 (UTC); (e) and (f): Over the disk edge area, 2010068 11:00 (UTC) ................................................................. 65

Figure 21 Sample SEVIRI false color images and water maps with two mistaken situations on the 68th Julian day in 2010: (a) and (b): snow/ice or land/sea edges, 2010068 09:30 (UTC); (c) and (d): undetected shadow pixels, 2010068 15:00 (UTC) ............................................................................. 67

Figure 22 The curves of reflectance for different land types (left) and ratio of spectral reflectance to SWIR (MODIS CH6) channel (right) (Spectral data is provided by Yuxiang Zhang in National Satellite meteorological Center, China Meteorological Administration)......................... 73

Figure 23 Algorithm flow chart of water fraction calculation with MODIS data .... 80

Figure 24 (a) MODIS Swath False color composite image (2008035 07:55GMT); (b) Image of classification result of (a); (c) A subset image of (a); (d) Water fraction map of (c) .................................................................................. 82

Figure 25 (1) MODIS false color composite image (17:10(UT) on June 5th, 2001), (2) Simulated MODIS false color composite image (3) Subsets images and Histogram plots of each band for MODIS and simulated MODIS data: (a) Histogram of MODIS band 6; (b) Histogram of simulated MODIS band 6; (c) Histogram of MODIS band 2; (d) Histogram of simulated MODIS band 2; (e) Histogram of MODIS band 1; (f) Histogram of simulated MODIS band 1 ..... 86

Figure 26 (a) Water fraction map from TM; (b) Water fraction map from the simulated MODIS; (c) A subset false color image from MODIS; (d) A subset water fraction map from TM; (e) A subset water fraction map from MODIS with DNNS method; (f) A subset water fraction map from MODIS with histogram method........................................................................ 88

Figure 27 (a) Scatter plot of water fractions between TM and MODIS using DNNS method; (b) Scatter plot of water fractions between TM and MODIS using histogram method................................................................. 90

Figure 28 Plots of seasonal average $\Delta T_{37GHZ}$ (top), $PR_{37GHZ}$ (middle) and $TR_{37_89GHZ}$ (bottom) over land cover types (Land cover data is from 1-km USGS Land Use/Land Cover System by calculating fractions of each land cover type at 25-km resolution, and samples are collected with 100% fraction of each type at 25-km.) ........................................................................................................... 124

Figure 29 Correlation test between water fraction and $\Delta T_{37GHZ}$, $PR_{37GHZ}$ and $TR_{37_89GHZ}$ over non-desert area (Left) and desert area (Right)....................... 125

Figure 30 Distribution of Correlation Coefficients between precipitation and $\Delta T_{37GHZ}$ (Left), $PR_{37GHZ}$ (Middle) and $TR_{37_89GHZ}$ (Right) ......................................................... 126

Figure 31 Scatter analysis of NDVI (Left) and tree cover (Right) on $\Delta T_{37GHZ}$ (Top), $PR_{37GHZ}$ (Middle) and $TR_{37_89GHZ}$ (Bottom)......................................................... 126

Figure 32 Correlation test between cloud fraction and $\Delta T_{37GHZ}$ (Top), $PR_{37GHZ}$ (Middle) and $TR_{37_89GHZ}$ (Bottom) over water surface ......................................... 127

Figure 33 Part of a Regression Tree for water fraction retrieval over vegetation land during night time with AMSR-E data................................................................. 128
Figure 49 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 28, 2013 ................................................................. 166
Figure 50 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 29, 2013 ................................................................. 167
Figure 51 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 30, 2013 ................................................................. 168
Figure 52 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 31, 2013 ................................................................. 169
Figure 53 VIIRS 375-m false color image and the corresponding flood detection map near Galena on June 01, 2013 ................................................................. 170
Figure 54 VIIRS 375-m false color image and the corresponding flood detection map along South Platte River on Sep. 14, 2013 .......................... 174
Figure 55 VIIRS 375-m false color image and the corresponding flood detection map along South Platte River on Sep. 17, 2013 .................. 175
Figure 56 A whole look at VIIRS 30-m water map on Sep. 17, 2013 along South Platte River overlapped on Google Earth (light purple color is flooding water) .................................................................................................................................................................................. 176
Figure 57 VIIRS 30-m water map on Sep. 17, 2013 near Evans overlapped on Google Earth (light purple color is flooding water) .................. 176
Figure 58 VIIRS 30-m water map on Sep. 17, 2013 near Garden City overlapped on Google Earth (light purple color is flooding water) .......... 177
Figure 59 VIIRS 30-m water map on Sep. 17, 2013 near Orchard overlapped on Google Earth (light purple color is flooding water) .......... 178
Figure 60 VIIRS 30-m water map on Sep. 17, 2013 near Weldona overlapped on Google Earth (light purple color is flooding water) .... 178
Figure 61 VIIRS 30-m water map on Sep. 17, 2013 near Snyder overlapped on Google Earth (light purple color is flooding water) .... 179
Figure 62 VIIRS 30-m water map near Evans (Left) and a photo taken on Sep. 16, 2013 covering the same area ...................................................... 180
Figure 63 VIIRS 30-m water map in 37th St near Evans (Left) and a photo taken on Sep. 16, 2013 covering the same area ...................................... 180
Figure 64 VIIRS 30-m water map in communities Evans (Left) and photos taken on Sep. 13, 2013 covering the same area ...................................... 181
Figure 65 VIIRS 30-m water map in the 37th St of Evans (Left) and a photo map taken on Sep. 16, 2013 covering the same area .................. 182
Figure 66 VIIRS 30-m water map near County Road 61 (Left) and a photo map taken on Sep. 16, 2013 covering the same area .................. 183
Figure 67 VIIRS 30-m water map near US 34 highway (Left) and a photo map taken on Sep. 16, 2013 covering the same area .................. 183
Figure 68 VIIRS 30-m water map near US 52 highway (Left) and a photo map taken on Sep. 16, 2013 covering the same area .................. 184
Figure 69 VIIRS 375-m false color image and the corresponding water fraction map around Near York City on Nov. 04, 2012 .................. 187
Figure 70 VIIRS 30-m water map around New York City on Nov. 04, 2012

Figure 71 Simulated inundation map from hydrological models (left), VIIRS 30-m water map on Nov. 04, 2012 (middle) and Google Flood Evacuation Map due to hurricane Sandy (right) around New York City

Figure 72 ATMS integrated 30-m water map around New York area on Nov. 01, 2012
Flood is the most frequent and also one of the costliest natural disasters in the world. During the 20th century, flood was the number-one natural disaster in the United States in terms of number of lives lost and property damage. In the long term, flooding kill more people than any other types of severe weather events in the United States. Therefore, dynamic flood detection and analysis is an important topic in research and applications.

The development of remote sensing technology has provided new data sources for flood detection and has made it possible to derive continuous and comprehensive information about floods. The high temporal resolution and large coverage of coarse- to moderate-resolution satellite imagery are very advantageous for flood monitoring; however, their coarse spatial resolution (0.250 km for EOS/MODIS and 0.375 km for Suomi-NPP/VIIRS) precludes useful delineation of flooded areas in small regions, like New York area during Hurricane Sandy time. Compared to NOAA/AVHRR and
EOS/MODIS, Landsat/TM data have much higher spatial resolution. Nevertheless, the long revisit interval and narrow swath coverage limit its applications in flash flood detection. In addition, all the optical sensors are restricted to clear conditions, while peak floods are usually associated with clouds, so cloud cover is a big issue for obtaining comprehensive real-time flood information from these satellite data. Compared to optical satellite data, microwave sensors (both passive and active), such as the AMSR-E and ATMS, can penetrate most of clouds, meanwhile with low cost, large swath coverage and good data availability, but they are at very coarse spatial resolution, usually larger than 10 km; while some other microwave-based data, such as the Radarsat, have high spatial resolution, but are usually with high cost, long revisit interval and narrow swath coverage.

Considering the characteristics of these satellites, if there is a way to fuse these data together to make up each other’s deficiencies, then it is very possible to derive a good flood map which benefits most of the advantages of these satellite data. Such flood product definitely helps improve the applications of satellite data in flood analysis significantly. Therefore, this dissertation focuses on the development of such an enhanced flood product with multi-source data including satellite data from EOS/MODIS or Suomi NPP/VIIRS, and AMSR-E or ATMS, high resolution digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM), precipitation data and river gauges’ observations. During the development, decision-tree approach is used as the main method to extract water distribution from MODIS or VIIRS data. Since cloud shadows are often spectrally similar to water and thus can be easily misclassified as water,
a new geometric method for automatic cloud shadow removal has been developed to remove cloud shadow from water maps. Moreover, a dynamic nearest neighbor searching method (DNNS) is developed to retrieve water fractions from MODIS and VIIRS data using shortwave infrared band based on the water detection results. The retrieval of water fraction from microwave sensors, such as the AMSR-E or ATMS, is mainly based on regression-tree approach using a series of features generated from polarizations around 37GHz and 89GHz considering land cover mixture, precipitation and cloud fraction. By combining with the 30-m DEM data from the SRTM, an integration method is developed to upgrade water fraction products derived from the course- to moderate-resolution satellite imagery, such as MODIS/VIIRS and AMSR-E/ATMS, to 30-m spatial resolution flood maps. As derived from both optical and microwave sensors, and high resolution DEM data, the final enhanced flood maps can fill the gaps due to clouds, and meanwhile have high temporal and spatial resolutions, and large swath coverage.
CHAPTER 1 Introduction

1.1 Background

A flood is an overflow of water that submerges or "drowns" land. Floods are some of the most serious and frequent natural disasters. Floods make up more than 50% of all natural disasters recorded from 2000 to 2010 (Fig. 1). More than 14 floods that have killed at least 15,000 people have occurred in each decade since 1950 (Table 1). Floods are also among the costliest natural disasters at a global scale. For example, at least 383 people were killed across seven states during the flood in the Mississippi River Basin in 2011. Thousands of homes were ordered evacuated, and the economic loss was more than $7 billion.
Figure 1 Pie-chart of natural disaster statistic between 2000 and 2010

Table 1 List of deadliest floods and death toll

<table>
<thead>
<tr>
<th>From (year)</th>
<th>To (year)</th>
<th>Number of deadliest floods</th>
<th>Death toll</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>2011</td>
<td>17</td>
<td>10324</td>
</tr>
<tr>
<td>2000</td>
<td>2009</td>
<td>31</td>
<td>23586</td>
</tr>
<tr>
<td>1990</td>
<td>1999</td>
<td>28</td>
<td>46238</td>
</tr>
<tr>
<td>1980</td>
<td>1989</td>
<td>17</td>
<td>20468</td>
</tr>
<tr>
<td>1970</td>
<td>1979</td>
<td>17</td>
<td>371295</td>
</tr>
<tr>
<td>1960</td>
<td>1969</td>
<td>16</td>
<td>15601</td>
</tr>
<tr>
<td>1950</td>
<td>1959</td>
<td>14</td>
<td>56384</td>
</tr>
</tbody>
</table>
Because floods are such significant disasters, determining how best to predict, detect and evaluate floods has been the focus of numerous studies since the early part of the last century. Determining how to obtain dynamic flood information at a regional scale is especially crucial for decision-makers to minimize losses. With the progress of remote sensing technology, satellite data, especially from meteorological and environmental satellites, have been shown to have unique advantages in dynamic flood detection because they have large swath coverage and high revisit frequencies. Moreover, they are not significantly affected by ground conditions. Before flooding, satellite data can be used to estimate precipitation and track storms. After flooding, they can be utilized to monitor floods and standing water and to evaluate losses. Thus, flood detection with satellite data has become an important area of research and a topic for remote sensing applications. In many countries, the monitoring of flooding and standing water bodies is now a daily occurrence and plays a significant role in early flood warning and loss assessment.

1.1.1 Flood detection with optical satellite data

Optical satellite data are widely applied in flood analysis because the data are straightforward to use. Since the early 1970s, NOAA/VHRR and NOAA/AVHRR data have played a significant role in flood/standing water detection and have been widely used for the analysis of large floods. Since the first EOS (Earth Observing Satellite) TERRA satellite was launched in 1999, one of the on-board sensors, the Moderate Resolution Imaging Spectroradiometer (MODIS), has gradually replaced the AVHRR in
operational flood detection due to its higher spatial resolution (250 to 500 m for water classification compared to AVHRR’s 1 km resolution). The VIIRS (Visible Infrared Imager Radiometer Suite) sensor, which is on the Suomi-NPP satellite that was launched in 2011, is even better than MODIS for environmental and natural disaster analysis. VIIRS provides five imagery bands at visible to thermal infrared channels at a spatial resolution of 375 m with much larger swath coverage than MODIS. Although MODIS and VIIRS both have much higher spatial resolution than AVHRR, it is still difficult for the data to meet the needs of flood detection. Landsat Thematic Mapper (TM) data has a 30 m spatial resolution, but the temporal resolution and narrow swath coverage limit the use of the data for dynamic flood monitoring because floods are generally short-term events.

Spatial resolution is not the only problem with optical satellite data. Another issue is that optical radiation is hindered by cloud cover. Floods are generally caused by intense rainfall that is accompanied by large amounts of clouds, and the clouds prevent the sensors from obtaining valid surface information and thus limit the application of optical satellite data in real-time flood detection.

1.1.2 Flood detection with microwave-based satellite data

To overcome the problem caused by clouds in optical satellite data, microwave-based satellite data are utilized in severe flood detection because microwave radiation can penetrate most clouds and obtain surface information during non-rainy weather conditions during both the day and the night. The SMMR sensor on the Nimbus-7
satellite, which was launched in 1978 and functioned for more than nine years, was the first microwave sensor to be successfully used in flood detection. SMMR is a panoramic instrument with five frequencies and two polarizations. The availability of the 37 GHz frequency allows the data to be used to detect water. Similar to SMMR, the SSM/I sensor on the DMSP (Defense Meteorological Satellite Program) and the TMI sensor on the TRMM (Tropical Rainfall Measuring Mission) have also been utilized in severe flood analysis since the early 1990s. AQUA, the second EOS satellite that was launched in 2002, not only has a MODIS sensor but also carries an AMSR-E microwave sensor. AMSR-E has more frequencies and a higher spatial resolution than SMMR, SSM/I and TMI. The new microwave sensor ATMS (Advanced Technology Microwave Sounder), which is carried by Suomi-NPP, also uses multiple frequencies around 50 GHz and 85 GHz at a spatial resolution of approximately 15 km. These microwave sensors provide abundant data for water detection and flood analysis.

Although microwave radiation is less affected by clouds, the spatial resolution is a significant issue for obtaining detailed flood information. The 37 GHz frequency of AMSR-E has a 12.5 km spatial resolution. Within any 12.5 km footprint, there can be a mixture of multiple land types, such as bare land, grass, trees, water and urban areas. The mixtures make it difficult to obtain very accurate results.

In recent years, LiDAR and radar data have also been used to detect water bodies. The Shuttle Radar Topography Mission was launched in 2000 to provide 90 m digital elevation models at a global scale and 30 m DEMs in the USA. Along with the elevation
data, the global static water mask SWBD (SRTM Water Body Dataset, which has 30-m resolution in the USA and 90-m resolution for the rest of the world) was generated as a by-product of the SRTM topography data and provides an excellent reference map for comparison and flood determination. Nevertheless, the availability of LiDAR and radar data is poor, and the data are expensive compared to data from other meteorological or environmental satellites.

1.1.3 Flood detection with hydrology models

Since the early 2000s, scientists and hydrologists have simulated the extent of flooding using the fluid characteristics of water. Inundation models have gradually developed from simple two-dimensional raster forms to complex three-dimensional forms. These models use either run-off data from ground observations or water level or water depth data from river gauges along with very high resolution DEM data based on hydrodynamic mechanisms (Hesselink et al., 2003; Zhenget al., 2008). Some of these models can obtain good results and can provide excellent approaches for flood analysis. Nevertheless, water level, water depth and run-off data are only observed at points. Because of the lack of regional water distribution information in grids, the simulations are highly sensitive to the inundation models and the DEMs. Simulations over small regions can take extensive amounts of time and consume many computing resources.
1.1.4 Flood detection with multi-source data

Although the data from different satellites and models individually contribute to flood detection and assessment at different scales, the use of combined datasets can significantly improve flood analysis. Therefore, in recent years, research has concentrated on flood detection using multi-source data. Optical and microwave-based satellite data are used together to either improve the product accuracy or retrieve more flood parameters. DEM data and river gauge observations are utilized together with satellite data to determine river discharge or run-off.

Previous studies on the application of multi-source data for flood analysis have been promising. However, the current state of research is not adequate and remains immature for obtaining satisfactory results for end users. In fact, the application of multi-source data can not only provide higher accuracy for water identification over complex atmospheric and ground conditions but can also obtain more details about floods. This is because water fractions can be retrieved at acceptable accuracies from both optical and microwave-based satellites data even though they have coarse-to-moderate resolution. The regional water extent or area can be determined from the water fractions. Based on water’s fluid characteristics and high spatial resolution DEM data, it is possible to determine a specific surface water distribution at the same resolution as the DEM using the estimated water extent and thus generate high resolution flood maps.

Based on this concept, this research focuses on the application of multi-source data to flood detection. The main objective is to derive a 30 m resolution flood product from
coarse-to-moderate resolution satellite data including MODIS and/or VIIRS, AMSR-E and/or ATMS, and SRTM 30-m DEM data. Compared to the original flood maps, which have spatial resolutions between 375 m and 15 km, the spatial resolution of the new flood maps are enhanced to approximately 30 m and thus provide much more detail. This improves the application of these coarse-to-moderate satellite data in hydrology and natural disaster analysis.

1.2 Literature review

1.2.1 Algorithms for flood detection with optical satellite data

Studies of flood detection using optical satellite data can be traced back to the 1970s. In the early 1970s, Wiesnet et al. (1974) and Williamson (1974) used NOAA-2 VHRR data to detect floods and laid the foundation for the field of flood detection using optical satellite data. In the 1980s, Cao et al. (1987), Xiao and Chen (1987), Barton and Bathos (1989) and Ali (1989) performed river flood analysis and flood extent identification using NOAA/AVHRR data and FengYun-1 data. Cao et al. used three-channel composite images to visually interpret floods. Xiao & Chen identified flood water by the difference between the NIR and VIS channels from NOAA/AVHRR data and determined the flooded areas. In the 1990s, Rasid and Paramanik (1990), Tao (1993) and Sheng et al. (1994) used visual interpretation of NOAA/AVHRR imagery to monitor floods by considering the local environmental settings. Verdin (1996) evaluated the feasibility of monitoring small ephemeral water bodies with NOAA/AVHRR thermal
infrared data and used TM data to validate the results. During this period, threshold segmentation served as the main approach for water detection with satellite imagery.

After EOS was launched in 1999, MODIS gradually replaced AVHRR in operational flood detection due to its higher spatial resolution (Gumley and King, 1995; R. Brakenridge and E. Anderson, 2006; Galina Wind et al., 2010). In 2009, MOD44W from MODIS and SWBD (SRTM Water Body Dataset) was released to provide a global 250 m water mask (Carroll, M et al., 2009). The product is considered to be an improvement over the existing MODIS land cover-based global land-water mask (Salomon et al., 2004) and provides a good reference map for flood determination. Although threshold segmentation is still used in water detection, other approaches, such as decision-tree techniques, have been introduced to identify water bodies more effectively with satellite imagery (Sun et al., 2011). In addition to NOAA/AVHRR and EOS/MODIS, Landsat TM data are also widely utilized in water body detection and delineation to obtain accurate water maps and evaluate post-flood damage due to its much higher spatial resolution (Frazier P. S. and Page K. J., 2000; Gianinetto M et al., 2006). Nevertheless, the revisit interval of Landsat is approximately 26 days, and the swath coverage is approximately 60 km. Because floods are often short-term disasters that occur over large areas, these satellite data are not as suitable as data from meteorological and environmental satellites, such as EOS/MODIS and NOAA/AVHRR, for dynamic flood detection.
However, data from meteorological and environmental satellites have spatial resolutions between 250 m and 1 km. Most pixels are mixed and contain a combination of vegetation, bare land, and water. Therefore, obtaining an accurate water fraction from a mixed pixel is a challenging but necessary task for monitoring flooding and standing water. In this situation, scientists have focused on developing algorithms for water fraction retrieval instead of simple water detection. Sheng et al. (2001) developed a quantitative method to calculate water fractions using NOAA/AVHRR data with a histogram method that is based on a linear mixing model. In their studies, pixels are assumed to be composed of land and water, and the visible and near-infrared AVHRR channels are used to calculate the water fraction using the histogram statistics. Their research marked an important milestone toward the quantitative calculation of water fractions. Recently, further research on hydrology and water resources has shown that water reflectance varies with water quality. Kallio (1999), Reinart and Herlevi (2003), Balkanov et al. (2003) and Ma et al. (2009) showed that the reflectance of water bodies in the VIS and NIR channels varies with the content of suspended matter. The reflectance of land pixels also varies with different vegetation fractions as well as vegetation and soil types. It is difficult to obtain accurate results using a simple histogram method based on the reflectance of the VIS and NIR channels without considering the mixture of water and land types, especially in complicated underlying conditions.

In addition to water fraction retrieval, another challenge and necessary task for accurate water detection in coarse-to-moderate optical satellite imagery is cloud shadow removal. Due to the high spectral and radiometric similarity between cloud shadows and
mixed water pixels, methods such as threshold segmentation or decision-tree approaches based on spectral features for water identification cannot effectively discriminate between cloud shadows and mixed water pixels. Therefore, numerous studies have focused on differentiating between these features. Wang and Ono (1999) used an image-fusion method to automatically remove cloud and cloud-shadow contamination in Landsat images. Giles (2001) developed a common confusion algorithm for delineating terrain shadows in a digital elevation model from Landsat images and was able to identify 86% of shadows. Melesse and Jordan (2002) compared the fuzzy technique and an augmented form of the iterative self-organizing data analysis (ISODATA) technique to discriminate low-altitude clouds from their shadows using Landsat Thematic Mapper (TM) images in the Econlockhatchee River Basin (Econ) of Central Florida. Rosin et al. (2002) used an image difference method to detect cloud shadows. Hutchison et al. (2008) identified cloud shadows using a logic-searching method based on cloud-classification results; this method is also used by the National Polar-orbiting Operational Environmental Satellite System (NPOESS) program. Luo et al. (2008) used a multi-channel threshold method based on a simple geometric relationship to detect cloud shadows in MODIS data. Meng et al. (2009) tested the performance of the closest spectral fit methodology to remove clouds and cloud shadows in satellite images. Wavelet analysis has also been applied to cloud shadow detection and contributed important information to this field (Chen F. et al., 2006). However, most of these methods are based on the spectral characteristics of shadows (i.e., their low reflectance). Because these properties are so similar to those of water bodies, especially mixed water bodies, cloud shadows are easily misclassified as
water pixels. This common misclassification is the primary obstacle of automatic water
detection with satellite imagery. Thus, to discriminate shadows from water, cloud-
shadow detection algorithms must include land-cover data to remove water bodies. Sheng
et al. (1998) differentiated water and floods from thin clouds and cloud shadows based on
the fact that the Normalized Difference Vegetation Index (NDVI) of water is negative in
shadowed and non-shaded areas. However, this method cannot identify mixed water
pixels because the NDVI of mixed water pixels might also be positive. The methods of
Luo et al. (2008) and Khlopenkov (2007) are primarily based on the spectral properties of
cloud shadows and water bodies. Additionally, these methods strongly depend on cloud
mask and classification products; thus, they fail to predict cloud shadows when cloud
mask and cloud classification products are unavailable or when clouds are not detected in
the cloud mask.

1.2.2 Algorithms with microwave data

Although optical satellites have good data availability, large swath coverage and
excellent balance between spatial resolution and revisit period, they are easily affected by
cloud cover. As floods happen simultaneously with a large amount of cloud cover, it is
always difficult for optical satellites to dynamically track floods and obtain
comprehensive flooding information. To fill the gap of floods under cloud cover,
scientists turn to use microwave radiation which can penetrate most clouds for flood
detection.
With SMMR 37GHz polarization frequencies, Sippel et al. (1994) developed an algorithm to derive the inundated area in the Amazon floodplain based on a linear combination model by dividing land types into dry land, open water, and inundated land with a stationary polarization difference for each type. In 1998, the 1994 algorithm was modified by assigning each land type polarization difference values aggregated from long-term SMMR datasets and river stage data (Sippel et al, 1998). Sippel’s algorithms showed promising results in the Amazon River floodplain and laid the foundation for water fraction retrieval with microwave data. As more microwave sensors such as SSM/I, TMI, and AMSR-E become available, research on flood detection using microwave data with the assistance of optical products is being undertaken on a wider scale (Frappart et al., 2005; Brakenridge et al., 2007; Gouweleeuw et al., 2010). In addition to brightness temperature difference between horizontal and vertical polarizations at 37 GHz, more variables such as polarization ratio (PR) and soil wetness index (SWI) between 37GHz and 89GHz were proposed for soil moisture and surface water detection (Njoku et al., 2003). By considering a time lag observed between flooded areas and measured discharge, Temimi et al. (2005) investigated the potential of a rating curve model for flood forecasting based on the use of a water surface fraction derived from SSM/I passive microwave images and discharge observations. Brakenridge et al. (2007) suggested a spatial ratio method which took the ratio between a river center pixel and a non-river pixel at 37GHz for river discharge and ice status measurement with AMSR- E data. Compared to Sippel’s algorithms, the spatial ratio method was no longer based on single-pixel retrieval. It focused on the ratio between a wet pixel and a nearby dry pixel to
reduce noises by assuming the relative contribution from the differences of emissivity or atmosphere moisture is constant over a large area. Based on Brakenridge’s spatial ratio method, Groeve et al. (2009) further developed the methodology using daily AMSR-E and TRMM passive microwave observations to detect flood magnitudes in real time for global major floods. Gouweleeuw et al. (2010) used AMSR-E data to monitor the 2010 flood in Queensland, Australia, with a 37GHz polarization ratio along with a rainfall mask and desert mask derived from 22GHz and 89GHz. In that research, SRTM DEM, river gauge and MODIS data were also utilized to account for the impacts from topography and vegetation. Ticehurst et al. (2009) merged ASAR GM, AMSR-E and MODIS data to monitor floods and calculate water volumes. Watts et al. (2012) used AMSR-E both horizontal and vertical polarized brightness temperatures at 19GHz and 22 GHz to estimate surface water fractions across the Arctic – Boreal region and analyzed surface inundation patterns from 2003 to 2010 with the retrieved results there. With a global dataset from SSM/I, ERS scatterometer and NOAA/AVHRR, Papa et al. (2010) quantified the monthly distribution of global surface water extent and its variations at 25 km sampling intervals from 1993–2004.

These studies have contributed significantly to water detection with microwave data and have provided an abundance of references for this research. However, further work is still required to derive a more robust result over a wider area. Although Sippel’s algorithms worked well in the Amazon River floodplain, there are still deficiencies in the algorithms. First, the land structures in the Amazon River floodplain are monotonic and
of two typical types: rainforest and open water. When applying the algorithms in other areas, such as the Mississippi River Basin, the ground conditions become more complex and the accuracy decreases substantially. Second, a series of tests shows on various land covers and soil types that the levels of precipitation, tree covers, cloud fractions, and vegetation indices cause the variables for water fraction retrieval to fluctuate, but these factors were not considered in the algorithm on the polarization difference at 37GHz. Although the long-term aggregation used in Sippel’s modified algorithm can indirectly reflect the seasonal changes in the vegetation index and precipitation, these factors, especially precipitation, can vary over short periods. Third, additional variables, such as the polarization ratio index at 37GHz and soil wetness index, are more strongly correlated to water fractions. Brakenridge’s spatial ratio method has its own advantages in tracking river discharge. However, the basic assumption of this method is that contribution from emissivity differences and atmosphere conditions is constant over a large area. In fact, the assumption might be true if a searched dry pixel is with similar land cover combinations and soil wetness to that of the land partial in a river-centered pixel and with similar cloud cover or atmosphere conditions to the whole river-centered pixel. Otherwise, uncertainties can sometimes be introduced because of microwave data’s coarse resolution. Nevertheless, these factors have not been mentioned yet in the method.

1.2.3 Algorithms with multi-source data

Flood detection with multi-source data was first applied in a series of pioneering studies that used high resolution satellite data, such as SPOT, Landsat MSS or TM, and
SAR imagery along with DEM or topographic data to obtain water surface elevations or establish the land-water contacts (Gupta and Banerji, 1985; Miller, 1986; Brakenridge et al., 1994; Smith, 1997). Wang et al. (2002) also proposed a method to simulate inundated water veiled by trees and urban structures, which are shown as land in Landsat TM imagery, with river gauge observations and DEM data. Schumann et al. (2007) established a Regression and Elevation-based Flood Information eXtraction model (REFIX) to derive 3D flood information based on SAR data for steady state floods.

All of these contributions provide techniques to apply satellite data to flood analysis and water management. However, most of these studies utilize high resolution satellite data, which are limited by long revisit intervals and narrow swath coverage. Data from meteorological or environmental satellites such as TERRA, AQUA or Suomi-NPP have not been applied to water simulations along with high resolution DEM data. However, data from these coarse-to-moderate resolution satellites, which carry both optical and microwave sensors and provide large swath coverage and high observing frequency, have even better potential for water simulations and dynamic flood analysis.

1.3 Research objective and content of this dissertation

1.3.1 Research objective

This research focuses on the development of an enhanced flood product using multi-source data, including optical satellite data from EOS/MODIS and/or Suomi-NPP/VIIRS, microwave data from AMSR-E or ATMS, Shuttle Radar Topography
Mission (SRTM) digital elevation models (DEM$s$), and other ground observations such as precipitation and river gauge data. The idea to develop such a flood product originates from the fact that water fractions derived from both optical and microwave-based satellite data provide information about the water extent or area, and that an integrated model can be developed based on the water area information to determine the water distribution using high resolution DEM data from SRTM. The integrated model is based on three hydrodynamic features of water: first, water always flows downhill in the optimally shortest path; second, the shapes of water bodies are controlled by the topography of the natural environment; and third, at the same surface level, the water level is the same everywhere. Based upon these fundamental properties, high resolution DEM data can be used to translate coarse-to-moderate water distributions identified through satellite image pixels to more detailed sub-pixel water distributions.

A series of algorithms are developed to accomplish this process. Among these algorithms, the water fraction retrieval is the most important step. First, a decision-tree approach is used to identify all possible water bodies (pure and mixed) from other land types in MODIS or VIIRS imagery; this will be described in Chapter 2. To obtain better results, a geometric method is developed in Chapter 3 to remove cloud shadows from the water maps. Based on the water detection results, a dynamic nearest neighbor searching method (DNNS) is introduced in Chapter 4 to retrieve the water fraction from MODIS and VIIRS data. The water fraction retrieval from AMSR-E and ATMS data is presented in Chapter 5 and is based on a regression-tree approach using 37 GHz and 89 GHz AMSR-E data and 50 GHz and 85 GHz ATMS data and considering the land cover.
mixture and the impact of precipitation and clouds. With the retrieved water fraction products from MODIS or VIIRS and AMSR-E or ATMS, an integrated model is established in Chapter 6 to combine the water fraction with SRTM 30-m resolution DEMs to derive 30-m resolution water maps. These algorithms are applied to several severe floods in the USA in Chapter 7, and Chapter 8 provides a discussion and the conclusion. Because it utilizes satellite data from both optical and microwave sensors and high resolution DEM data, the final enhanced flood product is expected to have large spatial coverage and high spatial and temporal resolution and works in cloudy conditions.

1.3.2 Study area

We choose Midwestern USA, centered on the Mississippi River Basin, as the study area for developing the algorithm (Fig. 2). We chose this study area because it contains abundant rivers and lakes and has a high risk of flooding from hurricanes or intense rainfall. A severe flood occurred over the Mississippi River Basin in 2011, and sufficient MODIS, Landsat TM/ETM and AMSR-E data were collected to allow flood analyses at a variety of spatial resolutions. The developed algorithms are further applied and evaluated to several other severe floods in Alaska, Colorado and New York.
1.3.3 Data preparation

More than 18 types of data from a variety of data source are processed in this study, which are listed as follows:

- MODIS/VIIRS level-1b reflectance and brightness temperature data.
- Geo-location data and geometric data (MOD03 and MYD03)
- VIIRS level-1b reflectance and brightness temperature data
- VIIRS geo-location data and geometric data
- AMSR-E 25-km global brightness temperature datasets
- ATMS brightness temperature datasets
• ATMS geo-location data
• Cloud mask data from MODIS and VIIRS
• MOD44W global 250-m water map
• 16-day normalized vegetation index and 16-day visible and near-infrared reflectance from MODIS and VIIRS
• MOD44B tree cover data
• SRTM digital elevation model
• 24-hour 4-km precipitation data in the United States
• SWBD 30-m water map
• Landsat TM/ETM images
• River gauge data of Mississippi river
• IGBP Land-cover data
• Water shed distribution map in the United States

1.3.4 Dissertation content

The dissertation is divided into eight chapters:

Chapter 1: Introduction. This chapter presents the advantages and disadvantages of using remote sensing techniques for flood analysis and reviews current progress on the topic. Previous studies related to this topic are also reviewed in detail.

Chapter 2: Water Detection with Optical Satellite Data. This chapter describes water detection using a decision-tree method with optical satellite data.
Chapter 3: Automatic Cloud Shadow Removal from Flood/Standing Water Maps. This chapter describes the development of a geometric method to remove cloud shadows from water maps.

Chapter 4: Water Fraction Retrieval with Optical Satellite Data. This chapter provides details on the dynamic nearest neighbor searching method (DNNS) for water fraction retrieval using MODIS or VIIRS data.

Chapter 5: Microwave-Based Surface Water Fraction Retrieval with a Classification and Regression Tree Technique. In this chapter, a water fraction retrieval algorithm for microwave radiation using the Classification and Regression Tree Technique (CART) is described in detail.

Chapter 6: Integration with SRTM DEM Data to Derive High Resolution Water Maps. This chapter presents the method to combine water fraction products from MODIS or VIIRS and AMSR-E or ATMS data with SRTM DEMs to derive high resolution water maps.

Chapter 7: Application of the Algorithm to Floods. This chapter lists applications of these algorithms to several severe floods.

Chapter 8: Discussion and Summary. This chapter discusses the developed methods and potential applications in the future and draws conclusions from this research.

CHAPTER 2 Water Detection with Optical Satellite Data
2.1 Introduction

As the base product of flood detection using optical satellite data, an accurate water extent map is crucial for the downstream water products because the errors of water detection can be inherited from the extent map. The water extent map requires high accuracy water detection or recognition from satellite imagery. Traditional methods of water detection include visual interpretation, threshold segmentation, histogram methods and other similar approaches (Wiesnet et al., 1974; Cao et al., 1987; Y. Sheng et al., 2001). All of these methods can discriminate water bodies from other land cover types such as vegetation and bare soil. However, most pixels in coarse-to-moderate resolution satellite imagery are mixed and contain more than one land cover type. The mixed structure causes fuzzy spectral measurements from the multi-channel data and complicates the classification. Therefore, it is difficult for traditional threshold segmentation or histogram methods to obtain reliable detection results over different ground or atmospheric conditions. In this chapter, a decision-tree approach, which is a classic prediction technique that is widely used in land cover classification, is introduced as the main technique for automatic water detection using optical MODIS satellite data. The method is also tested in later chapters with data from similar sensors, such as VIIRS and SEVIRI, to evaluate its performance.
2.2 Methodology

2.2.1 Basic Spectral Properties of Water Surface

In general, water detection with optical satellite data is mainly based on the different spectral characteristics of water and other land cover types in visible (VIS), near-infrared (NIR), short-wave infrared (SWIR) and thermal infrared channels. As shown in Fig. 3, water has a higher reflectance in the visible channel than in the near-infrared and short-wave infrared channels. In contrast, vegetation has a much higher reflectance in the near-infrared band than in the visible band. The reflectance of dry bare land increases with increasing wavelength and reaches the highest reflectance in the SWIR channel, whereas the reflectance of water is close to 0 in this channel. The spectral differences between water, vegetation and bare soil in these channels form the basis for automatic water detection with multiple channels from optical satellites.
2.2.2 Decision-tree approach

2.2.2.1 Simple description of decision tree algorithm

A Decision Tree (DT) algorithm, which is a supervised machine learning technique, is a classic prediction model used to support decision making (Han, 2001) by converting complex data into a relatively simple and direct viewing structure. The process of predicting unseen instances is the same for all decision tree algorithms, and the differences between decision tree algorithms are the methods used to create the tree structures. The splitting criterion, stop-splitting rule, class assignment rule, and pruning
method can be used in the tree generation (Quinlan, 1993). Many kinds of decision trees have been generated based on different splitting criteria. The most classic algorithm is the C4.5 algorithm, which is used in this research and is described in detail below.

The basic strategy of the C4.5 algorithm is to select an attribute that will best separate the samples into individual classes by an ‘Information Gain Ratio’. The objective is to produce the most accurate separation with the least amount of information (Han et al., 2001). The calculation of the ‘Information Gain Ratio’ is expressed from Equations (2.1) to (2.4).

Let \( S \) be the training set consisting of \( s \) data samples, and \( s (C_i) \) be the number records in \( S \) that belong to class \( C_i \) \((i = 1, 2 \ldots m)\) out of \( m \) classes. The information needed to classify \( S \) is:

\[
Info(S) = - \sum_{i=1}^{m} \frac{s(C_i)}{s} \log_2 \left( \frac{s(C_i)}{s} \right)
\]

(2.1)

Hence, the amount of information needed to partition \( S \) into \( \{S_1, S_2 \ldots S_v\} \) by attribute \( A \) (\( A \) has \( v \) distinct values) is:

\[
Info(A \mid S) = - \sum_{j=1}^{v} \frac{s_j}{s} \times Info(S_j)
\]

(2.2)

And, the gain to classify \( S \) by attribute \( A \) is:

\[
gain(A \mid S) = Info(S) - Info(A \mid S)
\]

(2.3)

Then the ‘Information Gain Ratio’ is computed as:

\[
gainRatio(A \mid S) = \frac{gain(A \mid S)}{Info(A \mid S)}
\]

(2.4)
2.2.2.2 Decision Tree Samples

Based on the principles of the decision tree generation, different decision trees can be derived with a sample collection. The following example shows a sample J48graft decision tree, which is based on the classic C4.5 algorithm for a simple classification of water, bare land, vegetation, wetlands and clouds from a collection of 36,262 samples with the variables ch1, ch2, ch3, NDVI, NDSI and NDWI.

```
ch3 <= 100
  | NDVI <= 0.104575
  |   | NDSI <= 0.517241
  |   |     | ch1 <= 47
  |   |     |     | ch3 <= 13.5: water (0.0|1482.0)
  |   |     |     | ch3 > 13.5
  |   |     |     |   | NDWI <= 0.456305
  |   |     |     |     | NDVI <= -0.303689: water (0.0|1404.0)
  |   |     |     |     | NDVI > -0.303689
  |   |     |     |     | ch2 <= 22.5: water (0.0|1220.0)
  |   |     |     |     | ch2 > 22.5: Wetland (50.0)
  |   |     |     | ch1 > 47: water (370.0/1.0)
  |   | NDSI > 0.517241: water (4283.0)
  | NDVI > 0.104575
  | ch1 <= 62
  |   | ch1 <= 39
  |   |     | ch1 <= 37: Vegetation (23.0)
  |   |     | ch1 > 37
  |   |     |     | ch3 <= 35.5: Wetland (0.0|167.0)
  |   |     |     | ch3 > 35.5
  |     | NDSI <= -0.428526: Vegetation (0.0|143.0)
  |     | NDSI > -0.428526
  |     |   | NDVI <= 0.655829
  |     |     | ch2 <= 214.5: Wetland (2.0)
  |     |     | ch2 > 214.5: Vegetation (0.0|143.0)
  |     | ch1 > 39
  |     | ch1 <= 57: Wetland (10325.0/9.0)
```
ch1 > 57
   NDSI <= -0.106897: Wetland (852.0/6.0)
   NDSI > -0.106897
      NDVI <= 0.215686: water (15.0/1.0)
      NDVI > 0.215686
         NDVI <= 0.410753
            ch1 <= 60
               ch1 <= 58
                  NDWI <= 0.289199
                     ch3 <= 63
                        ch3 <= 57.5: Wetland (0.0|625.0)
                        ch3 > 57.5
                           NDSI <= -0.072364: Wetland (0.0|430.0)
                           NDSI > -0.072364
                              ch2 <= 89.5: Wetland (0.0|354.0)
                              ch2 > 89.5
                                 NDVI <= 0.327076
                                    NDSI <= 0.032796
                                       NDWI <= 0.184105: Wetland (0.0|46.0)
                                       NDWI > 0.184105: water (2.0)
                                       NDSI > -0.032796: Wetland (0.0|338.0/2.0)
                                       NDVI > 0.327076: Wetland (0.0|321.0/1.0)
                                       ch3 > 63: Wetland (7.0/1.0)
                              NDVI > 0.327076: Wetland (0.0|321.0/1.0)
                           ch3 <= 60
                              NDVI <= 0.315927: Wetland (2.0)
                              NDVI > 0.315927: water (8.0/2.0)
               ch3 > 60: Wetland (4.0)
      ch2 <= 107: water (4.0)
      ch2 > 107
         NDWI <= 0.362162: Wetland (67.0/6.0)
         NDWI > 0.362162
            ch3 <= 60
               NDVI <= 0.315927: Wetland (2.0)
               NDVI > 0.315927: water (8.0/2.0)
            ch3 > 60: Wetland (4.0)
       ch1 > 60
          ch3 <= 87.5
             ch2 <= 155.5
                NDWI <= 0.259417: Wetland (0.0|66.0/4.0)
                NDWI > 0.259417
                   NDVI <= 0.260511: water (0.0|51.0/5.0)
                   NDVI > 0.260511: Wetland (40.0/8.0)
                      ch2 > 155.5: water (0.0|24.0)
                      ch3 > 87.5: Wetland (0.0|39.0)
             NDVI > 0.410753
ch1 <= 60
| ch3 <= 92.5
| NDWI <= 0.27531: Wetland (0.0|232.0/1.0)
| NDWI > 0.27531
| NDVI <= 0.412043: Wetland (0.0|20.0)
| NDVI > 0.412043: water (4.0)
| ch3 > 92.5: Wetland (0.0|137.0)
ch1 > 60
| ch1 <= 61: Wetland (2.0)
| ch1 > 61
| ch3 <= 82.5
| NDWI <= 0.28638: Wetland (0.0|70.0)
| NDWI > 0.28638: water (3.0)
| ch3 > 82.5: Wetland (0.0|76.0)
ch2 > 168
| NDWI <= 0.337209
| NDWI <= 0.310962
| NDWI <= 0.181527: Wetland (0.0|3120.0)
| NDWI > 0.181527
| NDSI <= -0.421821: Vegetation (0.0|3100.0/1.0)
| NDSI > -0.421821
| NDVI <= 0.61443
| ch2 <= 218: Wetland (23.0)
| ch2 > 218: Vegetation (0.0|3100.0/1.0)
| NDVI > 0.61443: Vegetation (0.0|3100.0/1.0)
| NDWI > 0.310962
| ch3 <= 93
| ch2 <= 170
| ch1 <= 54.5: Wetland (0.0|402.0)
| ch1 > 54.5
| ch3 <= 62.5: Wetland (0.0|432.0/5.0)
| ch3 > 62.5
| NDSI <= -0.063278
| NDVI <= 0.376614: Wetland (0.0|409.0/7.0)
| NDVI > 0.376614
| NDWI <= 0.328185
| NDWI <= 0.315122: Wetland (0.0|106.0/6.0)
| NDWI > 0.315122
| NDSI <= -0.221761: Wetland (0.0|8.0)
| NDSI > -0.221761
| NDVI <= 0.521176: water (3.0/1.0)
| NDVI > 0.521176: Wetland (0.0|7.0)
| NDWI > 0.328185: Wetland (0.0|140.0/3.0)
NDSI > -0.063278: Wetland (0.0|12.0/7.0)
| ch2 > 170: Wetland (6.0)
| ch3 > 93
| NDSI <= -0.392418: Vegetation (0.0|1045.0)
| NDSI > -0.392418
| NDVI <= 0.635156
| ch2 <= 222
| ch1 <= 53.5: Vegetation (0.0|904.0/2.0)
| ch1 > 53.5
| NDWI <= 0.311338: Vegetation (0.0|21.0)
| NDVI > 0.311338
| ch3 <= 94.5
| ch1 <= 60.5: water (2.0)
| ch1 > 60.5: Wetland (0.0|12.0)
| ch3 > 94.5: Wetland (0.0|18.0)
| NDVI > 0.635156: Vegetation (0.0|1045.0)
| NDWI > 0.337209
| ch3 <= 35.5: Wetland (0.0|124.0)
| ch3 > 35.5: water (15.0)
| ch1 > 62
| NDSI <= -0.108911
| NDVI <= 0.315126
| ch2 <= 142
| ch2 <= 140
| NDVI <= 0.384991
| NDSI <= -0.123549: Wetland (0.0|6555.0/4.0)
| NDSI > -0.123549
| ch3 <= 89.5
| NDVI <= 0.216687: Wetland (0.0|4743.0/12.0)
| NDVI > 0.216687
| NDSI <= -0.117382
| ch2 <= 137.5: Wetland (0.0|6538.0/38.0)
| ch2 > 137.5
| ch2 <= 139.5
| ch3 <= 79.5: Wetland (0.0|2934.0/29.0)
| ch3 > 79.5
| NDVI <= 0.369293: Wetland (0.0|2124.0/39.0)
| NDVI > 0.369293
| NDWI <= 0.266164: water (2.0)
| NDWI > 0.266164: Wetland (0.0|471.0/15.0)
| ch2 > 139.5: Wetland (0.0|135.0)
| NDSI > -0.117382: Wetland (0.0|206.0)
ch3 > 89.5: Wetland (0.0|1905.0/1.0)
NDVI > 0.384991: Wetland (0.0|3805.0)
ch2 > 140: Wetland (3.0)
ch2 > 142
NDWI <= -0.145003: Bare land (0.0|5069.0)
NDVI > -0.145003
NDVI <= 0.14763: Bare land (0.0|2074.0/2.0)
NDVI > 0.14763: Wetland (174.0/2.0)
NDWI > 0.315126: water (24.0/1.0)
NDSI > -0.108911
ch1 <= 70
ch2 <= 123
NDSI <= 0.416344
ch2 <= 54.5: water (0.0|105.0)
ch2 > 54.5
ch3 <= 21.5: water (0.0|86.0)
ch3 > 21.5
NDSI <= -0.097284: Wetland (0.0|253.0/6.0)
NDSI > -0.097284
NDVI <= 0.301223: water (70.0/4.0)
NDVI > 0.301223: Wetland (0.0|1586.0/59.0)
NDSI > 0.416344: water (0.0|113.0)
ch2 > 123
ch3 <= 48
NDVI <= 0.226214: water (0.0|45.0)
NDVI > 0.226214
NDSI <= 0.228473
NDWI <= 0.473389: Wetland (3.0)
NDWI > 0.473389: water (0.0|20.0)
NDSI > 0.228473: water (0.0|37.0)
ch3 > 48
ch1 <= 68
ch3 <= 95.5: water (55.0/7.0)
ch3 > 95.5: Wetland (0.0|64.0)
ch1 > 68
NDVI <= 0.323308: water (0.0|101.0)
NDVI > 0.323308
NDSI <= 0.088304
NDWI <= 0.400483
ch2 <= 177.5
ch3 <= 56.5: water (0.0|35.0)
ch3 > 56.5
NDVI <= 0.439851
ch2 <= 132.5: water (0.0|26.0|1.0)
ch2 > 132.5
ch3 <= 71.5: Wetland (2.0)
ch3 > 71.5: water (0.0|102.0|10.0)
NDVI > 0.439851: water (0.0|15.0)
ch2 <= 177.5: water (0.0|44.0)
NDWI > 0.400483: water (0.0|90.0)
NDSI > 0.088304: water (0.0|98.0)
ch1 > 70: water (191.0)
ch3 > 100
NDWI <= 0.05: Bare land (9965.0|5.0)
NDWI > 0.05
ch1 <= 64: Vegetation (6680.0|1.0)
ch1 > 64
ch2 <= 203
NDSI <= 0.23548
ch2 <= 123.5: water (0.0|1474.0)
ch2 > 123.5
NDVI <= -0.041828: water (0.0|1215.0)
NDVI > -0.041828
NDWI <= 0.471658
ch1 <= 232.5
ch3 <= 351.5
ch2 <= 197.5: water (28.0)
ch2 > 197.5: Bare land (0.0|473.0|4.0)
ch3 > 351.5: Cloud (0.0|563.0)
ch1 > 232.5: Cloud (0.0|943.0)
NDWI > 0.471658: water (0.0|1061.0)
NDSI > 0.23548: water (0.0|1496.0)
ch2 > 203
ch3 <= 198
ch1 <= 91
ch3 <= 165.5: Vegetation (0.0|5318.0)
ch3 > 165.5
NDVI <= 0.744795
NDWI <= 0.285616
NDSI <= -0.495664: Vegetation (0.0|5271.0|50.0)
NDSI > -0.495664
ch2 <= 355.5: Vegetation (14.0)
ch2 > 355.5: Cloud (0.0|71.0)
NDWI > 0.285616: Vegetation (0.0|5039.0|25.0)
NDVI > 0.744795: Vegetation (0.0|1737.0)
ch1 > 91
Part of the above binary tree is presented in Fig. 4.

Figure 4 Part of the sample J48Graft Decision Tree for water detection
The output of a decision tree also includes the statistical results of the training dataset with the current decision tree approach in three parts: general errors, detailed accuracy for each class and the confusion matrix.

General errors of the above decision tree:

Correctly Classified Instances 36152 99.6967 %
Incorrectly Classified Instances 110 0.3033 %
Kappa statistic 0.996
Mean absolute error 0.0017
Root mean squared error 0.033
Relative absolute error 0.5644 %
Root relative squared error 8.4438 %
Total Number of Instances 36262

Detailed accuracy for each class of the sample J48graft decision tree is shown in Table 2.

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare land</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>Bare land</td>
</tr>
<tr>
<td>Cloud</td>
<td>0.999</td>
<td>0.001</td>
<td>0.999</td>
<td>0.999</td>
<td>0.999</td>
<td>0.997</td>
<td>Cloud</td>
</tr>
<tr>
<td>Vegetation</td>
<td>1.000</td>
<td>0.002</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>Vegetation</td>
</tr>
<tr>
<td>Water</td>
<td>0.988</td>
<td>0.001</td>
<td>0.982</td>
<td>0.988</td>
<td>0.991</td>
<td>0.990</td>
<td>Water</td>
</tr>
<tr>
<td>Wetland</td>
<td>0.996</td>
<td>0.002</td>
<td>0.995</td>
<td>0.996</td>
<td>0.996</td>
<td>0.999</td>
<td>Wetland</td>
</tr>
<tr>
<td>Weighted Avg.</td>
<td>0.997</td>
<td>0.001</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.999</td>
<td></td>
</tr>
</tbody>
</table>
The Confusion Matrix of the sample J48graft decision tree is presented below as well:

\[
\begin{array}{cccccc}
& a & b & c & d & e \\
9959 & 1 & 0 & 0 & 0 & | a = \text{Bare land} \\
2892 & 2 & 0 & 0 & 0 & | b = \text{Cloud} \\
2 & 6714 & 0 & 0 & 0 & | c = \text{Vegetation} \\
5 & 0 & 0 & 5036 & 58 & | d = \text{water} \\
0 & 0 & 0 & 42 & 11551 & | e = \text{Wetland} \\
\end{array}
\]

From the statistics results, the sample tree has quite good discrimination ability among the five classes with the variables.

2.2.3 DT Application in water detection

2.2.3.1 Tree comparison

There are many kinds of decision trees, and different trees have different performances in classification. To determine the best decision tree algorithm for water detection, several decision tree algorithms are tested and compared based on the test accuracy (Table 3), including the J48graft or J48, which is based on the C4.5 algorithm and was originally proposed by Quinlan (1993), NBTree, which is a Naïve Bayes/Decision Tree hybrid (Kohavi, 1996), Random Tree (RT), Random Forest (Breiman, 2001), REP Tree, BFTree, Decision Stump (DS), FT (Final Tree), and CART (Classification and Regression Trees) (Breiman et al., 1984). Although all of the tested
DT algorithms have good discrimination capability, the J48graft/J48 based on the C4.5 algorithm and the CART have the best accuracy. When these trees are tested further and compared using samples without distinct differences, the accuracy of the classification results decreases substantially, but the J48graft tree still performs the best (Table 4). Therefore, the J48graft tree is chosen as the main DT algorithm to generate decision trees for automatic water detection in this study.

Table 3 TP Rate of Classified Instances in different types of Decision Tree Algorithms (Bare land, water, vegetation, cloud shade)

<table>
<thead>
<tr>
<th>Trees Types</th>
<th>NB</th>
<th>J48graft</th>
<th>RF</th>
<th>Random</th>
<th>REP</th>
<th>Cart</th>
<th>BF</th>
<th>DS</th>
<th>FT</th>
<th>J48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>87.2</td>
<td>88.1</td>
<td>88.8</td>
<td>87.4</td>
<td>87.6</td>
<td>86.8</td>
<td>87.6</td>
<td>61.5</td>
<td>87.7</td>
<td>88</td>
</tr>
<tr>
<td>Bare land</td>
<td>95.1</td>
<td>93.8</td>
<td>96</td>
<td>93.8</td>
<td>96</td>
<td>93.4</td>
<td>93.8</td>
<td>89.8</td>
<td>94.7</td>
<td>93.4</td>
</tr>
<tr>
<td>Water</td>
<td>87.1</td>
<td>89.7</td>
<td>91.5</td>
<td>88.2</td>
<td>87.5</td>
<td>86.8</td>
<td>88.6</td>
<td>90.9</td>
<td>88.9</td>
<td>89.3</td>
</tr>
<tr>
<td>Vegetation</td>
<td>97.3</td>
<td>97.3</td>
<td>95.1</td>
<td>96.2</td>
<td>96.2</td>
<td>97.3</td>
<td>97.3</td>
<td>0</td>
<td>96.8</td>
<td>96.8</td>
</tr>
<tr>
<td>Cloud shade</td>
<td>67.4</td>
<td>66.8</td>
<td>65.2</td>
<td>68.5</td>
<td>69</td>
<td>68.5</td>
<td>67.4</td>
<td>0</td>
<td>66.3</td>
<td>68.5</td>
</tr>
</tbody>
</table>

Table 4 TP Rate of Water Classified Instances with DT Algorithms

<table>
<thead>
<tr>
<th>Trees Types</th>
<th>NB</th>
<th>J48- graft</th>
<th>RF</th>
<th>Random</th>
<th>REP</th>
<th>Cart</th>
<th>BF</th>
<th>DS</th>
<th>FT</th>
<th>J48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>73.4</td>
<td>73.6</td>
<td>75.95</td>
<td>67.5</td>
<td>71.7</td>
<td>72</td>
<td>70.4</td>
<td>42.8</td>
<td>68.5</td>
<td>73.1</td>
</tr>
<tr>
<td>Water mixed with land</td>
<td>69.6</td>
<td>70.3</td>
<td>76.8</td>
<td>66.7</td>
<td>63.8</td>
<td>68.1</td>
<td>66.7</td>
<td>0</td>
<td>69.6</td>
<td>67.4</td>
</tr>
</tbody>
</table>
2.2.3.2 Decision Tree for automatic water detection

In addition to channel reflectance, a series of variables, such as NDVI (Normalized Difference Vegetation Index), NDSI (Normalized Difference Snow Index) and NDWI (Normalized Difference Water Index) are used to differentiate water bodies from other land cover types. NDVI, NDSI and NDWI are defined as follows:

\[
\text{NDVI} = \frac{R_{\text{NIR}} - R_{\text{VIS}}}{R_{\text{NIR}} + R_{\text{VIS}}} \quad (2-5)
\]

\[
\text{NDSI} = \frac{R_{\text{VIS}} - R_{\text{SWIR}}}{R_{\text{VIS}} + R_{\text{SWIR}}} \quad (2-6)
\]

\[
\text{NDWI} = \frac{R_{\text{NIR}} - R_{\text{SWIR}}}{R_{\text{NIR}} + R_{\text{SWIR}}} \quad (2-7)
\]

Fig. 5 presents a series of scatter plots of the variables \(R_{\text{VIS}}, R_{\text{NIR}}, R_{\text{SWIR}}, \text{NDVI},\) and NDSI for four classes, vegetation, bare land, water and cloud shadows, with a collection of samples from NOAA-17/AVHRR. Although all of these variables are effective in differentiating the classes from one another, none can independently discriminate them. Thus, a combination of these variables with a decision tree approach may be required for a more feasible and robust classification. One phenomenon shown in

| Pure River water | 82.1 | 82.1 | 84.6 | 73.2 | 82.1 | 82.4 | 79.7 | 0 | 80.5 | 82.1 |
| Pure Lake water | 80.7 | 82 | 84 | 81.3 | 79.3 | 86.9 | 82 | 89.3 | 80 | 82 |
| Wetland | 62.4 | 66.7 | 63.1 | 56 | 58.2 | 57.7 | 56 | 0 | 48.9 | 66.7 |
| Cloud shade | 72.8 | 69 | 72.8 | 67.5 | 75 | 68.5 | 68.5 | 98.4 | 65.2 | 69 |
Fig. 5 is that the cloud shadow samples have very similar characteristics to the water samples. This means that there is substantial confusion between the two objects and that no matter what approach is used, the two classes are difficult to differentiate from each other based on their spectral properties. Therefore, during the water detection with the C4.5 algorithm, cloud shadow samples are excluded from the sample collection, and cloud shadows are not classified as an independent type but are mixed with other classes, including the water class. After classification, a geometric method, which is described in detail in Chapter 3, is utilized to remove cloud shadow pixels that are mixed in the water class.
Figure 5 Series of scatter points images of samples from NOAA-17/AVHRR
Another issue arises from the difference in spatial resolution between channels. The cloud masks are produced from MODIS data with a 1 km resolution and from VIIRS data with a 750 m resolution, while the water detection is based on VIS, NIR and SWIR channels that have a 500 m resolution in the MODIS data or a 375 m resolution in the VIIRS data. New cloud pixels with smaller cloud fractions may be present in the 500 m and 375 m data that might not be detected with the 1 km and 750 m cloud masks. These new cloud pixels sometimes confuse the decision tree and reduce the accuracy of the water detection. To avoid this problem, clouds are classified as an independent class that is separate from vegetation, bare land and water.

Because of the existence of a mixed structure for most pixels in coarse-to-moderate resolution satellite data, the spectral properties of sub-pixel areas of land can significantly impact the total reflectance of mixed water pixels. Hence, water might be detected differently over desert and forest; a well-trained tree from a desert area may fail to correctly detect water in forests. To reduce the uncertainty from the variance of sub-pixel land areas, IGBP land cover data is introduced. For all clear-sky pixels filtered by the cloud and snow masks, a J48raft tree is applied to make a preliminary classification by dividing the pixels into water, vegetation, bare land and clouds. New samples of land cover types that have poor classification accuracy are then collected based on the preliminary classification results. With the newly collected samples, new decision trees are generated and utilized to reclassify classes with poor performance in specific land cover types. A total of seven decision trees are required to obtain robust results for different situations.
2.3 Results

The method was tested with large EOS/MODIS, Suomi-NPP/VIIRS and MSG/SEVIRI datasets. The results are reliable and promising. Fig. 6 shows a false-color composite map from EOS/MODIS data and the corresponding water detection results over the Mississippi River Basin. Figs. 7 to 11 show a series of water maps for the same area as Fig. 6, and Fig. 12 presents a VIIRS false-color image and the corresponding water detection results for Alaska on May 26, 2013.

Although pixels containing very small water percentages (less than 18% after the validation in Chapter 4) may not be detected, most of the water pixels shown in the false-color images are consistent with those identified by visual inspection. Because cloud shadows are not classified as a single class in the decision tree, most of the cloud shadow pixels are identified as water pixels. Thus, further removal of cloud shadows from the water classification results is required.

Figure 6 MODIS false-color composite image (Left) and water map (Right) in Mississippi river on May 04, 2011
Figure 7 MODIS false-color composite image (Left) and water map (Right) in Mississippi river on March 01, 2011

Figure 8 MODIS false-color composite image (Left) and water map (Right) in Mississippi river on April 10, 2011
Figure 9 MODIS false-color composite image (Left) and water map (Right) in Mississippi river on May 06, 2011

Figure 10 MODIS false-color composite image (Left) and water map (Right) in Mississippi river on May 20, 2011
Figure 11 MODIS false-color composite image (Left) and water map (Right) in Mississippi river on May 29, 2011

Figure 12 VIIRS false-color composite image (Top) and the corresponding water map (Bottom) in Alaska on May 26, 2013
2.4 Discussion and Summary

In this chapter, a decision-tree approach based on the C4.5 algorithm is used to discriminate water from other objects, including vegetation, bare land and clouds, by considering mixtures of land cover. To guarantee robust performance, seven decision trees are trained and applied. The method is tested with more than 500-granule MODIS data and 50-granule VIIRS data at a global scale and shows promising and stable results. The decision tree has advantages for complex ground conditions over traditional threshold segmentation or histogram methods. However, several problems remain unsolved.

First, because both MODIS data and VIIRS data are not atmospherically corrected, the reflectance of vegetation and bare soils at the granule edges in VIS, NIR and SWIR channels may vary substantially from that at the nadir area. Similarly, the surface reflectance in these channels varies with the season. To include these uncertainties in decision trees, samples covering different situations must be collected and thus may cause the trees to have too many branches and leaves. Fortunately, the water surface reflectance in these channels remains much more stable in all seasons and at the granule edges than vegetation or bare soil, which guarantees a constant accuracy of the water detection results with MODIS and VIIRS data.

Second, snow and ice over the forest can sometimes be confused with water surfaces because the thermal brightness band is not used as a variable in the decision tree.
Because a snow/ice mask is not used, the detection of water in winter or early spring may have lower accuracy.

Third, because a cloud mask is not used, a decision tree with variables from the visible to SWIR channels may not completely differentiate clouds from other land types.

Finally, most cloud shadows are detected as water pixels because they are spectrally similar in these channels. This requires a geometric method to discriminate cloud shadows from water, which will be addressed in Chapter 3.

In conclusion, the decision-tree technique based on the C4.5 algorithm (referred to as the J48graft tree in this chapter) has good predictive ability to identify water from other objects and thus can be used for automatic water detection using MODIS and VIIRS data instead of traditional threshold segmentation and histogram methods. Although residual problems remain unsolved, the results meet the requirements of water detection in most situations. The consistent performance also provides a good foundation for the remainder of the work in this study.
CHAPTER 3 Automatic Cloud Shadow Removal from Flood/Standing Water Maps

3.1 Introduction

Cloud shadows are a common phenomenon in optical satellite imagery and a significant concern for most land surface remote sensing applications, especially surface water detection. Cloud shadows are caused when any cloud casts its shadow over the ground or other lower clouds. The shadowed area is typically darker than the non-shadowed area because there is less irradiance. In optical satellite images, especially satellite images with coarse or moderate resolution, cloud shadows are easily confused with water or wetland area because they share a variety of spectral features that make it difficult to detect flood/standing water from satellite data with a high level accuracy.

As shown in chapter 2, cloud shadows and water cannot be differentiated from one and the other based on spectral properties. Therefore, in this chapter, we developed geometric relationships between clouds and their shadows both in ideal plane and spherical plane. Based on the developed geometric relationships, an iterative method is introduced to determine cloud and cloud shadow pairs considering the curvature of the Earth. The method is applied in MSG/SEVIRI continuous observations every 30 minutes and shows quite promising results. The application in MODIS and VIIRS data in the following chapters also presents high accuracy of the method in cloud shadow determination.
3.2 Methodology

3.2.1 Generation of the geometric relationship between clouds and cloud shadows

3.2.1.1 Parallax and shadow

A parallax is the apparent displacement or difference in the position of an object viewed along two different lines of sight and is measured by the angle or semi-angle of inclination between those two lines. As shown in Fig. 13, when viewing an object O with height $h$ from different viewing points ($V$ and $V_1$), the position of O is located at different points ($P$ and $P_{O1}$) in the image plane, where $P$ is the position obtained from a vertical viewing angle and represents the exact position of O without parallax. When O is viewed from point $V_1$ at elevation angle $\theta_V$, the corresponding position of O in the image plane is $P_{O1}$ instead of $P$. The displacement from $P$ to $P_{O1}$ is an example of parallax. The distance between $P_{O1}$ and $P$ is the parallax distance, which can be measured using the object’s height and elevation angle.

When an opaque object comes between an area and a source of radiation, thereby intercepting radiation to the area, a shadow occurs. Shadows differ from parallaxes in that they are not caused by a viewing obstacle. Instead, a shadow is formed because an obstacle is illuminated. However, shadows can be viewed as a type of parallax if we consider the light source to be a viewing point. As shown in Figure 13, when light source (S) has an elevation angle $\theta_S$, line object PO with height $h$ casts its shadow along line $PS_O$. The shadow position of O locates at $S_O$. Similar to the parallax distance, the object
height and light source elevation angle can also be used to measure shadow length $PS_O$.

Because parallaxes and shadows have a similar geometric relationship, they can be calculated in the same manner.

3.2.1.2 Geometric relationship of cloud shadows over an ideal plane

In this context, an ideal plane denotes a plane with no distortion or inclination in either the X or Y directions. If we extend the geometry shown in Fig. 13 to an ideal plane, then we obtain the shadow and parallax geometric relationship over the image.
plane (Fig. 14). Figure 14 shows that the object height $h$ and elevation angle $\theta$ determine the parallax or shadow distance, whereas the azimuth angle $\phi$ determines the direction.

The following formulas express the parallax or shadow distance in the X and Y directions of the ground plane respectively:

$$X = \frac{h}{\tan \theta} \cdot \sin \phi$$

(3-1)

$$Y = \frac{h}{\tan \theta} \cdot \cos \phi$$

(3-2)

In the image plane, the parallax or shadow distance in the $X'$ and $Y'$ directions are, respectively, expressed as

$$X' = \frac{h}{\tan \theta \cdot r} \cdot \sin \phi$$

(3-3)

$$Y' = \frac{h}{\tan \theta \cdot r} \cdot \cos \phi$$

(3-4)

Where,

$$r = \frac{O_1P_1}{P_0P}$$

(3-5)

According to Equations (3-3) through (3-5), object O exists at point $O_1$ in the image plane, with a parallax distance between $O_1$ and object O’s actual position $P_1$. The position of shadow point ($S_{O1}$) in the image plane is expressed as:

$$X_{S_{O1}} = X_{O_1} + \frac{h}{\tan \theta _v \cdot r} \cdot \sin \phi_v + \frac{h}{\tan \theta _s \cdot r} \cdot \sin \phi_s$$

(3-6)

$$Y_{S_{O1}} = Y_{O_1} - \frac{h}{\tan \theta _v \cdot r} \cdot \cos \phi_v + \frac{h}{\tan \theta _s \cdot r} \cdot \cos \phi_s$$

(3-7)
In Equations (3-6) and (3-7), $\theta_v$ is the viewing elevation angle, $\varphi_v$ is the viewing azimuth angle, $\theta_s$ is the light source elevation angle and $\varphi_s$ is the light source azimuth angle.

Given Equations (3-6) and (3-7), if the object height $h$ is known, then we can estimate a shadow position from an object’s image position when there is no distortion in either the X or Y directions. Regarding the satellite data nadir area, in which the distortion in the longitude and latitude is parallel and the curvature of Earth may be ignored, Equations (3-6) and (3-7) predict cloud shadows with a high level of accuracy.

Figure 14 Graph of parallax and shadow geometry over ideal plane
3.2.1.3 Geometric relationship of cloud shadows over a spherical surface

The geometric relationship over a plane surface described in Section 3.2.1.2 can be used in a nadir area of satellite imagery, where the ground can be viewed as an ideal plane that is relatively unaffected by the curvature of Earth, especially when the spatial resolution is high. However, Earth’s curvature must be taken into account outside of the nadir area.

Given Earth’s curvature, the geometric relationship of a parallax or shadow is more complicated over a spherical surface than over a flat surface. Fig. 15 shows the parallax and shadow geometry over a spherical plane; Object C represents a piece of a cloud, and the sun casts C’s shadow over the Earth onto A, where \( \theta_A \) is the solar elevation angle. Because C has a height \( h \), the position of the satellite imagery is located at B when the satellite views C. Arc \( \widehat{PB} \) is the parallax distance, and Arc \( \widehat{PA} \) is the shadow length. Angles \( \alpha \) and \( \beta \) are the shadow and parallax angles, respectively.

Using the triangle relationship, Shadow Angle \( \alpha \) and Parallax Angle \( \beta \) can be calculated via Equation (3-8) as follows:

\[
\delta = \cos^{-1} \left[ \frac{(R+h)^2 - (\sqrt{R \times R \times \cos^2 \theta + h \times (h + 2R) - R \times \cos \theta})^2 + R^2}{2.0 \times R \times (R+h)} \right] \quad (3-8)
\]

Where \( \delta \) represents Shadow Angle \( \alpha \) or Parallax Angle \( \beta \), \( R \) is Earth’s radius, \( h \) is the cloud height, and \( \theta \) is the zenith angle.

Using Equation (3-8), \( \widehat{PA} \) and \( \widehat{PB} \) can be calculated as an arc in a circle with \( \alpha \) and \( \beta \), respectively. Because \( \widehat{PA}, \widehat{PB} \) and C’s satellite image position in Image Plane B
(lon_B, lat_B) are known, Cloud Shadow Position P (lon_P, lat_P) can be located by calculating a circle using Sensor Azimuth Angle \( \varphi_B \).

\[
\text{lat}_P = \sin^{-1}[\sin(\text{lat}_B) \times \cos \frac{PB}{R} + \cos(\text{lat}_B) \times \sin \left( \frac{PB}{R} \right) \times \cos \varphi_B]
\]  (3-9)

\[
\text{lon}_P = \text{lon}_B + \tan^{-1}\left[\frac{\sin \varphi_B \times \sin \frac{PB}{R} \times \cos(\text{lat}_B)}{\cos \frac{PB}{R} \times \sin(\text{lat}_B) \times \sin(\text{lat}_P)}\right]
\]  (3-10)

Thus, Point A’s position can be calculated with Solar Azimuth Angle \( \varphi_P \).

\[
\text{lat}_A = \sin^{-1}[\sin(\text{lat}_P) \times \cos \frac{PA}{R} + \cos(\text{lat}_P) \times \sin \left( \frac{PA}{R} \right) \times \cos \varphi_P]
\]  (3-11)

\[
\text{lon}_A = \text{lon}_P + \tan^{-1}\left[\frac{\sin \varphi_P \times \sin \frac{PA}{R} \times \cos(\text{lat}_P)}{\cos \frac{PA}{R} \times \sin(\text{lat}_P) \times \sin(\text{lat}_A)}\right]
\]  (3-12)

In contrast, if Shadow Position A (lon_A, lat_A) is known, then Equations (3-8) through (3-12) can also be used to predict Cloud Position B (lon_B, lat_B) on the spherical surface. Because the positions of Points A, B and P are expressed as longitudes and latitudes, geo-location datasets should be used to transform these coordinate pairs into satellite imagery column and row pairs. This transformation is not discussed in this paper.
3.2.2 Cloud and cloud shadow determination

If a pixel is defined as a cloud or cloud shadow in a satellite image, then we can predict the corresponding cloud or cloud shadow pixel, respectively, using Equations (3-
8) through (3-12). Furthermore, if a cloud mask is available and the cloud top and base 
hights are calculated accurately, cloud shadows can be predicted correctly in satellite 
imagery using Equations (3-8) through (3-12). Nevertheless, because this study examines 
flood/standing water detection, cloud products (e.g., cloud mask, cloud top and base 
hights) are not available. Furthermore, generating coarse real-time water maps is much 
easier with a multi-threshold method to inversely remove cloud shadows from water 
maps compared with cloud product generation. In a MODIS, SEVIRI or VIIRS water 
map obtained with a decision-tree method using visible, near-infrared and shortwave-
infrared channels, we assumed that all of the water pixels were cloud-shadow pixels, 
except those with large negative NDVI scores. Because the cloud shadow positions were 
known, those that cast shadows could be predicted by providing a dynamic range of 
cloud heights $h$, which is the only unknown variable in Equation (3-8).

3.2.2.1 Cloud spectral properties compared with water and cloud shadows

While searching cloud pixels for the presumptive cloud-shadow pixels, some 
typical spectral properties shared between water/shadows and clouds can be used. In the 
daytime, the brightness temperature in the thermal-infrared channel for inland water and 
shadows is lower than that in illuminated land nearby but higher than that of clouds. This 
effect is due to water’s specific heat capacity and decreased irradiance over shadows. 
Furthermore, the reflectance in the visible and near-infrared channels of water without 
sun glint and cloud shadows might be lower than those of clouds or land. The spectral 
properties of water and cloud shadows enhance the efficacy of brightness temperature in
the thermal-infrared channel and reflectance in the visible and near-infrared channels when determining whether the pixel in question is a cloud. A pixel predicted using Equations (3-8) through (3-12) is a cloud when it meets the following criteria:

\[
BT_{11\mu m} < BT_{11\mu m_{\text{shadow}}} - \Delta BT \\
R_{\text{VIS}} > R_{\text{VIS}_{\text{shadow}}} \\
R_{\text{NIR}} > R_{\text{NIR}_{\text{shadow}}}
\]

Considering the variances of the surface temperatures of water and shadows, \( \Delta BT \) is set according to cloud types determined by the shortwave-infrared channel. In this study, we assign \( \Delta BT = 2K \) in most cases. With these conditions, most of the clouds are included, except those over water in winter.

### 3.2.2.2 Cloud height estimation

In Section 3.2.2.1, when a cloud pixel is determined using the iteration of an assumed cloud height from 0.5 - 14 km, which corresponds to a shadow pixel determined by its spectral properties, it casts a cloud shadow pixel nearby. This shadow pixel might be cast by the predicted cloud pixel or another shadow pixel. Because the cloud pixel is known, its shadow can be determined if a relatively accurate cloud height is provided using Equations (3-8) through (3-12).

Cloud height can be estimated when both the vertical atmosphere temperature profile and the environment temperature of the cloud are known. An accurate cloud height calculation requires a series of cloud products that are not available for many
sensors. Furthermore, many cloud pixels found via cloud shadows are fractional clouds mixed with land or shadows; as a result, the brightness temperature of the cloud pixels is reduced, although it remains close to that of the nearby land or shadows. This study did not seek out a method for accurately estimating cloud height. Instead, we recorded the temperature of the shadow or water pixel as the bottom temperature of atmospheric temperature profile. We also recorded the brightness temperature of the found cloud pixel or the lowest temperature of the clouds in a nearby 32x32 box when the found cloud pixel was considered as a fractional pixel, as the cloud-top temperature. Finally, we applied the recordings in a standard atmosphere temperature profile to estimate cloud height. Because estimating cloud height with this method only provides an approximate result, these data lie with an error range from -2 km~2 km.

3.2.2.3 Iterative method for determining cloud shadows

If cloud thickness (height between cloud base and cloud top) is ignored, then one cloud pixel has, at most, one cloud-shadow pixel on the ground or over another lower cloud because the vertical sunlight vector (but not the horizontal vector) tends to reduce and augment the size of the shadow on the ground. If one assumes that a small portion of a cloud has the same height at each point (i.e., level cloud sections), then there should be a one-to-one relationship between the cloud and its shadow. Thus, if the estimated cloud height is lower or higher than the cloud’s actual height, then some shadow points might not have corresponding cloud points, and some cloud points might fail to locate their corresponding shadow points. For example, three neighboring Cloud Points A, B and C
have a height $h$. Points A, B and C cast their shadow at Points $S_A$, $S_B$ and $S_C$ (Fig. 16 Left). When the cloud height is estimated to be $\Delta h$ less than the actual height (Fig. 16 Middle), Cloud-Shadow Points A and B are searched for in Points $S_B$ and $S_C$, leaving Shadow Point $S_A$ without a corresponding cloud point; thus, Cloud Point C has no shadow. When the cloud height is estimated to be $\Delta h$ higher than the actual height (Fig. 16 Right), Cloud-Shadow Points B and C are searched for in Points $S_A$ and $S_B$, leaving Cloud Point A with no shadow and Shadow Point $S_C$ without cloud. Only when cloud height is estimated accurately can this one-to-one relationship be established for a group of neighboring cloud or cloud-shadow pixels.

Based on the above concept, an iterative method is used to determine cloud shadows by estimating cloud heights until the chosen shadow pixels correspond to cloud pixels. For a presumptive cloud shadow pixel, the cloud height is estimated from 0.5 km to 14 km until a cloud pixel is found with the geometric relationship from Equations (3-8) through (3-12) over a spherical surface. The brightness temperature of this cloud pixel is subsequently used to calculate a more accurate cloud height using the method described in Section 3.2.2.2. Because the cloud height is an approximation, a new cloud-height iteration of 2 km above and below the current cloud height is performed. In this new cloud-height iteration, the found cloud pixel is used to predict a cloud-shadow pixel with the geometric relationship. If a shadow pixel is located, then the cloud height is applied to the nearby cloud and cloud-shadow pixels to test their one-to-one relationship. If all of the cloud and cloud-shadow pixels in the cloud-height iteration have corresponding cloud-shadow and cloud pixels, respectively, then the cloud height was estimated.
correctly. Thus, all of the cloud-shadow pixels determined by the cloud-height iteration are flagged as cloud shadows, and the cloud-height iteration terminates. Alternatively, if this one-to-one relationship is not established before the iteration is terminated, then the presumptive cloud-shadow pixel is flagged as water. In this manner, all of the “water pixels” in a water map are divided into water and cloud-shadow pixels.

Figure 16 Cloud and cloud shadow inter-determination with multiple points Left: normal cloud height; Middle: lower cloud height; Right: higher cloud height

### 3.3 Results

The proposed method was applied to the MSG/SEVIRI data because it provides continuous observations every 30 minutes, which is ideal to observe the change of cloud shadows. Longitude, latitude, surface altitude, sensor zenith angle and sensor azimuth angle were recorded from a static dataset. The solar zenith and solar azimuth angles were calculated in a general manner. SEVIRI visible, near-infrared and shortwave-infrared channels obtained water maps using a decision-tree method (Friedl et al., 1997; Dale et al., 2007). To keep water pixels at the maximum extent in a water map, the water pixels
with large negative NDVI values were flagged as water and not used to predict cloud shadows. A geo-location method for SEVIRI transforms the geodetic coordinates (i.e., longitude and latitude) to SEVIRI imagery coordinates (columns and rows) when the geodetic coordinates of cloud and cloud-shadow pixels are calculated using Equations (3-8) through (3-12).

This method was tested using ten-day SEVIRI data from 07:30 to 15:00 (UTC). Because there were no similar cloud-shadow products available from SEVIRI, the results were compared with SEVIRI false-color images via visual inspection. Figures 17, 18 and 19 show a series of MSG/SEVIRI false-color images (Left) in the Southern Africa with their corresponding water maps (Right) after cloud-shadow removal in the morning (Fig. 17, 07:30 to 10:00), at noon (Fig. 18, 11:00 to 13:00), and in the afternoon (Fig. 19, 14:00 to 15:00) on the 68th Julian day of 2010.
Figure 17 MSG/SEVIRI false color image series (Left) and the corresponding water maps (Right) in the morning from 07:30 to 10:00 (UTC) on the 68th Julian day in 2010. Top: 07:30, Middle: 08:30, Bottom: 10:00.
In Fig. 17, false-color images (Left) display shadows, which appear dark due to their low reflectance, distributed with the clouds from 07:30 to 10:00 (UTC). When the decision-tree method was used to detect water, these shadows were mistakenly detected as water pixels; thus, the precision of water detection is reduced. After using the method proposed in this paper, most of the shadows that were assumed to be water pixels were separated from the water bodies and show a consistent time series with the solar zenith and solar azimuth angles. Because the sun is in the east in the morning (Fig. 17, Top: 07:30; Middle: 08:30; Bottom: 10:00), shadows are cast from east to west. Shadows at 07:30 (when the largest solar zenith angle occurs) are the longest, and those at 10:00 are the smallest.

Because the solar zenith angle is at its smallest from 11:00 to 13:00 in southern Africa (Fig. 18, Top: 11:00; Middle: 12:00; Bottom: 13:00), the shadows are also at their smallest with few shadow pixels seen in these images.
Figure 18 MSG/SEVIRI false color image series (Left) and the corresponding water maps (Right) at noon from 11:00 to 13:00 (UTC) on the 68th Julian day in 2010. Top: 11:00, Middle: 12:00, Bottom: 13:00.
The shadows reappear when the sun moves to the west in the afternoon (Fig. 19, Top: 14:00; Bottom: 15:00). Unlike the shadows in the morning, however, the shadows in the afternoon are cast from west to east. More shadows pixels are detected as the solar zenith angle increases.

Figure 19 MSG/SEVIRI false color image series (Left) and the corresponding water maps (Right) in the afternoon from 14:00 to 15:00 (UTC) on the 68th Julian day in 2010. Top: 14:00, Bottom: 15:00.
This is a stable method for distinguishing cloud shadows from water pixels in different and complicated situations. When cloud shadows exist close to water bodies (the circled areas of Fig. 20 (a) and (b)), the decision-tree method classifies many land pixels as water bodies due to the low reflectance at the large solar zenith angle. However, the proposed method of removing cloud shadows does not consider all water pixels to be cloud shadows, although there are large clouds nearby. Point A in Fig. 20 (a) and (b) refers to a small lake surrounded by clouds; however, the cloud shadow pixels are detected along the clouds, allowing the water body to be represented as water pixels. Points B, C and D in Fig. 20 (c) and (d) show similar results with point A. The method also separates most of the shadow pixels from the water pixels both in the nadir area and at the edge of the SEVIRI imagery. In the false-color image, Fig. 20 (e), the elliptical area shows cloud shadows along the clouds, whereas most of the shadow pixels are detected in the elliptical area of the water map (Fig. 20 (f)).
Figure 20 Sample SEVIRI false color images and water maps with complicated situations on the 68th Julian day in 2010: (a), (b), and (c), (d): Over water and shadow mixing area, 2010068 09:00 (UTC); (e) and (f): Over the disk edge area, 2010068 11:00 (UTC)

However, some mistakes occurred in several typical situations. The first situation occurred when snow/ice mixes with water bodies because the brightness temperature of snow/ice is less than the nearby water bodies, and the reflectance of snow/ice in visible, near-infrared and shortwave-infrared channels is similar to that of clouds. Thus, snow/ice
is easy to mistake for clouds using the method for determining clouds described in Section 3.2.2.1. Thus, water pixels, along with snow/ice, are sometimes mistaken as cloud shadows when the difference between the brightness temperatures of water and snow/ice is large. As shown in Fig. 21 (a) and (b), the decision-tree method allows water pixels to be detected due to snow melting and wet surface conditions at the snow edges of areas B, C and D. When the cloud-shadow removal method is applied, some water pixels along the snow are detected as cloud shadows. The second situation involves edges of land and sea/lakes because the thermal capacity of water is larger than that of land; thus, the sea surface temperature decreases more slowly than the land surface temperature. With a large solar zenith angle (either in the early morning or late afternoon), the reflectance of land is much higher than that of the sea in land-sea edge areas because some land pixels are misclassified as clouds; thus, some water pixels are mistakenly detected as cloud shadows along the edge. In Fig. 21 (a) and (b), Areas A and E show the mistakes in this situation. Furthermore, although the iterative method shows promising results with regard to determining a one-to-one relationship between clouds and cloud shadows, not all of the shadow pixels are detected. Some water pixels remain undetected due to the failure to search for cloud pixels during the iteration. Although most cloud shadows were detected in the elliptical areas in Fig. 21 (c) and (d), some shadow pixels are still mistakenly as water pixels.
Figure 21 Sample SEVIRI false color images and water maps with two mistaken situations on the 68th Julian day in 2010: (a) and (b): snow/ice or land/sea edges, 2010068 09:30 (UTC); (c) and (d): undetected shadow pixels, 2010068 15:00 (UTC)

3.4 Discussion and summary

Although there were some mistakes, the results show that the proposed method effectively and automatically removes cloud shadows from a water map; thus, it can improve the precision of water products in satellite imagery. Because the method does not use any cloud products (including cloud mask), it can be applied when cloud products are unavailable.

Compared with the methods of Luo (2008) and Hutchison (2008), the methodology described in this chapter has some advantages in discriminating cloud shadows from water bodies. Luo only used geometric relationships over a flat surface to obtain possible cloud shadow areas. Subsequently, the reflectance thresholds of multiple channels, but especially the blue channel, were used to detect cloud shadows. Due to the spectral similarities of cloud shadows and mixed water bodies, this method may be difficult for
extracting shadow information from water pixels. Hutchison’s method is a logical searching method that focuses on cloud edges by comparing the reflectance of channels to nearby clear-sky land pixels in a 25x25 box. Similar to Luo’s method, Hutchison’s method primarily determines cloud shadows based on spectral properties. When cloud shadows mix with water bodies (e.g., during floods), the shadowed area may be enlarged because many water pixels are detected as shadows using spectral methods or image fusion or image difference methods. Furthermore, these methods significantly depend upon cloud products and easily inherit errors caused by these products. They may also be difficult to apply when cloud products are unavailable.

In comparison, the method presented in this chapter does not attempt to remove shadows by comparing the spectral properties of cloud shadows and water bodies. Instead, it uses geometry to distinguish these spectrally similar objects. The spherical geometric relationship between shadows and clouds is established and iteratively applied to construct the one-to-one relationship based on the assumption that one cloud pixel casts, at most, one cloud-shadow pixel. The spectral difference between water/shadow and clouds determines whether a non-water pixel is a cloud. This simple spectral difference may result in mistakes at snow/ice and land-sea edges; however, because there is no need to use a cloud mask, this method may be used with satellite data when a cloud mask is unavailable and of course does not inherit cloud-product-caused errors.

Assuming that one cloud pixel casts, at most, one shadow pixel, the iterations may be terminated before all of the shadow pixels are examined when cloud thickness cannot be ignored because there is little distortion in the horizontal direction during light
illumination. If cloud thickness cannot be ignored, then one cloud pixel might cast more
than one shadow pixel when the solar zenith angle is large; thus, a one-to-one
relationship cannot be established, and the iteration method might produce a confusing
cloud height in which some shadow pixels remain undetected. Alternatively, one may be
unable to obtain the corresponding cloud pixel given the cloud height. Furthermore, the
assumption that neighboring cloud pixels have the same cloud height at each point might
also cause confusion when several types of clouds mix with each other.

In a conclusion, most of the cloud shadows in a water map can be removed
effectively using the method described in this chapter. This method might be an effective
way to remove cloud shadows from water bodies, thereby improving the precision of
automatic water products derived from satellite imagery.

This chapter can be summarized as follows:

1. When a satellite views floating objects during the derivation of the
geometric relationship between clouds and cloud shadows, parallax must be
considered. Such parallax can be described as the function of satellite geometry and
an object’s height using a method similar to shadow casting.

2. The geometric relationship of clouds to cloud shadows over an ideal flat
surface may be used on the nadir area of satellite imagery; however, the curvature
and distortion of Earth cannot be ignored outside of the nadir area. Thus, a
spherical geometric relationship must be applied.

3. Based on the spherical geometric relationship, cloud shadows in a water
map can be determined using an iterative method based on the one-to-one
cloud/cloud-shadow relationship. During these iterations, cloud height is estimated until a one-to-one relationship is established. Although this iterative method may generate mistakes at snow/ice edges and land/sea edges, it successfully detects out most cloud shadows.

4. The proposed method does not use cloud products, even including a cloud mask. Thus, this method does not inherit any cloud-product-caused errors and can be applied to other satellite imagery when cloud products are unavailable.

5. The method removes cloud shadows from a water map. That means only after water pixels are identified can this method identify cloud shadows. In other word, it ignores cloud shadows that do not appear on a water map; thus, it cannot identify all of the cloud shadows in satellite imagery.
CHAPTER 4 Water Fraction Retrieval with Optical Satellite Data

4.1 Introduction

Because the spatial resolution of MODIS and VIIRS data are from 250m to 500m, many water pixels extracted from these data are mixing ones containing some combination of vegetation, bare land, and water. Therefore, water fraction retrieval is a necessary step to obtain water area information. In this chapter, we propose a quantitative method to use the SWIR channel (1.64 μm, MODIS channel 6, VIIRS Imager band 3) to calculate water fractions based on a multispectral linear mixture approach.

4.2 Data used

The method is developed based on MODIS data and validated by Landsat TM data. The data used in this chapter are listed as follows:

- MODIS L1B calibrated radiance or TOA reflectance data at 1 km resolution (MOD021KM). MOD02HKM data for land surface classification, water type classification, and water fraction calculation;
- MODIS geo-location data (MOD03);
- TM data from Landsat observations at 30-meter spatial resolution as the principal data to evaluate water fraction derived from MODIS.
4.3 Methodology

4.3.1 Spectral characteristics of water and land

The surface of the Earth is generally made up of different types of vegetation, bare land, and bodies of water. Different land types have different reflectance in the visible to short-wave infrared channels. As shown in Fig. 22 (left), vegetation has low reflectance in the visible wavelength range, a much higher reflectance in the NIR band, and moderate reflectance in the short wave infrared (SWIR) band. The reflectance of dry bare land is higher in the SWIR band than in the NIR and VIS bands, whereas the reflectance of water is close to 0 in the SWIR band.

Although vegetation, bare land, and water bodies each have associated general spectral characteristics, the optical properties may be different for specific types within each of these three main surface categories. For example, the reflectance of vegetation types, such as soybean, pine tree, grass, and wheat, are usually larger in the NIR band than in the VIS or SWIR band (Fig. 22 (left)). In addition, turbid lake water is more reflective than moderately turbid lake water.

Besides the channels’ reflectance, reflectance ratios among NIR, VIS and SWIR bands may also be related to land cover types (Fig.22 (right)). Similar to normalized difference vegetation index (NDVI), which is one of effective parameters to differentiate bare land and vegetation, reflectance ratios between VIS, NIR or SWIR channels, such as $\frac{R_{NIR}}{R_{SWIR}}$, $\frac{R_{VIS}}{R_{SWIR}}$ and $\frac{R_{NIR}}{R_{VIS}}$, may also characterize land cover types and thus can be used to
estimate the mixture of land types. In a mixed water pixel, reflectance includes contributions from both water and land. To estimate water fraction based on a multispectral linear combination model, pure land and water reflectance should be calculated first. The traditional method is to use histograms to take average reflectance of all the pure land and water pixels nearby to get the pure land reflectance and pure water reflectance respectively. As shown in Fig.22, however, different surface types, such as bare land and vegetation, reflect differently. Taking average to all nearby land pixels may introduce errors, especially when the mixture of land types is complicated. Therefore, a new way should be found to obtain robust results for water fraction calculation. Since the reflectance ratios between the two different channels’ reflectance suggest the land cover type information and their mixture, they can be used to search for land pixels with similar land types to the sub-pixel land part of a mixed water pixel.

Figure 22 The curves of reflectance for different land types (left) and ratio of spectral reflectance to SWIR (MODIS CH6) channel (right) (Spectral data is provided by
4.3.2 Derivation of the water fraction from a multispectral linear mixture model

The theory of multispectral linear mixture for mixed pixels has been proven by many researchers, such as Sheng et al, DeFries et al., and Jiang et al. Based on the theory the reflectance of a mixed water pixel combined with different land types in the visible to the SWIR channels can be expressed as:

$$ R = \sum_{i=1}^{n} f_i \cdot R_i $$

Or

$$ R_{ch_{mix}} = (1 - f_w) \cdot R_{ch_{land}} + f_w \cdot R_{ch_{water}} $$

Equation (4-1), $f_w$ represents the fraction of water, $R_{ch_{mix}}$ is the observed reflectance of a mixed pixel (comprised of water and land), $R_{ch_{land}}$ is the reflectance of a land pixel with similar mixture of land types to the land component of the mixed pixel in the visible, near infrared or short-wave infrared channels, and $R_{ch_{water}}$ is the reflectance of a water pixel with similar water type to the water component of the mixed pixel in each channel.

In principle, spectral reflectance in any visible, near infrared (NIR), or shortwave infrared (SWIR) channels can be used to solve Equation (4-1) to derive water fraction $f_w$:

$$ f_w = \frac{R_{ch_{land}} - R_{ch_{mix}}}{R_{ch_{land}} - R_{ch_{water}}} $$

(4-2)
By analyzing inland water characteristics, the optical properties of water in the SWIR channel, such as the MODIS channel 6, are much less variable than those in the visible and near infrared channels, and thus this channel may be applicable for water fraction calculation. In the visible channel, the reflectance of water is close to that of land and sometimes even higher than that of some land types (Fig. 22 (a)). The reflectance in this channel increases significantly with the increase of suspended matter. In the NIR channel, the reflectance of water is low and less than that in the visible channel. The reflectance in this band is also largely affected by water types as turbid water has a higher reflectance than clean water. When water is contaminated by blue-green algae, the reflectance increases significantly and is much higher than that in visible channel.

Water pixels (pure and mixed) which are extracted with the decision tree approach presented in chapter 2 and differentiated from cloud shadow by the geometric method in chapter 3, are further classified into clean water, moderately turbid water, turbid water, and blue-green algae contaminated water. For each water type, pure water pixels are identified by threshold values obtained from training with a decision tree method according to channel 6 reflectance and assigned with a water fraction value of 1.0. For a mixed water pixel, the determination of $R_{ch6_{-land}}$ and $R_{ch6_{-water}}$ becomes key in order to calculate $f_w$ from Equation (4-2).
4.3.3 Dynamic nearest neighbor searching (DNNS) method

From spectral data analysis, as shown in Fig. 22, in the SWIR channel, the reflectance of clean water and moderately turbid lake water is less than 0.5%. Although the increase of turbidity leads to slight increment in water reflectance, compared to the reflectivity of vegetation and bare land, the reflectance of water in this channel can still be ignored. Because reflected radiance of water in SWIR band (e.g., MODIS channel 6) is insignificant, when the percent surface water in a mixed water pixel is not too high, the reflectance contribution of the water portion in MODIS channel 6 \( (f_w \cdot R_{ch6\_water}) \) is also insignificant. Therefore, the reflectance of a mixed water pixel in this channel is mainly from land with a total fraction of \( (1 - f_w) \). Fig. 22 Right) further demonstrates that the spectral reflectance ratios between VIS (ch1) or NIR (ch2) regions and the SWIR channel (ch6) may be related to surface types. If the reflectance of a surface type is fixed, then VIS or NIR to SWIR reflectance ratios are affected by the proportions of surface types. For SWIR channels, Equation (4-1) can be approximately simplified to Equation (4-5).

\[
R_{ch1\_mix} = (1 - f_w) \cdot R_{ch1\_land} + f_w \cdot R_{ch1\_water} \quad (4-3)
\]

\[
R_{ch2\_mix} = (1 - f_w) \cdot R_{ch2\_land} + f_w \cdot R_{ch2\_water} \quad (4-4)
\]

\[
R_{ch6\_mix} \approx (1 - f_w) \cdot R_{ch6\_land} \quad (4-5)
\]

The ratios of channel 1 and 2 to channel 6 can be written as:

\[
\frac{R_{ch1\_mix}}{R_{ch6\_mix}} \approx \frac{R_{ch1\_land}}{R_{ch6\_land}} + f_w \cdot \frac{R_{ch1\_water}}{R_{ch6\_mix}} \quad (4-6)
\]
\[
\frac{R_{ch2_{mix}}}{R_{ch6_{mix}}} \approx \frac{R_{ch2_{land}}}{R_{ch6_{land}}} + f_w \frac{R_{ch2_{water}}}{R_{ch6_{mix}}}
\]

(4-7)

By transmutation, Equations (4-6) and (4-7) can be written as:

\[
\frac{R_{ch1_{land}}}{R_{ch6_{land}}} \approx \frac{R_{ch1_{mix}}}{R_{ch6_{mix}}} - f_w \frac{R_{ch1_{water}}}{R_{ch6_{mix}}}
\]

(4-8)

\[
\frac{R_{ch2_{land}}}{R_{ch6_{land}}} \approx \frac{R_{ch2_{mix}}}{R_{ch6_{mix}}} - f_w \frac{R_{ch2_{water}}}{R_{ch6_{mix}}}
\]

(4-9)

For a mixed pixel \( f_w \) should have a range of \( 0.0 < f_w < 1.0 \), and thus the ranges of the ratios of channel 1 and 2 to channel 6 have the following relationships:

\[
\frac{R_{ch1_{mix}}}{R_{ch6_{mix}}} - \frac{R_{ch1_{water}}}{R_{ch6_{mix}}} < \frac{R_{ch1_{land}}}{R_{ch6_{land}}} < \frac{R_{ch1_{mix}}}{R_{ch6_{mix}}}
\]

(4-10)

\[
\frac{R_{ch2_{mix}}}{R_{ch6_{mix}}} - \frac{R_{ch2_{water}}}{R_{ch6_{mix}}} < \frac{R_{ch2_{land}}}{R_{ch6_{land}}} < \frac{R_{ch2_{mix}}}{R_{ch6_{mix}}}
\]

(4-11)

The relationship described in Equations (4-10) and (4-11) provides the basis to find the nearby pure land pixels. This method is defined as a dynamic nearest neighbor searching (DNNS) method in this study. All the nearby land pixels found by this DNNS method are aggregated to calculate the average land reflectance \( R_{ch\_land} \). For a mixed water pixel, the land pixels that satisfy Equations (4-10) and (4-11) are searched in a dynamic window (for example, 100×100 pixels) centered on each mixed pixel. The average or median reflectance of all these pixels found by this way is taken as \( R_{ch\_land} \). The nearest pure water pixels with each identified water type are also located in the loop and the average or median channel reflectance of all the found water pixels is used as the
reflectance of pure water $R_{ch\_water}$. Once pure water and land-only reflectance values are calculated, water fraction can be calculated from Equation (4-2). As shown in Fig.23, land and water pixels are classified initially by using the decision tree method. Water pixels are then further classified as clean water, moderately turbid water, turbid water, and vegetation contaminated water. The threshold method in the SWIR channel for each water type is used to define pure water pixels and to assign a water fraction of 1.0. For the mixed water pixels, Equation (4-2) is used to estimate the water fractions for mixed water pixels with water types other than pure water. Once we get the land and pure water type classification information, we know the values of $R_{ch\_land}$ over the land pixels and $R_{ch\_water}$ over the pure water pixels. There may be many land pixels around a mixed water pixel, so Equations (4-10) and (4-11) are then used to determine the appropriate land pixels to get the right $R_{ch\_land}$ value for water fraction estimation for a mixed water pixel with Equation (4-2).

When water is the main portion in a mixed water pixel (for example, $f_w$ is larger than 0.9), the reflectance from the sub-pixel land portion in channel 6 is close to that of the sub-pixel water portion for the low fraction of land. In this case, compared to the decreased reflectance of the land portion, the reflectance of the water portion in this channel cannot be ignored and Equations from (4-5) to (4-11) may not be applicable. Furthermore, even if Equations from (4-5) to (4-11) are effective, the ratios of channels 1 and 2 to channel 6 are so large that no land pixels can meet the conditions of Equations (4-10) and (4-11). In this situation, the total reflectance of a mixed water pixel in channel
6 is also very low and close to that of a pure water pixel. Therefore, Equation (4-2) can also be used to calculate the water fraction because the observed reflectance of a mixed pixel $R_{\text{ch, mix}}$ will be mainly from the contribution from water ($f_w \times R_{\text{ch, water}}$), whereas the effect of land (($1 - f_w) \times R_{\text{ch, land}}$) will be insignificant. Under this condition

$$R_{\text{ch, land}} - R_{\text{ch, mix}} \approx R_{\text{ch, land}} - R_{\text{ch, water}}$$

is very close to $R_{\text{ch, land}} - R_{\text{ch, water}}$ and Equation (4-2) is much less relevant to $R_{\text{ch, land}}$, which means that any pure land pixel nearby can be used. Therefore, as shown in Fig. 23, the average reflectance of nearby pure land pixels in channel 6 will be used simply as $R_{\text{ch, land}}$ for water fraction calculation from Equation (4-2).
Figure 23 Algorithm flow chart of water fraction calculation with MODIS data
4.4 Results and validation

4.4.1 Results from the EOS/MODIS observations

The proposed method is applied to EOS/MODIS data at 500m resolution. Decision-tree approach described in chapter 2 is used to identify all possible water pixels including cloud shadow, and then the geometric method presented in chapter 3 is employed to remove cloud shadow pixels from water detection result. Water pixels are further classified into different water types. For each water type, pure water pixels are determined with threshold method and are assigned a water fraction 1.0. Water fraction is then calculated for all the rest mixed water pixels with the method described in this chapter.

More than 100 MODIS granules have been tested and show promising results. During the 2007-2008 rainy season, heavy rainfall brought floods to the Zambezi River in Southern Africa on February 4, 2008. Fig. 24 (a) shows a three-channel (ch6, ch2, and ch1) false color composite image of MODIS in Southern Africa at 07:55 (GMT) on February 4, 2008. The swollen Zambezi River runs wide over its banks in the image. Fig. 24(b) shows the derived surface classification image, in which vegetation is shown in green, bare land is in deep yellow, inland water is in blue, clouds or snow are in white, cloud shadow are in black, and sea water is in deep blue. Fig. 24 (c) and (d) shows a subset MODIS false color composite image of Fig. 24 and the derived water fraction map.
Figure 24  (a) MODIS Swath False color composite image (2008035 07:55GMT); (b) Image of classification result of (a); (c) A subset image of (a); (d) Water fraction map of (c)

4.4.2 Evaluations using Landsat/TM observations

Because there are no ground observations on water fractions, evaluation of water fraction derived from coarse-to-moderate resolution satellite data are challenging. The
utilization of higher resolution satellite data such as TM, ETM, and ASTER is a feasible way to solve this problem. In this chapter, TM data (p025r025 on June 5th, 2001) with 30m spatial resolution are used to evaluate water fraction estimates from MODIS observations on the same day. There are two evaluation methods. One is to compare water fractions in a pixel-to-pixel way between TM and MODIS data by simulating a MODIS image to spatially match with TM data. The other method is to compare the areas of the same lakes observed from TM and MODIS respectively.

4.4.2.1 Pixel-to-pixel comparison between TM and simulated MODIS data

4.4.2.1.1 Spatial matching between TM and MODIS

In order to perform a pixel-to-pixel comparison between TM and MODIS results, water pixels in the TM data with 30m resolution are detected and each water pixel in a TM image is assumed to be a pure water pixel with water fraction of 1.0. The water percentage in a 500m×500m grid is calculated to match the 500m resolution MODIS image and is then assumed to be the true grid water fraction for validation.

However, when directly applying TM data to evaluate water fractions derived from MODIS, two major problems need to be resolved. First, the geo-location accuracy between the TM and MODIS imagery is different. TM has a higher geo-location precision than MODIS. This makes the two kinds of imagery not match when re-sampling TM imagery to MODIS resolution. The other problem is that the viewing angles and imaging models between the two sensors are different. This results in different
shapes for the same scanned objects in the two kinds of satellite imagery. The difference in geo-location, viewing angle, and imaging model between the two sensors is always the bottleneck when using high resolution (i.e. Landsat-class) satellite data for evaluating coarse-to-medium resolution satellite products.

Based on the theory of multispectral linear mixture for mixed pixels described in Section 4.3, the reflectance of mixed pixels can be simulated when the fraction of each type of mixture is known. TM data has 30m resolution and thus most pixels can be assumed to be composed of three end members: certain type of vegetation, bare soil, and water. Using a decision-tree approach to perform the classification on TM data, fractions of different land types in a MODIS 500m grid can be calculated. To get the channel reflectance of each land type of MODIS, pixels in original MODIS data with a fraction of 1.0 for each type are clustered to calculate the average channel reflectance. Combining with fraction from TM and channel reflectance from MODIS of each land type in a 500m resolution pixel, simulated MODIS reflectance data for each pixel can be obtained by Equations (4-1) and (4-2). The simulated MODIS data both spatially matches with TM data and spectrally matches with MODIS data and can be a good data source to recalculate water fraction through the process shown in Fig. 23 and launch pixel-to-pixel validation with TM data instead of original MODIS data.

With the simulation method described above, sample data from TM and MODIS are applied to get a simulated MODIS image. Fig. 25 shows the simulated results. Fig. 25 (1) is an original three-channel false color composite image of MODIS and Fig. 25 (2) is the corresponding simulated false color composite image of MODIS simulated from TM.
Comparing Fig. 25 (1) with Fig. 25 (2), the simulated MODIS image looks very similar to the original MODIS image. Fig. 25 (3) from (a) to (c) are the subset images of MODIS, TM (aggregated in 500m resolution by nearest neighboring method from 30m resolution), and simulated MODIS respectively. In these three images, points A, B, and C are three small lakes. In the original MODIS image, the shapes of the three lakes are different from those in the TM image. This is mainly due to difference from geo-location, viewing angles and imaging model between the satellites. In contrast, the shapes of the three lakes in the simulated MODIS image are very similar to those in the TM image, indicating that the two images are spatially matched. Fig. 25 (3) from (d) to (i) are the corresponding channels’ reflectance histogram plots of the original MODIS data and the simulated MODIS data. Although there are small differences between the plots, analyzing the ranges of the reflectance in the VIS, NIR, and SWIR channels and the shapes reveal markedly similar plots. These histograms show that the simulated MODIS data match the original MODIS data in spectral reflectance. Therefore, the water fraction calculation method described in this chapter is applied to the simulated MODIS data and the calculated result is used for pixel-to-pixel comparison to that of TM.
Figure 25 (1) MODIS false color composite image (17:10(UT) on June 5th, 2001), (2) Simulated MODIS false color composite image (3) Subsets images and Histogram plots of each band for MODIS and simulated MODIS data: (a) Histogram of MODIS band 6;
4.4.2.1.2 Evaluation and analysis

By using the simulated MODIS data instead of the original MODIS data, water and land are first classified with a decision tree method. The water is then further classified into several water types: clean water, moderately turbid water, turbid water, and so on. Because the decision tree approach used in this study is limited to detect mixed pixels with very low water fractions as water pixels, pixels with water fractions less than 0.18 are regarded as land pixels and are not used to calculate water fraction. All the identified water pixels including pure and mixed water pixels are classified into different water types by decision tree approach as well. Pure water pixels in each water type are identified with a simple multiple-threshold method and then assigned a water fraction of 1.0. For the rest water pixels (mixed water pixels), which consist of 17,010 pixels in this case, the DNNS method described in Section 4.3 is used to calculate water fraction, and the result (Fig. 26 (a)) is compared with the true water fraction observations obtained from the TM 30m resolution image (Fig. 26 (b)). Fig. 27 (a) shows the scatter plot of water fractions between TM and MODIS using this method. From Fig. 27 (a), the two water fractions are close with a regression coefficient of 1.0347, an intercept of -4.0481, and correlation coefficient of 0.9812. The result is quite promising.
To show the advantage of the dynamic nearest neighbor searching (DNNS) method using SWIR channel, other methods, including the DNNS method using the NIR channel, and the histogram method using both SWIR and NIR channels, are also applied to
calculate water fractions for comparison. The results indicate that DNNS method with the SWIR method yields the best results with the lowest average standard deviation of the water fraction difference, the highest correlation coefficient, and the largest percentage (more than 96%) of samples with a water fraction error less than 0.1 (Table 5). Although the histogram method (Fig. 27 (b)) shows good results for most of the pixels, there are some fatal errors in pixels shown in the red circle. Moreover, pixels with water fraction differences larger than 0.2 between MODIS and TM data from the histogram method are far more than the cases using the DNNS method with the SWIR channel. Results of the NIR channel from either the DNNS method or the histogram method (Table 5) show less accurate results than that of the SWIR channel. In the DNNS method using the NIR channel, only about 85% of samples have a water fraction error less than 0.1. In contrast, in the histogram method, although there are fewer samples with water fraction difference beyond 0.3, the sample number with water fraction difference beyond 0.1 is much larger than that of SWIR channel.

Table 5 Analysis on water fraction of TM and simulated MODIS using different methods

<table>
<thead>
<tr>
<th>Method</th>
<th>DNNS</th>
<th></th>
<th>Histogram</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SWIR</td>
<td>NIR</td>
<td>SWIR</td>
<td>NIR</td>
</tr>
<tr>
<td>Comparison</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ</td>
<td>0.034</td>
<td>0.037</td>
<td>0.037</td>
<td>0.039</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.981</td>
<td>0.981</td>
<td>0.976</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>total</td>
<td>17010</td>
<td>17010</td>
<td>17010</td>
<td>17010</td>
</tr>
</tbody>
</table>

Annotation:

DF: Water fraction of MODIS minuses Water fraction of TM

|DF|: absolute value of DF

Figure 27 (a) Scatter plot of water fractions between TM and MODIS using DNNS method; (b) Scatter plot of water fractions between TM and MODIS using histogram method
From these evaluation results, we can see that using the DNNS method with SWIR channel shows the best and most reliable result among all methods investigated. In the histogram method, land types are not considered and therefore complicated underlying conditions may lead to severe errors in the calculation of $R_{ch\_\text{land}}$. For example, when the sub-pixel land portion in a mixed water pixel is combined only with bare land, but most of the pixels nearby are made up of vegetation, all the vegetation pixels nearby will be used to get $R_{ch\_\text{land}}$ in the histogram method. Because vegetation and bare land reflect differently in the SWIR channel, the histogram method will calculate $R_{ch\_\text{land}}$ with a much larger error than the DNNS method, and thus leads to large errors in the water fraction. For the DNNS method, by implementing Equations (4-10) and (4-11), only bare land pixels similar to the sub-pixel land portion in the mixed water pixel are selected to derive $R_{ch\_\text{land}}$ and can lead to a reasonable result and thus show a reliable and more accurate solution. The NIR channel is easily contaminated by suspended matter and other hydrophytic vegetation, such as algae. The reflectance in NIR channel over an inland water surface is not stable and may decrease the water fraction accuracy. Compared with the NIR channel, the SWIR channel is barely affected by water types and water depth and can get stable reflectance over water surface, thus yielding less error in the pure water reflectance calculation. The characteristics of the SWIR channel with less variation in reflectance with water types may make it a better channel for water fraction calculation from the linear mixture model than the NIR channel.
4.4.2.2 Validation with observed Lake Areas

Observed lake areas can be another way to validate water fractions from coarse-to-medium resolution satellite data. A lake’s area $S$ can be calculated by Equation (4-12):

$$S = \sum_{i=1}^{n} f_{wi} \cdot s_i,$$

(4-12)

In Equation (4-12), $f_{wi}$ is water fraction of pixel $i$ and $s_i$ is the area of pixel $i$.

From Equation (4-12), for a lake, since $s_i$ is fixed for each pixel, $f_{wi}$ is the only variable for each pixel to calculate $S$. That is to say, the more accurate $f_{wi}$ is, the higher precision $S$ can be derived.

Lake areas from MODIS data calculated by water fractions with different methods are compared to those from TM data which can be assumed as true observations for validation. In this case, Figs. 26 (c) to (f) are the subset images of MODIS false-color image and water fraction maps of TM and MODIS using DNNS method and histogram method respectively. Small lakes (A to G in Figs. 26 (c) to (f)) are chosen to compare the areas from MODIS data with the assumed true ground observations from TM data. As shown in Fig. 26, since pixels with water fraction less than 0.18 are not detected as water pixels in MODIS image there are more water pixels in the TM water fraction map than in the MODIS map. When comparing lake areas between TM and MODIS data, it is hard to identify isolated lakes from TM water maps among all the water pixels. Considering that pixels with water fraction of less than 0.18 contribute little to lake areas in a TM image,
only pixels with water fraction equal to or larger than 0.18 are chosen to calculate lake areas for the TM image (Fig. 26 (d)). In this way, lakes are isolated from the TM water fraction map though the shapes of these lakes are still different from those in the MODIS maps. Areas of lakes A to G in Fig. 26 (d), (e) and (f) are calculated respectively for comparison.

The result, showing a consistent tendency with that of simulation method described in Section 4.4.2.1, suggests again that the DNNS method is more robust than the histogram method. Table 6 shows the specific calculated areas of the seven lakes. From Table 6, for lakes B, C, D and F, the two methods show similar results while for lakes A, E and G, DNNS method performs much better than histogram method. When analyzing these lakes, in lakes B, C, D and F, the sub-pixel portion of mixed water pixels are vegetation and since most of the pixels nearby are covered with vegetation, the histogram method can get accurate results. However, for lakes A, E and G, the sub-pixel portion of some mixed water pixels are bare land. In this situation, the DNNS method aggregates most of the bare land pixels by Equations (4-10) and (4-11) to calculate water fractions for these pixels and then get more accurate results; whereas for the histogram method, the average reflectance of all pixels nearby, whether covered with vegetation or bare land, is used to calculate water fractions. Since most pixels nearby are vegetation, the results inevitably show less accuracy over these mixed water pixels.
Table 6 Comparison of lake areas calculated from MODIS by using DNNS and histogram methods with the SWIR channel

<table>
<thead>
<tr>
<th>Lake</th>
<th>TM (km²)</th>
<th>DNN-S (km²)</th>
<th>HIS (km²)</th>
<th>Error (%) (DNNS)</th>
<th>Error (%) (HIS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.733</td>
<td>0.537</td>
<td>0.241</td>
<td>-26.74</td>
<td>-67.11</td>
</tr>
<tr>
<td>B</td>
<td>3.097</td>
<td>2.410</td>
<td>2.441</td>
<td>-22.19</td>
<td>-21.17</td>
</tr>
<tr>
<td>C</td>
<td>2.039</td>
<td>2.041</td>
<td>2.082</td>
<td>0.10</td>
<td>2.12</td>
</tr>
<tr>
<td>D</td>
<td>1.145</td>
<td>1.098</td>
<td>1.094</td>
<td>-4.11</td>
<td>-4.45</td>
</tr>
<tr>
<td>E</td>
<td>1.261</td>
<td>1.345</td>
<td>1.059</td>
<td>6.63</td>
<td>-16.05</td>
</tr>
<tr>
<td>F</td>
<td>0.752</td>
<td>0.686</td>
<td>0.692</td>
<td>-8.81</td>
<td>-8.03</td>
</tr>
<tr>
<td>G</td>
<td>0.994</td>
<td>0.843</td>
<td>0.775</td>
<td>-15.16</td>
<td>-22.05</td>
</tr>
</tbody>
</table>

4.5 Discussion and Summary

4.5.1 Discussion

By using the short-wave infrared channel (1.64 μm) instead of the visible or near infrared channels, based on the linear mixture theory with DNNS method, water fraction can be calculated with a high accuracy. The method was applied to MODIS data and validated by TM data. The method outlined in this chapter considers mixture of different land and water types and uses the SWIR channel to decrease the uncertainty of reflectance in the visible and near infrared channels caused by the variation of land and water types. The DNNS method is also used instead of traditional histogram method.
considering the mixture of land types within the sub-pixel land portion in a mixed water pixel. Although the validation in different two ways yield promising results, errors and risks caused by some other factors may still exist.

First, the channels’ reflectance proportions $\frac{R_{ch1_{land}}}{R_{ch6_{land}}}$ and $\frac{R_{ch2_{land}}}{R_{ch6_{land}}}$ to search for land-only pixels nearby are restricted in a range shown in Equations (4-10) and (4-11) by assuming that $0 < f_w < 1.0$ rather than assigning an approximate value. Therefore, the assumed range of $f_w$ from 0 to 1.0 makes Equations (4-10) and (4-11) still likely to count in some irrelevant land pixels. If $f_w$ can be approximately predefined by the traditional histogram method, $f_w$ will have a narrower range than from 0 to 1.0 and thus count in less uninvolved pixels. The accuracy of DNNS method may be further improved, especially when water fraction is low.

Another risk lies in that Equations (4-10) and (4-11) are just necessary conditions but not sufficient and must conditions to search for pixels with similar mixture to the sub-pixel land portion in a mixed water pixel. That is to say, the pixels to search should satisfy Equations (4-10) and (4-11), but not all the pixels satisfying Equations (4-10) and (4-11) have similar mixture of land types. In some complicated situations, especially when pixels with water fraction less than 0.18 which are detected as pure land, some pixels with different mixture of land types can be mistakenly counted in and thus decrease the accuracy of DNNS method.

Since the DNNS method with the SWIR channel to calculate water fraction is based
on the result of water detection, the precision of water detection directly affects the result of water fraction. In this study, since the method we use to detect water fails to identify pixels with water fraction less than 0.18 as mixed water pixels, the DNNS method is invalid in water fraction calculation for these pixels as well. Additionally, these pixels, which are incorrectly detected as pure land pixels, affect the accuracy of water fraction from the DNNS method. However, water detection can be very complicated for coarse-to-moderate resolution satellite data over water surfaces contaminated by thin clouds, smokes, aerosols and sun glint and makes the DNNS method with the SWIR channel of limited use in these situations.

Although errors caused by the water type classification can be decreased by using the SWIR channel, the reflectance of pure water in this channel still varies with water type from 0.1% to about 3% and slightly affects the precision. Moreover, thresholds to identify pure water for each water type are difficult to determine accurately, which may also affect water fraction accuracy. The errors could be caused by water detection, by impure land (partially mixed with water) pixels found by the dynamic nearest searching method, and by water reflectance varying slightly with water types in the SWIR channel.

4.5.2 Summary

This chapter introduces the details of the DNNS method with the SWIR channel to derive water fraction from coarse-to-moderate resolution satellite data based on the linear mixture theory of surface reflectance from visible to short-wave infrared channels. The method is evaluated both by pixel-to-pixel comparison with results from high resolution
TM data and traditional lake area comparison. Both the two evaluation means show a consistent and promising result. For the pixel-to-pixel comparison means, more than 96% of the tested water pixels have a water fraction difference or error less than 0.1, a total mean difference of -0.021, and a standard deviation of about 0.0338. For the lakes area comparison way, the DNNS method gets more accurate results than histogram method when land partial of a mixed water pixel is different to the surroundings. Main conclusions of the chapter can be summarized as follows:

1. The short-wave infrared channel (1.64 μm) may be applicable for water fraction calculation. Theoretically, water radiance emitted in the SWIR channel is insignificant and can be ignored. The reflectance in the SWIR channel is also less affected by water types and makes the SWIR reflectance over inland water surface more stable.

2. Comparing with traditional histogram method, the dynamic nearest neighbor searching method with given restrictions is used to dynamically find land (or non-water) pixels nearby a mixed water pixel, to determine the average reflectance for land pixels: $R_{ch\_land}$. Compared with the histogram method, the method outlined here considers the varying sub-pixel land portion in a mixed water pixel and counts pixels with similar mixture ratios to avoid the errors that may often occur in the histogram method, especially under complicated surface conditions.
CHAPTER 5 Microwave-Based Surface Water Fraction Retrieval with Classification and Regression Tree Technique

5.1 Introduction

Because of the limitation of optical satellite on flood detection over cloud cover conditions, microwave-based satellite data instead are utilized to fill the gap where optical satellite fails. As frequencies suitable for water detection in microwave sensors such as AMSR-E and ATMS are with very coarse spatial resolution varying from 12.5km to 15km, water detection with these data always involves water fraction retrieval. Therefore, in this chapter, we developed an algorithm for AMSR-E or ATMS water fraction retrieval by using data around 37GHz, 89GHz for AMSR-E and 50GHz, 85GHz for ATMS. During the algorithm development, regression tree serves as the main approach by considering a series of factors including land cover mixture, precipitation and cloud fraction.

5.2 Data preparation

In this chapter, AMSR-E/AQUA daily global 25 km quarter-degree gridded brightness temperatures in 2011 are the main dataset used for water fraction retrieval (Knowles, 2006). This dataset, which was collected by the University of Colorado, is well projected and spatially matched among frequencies. We chose the 2011AMSR-E data as the main test data because there was a severe river flood covering large areas of
and lasting in the Mississippi River Basin from April to June. These data provides
enough samples at the AMSR-E scale for the development and evaluation of an
algorithm.

In addition to AMSR-E data, a series of data and products from both TERRA/MODIS and AQUA/MODIS was collected. TERRA/MODIS level-1B data over the Mississippi River Basin were used to calculate real-time water fractions at 500m resolution using the Dynamic Nearest Neighbor Searching (DNNS) method described in chapter 4 (Li et al., 2012). The MODIS 500m resolution water fraction results were used with the MOD44W global 250m water mask (Carroll et al., 2009) to collect water fraction samples at AMSR-E 25km resolution. AQUA/MODIS 16-day NIR and VIS reflectance (available in MYD13A) and an AQUA/MODIS daily cloud mask (MYD35) were collected to calculate the vegetation index and cloud fractions at 25km resolution. In addition to the vegetation index and cloud fractions, global tree cover data (MOD44B, version 2005) were also obtained to derive a 25km global tree cover map matching the AMSR-E 25km quarter-degree gridded dataset. Daily 4km precipitation in the USA from NWS/NOAA was downloaded and spatially matched with AMSR-E data to derive 25km precipitation. Global 1km Land Cover data from USGS and global soil type data were resampled by calculating fractions of each type. The land cover and soil types with the largest three fractions were stored in the three layers that were used in this study.

The specific datasets used in this study are listed as follows:

- AMSR-E/AQUA daily global 25 km quarter-degree gridded brightness temperatures at 22GHz, 37GHz and 89GHz;
• TERRA/MODIS L1B calibrated radiance at 500 m resolution (MOD02HKM);
• TERRA/MODIS geo-location and geometric angles (MOD03);
• AQUA/MODIS cloud mask (MYD35);
• AQUA/MODIS 16-day NIR reflectance and VIS reflectance (available in MYD13A);
• TERRA/MODIS 500 m tree cover data (MOD44B);
• USA 4 km precipitation data from National Weather Service (NWS);
• MODIS global 250 m static water maps (MOD44W);
• Global 1 km Land Cover data from USGS Land Cover Characteristics Data Base Version 2.0;
• Global soil type data from Harmonized World Soil Database version 1.2 (HWSD Version 1.2).

5.3 Methodology

5.3.1 Principle theory for water fraction retrieval with passive microwave data

5.3.1.1 Land Surface Radiation Properties around 37 GHz and 89 GHz frequencies

According to pioneering studies, 37 GHz (vertically and horizontally polarized frequency) is a standard frequency to map surface water because it provides balance between spatial resolution and atmospheric interference effects among microwave frequencies (Brakenridge et al., 2007). Variables, including the polarization difference at
$37\text{GHz}\ (\Delta T_{37\text{GHz}})$ and the polarization ratio at $37\ \text{GHz}\ (PR_{37\text{GHz}})$, as well as the polarization ratio between $37\ \text{GHz}$ and $89\ \text{GHz}$ (SWI), which is also known as soil wetness index, are very sensitive to the presence of surface water (Choudhury, 1991; Kerr et al., 1993; Njoku et al., 2003). Here, $\Delta T_{37\text{GHz}}$, $PR_{37\text{GHz}}$ and SWI are defined as follows:

\[
\Delta T_{37\text{GHz}} = T_{V,37\text{GHz}} - T_{H,37\text{GHz}},
\]

\[
PR_{37\text{GHz}} = \frac{T_{V,37\text{GHz}} - T_{H,37\text{GHz}}}{T_{V,37\text{GHz}} + T_{H,37\text{GHz}}}
\]

\[
SWI = \frac{T_{V,37\text{GHz}} - T_{V,89\text{GHz}}}{T_{V,37\text{GHz}} + T_{V,89\text{GHz}}}
\]

where, $T_{V,37\text{GHz}}$ is the vertical polarization at $37\ \text{GHz}$, $T_{H,37\text{GHz}}$ is the horizontal polarization at $37\ \text{GHz}$, and $T_{V,89\text{GHz}}$ is the vertical polarization at $89\ \text{GHz}$ frequencies.

Generally, at $37\ \text{GHz}$, the polarization differences of land surfaces vary from $1\text{K}$ to $80\text{K}$, with water having the highest values, ranging from $20\text{K}$ to $80\text{K}$ depending on salinity and depth. By contrast, forests have the smallest values – approximately $2\text{K}$. Barren land has a higher polarization difference at this frequency than vegetation cover. In particular, among barren land types, sand has the highest polarization difference range ($20$-$40\text{K}$), which is very close to that of water surfaces; thus, the two can be easily confused. The polarization ratio $PR_{37\text{GHz}}$ shows similar distributions for land surface. Although the soil wetness index SWI similarly shows the largest values over water surfaces, it is slightly different from the other two variables because it has the smallest value over deserts. This feature can be used to more easily differentiate deserts from water surfaces. Fig. 28, based on 2011 AMSR-E data, shows the seasonal average
distributions of these three variables over land cover types. In Fig. 28, all the samples were collected from each land cover type 100% at 25km resolution. From Fig. 28, water (land cover type: 16) has the largest values of the three variables. Desert (land cover type: 19) has much higher values in $\Delta T_{37\text{GHz}}$ and $\text{PR}_{37\text{GHz}}$, but the smallest value in SWI. Forest (land cover type: 11-15) has the smallest values in $\Delta T_{37\text{GHz}}$ and $\text{PR}_{37\text{GHz}}$, but slightly larger values in SWI than deserts (land cover type: 19) and grassland (land cover type: 7, 8).

However, most pixels at 25km resolution consist of multiple land cover types, so it is not feasible to solely use land cover types at 25km resolution for water fraction retrieval with microwave data. Since land cover can be represented by various combinations of vegetation types, the variables NDVI (Normalized Difference Vegetation Index) and tree cover, which provide vegetation’s dynamic distribution and fluctuation, can be used instead to quantitatively estimate the variations of the three microwave variables from land cover. The test on NDVI and tree cover shown in Fig. 29 over pure dry land, where there is no precipitation in the last 96 hours and water fractions are close to 0, indicates when NDVI is above 0.2 or tree cover is above 10%, $\Delta T_{37\text{GHz}}$ and $\text{PR}_{37\text{GHz}}$ decrease gradually and SWI increases slowly with the incremental NDVI and tree cover.

The different radiation properties of land surfaces at 37GHz and 89GHz provide the possibility to differentiate water from other land objects.
5.3.1.2 Correlation between water fraction and microwave variables

According to Sippel’s study, the polarization difference at 37GHz ($\Delta T_{37GHz}$) is a linear combination of all land types with respective fractions within any footprint: water, non-inundated land and inundated land (Sippel et al., 1994). Fig. 28 shows water has higher values than other land cover types for every variable, indicating the high sensitivity of microwave radiation to water surface at these frequencies. In other words, a small fraction of water surface within the footprint might change the total values of these three variables substantially, providing the possibility to estimate water fractions from these frequencies at as low as 25-km resolution. The correlation test shown in Fig. 30 proves the feasibility of such estimation. Fig. 30 (Left) provides a general look at the correlation between water fractions and the three variables from a total sample collection of approximately 20,000 over non-desert and non-grassland land cover types. From Fig. 30 (Left), there is a strong linear relationship between water fractions and the three variables. The correlation coefficients are 0.9277, 0.94 and 0.943, respectively. Among the three variables, SWI shows the least variation and the strongest relationship. The correlation test over desert and grassland areas, which is shown in Fig. 30 (Right), also presents an approximate linear relationship between water fractions and the three variables. However, the variations, especially of $\Delta T_{37GHz}$ and $PR_{37GHz}$, are much larger than those for the non-desert area. The correlation coefficients of the three variables with water fractions decrease to 0.63, 0.729 and 0.831, respectively. Similar to Fig. 30 (Left), SWI shows the strongest correlation and the least variation among the three variables.
5.3.2 Variations from other factors on land surface polarization temperatures

Both the radiation properties and the correlation analyses above prove that it is feasible to calculate water fractions from microwave data. However, as proven by a series of tests, land surface polarization temperatures vary according to several factors, including precipitation and cloud cover. To reduce the variation, these factors have to be considered and quantified during retrieval.

5.3.2.1 Precipitation

Precipitation is the first factor that demonstrates a strong variation on land surface polarized brightness temperatures at 37 GHz and 89 GHz. Precipitation changes ground wetness and vegetation water content. Because the polarized brightness temperatures at these two frequencies are the most sensitive to water surface, fluctuations in soil wetness or water content within any footprint substantially change the polarized brightness temperatures, especially the horizontal.

The evaporation of soil water is a gradual process, although soil wetness is mostly affected by precipitation on the day of, and continues to be affected by the precipitation of the preceding few days. A test using AMSR-E data shows that medium rainfall can increase \( \Delta T_{37\text{GHz}} \) in the next three days during early summer in the Mississippi River Basin. Weighted precipitation from the last 96 hours is utilized by assuming precipitation impact declines linearly with time. Thus, weighted precipitation is used, with
precipitation in the last 96 hours having the least weight, 0.25, and precipitation in the last 24 hours having the largest weight, 1.0.

\[ P = 1.0 \times P_0 + 0.75 \times P_1 + 0.5 \times P_2 + 0.25 \times P_3, \]  

(5-4)

where \( P \) is the weighted precipitation and \( P_0, P_1, P_2, \) and \( P_3 \) are the precipitations in the last 24, 48, 72, and 96 hours, respectively.

A coarse correlation test between the weighted precipitation and the three variables is performed pixel to pixel using the data from Julian day 102 to 270 in 2011. The results are shown in Fig. 31, where all three variables have many pixels with correlation coefficients greater than 0.5; some of the pixels even have correlation coefficients greater than 0.75. Errors in the sampling collection and precipitation data aside, an interesting phenomenon is found in Fig. 31: SWI has a strong-impact area different from that of \( \Delta T_{37\text{GHz}} \) and \( PR_{37\text{GHz}} \). From Fig. 31, pixels with high correlation coefficients for SWI are concentrated in the Midwest, where the land cover is mainly grassland with less vegetation; whereas pixels from the other two are concentrated in the middle eastern area, where the land is covered with dense vegetation. This finding might be related to the different sensitivities of the three variables over different ground conditions.

### 5.3.2.2 Cloud fraction

Although microwave radiation can penetrate some clouds, it cannot penetrate all and can still be affected by clouds (CHO and Nishiura, 2010). First, microwave radiation cannot penetrate precipitating clouds. Second, microwave radiation can be absorbed or
scattered by clouds or water droplets. Therefore, weather filters are necessary to reduce the atmospheric effect.

Precipitating clouds might be defined as those with lower vertical polarization temperatures at 89GHz than those at 22 GHz or 37 GHz (Ferraro et al., 1998). Here, we define a pixel in AMSR-E imagery as a cloud-contaminated pixel if it meets the following conditions:

\[
T_{89v} - T_{37v} < -1K \quad (5-5)
\]

\[
T_{89v} - T_{22v} < -1K \quad (5-6)
\]

Equations (5-5) and (5-6) can remove most precipitating cloud pixels. However, pixels contaminated by clouds with large water droplets can still impact polarization temperatures. To minimize the impact from weather systems, pixels with precipitation within the last 24 hours greater than 10 mm are defined as cloud-contaminated and are not used for sampling or water fraction retrieval.

After these two types of pixels are removed from the sample collection, a test on cloud fractions is performed at water surfaces over the Great Lakes (Fig. 32). Samples are collected from continuous three-month AMSR-E data over the middle of Lake Michigan, where there is no salinity and the water depth is steady. Cloud fractions are calculated at a 25-km resolution from AQUA/MODIS 1-km cloud mask data (MYD35). The result in Fig. 32 shows a moderate negative correlation between cloud fractions and the three variables. Under overcast conditions (cloud fraction 100%), the situation becomes more complex, with larger variations in the three variables indicating that the impacts might vary by cloud types and optical thickness.
5.3.3 Application of Classification and Regression-Tree

5.3.3.1 Regression-Tree and CART

The Regression Tree (RT), which was first introduced by Morgan et al. in 1963 and further developed by Breiman et al. (1984), is a supervised learning method that addresses multiple regression problems. RT technology integrates machine learning, artificial intelligence, pattern recognition, and statistics to extract information from a large database (Hand et al., 2001), search for hidden patterns, and find association rules for target datasets. Thus, it can provide a basis for knowledge-based decision making and simulate non-linear relationships existing in the database (De’ath and Fabricius, 2000). Because it combines decision trees (DT) and traditional regression analysis, the RT approach helps determine regression relationships under various conditions when dealing with massive, dynamic, ambiguous, and possibly conflicting digital data.

Among all the Regression Trees, the Classification and Regression Tree (CART) is considered the most successful classification tree approach and has been widely used in the field of advanced analytics and data mining (Breiman et al., 1984; Lewis, 2000). In this study, CART serves as the main RT technique to establish water fraction retrieval models.
5.3.3.2 Application of Regression-Tree Approach in Water Fraction Retrieval

Based on the analysis in Section 5.3.1 and Section 5.3.2, water fraction is strongly correlated with the $\Delta T_{37\text{GHz}}$, $PR_{37\text{GHz}}$, and SWI, and over pure land surface, these three variables are varied by multiple factors, including NDVI, tree cover, precipitation, and cloud fractions. Thus, the relationship between water fraction and these variables can be expressed in the following format:

$$ WF = f(\Delta T_{37\text{GHz}}, PR_{37\text{GHz}}, \text{SWI}, P, \text{NDVI}, \text{TF}, \text{CF}) $$  \hspace{1cm} (5-7)

From Equation (5-7) to (5-13), P is the weighted precipitation, TF is tree cover, and CF is the cloud fraction.

With the Regression Tree approach, Equation (5-7) can be expanded in multiple linear functions with restricted conditions under various ground and atmospheric situations, as shown in the following format:

$$ WF = A_i \Delta T_{37\text{GHz}} + B_i PR_{37\text{GHz}} + C_i \text{SWI} + D_i P + E_i \text{NDVI} + F_i \text{TF} + G_i \text{CF} + H_i $$  \hspace{1cm} (5-8)

where $A_i$, $B_i$, $C_i$, $D_i$, $E_i$, $F_i$, and $G_i$ are coefficient vectors under different conditions, and $H_i$ is the constant vector. The restricted conditions are taken from classification analysis.

Both the multiple linear functions and the restricted conditions compose a classification and regression tree.

To adjust the algorithm to data from a series of passive microwave sensors including the new sensor ATMS which only have vertical polarization at 20 GHz and 30GHz, two strategies are used and compared in this study. The first strategy (hereafter called WF1) uses all the three variables and impact factors in the format of Equation (5-
and the second strategy (hereafter called WF2) only makes use of SWI, which is the only available variable in ATMS data and shows the best performance from the correlation analysis in Section 5.3.2. Note although microwave emission is more determined by the object's physical properties, such as atomic composition and crystalline structure, rather than the temperature of an object, compared to infrared, there is still slight difference in radiation between day and night. From Brakenridge’s research, AMSR-E data in night time are more stable than that in day time (Brakenridge et al., 2007). Therefore, for each strategy, samples are classified into four categories based on analyses from Fig. 28, Fig. 29, and the consideration of differences from both the environment and objects’ physical properties in the day and night: barren land (desert and grassland) in the daytime, barren land in the nighttime, vegetation land (non-desert and non-grassland) in the daytime, and vegetation land in the nighttime. AMSR-E data covering the entire Mississippi River Basin on Julian days 102, 107, 125 and 157 are used for sample collection for vegetation land; desert and grassland samples are collected from AMSR-E data in the Southwest from Julian days 133, 138, 150 and 158. A regression tree is built based on each sample collection. Altogether, 4 trees are generated for each strategy. For the training samples, the correlation coefficients of vegetation land and desert land are approximately 0.98 and 0.93 respectively; whereas the mean absolute errors are approximately 2.0 and 0.7 respectively. Fig. 33 shows a sample regression tree during the night for vegetation land using the second strategy. The leaves of the tree consist of regression functions, parts of which are shown in Equations (5-9) to (5-13). Conditions: SWI <= 0.01, TF <= 59.5, CF <= 0.5
\[ WF = 74.882SWI - 0.8268NDVI - 0.0003P + 0.0003TF + 0.0003CF + 0.4372 \quad (5-9) \]

Conditions: \( SWI \leq 0.01, TF \leq 59.5, CF > 0.5 \)

\[ WF = 93.7977SWI - 4.3424NDVI - 0.0162P + 0.0208TF + 0.0003CF + 2.8103 \quad (5-10) \]

Conditions: \( SWI \leq 0.01, TF > 59.5 \)

\[ WF = 6.2102SWI - 0.0634NDVI - 0.0003P + 0.0001TF + 0.0002CF + 0.3318 \quad (5-11) \]

Conditions: \( SWI > 0.01, CF \leq 0.5, NDVI \leq 0.709, TF \leq 11.5 \)

\[ WF = 13.8724SWI - 0.0972NDVI - 0.001P + 0.0021TF + 0.0002CF + 0.4915 \quad (5-12) \]

Conditions: \( 0.01 < SWI < 0.011, CF = 0, NDVI \leq 0.709, TF > 11.5 \)

\[ WF = 17.9186SWI - 0.0972NDVI - 0.0018P + 0.0021TF + 0.0002CF + 0.8158 \quad (5-13) \]

### 5.4 Results and Evaluation

#### 5.4.1 Results

With the generated regression trees, water fractions are calculated with AMSR-E 37 GHz and 89 GHz data along with a series of spatially matched datasets, including precipitation, NDVI, tree cover, and cloud fractions. Because the 4 km precipitation from NWS/NOAA covers only the US mainland, water fractions are calculated in the same area. Among all the available pixels, the pixels satisfying Equations (5-5) and (5-6) or
pixels with precipitation more than 10 mm within the last 24 hours are identified as cloud-contaminated pixels and are not used for water fraction retrieval. The rest of the pixels are applied with both the strategies to obtain water fractions from AMSR-E data.

Fig. 34 presents a series of AMSR-E water fraction retrieval maps from the second strategy, which uses only SWI, on Julian days 119, 128, 130, 137, 148, and 155 in 2011 from the top left to the bottom right respectively, with blue colors indicating the different water fraction distributions.

From the pictures shown in Fig. 34, the overall distribution of water fractions is reasonable. The large lakes and coastal area have the largest water fractions, with good continuity at the edges. Over the Mississippi River Basin, from late April to early June in 2011, a severe flood occurred which was recorded as the largest flood in the last 30 years. The six maps shown in Fig. 34 track the dynamic flood process. Around April 29 (Julian day 119) was the beginning and included the development period of the flood after intensive rainfall beginning around April 15, 2011, and snow/ice melting from upstream during April. The top left map presents large water fractions in the Mississippi River Basin. The flood reached its peak in early May 2011; the map for May 08, 2011 (Julian day 128, top right) shows the largest water fractions in this area. Afterwards, the flood waters began to retreat, but based on the map on May 10 (Julian day 130), the Mississippi River Basin still maintained large water fractions. Eventually the flood came to an end, and the maps on May 17, May 28, and June 04 (Julian days 137, 148 and 155, respectively) show water fractions gradually decreasing. This continuous change in water fractions in the Mississippi River Basin indicates promising results from the method.
5.4.2 Evaluation

5.4.2.1 Visual comparison with MODIS

The AMSR-E water fraction results estimated with the RT approach in this chapter are compared with MODIS aggregated 25-km water-fraction products. Here, the MODIS water-fraction product is a combination of a static MOD44W global 250-m water map and a real-time 500-m water-fraction map retrieved with the DNNS method in chapter 4 over the Mississippi River Basin. Although the MODIS water-fraction product here does not represent real ground truth, it still provides a general water body distribution. Thus, a visual comparison with it at least demonstrates the overall situation of the RT estimates.

Fig. 35 shows a comparison between the water fractions from AMSR-E with the two strategies (Left: WF2, Right: WF1) and MODIS (Middle) on Julian days 119, 128, and 155 respectively. From Fig. 35, the entire distribution of the water bodies in AMSR-E maps with the two strategies is consistent with that in the MODIS images, although variations still exist. In the Midwest area, AMSR-E with WF2 shows a much larger water area (bluish-white color) with a water fraction of 1%; these areas are sometimes not steady. Compared to the WF2 water maps, the WF1 maps (Fig. 35 Right) show a smaller water area, which has a 1% water fraction, and appear more consistent with MODIS maps. In the northeast area, WF1 shows a water fraction distribution similar to that of the MODIS maps, whereas WF2 has a much smaller water area. Another area with large variations is in the northern Midwest area (North Dakota, Minnesota and northern Wisconsin), where water fractions fluctuate substantially on both the WF1 and WF2
maps. This finding might be not only related to the impact from snow/ice cover and melting, but also related to a severe flood due to ice jam in Red River. Because snow/ice maps are not used in this study, the existence of snow/ice may not only be classified as water pixels but also enlarge water fractions due to the similar radiation properties at microwave frequencies. Over the Mississippi River Basin, where a severe flood occurred during that period, both the WF1 and WF2 maps show similar tendencies to the MODIS maps, though the WF2 results are more consistent with the MODIS maps.

From the comparison between the AMSR-E results and MODIS maps, although there are large differences in the 1% water fraction area, the overall distribution of water fractions from AMSR-E data is reasonable, and consistent with the MODIS results. Actually, the 1% water fraction difference is acceptable and can be ignored. A comparison of the two strategies shows that in the Midwest area, where the types of land cover are mainly desert and grassland, WF2 has larger water fractions than WF1. However, in the mid-eastern area, where the types of land cover are forest and dense vegetation, WF1 shows larger water fractions. This finding might be related to the sensitivity of 37 GHz to vegetation (Kerr et al., 1993).

5.4.2.2 Quantitative Comparison

To further evaluate the accuracy of the AMSR-E water fraction results with the RT approach, water fractions are quantitatively compared pixel to pixel between AMSR-E and MODIS from Julian day 103 to 207, excluding the sampling days. $|\Delta WF|$ or $|D_{WF}|$, which is defined as the absolute water fraction bias, is calculated as follows: $|\Delta WF| =$
\(|WF_{AMSRE} - WF_{MODIS}|\) (Hereafter, \(|\Delta WF1|\) or \(|D_{WF1}|\) refers to the absolute water fraction bias from the first strategy, and \(|\Delta WF2|\) or \(|D_{WF2}|\) refers to the absolute water fraction bias from the second strategy). Let \(N\) be the total pixel number and \(N_5\) and \(N_{10}\) be pixel numbers with \(|\Delta WF|\) within 5% and 10% respectively. Percentage (P) is calculated according to Equation (5-14):

\[
P_i = \frac{N_i}{N} \times 100\%, \quad i = 5, 10
\]  

Fig. 36 shows the distribution of \(P_i\) for all available pixels (Left) and water-only (water fraction \(\geq 1\%\)) pixels (Right) in the day (Top) and night (Bottom), respectively. Black solid curves represent \(P_5\) for WF1, the black dotted line represents \(P_{10}\) for WF1, the red solid line represents \(P_5\) for WF2, and the red dotted line represents \(P_{10}\) for WF2.

In Figure 36, \(P_5\) and \(P_{10}\) for both WF1 and WF2 for all pixels during the daytime are concentrated approximately 90\%. During the night, \(P_5\) and \(P_{10}\) for WF2 show percentages similar to those during the day, but \(P_5\) for WF1 has an approximate 10\% decrement. For water-only pixels (Right), \(P_5\) and \(P_{10}\) for WF1 and WF2 show approximately 20\% lower average percentages, and the fluctuations also become larger.

During the day, WF1 is more accurate than WF2, whereas at night, WF2 is more accurate than WF1. In particular, \(P_5\) for WF1 performs the worst, with an approximate average percentage of 45\%. Compared to WF1, WF2 performs more steadily, especially at night.

The scatter plots shown in Fig. 37 demonstrate the distribution of AMSR-E water fractions with the RT approach from the two strategies for three selected days. In Fig. 37, both strategies present highly consistent results with that from MODIS, and the
correlation coefficients are approximately 0.99 for these three days. Compared to WF1, WF2 also performs slightly better, with larger correlation coefficients and less variation.

From the quantitative comparison, the two strategies with CART show promising performance in AMSR-E water fraction retrieval, indicating the high feasibility of CART for water fraction retrieval with AMSR-E. By using only SWI, the second strategy, WF2, presents a steadier result than WF1, which makes use of all the three variables: $\Delta T_{37GHz}$, $PR_{37GHz}$ and SWI. This finding suggests there is a good chance that the algorithm can be applied to ATMS data as well.

5.4.2.3 Continuity test

Both the visual and quantitative comparisons provide an overall evaluation of water fraction results with the RT approach. To further validate the continuity of the RT approach in AMSR-E water fraction retrieval, four pixels located in four representative areas are chosen to compare their water fractions with MODIS water fractions during the entire evaluation period.

The first pixel is located at a point (Longitude: -89.25, Latitude: 36.5), which is in the Mississippi River near the New Madrid station of Missouri, where there was a severe flood from April to June 2011. During the flooding period, near-real-time water fractions were obtained from TERRA/MODIS data. For both before the flood and after the flooding retreated, water fractions at this point are calculated using MOD44W. They show an average water fraction of 12%. The near-real-time and static water fractions are both utilized for comparison. The results are shown in Fig. 38 (Top Left), where the black
line is the MODIS water fraction, the green is the water fraction from WF1, and the blue is the water fraction from WF2. From the plot, although there are differences between the water fractions from AMSR-E and MODIS, the two strategies show good accordance and consistent fluctuation tendencies with the MODIS result. The change of AMSR-E water fractions fits the flood process well. After the flood began retreating (Julian day 160), WF2 showed fewer fluctuations than WF1 and was more consistent with the MODIS results.

The second pixel, which is located at a point (Longitude: -87, Latitude: 43) in the middle of Lake Michigan, is a pure water pixel with a water fraction of 100%. The two strategies calculate water fractions of 100% for all but two days, with water fractions of 98% and 99% respectively (Fig. 38 (Top Right)).

The third pixel is chosen at the point (Longitude: -82.5, Latitude: 41.5), which is located at the edge of Lake Erie. MODIS shows a water fraction of approximately 96% at this pixel (Fig. 38 (Bottom Left)). WF1 mostly results in 100% water fractions, entraining 6-day water fractions varying from 55% to 85%. WF2 presents water fractions from 96% to 100% most of the time, with 5-day low water fractions from 60% to 82%. Verification of the data shows days with low water fractions might be due to residual cloud impacts, as cloud fractions were 100% and there were weather systems present around those dates. Disregarding the dates, the two strategies actually show good continuity at this point.

The last pixel has geo-location (Longitude: -91.5, Latitude: 31) and is located downstream in the Mississippi River. The average water fraction from MOD44W is approximately 12% with fluctuations within 10% during the 2011 Mississippi river flood.
(Fig. 38 (Bottom Right)). The two strategies show continuous results with fluctuations of approximately 5% during the evaluation period from Julian day 103 to 207, 2011.

Although the results can be severely impacted by weather systems, the continuity test still shows that the RT approach with both strategies results in continuous water fractions for AMSR-E retrieval.

5.4.2.4 Severe error analysis

The above evaluation tests prove water fraction retrieval with the RT approach using AMSR-E data shows promising results. However, the evaluation results demonstrate severe errors in the retrievals. Here, we define a pixel as having severe errors if $|\Delta WF|$ is greater than 30%.

To further analyze severe errors from RT retrievals, a statistic is arrived at by counting the frequency of a pixel with $|\Delta WF|$ greater than 30% from Julian day 103 to 207 and then mapping the frequency to analyze the pattern (Fig. 39, Left: frequency of WF1, Right: frequency of WF2). In Fig. 39, pixels with severe errors are concentrated mostly on lake shores and ocean coasts. This is expected, because these pixels have large water fractions and weather systems can cause substantial decrement. There is one large area, located in North Dakota, Minnesota and northern Wisconsin that has high severe error frequency and many small water bodies, but most of the pixels have water fractions less than 30%. This is caused by two factors. First, because of the severe flood occurred along Red River from April to late May in 2011, the actual water fractions were very large but they could not be reflected by MOD44W static water map which only shows
normal water fractions. This “false” error actually proves the stability of water fraction retrieval algorithm in this study. The other reason is related to snow/ice melting from April to early May in this area. Because a snow/ice mask is not used in this study, snow/ice pixels are easily classified as water pixels resulting in larger water fractions due to similarities in the radiation properties at 37 GHz and 89 GHz between water and snow/ice and thus bring about large water fraction difference compared to that from MOD44W.

Compared with WF1, WF2 shows a similar overall distribution of pixels with severe errors. However, it has fewer pixels with severe errors in the west area and more pixels in the east area. This phenomenon might indicate that WF2 is more suitable for water fractions in deserts and grassland with less vegetation cover, whereas WF1 performs better in forest and dense vegetation areas.

Based on the above evaluation analysis, the RT approach performs steadily and continuously in AMSR-E water fraction retrieval by considering the impact from precipitation, NDVI, tree cover and cloud fractions. Both strategies developed in this chapter have their own advantages: WF1 has better results in dense vegetation areas, whereas WF2 performs better in barren land. A combination of WF1 and WF2 is suggested for AMSR-E water fraction retrieval.
5.5 Discussion and Summary

5.5.1 Discussion

With the method presented in this study, the water fractions retrieved using 37 GHz and 89 GHz from AMSR-E data present promising results. Although the method considers many possible impact factors and was tested in two strategies, there are still many problems that could introduce uncertainty and errors in the water fraction results. These problems include those listed below.

The first problem is cloud contamination. Although equations (5-5) and (5-6) and precipitation data within the last 24 hours can help remove many cloud-contaminated pixels, not all cloud-contaminated pixels can be removed. From the evaluation test, there are still pixels contaminated by clouds, which can result in large water fraction errors. A test on cloud fractions shows that cloud fractions can decrease the three microwave variables’ values. However, the decrement still varies substantially when the cloud fraction is 100%, which means cloud types and thickness might pose greater impacts on microwave variables for water fraction retrieval. Unfortunately, different types of clouds and thicknesses are not considered in this study yet.

The second problem comes from snow/ice cover or melting. Because a snow/ice mask is not used in this study, pixels with snow/ice cover are mostly classified as water pixels and calculated in large water fractions with the $\Delta T_{37\text{GHz}}$, $P_{R37\text{GHz}}$, and SWI used in this study. This problem is more serious during the night, which explains why the RT approach results in lower accuracy during night.
Third, there are uncertainties from AMSR-E data processing. Because AMSR-E global 25 km quarter-degree gridded brightness temperatures are used as the main dataset for the development and evaluation of the microwave water fraction algorithm in this study, the re-projected data in a quarter-degree projection from conic footprints sometimes can result in low confidence, especially when the emissivity within the footprint varies. Additionally, the original 37GHz polarized data have a 12.5-km resolution, and the 89GHz data have a 6.5-km resolution. Resampling from a high spatial resolution to a lower resolution can also bring about uncertainties for the retrievals due to the “beam-filling” effect, which has been discussed in a series of studies (Chiu et al., 1990; Rapp et al., 2009).

Because precipitation data are used as an important impact factor, the algorithm can only be applied in areas where and when precipitation data are available. The quality of precipitation data can also affect the accuracy of retrievals. Related to precipitation, over land after intensive rainfall, wet land might be estimated with an approximately 1% or 2% water fraction, which is more common with the second strategy WF2.

Finally, it is difficult to obtain ground truths for both sample training and evaluation. In this study, we use TERRA/MODIS water fractions over the Mississippi River Basin calculated with the DNNS method and MOD44W static water map as the ground truth for regression tree generation and evaluation. Although the DNNS method provides near-real-time water fractions, it is only 80% accurate and affected by cloud and vegetation cover. MOD44W is only a static water map and does not reflect any information on the fluctuations of water bodies. Additionally, MODIS for water detection
has 250-500 m resolution; there are many small water bodies that are not detected. In fact, both products cannot be viewed as ground truth. Therefore, the regression trees are generated based on controversial samples, and the results are validated with non-ground-truth data. Using controversial samples for the regression tree generation can increase tree leaves and make the generated trees more complex, thereby decreasing the tree’s performance capabilities, and sometimes bringing about large errors.

5.5.2 Summary

In this study, CART serves as the main approach to develop a water fraction retrieval algorithm using 37 GHz and 89 GHz from AMSR-E data by considering a series of impact factors, including precipitation, NDVI, tree cover and cloud fractions. To make the algorithm applicable for data from a series of passive microwave sensors including the new sensor ATMS, regression trees are generated in two strategies: one uses all three variables calculated from 37 GHz and 89 GHz, and the other only uses the soil wetness index from 37 GHz and 89 GHz. Although it is difficult to obtain highly accurate ground truth, the evaluation results still show CART performs quite steadily and promisingly in water fraction retrieval. Thus, the application of low-resolution microwave data in flood detection to fill the gap in optical data under cloud cover or data unavailability becomes feasible. The chapter can be summarized as follows:

1. The variables derived from 37 GHz and 89 GHz, $\Delta T_{37\text{GHz}}$, $\text{PR}_{37\text{GHz}}$ and SWI, show strong correlation with water fractions. Among the three variables, $\text{PR}_{37\text{GHz}}$ performs better than $\Delta T_{37\text{GHz}}$, and SWI performs the best with the least
variation, especially over barren land.

2. The three variables are varied by a series of factors including NDVI, tree cover, precipitation and cloud fractions. Since these factors can be dynamically obtained from optical satellite data or ground/model outputs, it provides a way to combine microwave data with optical satellite data for water fraction retrieval.

3. Compared with the physical calculation method used in Sippel’s algorithms, the Regression Tree approach is uniquely advantageous because it integrates all the factors and variables to simulate non-linear models over various conditions. From the evaluation analysis, the results from Regression Tree appear to be steady and continuous, indicating the strong possibility of water fractions being retrievable over a wide range with microwave data.

4. Both of the two strategies obtain reasonable water fraction distributions. The first strategy, which uses all variables, shows better results with regard to differentiating pure land from water pixels and, in particular, obtains better retrievals over forest and dense vegetation areas. However, the second strategy, which only uses the soil wetness index, performs more steadily during night and results in better retrievals over barren land. Therefore, a method for AMSR-E data combining only the soil wetness index for barren land and all variables for dense vegetation land is suggested.

5. Although a number of factors have been considered in this chapter, there are still some unsolved problems. Further work is required to remove pixels
contaminated by clouds. Instead of the cloud fraction, cloud types and cloud optical thickness might be more closely related to the microwave variables. A snow/ice mask should also be utilized to remove snow/ice pixels. To make the algorithm applicable in other areas, a replacement of current precipitation data may be required. To obtain more robust water fraction samples, Landsat/TM 30 m imagery instead of a MODIS water fraction product might be better for both sampling and evaluation, although collecting enough samples can be difficult and would require a great deal of work.
Figure 28 Plots of seasonal average $\Delta T_{37\text{GHz}}$ (top), $\text{PR}_{37\text{GHz}}$ (middle) and $\text{TR}_{37,89\text{GHz}}$ (bottom) over land cover types (Land cover data is from 1-km USGS Land Use/Land Cover System by calculating fractions of each land cover type at 25-km resolution, and samples are collected with 100% fraction of each type at 25-km.)
Figure 29 Correlation test between water fraction and $\Delta T_{37\text{GHz}}$, $PR_{37\text{GHz}}$ and $TR_{37,89\text{GHz}}$ over non-desert area (Left) and desert area (Right)
Figure 30 Distribution of Correlation Coefficients between precipitation and $\Delta T_{37\text{GHz}}$ (Left), PR$_{37\text{GHz}}$ (Middle) and TR$_{37_{-}89\text{GHz}}$ (Right)

Figure 31 Scatter analysis of NDVI (Left) and tree cover (Right) on $\Delta T_{37\text{GHz}}$ (Top), PR$_{37\text{GHz}}$ (Middle) and TR$_{37_{-}89\text{GHz}}$ (Bottom)
Figure 32 Correlation test between cloud fraction and $\Delta T_{37GHz}$ (Top), $PR_{37GHz}$ (Middle) and $TR_{37,89GHz}$ (Bottom) over water surface
Figure 33 Part of a Regression Tree for water fraction retrieval over vegetation land during night time with AMSR-E data
Figure 34 Water fraction retrieval maps from AMSR-E in 2011 (From Top left to down right: April 29, May 08, May 10, May 17, May 28 and June 04 (Julian day 119, 128, 130, 137, 148 and 155)
Figure 35 Water fraction comparison between AMSR-E and MODIS in 2011 (From Top to bottom: April 29, May 08, June 04 (Julian day 119, 128 and 155)). Left: WF2 from AMSR-E, Middle: MODIS, Right: WF1 from AMSR-E
Figure 36 Plot of $P_5$ and $P_{10}$ for all pixels and water-only pixels in day and night time from the two RT strategies
Figure 37 Scatter Plots between water fractions from AMSR-E and MODIS over Mississippi River Basin on April 29, June 06 and June 26, 2011 (Julian day 119, 158 and 207). Left: Results of WF1, Right: Results of WF2.
Figure 38 Continuity test on single pixel’s water fractions, Top Left: Pixel around New Madrid station, Top Right: Pixel in Michigan Lake, Bottom Left: Pixel at the edge of Illinois Lake, Bottom Right: Pixel at the downstream of Mississippi River

Figure 39 Map of frequency distribution of pixels with |ΔWF| larger than 30%
CHAPTER 6 Integration with SRTM DEM Data to Derive High Resolution Water Maps

6.1 Introduction

From chapter 4 and 5, water fractions can be retrieved from both optical and microwave satellite data with reliable accuracy, and then regional water extent or area can be obtained at a coarse or moderate scale, rather than point observations on the water level, depth and discharge. Based on water’s hydrodynamic mechanism, it is highly possible to derive more detailed information with high resolution DEM. This comes to the idea to integrate the water fraction products from MODIS or AMSR-E with SRTM 30-m DEM to derive water maps at a 30-m resolution. Therefore, in this chapter, we present a method using MODIS water fraction product and SRTM 30-m elevation data to generate MODIS 30-m flood maps.

6.2 Data preparation

In this chapter, level-1b data from TERRA/MODIS are used to obtain water fraction product using method described in chapter 4. Data from TM onboard Landsat are used for validation. Static data, including SRTM 30-m elevation data, MOD44W 250-m water maps and the IGBP land cover product, are also utilized for water classification and integration. The specific data used for the study in this chapter include the following:

- TERRA/MODIS L1B calibrated radiance at 500-m resolution MOD02HKM
data are used for water detection and water fraction calculation;

- TERRA/MODIS geo-location and geometric angles (MOD03) for swath projection with the MRTSWATH tool;
- TERRA/MOD35 cloud mask to remove clouds;
- SRTM 30-m elevation data to integrate the TERRA/MODIS 500-m water fraction with 30-m water maps.
- TM data from Landsat observations at a 30-meter spatial resolution as the principal data to evaluate integrated 30-m water maps derived from the MODIS and SRTM elevation.
- Worldview-2 2-m spatial resolution imagery serving as ground truth for validation.
- MOD44W 250-m water maps are used for background water bodies before flooding;
- IGBP land cover product as assistant data for land classification with a decision tree method.
- 30-m National Land Cover Data (NLCD 2006 version) for validation analysis.

### 6.3 Method

Because the MODIS water fraction product provides water area information in 500-m pixels, the area and geo-location information can be used to locate the distribution of water bodies with 30-m SRTM DEM data after precise co-registration between MODIS and SRTM DEM, thus making it possible to derive 30-m-resolution water maps.
6.3.1 The principle inundation machine

According to water’s hydrodynamic properties, when a flood occurs, inundation always happens from the lowest to the highest point in elevation. If the topography is not very steep around the water bodies, the higher the water level, the larger the inundated area. This inundation mechanism can be expressed as:

\[ A = \int_{\text{min}_h}^{\text{max}_h} f(h) \, dh \]  \hspace{2cm} (6-1)

Where, \( A \) is the total water area between the minimal surface elevation \( \text{min}_h \) and maximal inundated surface elevation \( \text{max}_h \), and \( f(h) \) is the increment of water area with the change of surface elevation \( h \).

Equation (6-1) indicates the relationship between inundated area and surface elevation. Because both the minimal surface elevation \( \text{min}_h \) and the increment function of water area \( f(h) \) in a specific region can be derived from the DEM, the maximal inundated surface elevation \( \text{max}_h \) decides the total water area. Consequently, the maximal inundated surface elevation \( \text{max}_h \) can also be calculated if the total water area \( A \) is known. Because the water fraction can be calculated with the DNNS method as described in chapter 4 with MODIS multi-channel data at 500-m resolution, using the 30-m SRTM digital elevation model, Equation (6-1) can be written as Equations (6-2) and (6-3) with the water fraction:

\[ f_w = \int_{\text{min}_h}^{\text{max}_h} g(h) \, dh \]  \hspace{2cm} (6-2)

\[ f_w < 1.0 \] \hspace{2cm} (6-3)

where \( f_w \) is the MODIS water fraction in a 500-m pixel, \( g(h) \) is the increment of
inundated water fraction with the change of water surface elevation $h$, $\text{min}_h$ is the
minimal elevation in the SRTM 30-m cells matching the MODIS 500-m pixel, and $\text{max}_h$ is the maximal elevation in the SRTM 30-m cells.

Equations (6-1) to (6-3) can be explained as follows: before the whole pixel is
submerged ($f_w < 1.0$) under the rising water surface level, the land area around the water
bodies between the minimal surface elevation and the increased surface elevation is
gradually inundated. When the inundated water area/fraction is equal to the water
area/fraction calculated from MODIS, the iteration is terminated with the final surface
elevation being the maximal water surface level $\text{max}_h$, and all 30-m cells with a surface
elevation between $\text{min}_h$ and $\text{max}_h$ within the 500-m MODIS pixel are assigned as
water cells. When $f_w$ is equal to 1.0, all 30-m cells within the MODIS 500-m pixel are
assigned as water cells.

6.3.2 Uniformity process on water levels

Although the MODIS water fraction product can be used for integration with
Equation (6-2), the errors of the product cannot be ignored because the existence of errors
may cause a concavo-convex water surface in an adjacent standing water body. Generally, for an adjacent inundated water body, especially in a floodplain, the water
level should be uniform and the difference can be ignored. If water fractions are
calculated strictly and correctly, then the maximal water surface levels in the adjacent
water pixels are similar. However, from the validation result of the DNNS method shown
in chapter 4, the accuracy of the water fraction is approximately 80% (Li. et al., 2012). Water fraction errors varying from -0.2 to 0.2 can sometimes cause several meters of difference in maximal water levels for neighboring adjacent water pixels, thus decreasing the accuracy of integration.

According to water’s fluidal properties, in the same container, water level will stay the same everywhere. Thus, for a single adjacent standing water body, the variance of water levels should be close to 0. In other words, the maximal elevation $\max_h$ should be the same for all the pixels covering that water body except for pixels with a water fraction equal to 1.0 (water levels are greater than the maximal surface elevation for immersed water pixels). We then obtain:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (\max_h - \overline{\max_h})^2}{N}} = 0 \quad (6-4)$$

From Equation (6-4), we obtain:

$$\max_h_i = \overline{\max_h_i} \quad (6-5)$$

In Equations (6-4) and (6-5), $N$ is the number of the total pixels covering an adjacent standing water body with a water fraction less than 1.0, and $\overline{\max_h_i}$ is the average maximal elevation for the $N$ pixels.

Equation (6-5) suggests that for a group of neighboring adjacent water pixels, the maximal water level in any pixel calculated from equation (6-2) is equal to the group average water level. From Equation (6-2), because $\min_h$ and $g(h)$ are known and static for any specific pixel, $f_w$ and $\max_h$ are the only two variables related to each other. Therefore, the average of $\max_h$ actually makes the average of water fractions as well. If
we assume the errors of $f_w$ are distributed randomly from -0.2 to 0.2, then the errors caused by the water fraction in each pixel can be reduced to the minimum after the uniformity process. Thus, the uniformity process on $\text{max}_h$ for a single adjacent standing water body actually reduces the errors caused by the water fraction retrieval and can help obtain a better integration result.

6.3.3 Integration with the SRTM data

Following up on the previous two sections, we can proceed integrating the MODIS water fraction product with SRTM 30-m DEM data. First, the minimal elevation $\text{min}_h$ is searched among SRTM 30m cells within a MODIS 500-m pixel, and the pixels’ maximal water level $\text{max}_h$ is calculated with Equations (6-2) and (6-3). Then, MODIS 500-m water pixels are clustered in different water polygons with the $N_4(P)$ adjacency rule (only considering 2 horizontal and 2 vertical neighbors) in a recursion method. In each water polygon, pixels with a water fraction less than 1.0 are used to calculate the average maximal water level $\overline{\text{max}_h}$, which is taken as the polygon’s maximal water level for integration. When a water polygon is too large, the maximal water level may not be similar for all the pixels in the polygon due to the fluidity of the water. In this situation, pixels in an $m \times m$ neighboring box in the same water polygon are used to derive $\overline{\text{max}_h}$ instead of all the pixels in the water polygon (in our study, $m$ is 25 for 500m MODIS data). Another exception is when a water polygon is too small, for example, less than three 500-m water pixels, and the average tends to be less significant. In this situation,
water pixels with a water fraction less than 1.0 in a neighboring 3×3 box are used to calculate \( \overline{\text{max}_i h_i} \). After deriving \( \overline{\text{max}_i h_i} \), a cell \( P_j \) in a 30-m resolution mask can be determined as water if the elevation \( h \) in SRTM 30-m DEM is between \( \min_i h_i \) and \( \overline{\text{max}_i h_i} \) or as non-water if the elevation \( h \) is larger than \( \overline{\text{max}_i h_i} \), which is expressed in Equation (6-6).

\[
P_j = \begin{cases} 
1 \text{ (water), } & h \in [\min_i h_i, \overline{\text{max}_i h_i}] \\
0 \text{ (nonwater), } & h \geq \overline{\text{max}_i h_i}
\end{cases} \tag{6-6}
\]

In this manner, the MODIS 500-m water fraction product is integrated into a 30-m resolution water map with the SRTM data. Fig. 40 shows the flow chart of the method.
Figure 40 Process flow chart of the integration method
6.4 Results and validation

6.4.1 Application to the Mississippi River flood of 2011

We applied this method to the severe flood of the Mississippi River Basin (April to June 2011) in the USA, in which at least 383 people were killed across seven states, thousands of homes were evacuated, and the economic losses exceeded $7 billion. MODIS tracked the flood process during the flooding period, providing an abundance of real-time data. However, compared with Landsat/TM data, the 250-m to 500-m MODIS resolution for water detection is too coarse for users to obtain detailed information about the flood.

With MODIS multiple-channel data and SRTM 30-m DEM data, the method described in Section 6.3 is applied to this flood to derive 30-m-resolution water maps. More than 10 scenes of TERRA/MODIS data are tested covering the whole Mississippi River Basin varying in 10 different dates from April to June, 2011 and the results turn to be steady. First, the decision-tree method described in chapter 2 is used to obtain water cover maps at 500-m resolution, and cloud shadows are removed with a geometric method from water maps described in chapter 3. Then, the water fraction is calculated with the DNNS method presented in chapter 4 based on the water cover maps. Fig. 41 (Left) shows a MODIS false color composite image over the Mississippi River basin on May 04, 2011, and its corresponding water fraction result is shown in Fig.41 (Right). MOD44W water maps are used to determine normal water bodies, which are shown in blue in Fig.41 (Right). Then, after spatially matching the MODIS water fraction map with
SRTM 30-m DEM data, the integration method presented in section 6.3 is used to derive MODIS 30-m water maps. Figs. 42 to 43 show subsets of the original 500-m water fraction maps (Fig.42 (Left) and Fig. 43 (Left)) and the corresponding integrated 30-m results (Fig. 42 (Right) and Fig. 43 (Right)) on May 04, 2011, and May 29, 2011, respectively. Compared with the original MODIS 500-m water maps, the integrated 30-m resolution maps not only maintain the same large coverage with MODIS but are also enlarged approximately 18 times to the original size, providing substantial detail inside the inundated area. In Fig. 42 (Right) and Fig. 43 (Right), rivers and lakes are accurately located from 500-m pixels to 30-m cells, indicating a robust result from the method. Large inundated areas are distributed along the Mississippi River, entraining a great deal of detailed non-inundated information, which helps decision-makers “see” the flood more clearly.

Figure 41 Left: MODIS false color composite image on May 04, 2011; Right: The corresponding water fraction map over the Mississippi River Basin on May 04, 2011
Figure 42 Left: TERRA/MODIS 500-m water fraction map on May 04, 2011; Right: The corresponding TERRA/MODIS 30-m integrated water map on May 04, 2011
Figure 43 Left: TERRA/MODIS 500-m water fraction map; Right: The corresponding TERRA/MODIS 30-m integrated water map on May 29, 2011
6.4.2 Validation

6.4.2.1 Comparison with Landsat TM data

6.4.2.1.1 Spatial distribution comparison

The most direct method of validation is to compare the integrated MODIS 30-m water maps with simultaneous Landsat TM/ETM 30-m water maps. In this chapter, two good-quality scenes from Landsat TM/ETM over the Mississippi River basin are used for validation:

- one, on May 10, 2011, compares TERRA/MODIS data with the closest date on May 11, 2011;
- the other, on June 11, 2011, is compared with the TERRA/MODIS data on June 12, 2011.

The comparison is performed pixel-to-pixel. Fig. 44 (Left) presents a Landsat TM false color image on May 10, 2011, and the corresponding TERRA/MODIS integrated 30-m water map on May 11, 2011 is compared with the TM result on May 10, 2011 (Fig. 44 (Right)). In the comparison figure between TERRA/MODIS 30-m water map and Landsat/TM 30-m water map (Fig. 44 (Right)), the pixels in blue are water pixels matched with the TM result, the pixels in yellow are water pixels appearing only in the MODIS water map, the pixels in red are water pixels only in TM, and the pixels in cyan are water pixels that are not detected in the MODIS 500-m water fraction product. From Fig. 44 although there is a slight difference, the distribution of water bodies between TM
and TERRA/MODIS are highly matched with each other. The normal water bodies, such as rivers and lakes, show good consistency in shapes and sizes, and the inundated water bodies indicate similar details inside.

Figure 44 Left: Landsat/TM5 false-color composite image on May 10, 2011; Right: 30-m water map comparing between TERRA/MODIS on May 11, 2011 and Landsat TM on May 10, 2011

Fig. 45 shows the other validation case during the flood retreat period over the Mississippi River basin. In the TM false color composite image on June 11, 2011 (Fig. 45 (Left)), not many inundated water bodies can be observed along the Mississippi River. Because there are many small water bodies, the undetected water bodies in the MODIS 500-m water fraction map increase dramatically. In the integrated 30-m water map (Fig. 45 (Right)), however, the water bodies also present a consistent distribution in both shape and size.
6.4.2.1.2 Quantitative comparison

Based on the two validation cases, quantitative statistics are also performed for further validation on the integration method. The quantitative variables for the statistics include the total number of water pixels in both TM ($N_{total, TM}$) and MODIS ($N_{total, MODIS}$), the number of matched water pixels ($N_{matched}$), the number of unmatched water pixels both in TM ($N_{TM, unmatch}$) and MODIS ($N_{MODIS, unmatch}$), the number of water pixels covered by cloud ($N_{cloud}$) and the number of undetected water pixels in the MODIS 500-m water fraction maps ($N_{undetect}$). As the undetected water pixels in the MODIS 500-m water fraction maps are not caused by the integration method but by
water detection, $N_{undetect}$ should be eliminated from the total pixels. Considering the impact from the cloud cover, the matched rate of integration is calculated in Equation (6-7):

$$P = \frac{N_{matched}}{N_{total,TM} - N_{undetect} - N_{cloud}} \times 100\% \quad (6-7)$$

The commit error caused by integration can be quantified from the two parts: 1) water pixels that only appear in TM except for pixels that are not detected in MODIS 500-m water fraction maps and pixels covered by cloud, and 2) water pixels that only appear in MODIS. Thus, the commit error can be calculated with Equation (6-8):

$$E = \frac{N_{MODIS\ unlatch} + N_{TM\ unlatch\ -\ N_{cloud\ -\ N_{undetect}}}}{N_{total,TM}} \times 100\% \quad (6-8)$$

However, if water detection error is considered, then $N_{undetect}$ should be counted in the total error statistics. The total omit error can be calculated with Equation (6-9):

$$E' = \frac{N_{MODIS\ unlatch} + N_{TM\ unlatch\ -\ N_{cloud\ -\ N_{undetect}}}}{N_{total,TM}} \times 100\% \quad (6-9)$$

Table 7 lists the statistical results. For May 10, 2011 (Fig. 44) data, we have 1101701 water pixels altogether from TM data compared with 1116376 water pixels in the MODIS integrated 30-m water map. Among all the 1116376 pixels in the MODIS 30-m water map, 993753 are matched in the TM result, 2394 are affected by clouds in the MODIS 500-m water fraction map, 45088 are not detected in the MODIS 500-m water fraction map and 122623 are not shown in the TM result. According to Equations (6-7) to (6-9), the matched rate is approximately 94.3%, while the commit error is approximately 16.4% and the total omit error is about 20.5%. The case on June 11, 2011, also shows a highly matched rate of 93.9%, the commit error 26.9% and the total omit error about
48.8%. Considering there is a one-day time difference between TM and MODIS observations for the two cases, the actual commit errors might be lower than the current results.

Comparing the total omit errors from the two cases, it is shown that water detection with moderate resolution satellite imagery results in much larger errors over small water bodies. On May 11, 2011, the undetected water pixels in MODIS data are only about 4.1% because of larger water bodies in the image, while in the image on June 12, 2011, the undetected rate reaches approximately 22% because of a lot of small water bodies in this area which show very weak water signals in MODIS 500-m pixels and are usually misclassified as pure land. Therefore, independent small lakes or rivers disappear in the integrated images. Moreover, the edges of some water bodies look quite artificial with straight sides like squares or serrations.

Table 7 Quantitative statistics on validation cases

<table>
<thead>
<tr>
<th>Number of water pixels</th>
<th>TM (May 10)</th>
<th>MODIS (May 11)</th>
<th>TM (June 11)</th>
<th>MODIS (June 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1101</td>
<td>1116376</td>
<td>10696</td>
<td>1021154</td>
</tr>
<tr>
<td>Matched</td>
<td>701</td>
<td>993753</td>
<td>77671</td>
<td>776715</td>
</tr>
<tr>
<td>Unmatched</td>
<td>9937</td>
<td>122623</td>
<td>28531</td>
<td>244439</td>
</tr>
<tr>
<td>Cloud-cover</td>
<td>54</td>
<td>2394</td>
<td>8</td>
<td>7582</td>
</tr>
<tr>
<td>Undetected</td>
<td>0</td>
<td>45088</td>
<td>0</td>
<td>234818</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Matched rate (P)</th>
<th>Commit error (E)</th>
<th>Total omit error (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM (May 10)</td>
<td>94.3%</td>
<td>16.4%</td>
<td>20.5%</td>
</tr>
<tr>
<td>MODIS (May 11)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM (June 11)</td>
<td>93.9%</td>
<td>26.9%</td>
<td>48.8%</td>
</tr>
<tr>
<td>MODIS (June 12)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When the matched and unmatched water pixels are further analyzed with land
cover types, the matched rates vary substantially with land cover types. A test with 30-m land cover dataset (NLCD 2006 Land Cover) shows open water, barren land, cultivated crops, and herbaceous grassland have higher and steadier matched rates (larger than 80%), while woody wetlands, forests (deciduous, ever green, mixed), and urban land (Developed Low Intensity, Developed medium Intensity and Developed high Intensity) have low matched rates (less than 55%). For the total unmatched water pixels, nearly 50% come from woody wetlands and deciduous forest (Table 8). The difference over land cover types reflects the impact from vegetation covers and urban structure on both Landsat TM and MODIS. Over woody wetlands, forest and urban land, both Landsat TM and MODIS might fail to obtain flooding water information under vegetation covers from trees or crops or urban buildings. In this situation, Landsat TM data cannot be used as ground truth over these land cover types, and water fraction retrieval from MODIS might be under-estimated and thus results in lower maximal water surface elevations which further impact the integration results.

Table 8 Analysis on the unmatched and matched water pixels with land cover

<table>
<thead>
<tr>
<th>Land cover</th>
<th>Unmatched</th>
<th></th>
<th></th>
<th>Matched</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11-May</td>
<td>12-Jun</td>
<td></td>
<td>11-May</td>
<td>12-Jun</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MODIS TM</td>
<td>MODIS TM</td>
<td>Pixels</td>
<td>P (%)</td>
<td>Pixels</td>
<td>P (%)</td>
</tr>
<tr>
<td>Woody Wetlands</td>
<td>49255</td>
<td>22156</td>
<td>134263</td>
<td>27983</td>
<td>51836</td>
<td>42.1</td>
</tr>
<tr>
<td>Cultivated Crops</td>
<td>44255</td>
<td>27044</td>
<td>75184</td>
<td>17727</td>
<td>62960</td>
<td>89.8</td>
</tr>
<tr>
<td>Deciduous Forest</td>
<td>11695</td>
<td>4343</td>
<td>2245</td>
<td>442</td>
<td>12408</td>
<td>43.6</td>
</tr>
<tr>
<td>Developed Open Space</td>
<td>6969</td>
<td>2081</td>
<td>5885</td>
<td>1464</td>
<td>18772</td>
<td>67.5</td>
</tr>
<tr>
<td>Open Water</td>
<td>3518</td>
<td>4886</td>
<td>13153</td>
<td>7168</td>
<td>264160</td>
<td>96.9</td>
</tr>
<tr>
<td>Emergent Herbaceous Wetlands</td>
<td>5711</td>
<td>1650</td>
<td>6164</td>
<td>891</td>
<td>10131</td>
<td>57.9</td>
</tr>
<tr>
<td>Developed Low Intensity</td>
<td>373</td>
<td>121</td>
<td>238</td>
<td>89</td>
<td>2481</td>
<td>83.4</td>
</tr>
<tr>
<td>Barren Land</td>
<td>141</td>
<td>328</td>
<td>602</td>
<td>358</td>
<td>2114</td>
<td>81.8</td>
</tr>
<tr>
<td>Shrub</td>
<td>236</td>
<td>99</td>
<td>3021</td>
<td>822</td>
<td>776</td>
<td>69.8</td>
</tr>
</tbody>
</table>

151
<table>
<thead>
<tr>
<th>Developed Medium Intensity</th>
<th>145</th>
<th>77</th>
<th>92</th>
<th>128</th>
<th>146</th>
<th>39.7</th>
<th>73</th>
<th>24.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grassland Herbaceous</td>
<td>183</td>
<td>21</td>
<td>75</td>
<td>32</td>
<td>1191</td>
<td>85.4</td>
<td>283</td>
<td>72.6</td>
</tr>
<tr>
<td>Developed High Intensity</td>
<td>80</td>
<td>21</td>
<td>22</td>
<td>111</td>
<td>32</td>
<td>24.1</td>
<td>30</td>
<td>18.4</td>
</tr>
<tr>
<td>Evergreen Forest</td>
<td>32</td>
<td>7</td>
<td>189</td>
<td>78</td>
<td>50</td>
<td>56.2</td>
<td>52</td>
<td>16.3</td>
</tr>
<tr>
<td>Mixed Forest</td>
<td>24</td>
<td>10</td>
<td>2762</td>
<td>418</td>
<td>32</td>
<td>48.5</td>
<td>844</td>
<td>21.0</td>
</tr>
<tr>
<td>Pasture Hay</td>
<td>6</td>
<td>16</td>
<td>544</td>
<td>371</td>
<td>16</td>
<td>42.1</td>
<td>508</td>
<td>35.7</td>
</tr>
</tbody>
</table>

### 6.4.2.2 Visual Comparison with Worldview-2 imagery

Landsat TM data undoubtedly is a good data source for validation on the integration method. However, the validation might be controversial due to the uncertainties over vegetation covers and urban area. An image from Worldview-2 at 2-m resolution over Cairo, Illinois, which can be served as ground truth, is obtained for a visual comparison with the MODIS integrated 30-m water map (Fig. 46). In Fig. 46, the middle left is a resampled Worldview-2 image on May 05, 2011 around Cairo (blue is water, yellow is vegetation and brown is dry land), and the middle right is an integrated MODIS 30-m water map on May 04, 2011 covering the similar area. From an overall comparison, the two middle images demonstrate consistent shapes and sizes of water bodies. Points A, B, C and D in the two middle images indicate four small representative areas which are shown in the four corner Worldview-2 images at actual 2-m resolution (Upper left: Point A, Upper right: Point B, Down left: Point C, Down right: Point D). In the upper left image (Point A), flooding water can be seen through thick woods in the lower middle of the image, adjacent to un-submersed Cairo downtown. The water extent is similar to that around the corresponding area of point A in the integrated map. The
upper-right and down-right images also show consistent water distribution with that around point B and point D in the integrated map. While in the down-left image which is related to the area around point C in the integrated water map, there is dense vegetation centering in the image with flooding water flowing below and vegetation tops unsubmerged. However, in the integrated water map, this dense vegetation is recognized as land shown in point C.

From Fig. 46, dense woods exist in all of the areas around points A, B, C and D. Nevertheless, only the dense wood around point C is recognized as land. When SRTM DEM is utilized for a more detailed analysis, the dense wood located around Point C has higher surface elevation (100 meter as average) than the surrounding area (93 meter as average), while the surface elevations of dense woods located around the other three points are similar to the surroundings. This reveals one possible error source of the integration method in woody or urban area with flooding water flowing below and higher surface elevation than the surrounding. Because of the land signals from un-submersed vegetation or urban construction surfaces, water fraction is under-estimated from MODIS DNNS method, resulting in pixels’ maximal water surface elevations from the integration method are lower than actual water levels. With the elevation difference from the surrounding open flooding area, uniformity process also reaches an under-estimated average surface water level. Therefore, cells with higher surface elevation within the woods or urban area are inevitably assigned as land.

Although there is only one image obtained from Worldview-2 for validation, the visual comparison still proves promising result of the integration method and reveals
interesting information on the error sources, which provides an additional supplement to the validation with Landsat TM data.

Figure 46 Image for comparison around Cairo between MODIS integrated water map on May 04, 2011 and World-view 2-m imagery on May 05, 2011 (Upper left: World-view 2-
6.4.2.3 Comparison with river gauge observations

The maximal water surface elevation which is calculated from water fraction and used for integration can actually be viewed as water levels. Therefore, MODIS water surface level retrieved from the integration method can be validated against ground observed water level from river gauges.

In this study, river gauges data are collected during Mississippi flood in 2011 (http://www2.mvr.usace.army.mil/WaterControl/new/layout.cfm). Table 9 lists the comparison at 7 stations between MODIS and ground daily observations at 8:00am on May 04, May 11, May 29 and June 12, 2011. From Table 9, among the 7 stations, Hickman, New Madrid, Tiptonville, Caruthersville and Vicksburg show similar water levels between MODIS and ground observations, while water levels from MODIS show 2-3m lower than ground truth at Cape Girardeau and Thebes. Topography analysis from SRTM DEM shows both Cape Girardeau and Thebes are located in hilly area, while the other 5 stations are located in plains. Around Thebes, the channel of Mississippi River is narrow flowing through steep topographic conditions. Water fraction varies little with the change of water levels, resulting in larger error on water levels retrieved from MODIS.

Another issue is that SRTM DEM has 1-meter vertical accuracy. That means it does
not reflect elevation difference within one meter. This explains why MODIS retrieved water levels are all shown in integers, while water levels from river gauge observations are with decimal digits. Over hilly area, 1-meter accuracy does not matter much on the integration result. However, over plain area, 1-meter change in elevation might bring about substantial difference to the inundated area.

Table 9 Water surface elevation comparison between MODIS and ground observations (Unit: meter)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cape Girardeau</td>
<td>92.86</td>
<td>103</td>
<td>106.39</td>
<td>--</td>
<td>104.37</td>
<td>102</td>
<td>104.06</td>
<td>--</td>
<td>103.89</td>
</tr>
<tr>
<td>Thebes</td>
<td>91.44</td>
<td>102</td>
<td>104.5</td>
<td>--</td>
<td>102.88</td>
<td>98</td>
<td>102.28</td>
<td>--</td>
<td>101.99</td>
</tr>
<tr>
<td>Hickman</td>
<td>80.69</td>
<td>96</td>
<td>96.87</td>
<td>95</td>
<td>96.27</td>
<td>92</td>
<td>92.69</td>
<td>--</td>
<td>90.08</td>
</tr>
<tr>
<td>New Madrid</td>
<td>77.87</td>
<td>92</td>
<td>92.28</td>
<td>90</td>
<td>92.07</td>
<td>89</td>
<td>88.3</td>
<td>--</td>
<td>85.83</td>
</tr>
<tr>
<td>Tiptonville</td>
<td>74.72</td>
<td>89</td>
<td>89.21</td>
<td>88</td>
<td>88.94</td>
<td>85</td>
<td>85.68</td>
<td>--</td>
<td>83.41</td>
</tr>
<tr>
<td>Caruthersville</td>
<td>71.78</td>
<td>85</td>
<td>85.87</td>
<td>--</td>
<td>85.9</td>
<td>82</td>
<td>82.25</td>
<td>--</td>
<td>80.05</td>
</tr>
<tr>
<td>Vicksburg</td>
<td>14.09</td>
<td>28.39</td>
<td>--</td>
<td>30.4</td>
<td>--</td>
<td>30.52</td>
<td>29</td>
<td>28.2</td>
<td></td>
</tr>
</tbody>
</table>

Annotation: -- represents no data. Ground observation is daily observed at 8:00am (rivergauge.com).

From the validation with Landsat TM data, Worldview 2-m imagery and river gauge data, although there are uncertainties over vegetation covers, urban and hilly area, the integration method described in section 6.3 shows promising results. The consistency in shapes and sizes with TM and World-view high resolution water bodies, highly matched rates and similarity in water levels to river gauges indicate the high feasibility of integrating the MODIS 500-m water fraction maps to 30-m-resolution water maps with SRTM 30-m DEM data. Compared with the original 500-m-resolution water fraction maps, the 30-m-resolution integrated water maps are enlarged 18 times, showing far more details in the inundated area. In contrast with high resolution satellite data, the integrated water maps not only share the same spatial resolution but also maintain a more frequent temporal resolution and much larger spatial coverage. These advantages help make the
MODIS data more significant for flood/standing water monitoring and assessment.

6.5 Discussion and Summary

6.5.1 Discussion

Although the integration method shows promising results, errors still exist in the validation results. These errors can be due to the following reasons:

1. *Water detection errors*. First, pixels with a water fraction less than 18% cannot be detected as water with the current decision-tree method, and these missing water pixels are also missing in water fraction maps and integrated water maps (Li et al., 2012). From the validation result on June 12, approximately 22% water pixels went undetected in the MODIS water maps. Second, for flood water, obstacles such as trees, crops or urban construction also decreased the water detection accuracy. Over complicated conditions, such as sun glint or cloud contamination, water detection might result in a low precision, which impacts the water fraction estimation and integration dramatically.

2. *Water fraction calculation errors*. Among all the errors, error from water fraction retrieval is fatal to the integration. Although the errors from the water fraction may be reduced by the uniformity process, the errors still have a strong impact on the average maximal water level, which is the key variable in integration. The integration errors from water fraction can be described in three aspects: first, because the water fraction is calculated based on the water detection result, it
inherits all the errors from the water detection. Pixels with low water fraction which are not detected in water detection maps disappear in water fraction map too. Consequently, these pixels disappear in the integrated water maps, resulting in serration sides of water bodies. Secondly, for woody or urban pixels with flooding water flowing below, the under-estimation of water fractions might not be smoothed by uniformity process if the surface elevations of the woody or urban area are significantly higher than the surrounding open inundated area. Thirdly, over hilly area, water fraction shows slight changes with the water levels, which brings about larger variations on the average maximal water surface levels for integration.

3. The geo-location errors of MODIS data. Because the integration is from moderate resolution to high resolution, a strict co-registration between MODIS data and SRTM DEM is required, and thus MODIS data is demanded in high geo-location accuracy. Generally, MODIS has 50m accuracy at nadir and up to 385m at 55 degree (Wolfe et al., 2002). Because MODIS 500-m pixels spatially match 18×18 30-m cells in the SRTM DEM data, a 1-pixel displacement can cause a substantial shift in the integrated maps and thus decrease the accuracy.

4. The resolution and accuracy of the DEM. The resolution and accuracy of DEM are crucial to the integration result. In low-resolution DEM data, some arbitrary construction works, such as dams and banks, which substantially change the water’s flow direction and accumulation, may not be reflected and thus cause errors in integration. In addition, the average precision of SRTM 30-m DEM
data is one meter. Over plain area, a 1-meter difference in water level can sometimes cause substantial differences in simulated inundated area.

5. The uniformity process is based on the assumption that in the same water polygon within an \( m \times m \) box (\( m = 25 \) in this study), the water level is similar and the difference can be ignored. In some flood cases, however, the difference of water levels cannot be ignored even in the same water polygon, which may produce errors in the calculation of maximal water surface elevation, thus reducing the integration accuracy.

Since microwave satellite data also provides reliable water fraction product with the method in chapter 5, the method developed with MODIS data in this chapter can also be applied in water fraction product from AMSR-E or ATMS in a similar way to derive high resolution water maps. The slight difference lies on that because microwave data has coarse resolution varying from 12.5km to 25km, the uniformity process may be disabled to get average water levels. Additionally, within a microwave coarse resolution pixel, water level might vary as well. Therefore, the integration not only depends more critically on water fraction retrieval accuracy, but also requires consideration of topography change within the pixel.

6.5.2 Summary

In this chapter we demonstrated that it is quite feasible to use an integration method to derive high-resolution water maps from MODIS 500-m water fraction maps and 30-m SRTM DEM data. The validation analysis shows that the matched rates of the integrated
results can be more than 93% compared with TM 30-m water maps, and the maximal water surface elevations calculated from MODIS for integration are very similar to the ground observations from river gauges, especially in plains. Because the new integrated water maps not only have high spatial resolution but also maintain the MODIS high-temporal-resolution and large spatial coverage, they can be more advantageous for users than original moderate-resolution water maps in flood analysis. This chapter can be summarized as follows:

1. The water fraction products from coarse or moderate resolution satellite data provide a potential way to obtain high-resolution water maps with high-resolution DEM data by making use of water’s special fluid features. The precision of water fraction products is crucial to the final integrated result. As he DNNS method described in chapter 4 is used to calculate the MODIS water fraction, the promising integration results indirectly prove the method enjoys high accuracy. However, the retrieval of water fraction is still affected by water detection result. Water pixels with low fractions might be detected as land with decision-tree approach in MODIS imagery. Besides that, shadows from clouds or terrains might also be sources of false positive water detections. The contamination of sun glint not only decreases water detection accuracy, but also disables the assumption that water reflectance in SWIR channel is close to 0 in the DNNS method. All these error sources reduce the accuracy of the integration results.

2. The DEM data lay the foundation for the integration. The resolution and
precision of the DEM not only decides the final resolution of the integrated water maps but also affects the integration accuracy. Generally, the higher the quality of DEM data, the better the integrated result it can derive.

3. The integration method successfully combines MODIS 500-m water fraction maps with SRTM 30-m DEM data to derive high-resolution water maps. The validation show promising results, which suggests the high feasibility of this method to be used in flood analysis. Errors are mainly caused by water detection, water fraction retrieval, the geo-location of satellite data, the quality of the DEM data and the uniformity process on maximal water surface elevation.

Despite the errors in the presented method, this study still succeeds in obtaining high-resolution water maps from coarse- or moderate-resolution satellite data. Because these satellites generally have high temporal resolution and large coverage, it is very possible to obtain real-time high-spatial-resolution water maps over a large area, which is highly significant in flood dynamic monitoring and loss assessments for decision-makers and downstream users. The new application in flood detection provides new ways to make better use of moderate-resolution satellite data in natural disaster analysis.
CHAPTER 7 Algorithm Application in Flood Cases

In this chapter, the algorithms developed in this research are applied to several floods using MODIS, VIIRS and ATMS data. These cases include the severe flood in Galena, Alaska in May, 2013, the Colorado flood in September, 2013 and the disastrous flood in New York due to Hurricane Sandy in 2012. The results are compared with the corresponding false-color satellite images by visual inspection, and some of the results are validated with ground data.

When applying the algorithms to the floods, the decision tree described in Chapter 2 is applied to optical satellite data to extract water pixels to obtain water cover maps, and cloud shadows are then removed from the water maps using the geometric method described in Chapter 3. Based on the water detection results, the water fractions are determined using the DNNS method presented in Chapter 4. The retrieved water fractions are compared with the MOD44W water map to determine the flooding area. For the microwave satellite data, the regression tree approach discussed in Chapter 5 is applied directly to calculate the water fractions. Based on the water fractions, the integrated model developed in Chapter 6 is applied to SRTM DEM data to derive high resolution water maps, parts of which are compared with 30-m SRTM SWBD data to determine the areas of flood water.
7.1 Application in Alaska flood

7.1.1 Background

A severe flood occurred along the Yukon River in Alaska from May 27 to early June, 2013 due to an ice jam. The city of Galena was affected the most by this flood. Most of the town was flooded, and hundreds of people were forced to evacuate. After the flood, Galena was unsafe for a long time due to the destroyed infrastructure, such as power lines. Fig. 47 shows two images of the submerged community of Galena during this flood.

Figure 47 Submerged communities during Galena flood in May, 2013

7.1.2 Near real-time dynamic flood monitoring with VIIRS data

The entire flood was monitored dynamically with Suomi-NPP/VIIRS data using the algorithms developed in this research. Because no high resolution DEM data are available for this area, the flood detection is performed only with the water detection and fraction products. Figs. 48 to 53 show a series of VIIRS false-color images and the
corresponding flood detection maps from May 27 to June 01, 2013. In the flood detection
maps, blue represents normal water from MOD44W, white represents snow/ice, light
gray represents clouds, dark gray represents cloud shadows, and colors from cyan to red
represent different water fractions of flooding.

Fig. 48 shows a long section of the Yukon River near Galena that was covered with
ice from the flood detection map on May 27, 2013. Ice in the eastern section was mostly
melted. Water flowed out of the riverbed to the east of Galena due to the ice jam. The
flood could be identified from the VIIRS false color images. With the flood detection
algorithms developed in this research, flood water was detected at water fractions from
60% to 100%. At this time, the flood water was confined to a small area, and city of
Galena was still safe.

The flooding progressed rapidly. Late on the night of May 27, the residents of
Galena were forced to evacuate, and most of Galena was under water by the morning of
May 28. Fig. 49 shows a much larger area of flood water detected from the VIIRS flood
detection map. Most of the water fractions near Galena are close to 100%. The largest
flooding occurred on May 29. Fig. 50 shows VIIRS data with large areas of flood water
near Galena. The largest area of the flood was estimated to be approximately 18 miles
long. In addition to the flooding along the Yukon River, flooding also occurred along the
Koyukuk River because of an ice jam. Afterward, the ice downstream gradually melted,
and the flood water began to retreat. The VIIRS images from May 30 (Fig. 51) show that
although a large area of flood water remained near Galena, it had decreased substantially
and showed smaller water fractions. On May 31 (Fig. 52), the VIIRS data did not obtain
much ground information near Galena due to clouds. However, most of the flood water had retreated by June 1 (Fig. 53), and only very small residual flooding was caught in the VIIRS images.

Figure 48 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 27, 2013
Figure 49 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 28, 2013
Figure 50 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 29, 2013
Figure 51 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 30, 2013
Figure 52 VIIRS 375-m false color image and the corresponding flood detection map near Galena on May 31, 2013
Although no high resolution Landsat satellite images were obtained for comparison, the VIIRS flooding detection results were validated with photos and news reports. The spill and retreat times and locations recorded in the VIIRS observations and the media reports were consistent. The comparisons of visual inspections with the VIIRS false color images also showed good accuracy and the consistency of the flood detection results. The VIIRS data had an advantage over visual inspections in the detection of this kind of
7.2 Application in Colorado flood case

7.2.1 Background

Starting on September 9, 2013, catastrophic flooding occurred along Colorado's Front Range from Colorado Springs north to Fort Collins due to intense rainfall from a slow-moving cold front that stalled over the region. The situation intensified on September 11 and 12. The South Platte River, which is a very small river and is mostly invisible in MODIS 250-m or VIIRS 375-m imagery, suffered the most severe flooding of the past 30 years.

The flood covered a large area and caused at least eight deaths, with two more missing and presumed dead and hundreds remaining unaccounted for. More than 11,000 people were evacuated. Nearly 19,000 homes were damaged, and more than 1,500 were destroyed. The Colorado Department of Transportation estimated that at least 30 state highway bridges were destroyed and that an additional 20 were seriously damaged. Low-lying agricultural land in northeast Colorado was severely affected with huge economic losses. The flood also caused hazardous impacts, including oil spills and the shutdown of hundreds of oil and gas wells.

7.2.2 Flood detection in Colorado using VIIRS data

Flood detection was performed with the developed algorithms in near real-time to
provide flood monitoring products to the National Oceanic and Atmospheric Administration (NOAA) using Suomi-NPP/VIIRS data.

Fig. 54 and Fig. 55 show two sets of images along the South Platte River on September 14 and September 17, 2013, respectively. Fig. 54 shows water pixels that represent the western section of the South Platte River in a VIIRS 375-m false color image (Fig. 54 top). The flood detection algorithms developed in this research identified most of these water pixels as flood water. On September 17, 2013, the flood had continued much further downstream. In the VIIRS false color image (Fig. 55, top), most of the South Platte River became visible in more water pixels. The flood detection map (Fig. 55, bottom) showed consistent results with the flood water along most of the South Platte River.

Although the 375-m VIIRS data showed good results for this flood, the resolution is still too coarse to define the flood clearly. To obtain additional details of this flood, the integrated model developed in this research was applied with the SRTM 30-m DEM of the area. A series of 30-m flood maps were generated and overlapped on Google Earth with a transparency process to make the underlying background visible. Figs. 56 to 61 present a series of 30-m VIIRS water maps from the integrated model along the South Platte River using the VIIRS water detection results from September 17, 2013 and the SRTM 30-m DEM data. Compared to the original 375-m VIIRS flood detection results, these high resolution water maps showed many more details of the submerged area and clarified the flood area. These images received much attention from end users and were used by NOAA in reports of the flood to the federal government.
The integrated 30-m VIIRS water maps were also compared and validated with aerial photos from the website http://www.denverpost.com/2013coloradofloods/ci_24132375/2013-colorado-flood-photo-video-map. Figs. 62 to 68 show a series of VIIRS 30-m water maps and photos and photo maps taken at the same time and covering the same area. Although small differences are present between the two sets of images, the flood water bodies in the VIIRS 30-m water maps are consistent with those in the photos and photo maps. This indicates that the integrated model used to derive the high resolution water maps from moderate resolution satellite data has promising accuracy and stability.
Figure 54 VIIRS 375-m false color image and the corresponding flood detection map along South Platte River on Sep. 14, 2013
Figure 55 VIIRS 375-m false color image and the corresponding flood detection map along South Platte River on Sep. 17, 2013
Figure 56 A whole look at VIIRS 30-m water map on Sep. 17, 2013 along South Platte River overlapped on Google Earth (light purple color is flooding water)

Figure 57 VIIRS 30-m water map on Sep. 17, 2013 near Evans overlapped on Google Earth (light purple color is flooding water)
Figure 58 VIIRS 30-m water map on Sep. 17, 2013 near Garden City overlapped on Google Earth (light purple color is flooding water)
Figure 59 VIIRS 30-m water map on Sep. 17, 2013 near Orchard overlapped on Google Earth (light purple color is flooding water)

Figure 60 VIIRS 30-m water map on Sep. 17, 2013 near Weldona overlapped on Google Earth (light purple color is flooding water)
Figure 61 VIIRS 30-m water map on Sep. 17, 2013 near Snyder overlapped on Google Earth (light purple color is flooding water)

The integrated 30-m VIIRS water maps were also compared and validated by photos using as ground truth taken by planes which were from the website: http://www.denverpost.com/2013coloradofloods/ci_24132375/2013-colorado-flood-photo-video-map. Fig. 62 to Fig. 68 demonstrate a series of VIIRS 30-m water maps and photos or photo maps taken at the similar time and covering the similar area. Although small difference was still seen between the two sets of images, flooding water bodies in the VIIRS 30-m water maps were distributed and located in high consistency to that in the photos or photo maps. This indicates the integration model to derive high resolution water maps from moderate resolution satellite data is with promising accuracy and stability.
Figure 62 VIIRS 30-m water map near Evans (Left) and a photo taken on Sep. 16, 2013 covering the same area.

Figure 63 VIIRS 30-m water map in 37th St near Evans (Left) and a photo taken on Sep. 16, 2013 covering the same area.
Figure 64 VIIRS 30-m water map in communities Evans (Left) and photos taken on Sep. 13, 2013 covering the same area
Figure 65 VIIRS 30-m water map in the 37th St of Evans (Left) and a photo map taken on Sep. 16, 2013 covering the same area
Figure 66 VIIRS 30-m water map near County Road 61 (Left) and a photo map taken on Sep. 16, 2013 covering the same area.

Figure 67 VIIRS 30-m water map near US 34 highway (Left) and a photo map taken on Sep. 16, 2013 covering the same area.
Figure 68 VIIRS 30-m water map near US 52 highway (Left) and a photo map taken on Sep. 16, 2013 covering the same area

7.3 Application in New York flood

7.3.1 Background

Hurricane Sandy was a tropical cyclone that devastated portions of the Mid-Atlantic and Northeastern United States in late October 2012. The eighteenth named storm and tenth hurricane of the 2012 Atlantic hurricane season, Sandy was the largest Atlantic hurricane on record as measured by diameter, with winds spanning 1,100 miles (1,800 km). Sandy was estimated to have caused at least $20 billion of damage. Preliminary estimates of losses that include business interruptions exceed $50 billion (2012 USD), which, if confirmed, would make Sandy the second-costliest Atlantic hurricane in history, behind only Hurricane Katrina.

The powerful storm transformed some of Atlantic City’s streets into rivers and
inundated parts of Lower Manhattan, forming whitewater cascades at Ground Zero and swamping New York’s financial district.

7.3.2 Flood detection in New York City with VIIRS data

During Hurricane Sandy, the flood was closely monitored with both MODIS and VIIRS data. However, New York City had overcast conditions for most of the storm, and the first clear-sky image was obtained both from MODIS and VIIRS on November 4, 2012. By this time, most of the flood had retreated, and not much flood water was observed.

Fig. 69 presents a VIIRS 375-m false color image and the corresponding water fraction map developed using the algorithms developed in this research. A comparison of Fig. 69 with water maps from before the flooding shows that higher water fractions were only found along Hudson Bay and along the southern coast of Long Island. To obtain additional details, the integrated model was applied. The 30-m VIIRS water map was generated and is shown in Fig. 70, with light blue indicating normal water from the SWBD product. Fig. 70 shows only a few areas of flood water (shown in red). The main area of flood water at this time was mostly distributed along the southern coast of Long Island. The integrated water map was compared with the flood map simulated with hydrological models and the evacuation map provided by Google. Although not much flood water was found in the VIIRS 30-m water map, the overall distributions were consistent.

However, the flood detection using optical satellite data in the New York area
highlights the shortage of data during cloudy conditions. Microwave-based satellite data can fill this gap. Water fractions were calculated from 15-km ATMS data collected on November 1, 2012 around New York City. It is difficult to obtain detailed information about the flood from the original water fraction results because of the coarse resolution. Therefore, the retrieved water fractions were applied to the integrated model and the SRTM DEM data to generate a 30-m resolution water map. Fig. 72 shows the integrated 30-m resolution flood map for New York City from the ATMS data, which provided exciting results. The flood was ongoing in New York City at this time. Large areas of flood water are present along Hudson Bay and are shown in great detail. The distributions of the flood water are consistent with the flood map from hydrological models.

The good results demonstrate the ability to derive high resolution water maps using the integrated model that is based on the retrieved water fraction results even from very coarse resolution microwave data such as ATMS. In the future, the combination of optical satellite data and microwave satellite data might be a good tool to improve the applications of satellite data to hydrology.
Figure 69 VIIRS 375-m false color image and the corresponding water fraction map around Near York City on Nov. 04, 2012
Figure 70 VIIRS 30-m water map around New York City on Nov. 04, 2012

Figure 71 Simulated inundation map from hydrological models (left), VIIRS 30-m water map on Nov. 04, 2012 (middle) and Google Flood Evacuation Map due to hurricane Sandy (right) around New York City
7.4 Discussion and Summary

The algorithms developed in this research were also applied to other floods, such as the floods in Russia and China between late August and September, 2013 and have shown consistent and promising results. The successful application to floods demonstrates that the algorithms have good accuracy and are highly feasible for operational use. However, the specific applications also revealed potential problems with this research.

First, optical satellite data still serve as the main data for flood detection, and the application of microwave-based data requires improvement in both the microwave water fraction retrieval algorithm and the integrated algorithm for microwave data.

Second, although the integrated model provided exciting and promising results, the algorithm can be improved for even better performance. A water level product may be
generated prior to generating the high resolution water map.

Third, the problem of how to combine optical and microwave-based satellite data or products to obtain final flood detection products still remains unaddressed. This will require further research in the future.
CHAPTER 8 Summary and Future Work

8.1 Dissertation Summary

This dissertation focuses on the development of an enhanced high resolution flood detection product using coarse-to-moderate resolution satellite data along with a series of secondary datasets. Several algorithms are developed to obtain a series of flood detection products. The research can be summarized as follows:

First, an automatic water identification algorithm was developed that uses a decision tree to discriminate water from other objects using optical satellite data, including MODIS and VIIRS. Large samples were collected to cover various ground and atmospheric conditions. Seven decision trees were generated and applied to water detection.

Second, because of the spectral and radiometric similarity between cloud shadows and mixed water pixels, a geometric method was developed to remove cloud shadows from the water maps. This substantially improved the accuracy of water detection from optical satellite data.

Third, because water fraction is a key parameter used to derive high resolution water maps, accurate and steady water fraction retrieval is crucial to guarantee the accuracy of the final product. Therefore, water fraction retrieval with data from both optical and microwave-based satellites is an important part of this research. For optical satellite data, including MODIS and VIIRS, a quantitative method (DNNS method) that used a SWIR channel instead of a traditional histogram method was developed by
considering the mixture of water pixels. For microwave-based satellite data, such as AMSR-E and ATMS, a water fraction algorithm was developed using a regression-tree approach and considering land mixture and the impacts of precipitation and cloud fractions. Both algorithms show promising and consistent results.

Based on the water fraction products from the MODIS, VIIRS, AMSR-E and ATMS data, an integrated model was developed to derive 30-m resolution water maps using SRTM 30-m DEM. The final high resolution water maps which were validated and have high accuracy are more detailed than the original water fraction products.

Finally, the flood products developed in this research were applied to several severe floods, including the New York flood due to Hurricane Sandy in late October 2012, the Galena flood in late May 2013, and the Colorado flood in September 2013. These applications of the algorithms provided many good products for decision-makers and also had good responses from end-users.

8.2 Main Innovation of this Dissertation

The research in this dissertation provides good solutions to several key problems for automatic flood detection with coarse-to-moderate satellite data. This dissertation is innovative in four main areas.

The first problem that was solved by this research is cloud shadow removal. Most cloud shadow removal algorithms in previous studies are based on the spectral difference between cloud shadows and the water surface. However, the two have similar spectral and radiometric properties, so it is impossible to completely discriminate one from the
other based on the spectral or radiometric characteristics. In this dissertation, geometric relationships over both an ideal plane and a spherical surface are established between the clouds and the cloud shadows by considering parallax and earth curvature. The algorithm accurately removes cloud shadows under different conditions and can be applied to different types of satellite imagery, including MSG/SEVIRI, EOS/MODIS and Suomi-NPP/VIIRS. In particular, the algorithm can be used without an existing cloud mask. The development of a highly accurate cloud shadow removal algorithm substantially improves the water detection accuracy, making it possible to detect mixed water pixels with low water fractions.

The second problem that was solved is water fraction retrieval with optical satellite data. For optical satellite data, the traditional histogram method ignores the mixture in a mixed water pixel and thus results in significant errors when the sub-pixel land portion is different from the surrounding land pixels. The dynamic nearest neighbor searching method (DNNS) is used to correct this problem and uses the SWIR channel to obtain the relationship between the sub-pixel land portion of a mixed water pixel and the surrounding land pixels. The method shows more consistent results than the traditional histogram method in water fraction retrieval using MODIS and VIIRS data and has a high potential for operational applications.

The third problem that was solved is the determination of water fraction from microwave satellite data. The algorithm developed in this research was based on the regression tree approach and considers a series of factors, including land mixture, precipitation and cloud fraction, which have generally been ignored in previous studies.
The algorithm successfully discriminates water pixels from pure land automatically and continuously retrieved the water fractions.

The last and most important problem is the development of an integrated model to derive high resolution water maps from water fraction products generated from coarse-to-moderate resolution satellite data. This integrated model successfully links these satellite data with high resolution DEM data and overcomes the limitation of these satellite data in hydrology applications due to their low spatial resolution. The derived water maps not only have high spatial resolution but also benefit from the large swath coverage and high temporal resolution of the coarse-to-moderate resolution satellite data. In addition, the integrated model can estimate water levels from these satellite data. Because water level is an important variable in hydrological models, the algorithm improves the applicability of coarse-to-moderate resolution satellite data in hydrology.

8.2 Future work

Although this research has solved several problems associated with automatic flood detection with coarse-to-moderate resolution satellite data, several problems remain unsolved and require further work. Future research will address the following topics.

First, the algorithm for water fraction retrieval with microwave-based radiation requires improvement. The developed algorithm using the regression tree provides consistent results, but the overall accuracy is not satisfactory. The impact of cloud cover can affect the results substantially, and clouds are not currently considered in sufficient detail. In the future, cloud types and cloud phases will be considered instead of simple
cloud fractions.

Second, the integrated model can be improved. The current algorithm works well in most situations. However, the algorithm depends greatly on the water level or the maximum water surface height. When calculating this variable, the original topography of rivers was not considered. The integrated model also suffers from a disconnection problem; when rivers are invisible in the satellite imagery, the corresponding water bodies are also invisible in the high resolution water maps. However, the calculation of an accurate water level will solve this problem.

Third, the integrated model with the microwave data is more complicated than the model with the optical satellite data due to the very coarse spatial resolution. However, the current algorithm has not considered this problem sufficiently. More work is required to achieve consistent performance under various conditions.

Finally, the combined use of optical and microwave satellite data or products can be more significant. Currently, the flood products are generated separately. In the future, more work will be done to combine the two results into one high resolution water map. Such a product can be use microwave satellite data over areas of cloud cover and optical satellite data for clear-sky conditions.
REFERENCES


Cao, S., W., Dong, and Q., Xiao, 1987. Applying meteorological satellite imagery to monitor Laohe River flood, Remote Sensing Information, 12–16


201
International Journal of Applied Earth Observation and Geoinformation, 9, 247-263


and Western Pacific, Meteorology and Climate (Hong Kong: Hong Kong
Meteorological Society), pp. 46

optical remote sensing to monitor flood inundation in support of hydrologic
modeling, 18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July

Remote Sens., vol.17, pp. 733–748

Detection and Removal of Clouds and their Shadows from Landsat TM Images.

extent in a coastal floodplain using Landsat TM and DEM data. International
Journal of Remote Sensing, 23:18, 3681-3696

Satellite Microwave remote sensing of contrasting surface water inundation
changes within the Arctic-Boreal Region, Remote Sensing of Environment, 127
(2012): 223-236

Mississippi river floods by the NOAA-2 Satellite. Water Resources Bulletin,
10(5), 1040-1049

Resources Bulletin, 10(5), 1050-1059


Sanmei Li grew up in China. She attended Nanjing University in China, where she received her Bachelor of Geography Information System in 2001. From 2001, she worked in China Meteorological Administration. In 2004, she began her Master degree study in Peking University of China, where she received her Master of Atmospheric Physics and Atmospheric Environment in 2008. In 2011, she came to the USA for her PhD study in Department of Geography and GeoInformation Science of George Mason University. She will receive her Doctorate in 2013 and continue to work on the JPSS project on flood detection.