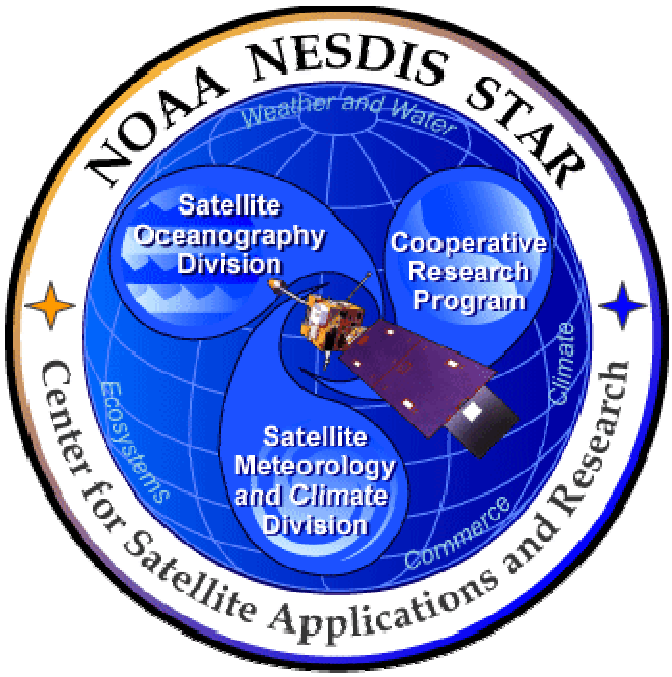

NOAA NESDIS
CENTER for SATELLITE APPLICATIONS and RESEARCH

GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document For Flood/Standing Water

Donglian Sun, George Mason University
Rui Zhang, George Mason University
Sanmei Li, George Mason University
Yunyue Yu, NOAA/NESDIS/STAR

Version 1.0

September, 2010

UNIT TEST PLAN DOCUMENT GUIDELINE
VERSION HISTORY SUMMARY

Version	Description	Revised Sections	Date
0.1	New ATBD Document according to NOAA /NESDIS/STAR Document Guideline		8/31/2008
0.2	ATBD Document according to NOAA /NESDIS/STAR Document Guideline		9/30/2008
1.0	ATBD Document 80% readiness	Section 2.4, Section 3.4, Section 4.3, Section 4.4, Section 5.4, Section 5.5, Section 6.3	6/30/2010

TABLE OF CONTENTS

Page

TABLE OF CONTENTS 5

LIST OF FIGURES 7

LIST OF TABLES 9

LIST OF ACRONYMS 11

ABSTRACT 13

1. INTRODUCTION 15
   1.1. Purpose of This Document 15
   1.2. Users of This Document 15
   1.3. Scope 15
   1.4. Related Documents 16
   1.5. Revision History 16

2. SYSTEM OVERVIEW 17
   2.1. Objectives of Standing Water Algorithm Development 17
   2.2 Instrument Characteristics 18
   2.3 Mission Requirement ..... **Error! Bookmark not defined.**
   2.4 Retrieval Strategies ..... **Error! Bookmark not defined.**

3 ALGORITHM DESCRIPTION 21
   3.1 Algorithm Overview 21
   3.2 Processing Outline 22
   3.3 Algorithm Input 23
      3.3.1 Primary Sensor Data 23
      3.3.2 Derived Sensor Data 23
      3.3.3 Ancillary Data 24
   3.4 Theoretical Description 24
      3.4.1 Physics of the Problem 24
      3.4.2 Mathematical Description of the FSW Algorithm 31
      3.4.3 The GOES-R ABI Forward Simulations 33
      3.4.4 Algorithm Selection ..... **Error! Bookmark not defined.**
      3.4.5 Algorithm Challenges 35

3.4.5.1 Cloud shadow effect 35
3.4.5.2 Anisotropic effect 36
3.4.5.3 Sun glint contamination 37
3.4.6 Cloud Shadow Detection ..... **Error! Bookmark not defined.**
3.4.6.1 Basic knowledge of shadows ..... **Error! Bookmark not defined.**
3.4.6.2 Basic knowledge of viewing geometry . **Error! Bookmark not defined.**
3.4.6.3 Geometrical relationship for cloud shadow **Error! Bookmark not defined.**
3.4.6.4 Some results for cloud shadow detection **Error! Bookmark not defined.**
3.5 Algorithm Output 40
4 TEST DATA SETS AND OUTPUTS 43
4.1 Simulated and Proxy Input Data Sets 43
4.2. Algorithm Testing with the Simulated Data 43
4.3 Algorithm Testing with the Proxy Observation Data 44
4.3.1 Test with the Proxy MODIS data 44
4.3.1.1 Data used for training 44
4.3.1.2 Results from the MODIS training data 45
4.3.1.2 DT training with additional Surface Temperature (ST) data 49
4.3.1.4 “Future prediction” with the rules obtained from training 52
4.3.2 Test with the proxy SEVIRI data 57
4.3.2.1 Data used for training 57
4.3.2.2 “Future prediction” with the rules obtained from training 58
4.4 Algorithm Validations 62
4.5 Cloud Detection on GOES-R ABI Data 65
5. PRACTICAL CONSIDERATIONS 67
5.1 Numerical Computation Considerations 67
5.2 Programming and Procedural Considerations 67
5.3 Configuration of Retrieval 67
5.4 Quality Assessment and Diagnostics 67
5.5 Exception Handling 68
6. ASSUMPTIONS AND LIMITATIONS 69
6.1 Assumptions 69
6.2 Limitation 69
6.3 Pre-launch Product Improvements 69
6.3.1 Improved Validation Methods 69
6.3.2 Algorithm Improvement 70
7. REFERENCES 71

LIST OF FIGURES

Page

Figure 3.1. Products and dependencies of the land algorithm module. 21

Figure 3.2. Flowchart of GOES-R ABI Flood/Standing Water algorithm. 22

Figure 3.3. Reflectance spectra of some land surface species. 25

Figure 3.4. MODIS 1km RGB (7, 2, 1) composite images in Namibia of Africa on December 26, 2007 before (upper) and on February 4, 2008 after (lower) flooding..... **Error! Bookmark not defined.**

Figure 3.5. Histograms of MODIS surface reflectance at CH1, CH2, and CH6, (CH2-CH1), CH2/CH1, NDVI, NDWI, and Land/Water samples (Sun and Yu, 2010). 29

Figure 3.6. Radiative transfer simulation procedure**.**..... 33

Figure 3.7. Scatter plots of NDVI vs. MODIS CH1 (left), and CH2 (right). ..... **Error! Bookmark not defined.**

Figure 3.9. Three- channel composite images in the same area at different time on December 26,2007..... 37

Figure 3.10. Three- channel composite images in the same area at different time on December26, 2007..... 38

Figure 3.11 Illustration of sun glint in an optical imagery, showing sun glint on the right hand side (upper), and sensor radiance plotted for 4 wavebands along the line marked in (upper) (from Kay et al., 2009)..... 39

Figure 4.1. MODIS Channel 7, 2, and 1 RGB composite images**. Error! Bookmark not defined.**

Figure 4.2. MODIS land cover at 1 km resolution and reflectance data at 0.648 μm (MODIS CH1) and 0.858 μm (MODIS CH2) on May 28, 2008 before flooding were used for training. 45

Figure 4.3. An example of decision tree structure derived from the MODIS observations with the C4.5 algorithm. 46

Figure 4.4. An example of Decision Tree structure derived from the BRDF adjusted surface reflectance data with the samples around the Mississippi River using the C4.5 algorithm. 48

Figure 4.5. MODIS land surface temperature over the training area on sate date (May 28, 2008). 49

Figure 4.6. Spatial distributions of water (blue)/land (green) identifications over time..... 53

Figure 4.7. The maximum flood extent map during the flood period (original water body isshown in blue, flood is shown in red, and land is shown in green)..... 54

Figure 4.8. MODIS RGB composite images on August 27 (a) and 30 (b), 2008..... 55

Figure 4.9. Water identification map on August 27 (upper) and 30 (lower), 2005 (water: blue; Green: Land or Non-water; white: clouds) 56

Figure 4.10. Flood map on August 30, 2005 shown as the water difference between August 30 (after flooding) and August 27 (before flooding) (original water body is shown in blue, flood is shown in red, and land is shown in green)..... 57

Figure 4.11. An example of decision tree structure derived from the SEVIRI observations with the C4.5 algorithm. 58

Figure 4.12. Water (blue)/Land (green) derived from SEVIRI observations at 15:15 UTC on12/25/2007. 59

Figure 4.13 Flood (Red), Water (Blue) and Land (Green) derived from SEVIRI observations at 12:00 UTC on 3/8/2010 60

Figure 4.14 MODIS RGB composite images on March 8 (upper) and March 12 (lower) 61

Figure 4.15 Flood map on 12:00 UTC, March 8, 2010. (original water body is shown in blue, flood is shown in red, land is shown in green and white pixels are cloud). 62

Figure 4.16. Water/land classification from the C4.5 DT algorithm with the MODIS surface reflectance data on 05/20/2008 (left) and ground truth of water/land classification (right) 63

Figure 4.17. Water/land classification from the C4.5 DT algorithm with the SEVIRI data at 11:45 UTC on 12/25/2007 (left) and Ground truth of water/land classification (right) 63

LIST OF TABLES

Page

Table 2.1. Spectral characters of Advanced Baseline Imager. 19

Table 2.2. GOES-R mission requirements for Standing Water. ....**Error! Bookmark not defined.**

Table 3.1. Input list of sensor data. 23

Table 3.2. Input list of derived sensor data. 24

Table 3.3 Input of ancillary data. 24

Table 3.4. Vegetation classification according to typical NDVI values. 27

Table 3.5. Solar zenithangle bins used for ABI forward simulations. 34

Table 3.6. Satellite zenith angle bins used for ABI forward simulations. 34

Table 3.7. Azimuth angle bins used for ABI forward simulations. 34

Table 3.8 Algorithm output data. 40

Table 3.9. Definition of the FSW product Quality Control flags. 40

Table 3.10. Metadata defined for the FSW product file 40

Table 4.1. Similarity of MODIS and SEVIRI Imagers to ABI in channel spectrum. 43

Table 4.3. The confusion matrix from the C4.5 test. 47

Table 4.4. The confusion matrix from the C4.5 test. 52

Table 4.5. Validation results from the proxy MODIS data..... 64

Table 4.6. Validation results from the proxy SEVIRI data..... 65

LIST OF ACRONYMS

ABI	Advanced Baseline Imager
ACM	ABI Cloud Mask
AIT	Algorithm Integration Team
ATBD	Algorithm Theoretical Base Document
AWG	Algorithm Working Group
CONUS	Continental United States
EOS	Earth Observation Systems
FD	Full Disk
GFS	Global Forecast System
GOES	Geostationary Operational Environmental Satellite
GS-F&PS	Ground Segment Functional and Performance Specification
IMS	Ice Mapping System
IPT	Integrated Product Team
IR	Infrared
ST	Surface Temperature
MODIS	Moderate Resolution Imaging Spectroradiometer
MRD	Mission Requirement Document
MSG	Meteosat Second Generation
NCEP	National center for Environmental Prediction
NESDIS	National Environmental Satellite, Data, and Information Service
NOAA	National Oceanic and Atmospheric Administration
NPOESS	National Polar-orbiting Operational Environmental Satellite System
OCD	Operations Concept Document
QA	Quality Assurance
RTM	Radiative Transfer Model
SEVIRI	Spanning Enhanced Visible and Infrared Imager
SRF	Spectral Response Function
SST	Sea Surface Temperature
STAR	Center for Satellite Applications and Research
SWA	Software Architecture Document
NIR	Near Infrared
TOA	Top Of Atmosphere
UMD	University of Maryland
GMU	George Mason University
VIIRS	Visible/Infrared Imager /Radiometer Suite
VVP	Verification and Validation Plan

ABSTRACT

This Algorithm Theoretical Basis Document (ATBD) describes in detail the procedures for developing and using a flood/standing water (FSW) algorithm designed for the GOES-R Advanced Baseline Imager (ABI). It includes a description of the requirements and specifications of the FSW products and some specific information about the ABI that is relevant to the derivation of the FSW products. The main part of the ATBD is a description of the science of the proposed ABI FSW algorithm. The process of algorithm selection is documented and includes a review of satellite FSW research. The simulated radiances were calculated using sensor spectral response functions (SRF) that are expected from the actual ABI instrument. A description of the expected implementation of the FSW algorithm is provided and ancillary data sets needed for the FSW calculation are listed.

In order to find the best DT approach, eleven decision tree (DT) algorithms, including the J48graft or J48, which is based on the C4.5, NBTree (a Naïve Bayes/Decision Tree hybrid), Random Tree, Random Forest, REP Tree, BFTree, Decision Stump, FT (final tree), and CART (Classification and Regression Trees) were adapted from the literature for evaluation as the Day 1 GOES-R FSW algorithm. The algorithm with the best accuracy and easiest implementation, the J48 or C4.5, was selected as our baseline algorithm.

The selected algorithm was applied to the proxy MODIS and SEVIRI data. The retrieved FSW preliminary product: yes/no water detections were compared against independent ground truth data and the results were analyzed. The properties of the algorithm were examined for selected cloud shadow conditions, time of day and illumination/observation geometry effects, and a variety of surface types. The algorithm was found to meet specs with the test data sets. Perfectly cloud free data is assumed in all testing of the ATBD research. A process for routine evaluation of the operational GOES-R FSW is described; this includes an automated cloud shadow detection algorithm, routine matchups against ground truth and methodology for product evaluation. Finally, practical matters, such as computer resources, instrument performance and its effects on the product are considered.

1. INTRODUCTION

The purpose, users, scope, related documents and revision history of this document are briefly described in this section. Section 2 gives an overview of the standing water retrieval objectives and operations concept. Section 3 describes the baseline algorithm, its input data requirements, the theoretical background, sensitivity analyses, and error budgeting. In section 4, testing cases were presented using MODIS data and SEVIRI data as proxies. Some practical considerations are described in Section 5, and some assumptions and limitations associated with the algorithm are described in section 6. Finally, Section 7 presents the references cited.

1.1. Purpose of This Document

This Algorithm Theoretical Basis Document (ATBD) explains the physical and mathematical background for an algorithm to derive the standing water/flood product as a part of the requirements for the Advanced Baseline Imager (ABI). ABI is the primary visible and infrared instrument to be flown aboard the platform of the Geostationary Environmental Operational Satellite (GOES) R series (GOES-R) of NOAA meteorological satellites. This document provides an overview of the required input data, the physical and mathematical backgrounds of the described algorithm and its predicted performance, sensitivity study of the algorithm, practical considerations, and assumptions and limitations.

1.2. Who Should Use This Document

The intended users of this document are those interested in understanding the physical basis of the standing water algorithm and how to use the output of this algorithm for a particular application. This document also provides information useful to anyone maintaining or modifying the original algorithm.

1.3. Inside Each Section

This document covers the theoretical basis for the derivation of the Standing Water product from ABI data. It is broken down into the following main sections:

- **System Overview**: provides objectives of the Standing Water algorithm, relevant details of the ABI instrument, and a brief description of the product requirements.

- **Algorithm Description**: provides all the detailed description of the algorithm including its physical basis, its input and its output.

- **Assumptions and Limitations**: provides an overview of the current limitations of the approach and gives the plan for overcoming these limitations with further algorithm development.

1.4. Related Documents

This document may contain information from other GOES-R documents listed in the website provided by GOES-R algorithm working group (AWG):
http://www.orbit2.nesdis.noaa.gov/star/goesr/index.php.

In particular, readers are directed to read these documents for a better understanding of this ATBD:

GOES-R Series Ground Segment Functional and Performance Specification
GOES-R Series Mission Requirements Document
GOES-R Algorithm Theoretical Base Document for ABI Cloud Mask
GOES-R Land Surface Team Critical Design Review

Other related references are listed in the Reference Section.

1.5. Revision History

Version 0.1 of this document was created by Dr. Donglian Sun of GMU, with its intent being to accompany the delivery of the version 0.5 algorithms to the GOES-R AWG Algorithm Integration Team (AIT)**.** The document was then revised following the document guideline provided by the GOES-R Algorithm Application Group (AWG) before the version 0.2 delivery. In 2010 summer, version 1.0 of the document was prepared by Dr. Donglian Sun and Dr. Rui Zhang, which includes some new results conducted from the algorithm Critical Design Review (CDR) and the Test Readiness Review (TRR), as the algorithm 80% readiness document.

2. SYSTEM OVERVIEW

This section will describe objectives of the Standing Water algorithm, details of the ABI instrument, and the product requirements.

2.1. Products generated

Floods are usually disastrous events occurring in short period of time. For this reason, satellite-derived flood maps, available in near-real time, are invaluable to stake holders and policy makers for disaster monitoring and relief efforts. Precise mapping of the floods/standing water is also required for detecting deficiencies in existing flood control and for damage claims afterwards.

Satellite sensors used in river/flood studies may be classified into two types: (1) passive, in which the sensor receives energy naturally reflected by or emitted from the earth's surface; and (2) active, in which the sensor provides its own illumination and records the amount of incident energy returned from the imaged surface (Smith, 1997). Passive sensors include all of the visible and infrared instruments such as the Landsat Thematic Mapper (TM) and Multi-Spectral Scanner (MSS), the Advanced Very High Resolution Radiometer (AVHRR), the Satellite Pour l'Observation de la Terre (SPOT) and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Moderate-Resolution Imaging Spectroradiometer (MODIS) and Landsat-7 sensors. The Advanced Baseline Imager (ABI) aboard future GOES-R belongs to this type of sensor.

In microwave spectrum, passive sensors such as the Special Sensor Microwave/Imager (SSM/I) aboard the defense meteorological satellites and active (radar) sensors such as RADARSAT (Bonn and Dixon, 2005) are excellent tools for monitoring floods since they can penetrate clouds, which usually occur during flood periods, and measure the microwave energy naturally emitted from the Earth's surface. If absent of vegetation or tress, radar returns are usually low over the smooth open water surface. This characteristic allows flood extent to be determined with good accuracy under multiple conditions. However, turbulence, wind-induced waves, vegetation and/or tress mixed with flooded water, can all cause significant increases in radar back-scattering, making inundation extent difficult or impossible to determine. Moreover, interpretation of SAR imagery is less straightforward than it is for the visible/infrared range.

As an advanced visible and infrared imager aboard the next generation GOES-R system, the potential of GOES-R ABI data in large area flood monitoring should not be ignored. Although the ABI offers coarser spatial resolution than many polar-orbiting sensors such as MODIS and LANDSAT, the high temporal resolution of geostationary satellites (5 minutes for the ABI), make them very useful for dynamic monitoring of flood events, because they usually occur quickly. Compared to the previous GOES series, the new near-infrared channel 3 (0.86 µm) of the GOES-R ABI makes it suitable for monitoring Standing Water/Floods. In the age of climate change, severe floods appear to be occurring more frequently than in years past. This makes the GOES-R ABI observations more attractive in dynamic flood monitoring.

In the GOES-R program, the Flood/Standing Water Identification accuracy requirement is 60% for all the four ABI scanning modes (i.e., full disk, hemisphere, CONUS, and mesoscale). A primary objective of the GOES-R Flood/Standing Water development team is to provide a state-of-the-art Flood/Standing Water identification algorithm that meets the GOES-R mission requirement.

The Flood/Standing Water requirements defined by the mission requirement document (MRD) and the Ground Segment Functional and Performance Specification (GS-F&PS) are listed in Table 2.2.

2. GOES-R mission requirements for Standing Water.

Observational Requirement	LEVEL[1]	Geographic Coverage[2]	Horiz. Res.	Mapping Accuracy	Msmnt. Range	Msmnt. Accuracy	Refresh Rate	Data Latency	Long-term Stability	Extent Qualifier
Flood/Standing Water: Hemisphere	T	FD	1 km	5 km	0-1	Probability of correct classification to 60%	60 min	3 min	TBD	LZA <67
	G	*FD*	*1 km*	*1 km*	0-1	*60%*	*15 min*	*1 min*	*TBD*	LZA <70
Flood/Standing Water: Mesoscale	T	M	1 km	1 km	0-1	Probability of correct classification to 60%	60 min	3 min	TBD	LZA <67
	G	*M*	*1 km*	*0.5 km*	0-1	*60*	*15 min*	*1 min*	*TBD*	*LZA <70*

[1] T=target, G=goal, LZA=Local Zenith Angle
[2] C=CONUS, FD=full disk, H=hemisphere, M=mesoscale

First of all, the ABI cloud mask will be used for cloud detection. Yes/no Flood/Standing Water retrieval in each scanning mode will be performed on each cloudless (i.e. "clear" and "possible clear" indicated by the cloud mask) pixel, during the day-time. Decision Tree technique will be applied to derive the rules and threshold values through the training (learning) process. Then the rules obtained from the training process will then be applied to "predict" or model "future" standing water. The resultant map is made by comparing to a reference non-flooding water mask to identify deviations which are then labeled as flood. Finally, the Yes/No Flood/Standing Water detection quality will be indicated with a set of quality control flags which are either generated in the Flood/Standing Water retrieval process or passed from the input data. The quality flags are assigned to each pixel.

Finally, algorithm simplicity and robustness is also a concern in order to produce the Flood/Standing Water identification product every fifteen minutes, which is the goal of ABI FSW product refresh rate.

2.2 Instrument Characteristics

The ABI will be a critical mission on GOES-R, providing over 60% of all the mission data products currently defined. Similar to the current GOES imager, ABI will be used for a wide

range of qualitative and quantitative weather, oceanographic, climate, and environmental applications. ABI will offer more spectral bands, higher spatial resolution, and faster imaging rate than the current GOES imager. Its spatial resolution will be nominally 2 km for the infrared bands and 0.5 km for the 0.64 µm visible band. While the instrument will allow a flexible scanning scenario, two basic modes are envisioned. One mode is that every 15 minutes ABI will scan the full disk (FD), plus 3 times continental United States (CONUS), plus a selectable 1000 km ×1000 km area every 30 seconds. The second mode is that the ABI can be programmed to scan the FD iteratively. The FD image can be acquired in approximately 5 minutes. The current GOES imager takes approximately 25 minutes for a FD; GOES-R will provide a fivefold increase in the coverage frequency (Schmit et al., 2004, 2007).

ABI has 16 spectral bands; five are similar to the 0.6-, 4-, 11-, and 12- µm windows and the 6.5-µm water vapor band on the current GOES-8/-9/-10/-11 imagers (Menzel and Purdom, 1994; Ellrod et al., 1998), and another is similar to the 13.3 µm on the GOES-12/-N/-O/-P imagers and the GOES-8/-P sounders (Hillger et al., 2003; Schmit et al., 2002). Additional bands on the ABI are 0.47 µm for aerosol detection and visibility estimation; 0.865 µm for aerosol detection and estimation of vegetation index and health, and Standing Water/Floods; 1.378 µm to detect very thin cirrus clouds; 1.6 µm for snow/cloud discrimination; 2.25 µm for aerosol and cloud particle size estimation, vegetation, cloud properties/screening, hot-spot detection, moisture determination, and snow detection; 7.0 and 7.34 µm for midtropospheric water vapor detection and tracking and upper-level sulfur dioxide (SO_2) detection; 8.5 µm for detection of volcanic dust clouds containing sulfuric acid aerosols and estimation of cloud phase; 9.6 µm for monitoring atmospheric total column ozone and upper-level dynamics (Steinbrecht et al.1998); and 10.35 µm for deriving low-level moisture and cloud particle size. Each of these bands is often used in conjunction with other bands in a multiple spectral approach for product generation. Channel specification of the ABI is given in Table 2.1. The advanced design of ABI will provide users with twice the spatial resolution, five times the scan rate, and more than three times the number of spectral channels compared to the current GOES imager (Schmit et al., 2007). These improvements will allow future meteorologists and climatologists to significantly improve the accuracy of their products, both in forecasting and nowcasting.

Table 2.1. Spectral characters of Advanced Baseline Imager.

Channel Number	Wavelength (µm)	Bandwidth (µm)	NEDT/SNR	Upper Limit Of Dynamic Range	Spatial Resolution
1	0.47	0.45 – 0.49	300:1[1]	652 W/m^2/sr/µm	1 km
2	0.64	0.59 – 0.69	300:1[1]	515 W/m^2/sr/µm	0.5 km
3	0.86	0.8455 – 0.8845	300:1[1]	305 W/m^2/sr/µm	1 km
4	1.38	1.3705 – 1.3855	300:1[1]	114 W/m^2/sr/µm	2 km
5	1.61	1.58 – 1.64	300:1[1]	77 W/m^2/sr/µm	1 km
6	2.26	2.225 – 2.275	300:1[1]	24 W/m^2/sr/µm	2 km
7	3.9	3.8 – 4.0	0.1K[2]	400K	2 km
8	6.15	5.77 – 6.60	0.1K[2]	300K	2 km
9	7.0	6.75 – 7.15	0.1K[2]	300K	2 km
10	7.4	7.24 – 7.44	0.1K[2]	320K	2 km

11	8.5	8.30 – 8.70	0.1K[2]	330K	2 km
12	9.7	9.42 – 9.80	0.1K[2]	300K	2 km
13	10.35	10.10 – 10.60	0.1K[2]	330K	2 km
14	11.2	10.80 – 11.60	0.1K[2]	330K	2 km
15	12.3	11.80 – 12.80	0.1K[2]	330K	2 km
16	13.3	13.0 – 13.6	0.3K[2]	305K	2 km

[1]100% albedo, [2]300K scene. Shaded channels are used for Water Identification**.**

The water identification will be produced for each pixel observed by the ABI sensor. The FSW retrieval will rely on channels 2, 3, and 5 of the ABI data using decision tree technique and post-class change detection method.

3 ALGORITHM DESCRIPTION

A complete description of the algorithm at the current level of maturity (which will improve with each revision) is given in this section.

3.1 Algorithm Overview

Flood/Standing water is an important product in the GOES-R ABI processing system, and is on the development list of the GOES-R algorithm working group (AWG). The standing water algorithm is developed by the GOES-R AWG land team within the land module processing subsystem (Figure 3.1). The right side of the land module includes all land products or algorithms, such as land surface temperature, active fire, flood/standing water, surface reflectance/albedo, and vegetation index. The left side includes some common inputs for land product algorithms.

The ABI Flood/Standing Water product is based on a decision tree algorithm to determine the presence of water, and applies visible and near-infrared observations and other auxiliary information. The resultant map is compared to a normal reference water mask to identify deviations, which are then labeled as flood. The algorithm will be tested using a radiative transfer model simulation dataset, and evaluated using a MODIS/SEVIRI proxy dataset.

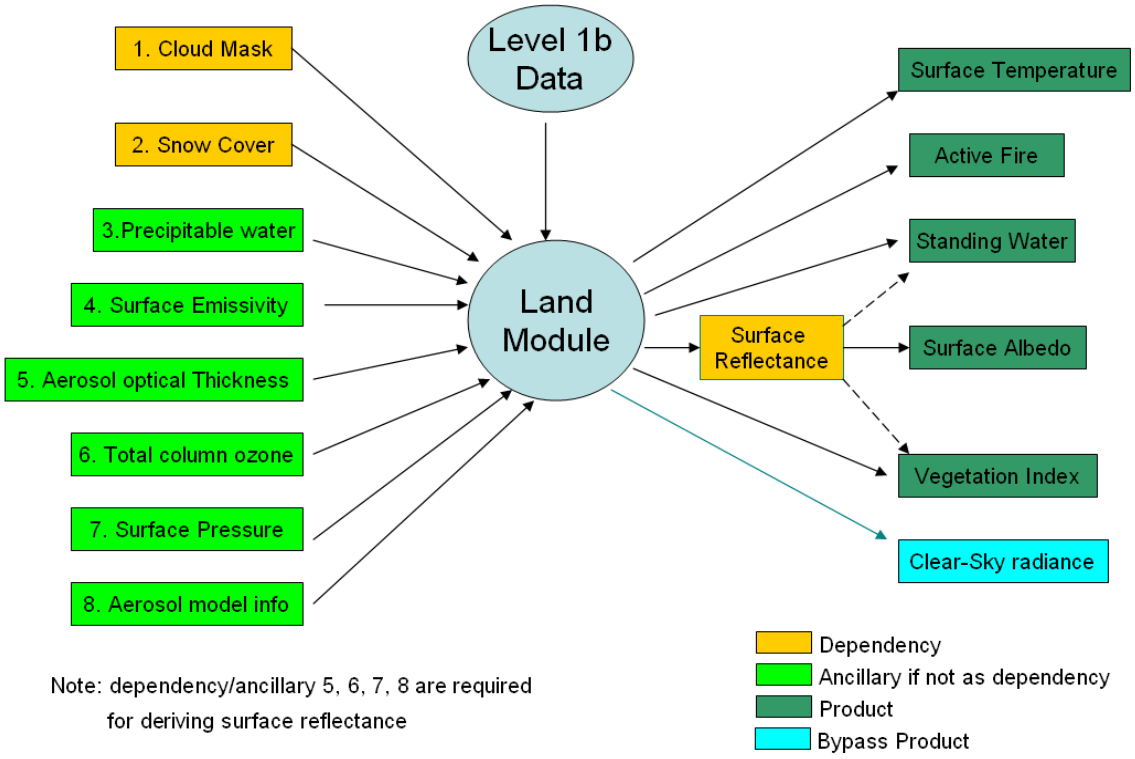

Figure 3.1. Products and dependencies of the land algorithm module.

Note that 1) Surface reflectance may be derived as an intermediate product in the surface albedo derivation (if the MODIS-like algorithm is applied); 2) Aerosol optical thickness and surface albedo may be dependent in the aerosol algorithm and the albedo algorithm; 3) Surface reflectance may be required for deriving standing water and top-of-canopy vegetation index. In addition, the AWG cryospheric team requires the surface reflectance as a dependency of its products.

3.2 Processing Outline

The processing outline of the standing water product is summarized in Figure 3.2. The standing water retrieval is started by extracting ABI sensor datasets including channel 2, 3, and 5 reflectances, pixel geolocation and the sensor data quality control flags. Afterwards, the process extracts ancillary datasets which can be categorized as ABI and non-ABI related datasets. The ABI related ancillary datasets include the ABI cloud mask and snow/ice mask, vegetation index and surface temperature, which are level 2 ABI products and were listed as dependencies in Figure 3.1. The non-ABI related datasets may include precipitation. More information on input datasets will be provided shortly in the Algorithm Input sub-section. Next, the ancillary datasets (precipitation) are mapped to the ABI pixel location. Then, the ABI sensor data is filtered using the cloud mask to ensure that only the cloud clear and probably clear pixels are processed for the standing water retrieval. Finally, the calculated water identification result and their associated quality control flags are combined with the flood/standing water product package and are written to files for user access.

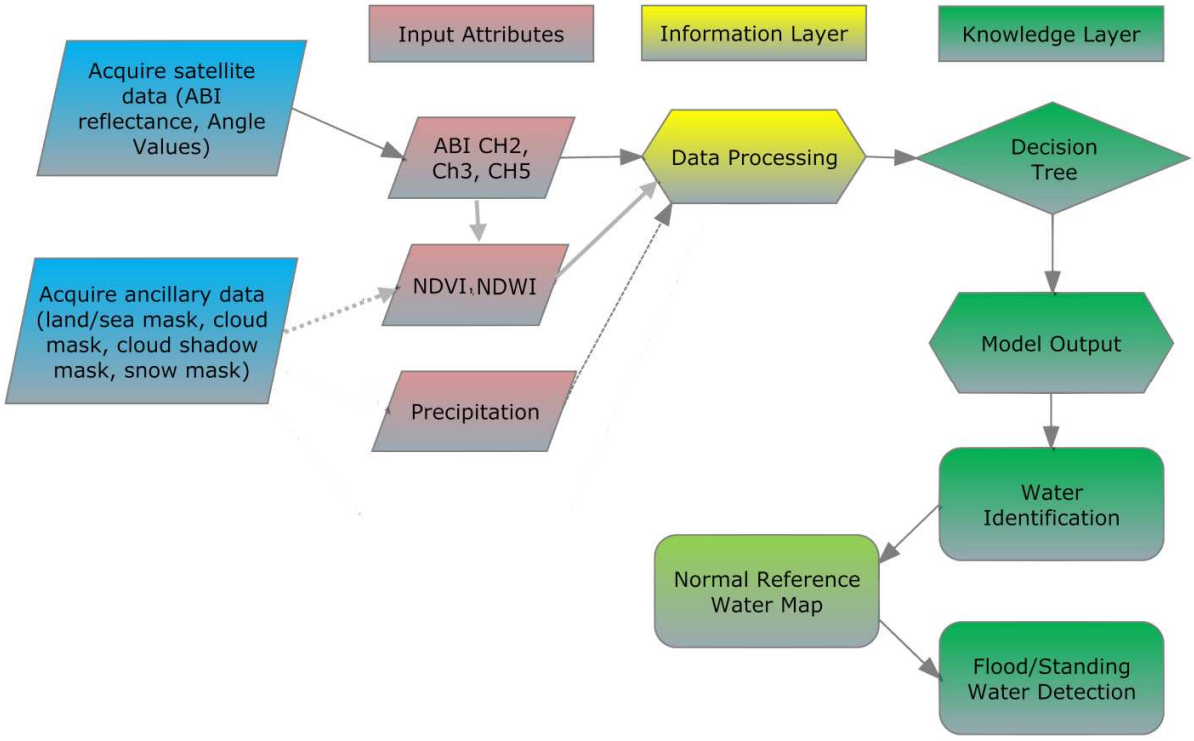

Figure 3.2. Flowchart of GOES-R ABI Flood/Standing Water algorithm.

There are two categories of change detection methods, prior-class and post-class. The prior-class detection method usually try to find the maximum changing extent during a period of time, and may be difficult to be applied to operational applications for every 60 minutes like the GOES-R FSW product. Due to the strong water and land discrimination in the visible and infrared region, like the ABI sensor, we selected post-class change detection method. The flood/standing water is detected by comparing water classification during the times of flood with reference water mapped during normal conditions. From the first beginning, we made our selection from the perspective of operational implementation. For water classification, we also chose the well-established and mature decision tree algorithm.

3.3 Algorithm Input

This section describes the input needed to process the standing water/flood product. While the standing water is derived for each pixel, ancillary datasets are required as well as the upstream ABI data.

3.3.1 Primary Sensor Data

The list below contains the primary sensor data used by the standing water retrieval. Primary sensor data are derived solely from the ABI observations and geolocation information, or the level 1b data. Table 3.1 lists those input sensor data and their descriptions.

Table 3.1. Input list of sensor data.

Name	Type	Description	Dimension
Ch2 reflectance	input	Calibrated ABI level 1b reflectance at channel 2	grid (xsize, ysize)
Ch3 reflectance	input	Calibrated ABI level 1b reflectance at channel 3	grid (xsize, ysize)
Ch5 reflectance	input	Calibrated ABI level 1b reflectance at channel 5	grid (xsize, ysize)
Latitude	input	Pixel latitude	grid (xsize, ysize)
Longitude	input	Pixel longitude	grid (xsize, ysize)
Solar zenith	input	ABI solar zenith angles	grid (xsize, ysize)
View zenith	input	ABI view zenith angle	grid (xsize, ysize)
QC flags	input	ABI quality control flags with level 1b data	grid (xsize, ysize)

3.3.2 Derived Sensor Data

There are two ABI derived sensor data sets (or the ABI related ancillary dataset quoted earlier) used by the LST retrieval: 1) the ABI cloud mask (ACM) product, which indicates four cloudiness conditions for each pixel: clear, probably clear, probably cloudy, and cloudy, and 2) snow/ice mask which indicates if the pixel is snow or ice covered. Table 3.2 briefly describes input of the derived sensor data.

Table 3.2. Input list of derived sensor data.

Name	Type	Description	Dimension
Cloud mask	input	ABI level 2 cloud mask data	grid (xsize, ysize)
Snow/Ice mask	input	ABI level 2 Snow/Ice mask data	grid (xsize, ysize)

In case the ABI snow/ice mask is not available at the GOES-R operational, the Interactive multi-sensor snow and ice Mapping System (IMS) will be used for the snow/ice mask.

3.3.3 Ancillary Data

The following table lists and briefly describes the ancillary data required to run the Standing Water/Flood which are the non-GOES-R data that provide information not included in the primary sensor data or the previously computed GOES-R data. The static standing water map will be used as the reference non-flood map to identify flood pixels. The non-flood map could be generated from ABI data by our decision tree algorithm per month or per season when no flood occurs. Before the ABI data are available, the yearly static water/land map could be used, which usually compiled from the IGBP land cover map.

Table 3.3 Input of ancillary data.

Name	Type	Description	Dimension
Reference Normal or non-flooding water map	input	Water/land reference map or non-flood standing water classification map	grid (xsize, ysize)
Land/sea mask	input	A land-ocean and coast mask	grid (xsize, ysize)
IMS snow/ice mask*	input	Interactive multi-sensor snow and ice Mapping System	grid (xsize, ysize)
Cloud shadow mask	input	Cloud shadow mask	grid (xsize, ysize)

** IMS snow/ice mask is an alternative in case the ABI snow/ice mask is not available at the GOES-R operation.*

3.4 Theoretical Description

3.4.1 Physics of the Problem

The GOES-R ABI is a visible and infrared imager. The interpretation of water identification by using visible/infrared remote sensing is relatively straightforward. According to the spectral characteristics, in the near-infrared (NIR) range (0.7-1.1μm), water has lower reflectance than vegetation and other land covers (Figure 3.3). On the contrary, water has slightly higher reflectance values than land features in the visible red (RED) band (0.5-0.7 μm). Therefore, the ratio image and the difference image between *NIR* and *RED* can be used to enhance the difference between water and land. In the ratio or difference image, water has extremely low value, while land has relatively high value.

Much of the pioneering work on the remote sensing of floods was accomplished using the Multi-Spectral Scanner (MSS) sensor on the First Earth Resources Technology Satellite, later renamed Landsat-1. With a spatial resolution of about 80 m, MSS data were used to map the extent of flooding in Iowa (Hallberg et al., 1973; Rango and Salomonson, 1974), Arizona (Morrison and Cooley, 1973), Virginia (Rango and Salomonson, 1974) and along the Mississippi River (Deutsch et al., 1973; Deutsch and Ruggles, 1974; Rango and Anderson, 1974; McGinnis and Rango, 1975; Deutsch, 1976; Morrison and White, 1976). In all studies, MSS band 7 (0.8-1.1 μm) was highly useful for separating water from dry soil or vegetated surfaces owing to the strong absorption by water in the near-infrared range, as shown in Figure 3.3. This was further confirmed by analyzing MSS band 5 (0.6-0.7 μm), band 7 and field spectral radiometer data along shoreline water-wet soil-dry soil transitions (Gupta and Banerji, 1985). The flood areas were delineated based on the sharp contrast between inundated regions and adjacent areas. The standing water areas appeared as dark blue to light blue depending upon the depth of water, while the receded water/wet areas appeared as dark to light gray (e.g., Figure 3.4).

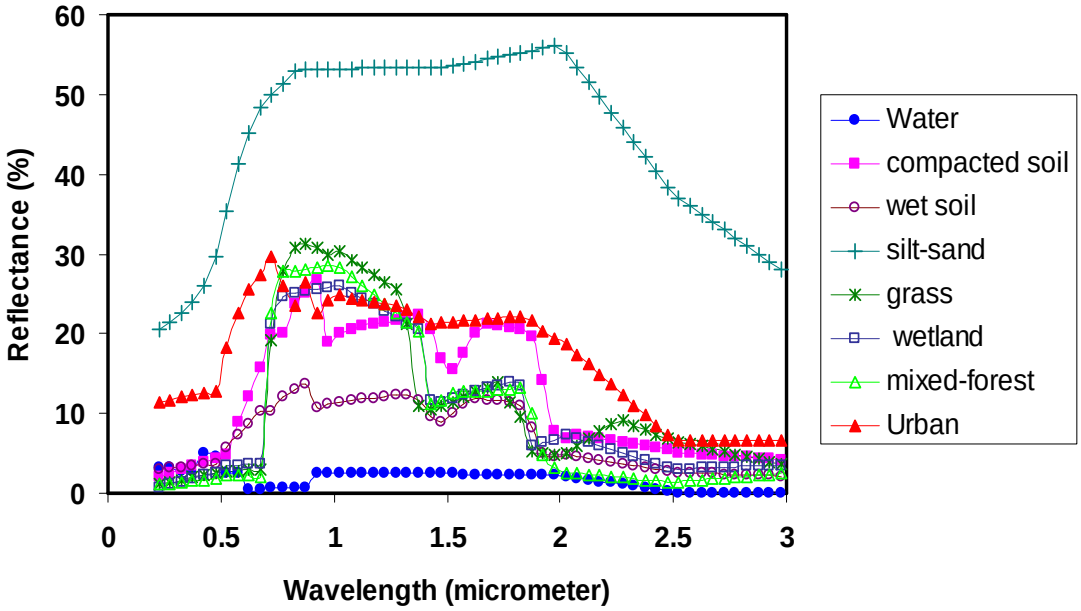

Figure 3.3. Reflectance spectra of some land surface species.

Other studies have continued the methodology first developed with MSS, using Landsat TM and SPOT data (France and Hedges, 1986; Jensen et al., 1986; Watson, 1991; Blasco et al., 1992; Pope et al., 1992; Vila da Silva and Kux, 1992). The coarser resolution (ca. 1 km) sensors, such as the Advanced Very High Resolution Radiometer (AVHRR) aboard the NOAA series, has been successfully used for studying very large river floods (Ali et al., 1989; Barton and Bathols, 1989; Gale and Bainbridge, 1990; Rasid and Pramanik, 1993).

Sheng *et al.* (2001) summarized the spectral characteristics of the main features (i.e. water, vegetation, soil, and clouds) during floods at the observation scale of NOAA satellites. Although AVHRR data can be displayed in 3-channel color composites for visual analysis (flood/standing water absorbs infrared wavelengths of energy and appears as blue/black in the RGB composite

imagery), water body identification in AVHRR imagery evolved from qualitative visual interpretation to automatic quantitative extraction. The reflectance of AVHRR channel 2 (0.73-1.1μm, similar to MSS band 7), the reflectance difference (CH_2-CH_1) and ratio (CH_2/CH_1) between AVHRR channel 2 and 1 (0.58-0.68 μm, similar to MSS band 5) are used to discriminate water from land if these parameters are less than the predetermined threshold values.

The methods used in the literature to discriminate water from land can be summarized as the following:

1) Channel 2 model (Lin 1989; Sheng et al., 1998)

$$\begin{cases} Water, & if\ NIR \le T_0 \\ Land, & if\ NIR > T_0 \end{cases} \tag{3.1}$$

Where CH_2 is the reflectance of channel 2, and T_0 is a threshold.

2) Differential model between channels 2 and 1 (Xiao and Chen, 1987)

$$\begin{cases} Water, & if\ (NIR - RED) \le T_0 \\ Land, & if\ (NIR - RED) > T_0 \end{cases} \tag{3.2}$$

3) Ratio model of channels 2 and 1 (Sheng and Xiao, 1994)

$$\begin{cases} Water, & if\ \ NIR / RED \le T_0 \\ Land, & if\ \ NIR / RED > T_0 \end{cases} \tag{3.3}$$

Sheng *et al.* (2001) found that the ratio (NIR/RED) image had a better discriminating capability than the difference (NIR-RED).

4) NDVI model

Normalized Difference Vegetation Index (NDVI):

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{3.6}$$

Where NIR and RED stand for the spectral reflectance measurements acquired in the near-infrared and red regions, respectively. Domenikiotis et al. (2003) used NDVI to discriminate water from land.

Table 3.4. Vegetation classification according to typical NDVI values.

Soil Type	NDVI
Dense vegetation	0.500
Intermediate green vegetation	0.140
Sparse vegetation	0.090
Bare soil	0.025
Snow and ice	-0.046
Water surface	-0.257

It is apparent from its mathematical definition of the NDVI (Rouse et al., 1975 and Tucker, 1979) that an area containing a dense vegetation canopy will tend to have positive values (say 0.3 to 0.8), while standing water, which has a rather low reflectance in both visible (VIS: from 0.4 to 0.7 µm) and near-infrared (NIR: from 0.7 to 1.1 µm) spectral bands (Figure 3.3 and Table 3.4), results in very low positive or even slightly negative NDVI values.

5) Brightness temperature model

During daytime (Verdin, 1996):

$$\begin{cases} Water, & if\ Tb_{11/12} \le T_0 \\ Land, & if\ Tb_{11/12} > T_0 \end{cases} \tag{3.4}$$

During nighttime (Barton and Bathols, 1989):

$$\begin{cases} Water, & if\ Tb_{11/12} \le T_0 \\ Land, & if\ Tb_{11/12} > T_0 \end{cases} \tag{3.5}$$

Where $Tb_{11/12}$ is the brightness temperature of the 11 or 12 µm channels, and T_0 is a threshold.

The temperature model using AVHRR channel 4 or 5 (at 11 and 12 microns) brightness temperature usually works well in areas where floods resulting from snow melt, because there is usually a temperature discrepancy between water and land. However, it may not work with floods caused by heavy rainfall during rainy seasons in the summer when there is relatively low or no temperature difference between land and water.

6) Surface temperature model

Domenikiotis et al. (2003) also tried to use surface temperature to discriminate water from land surfaces.

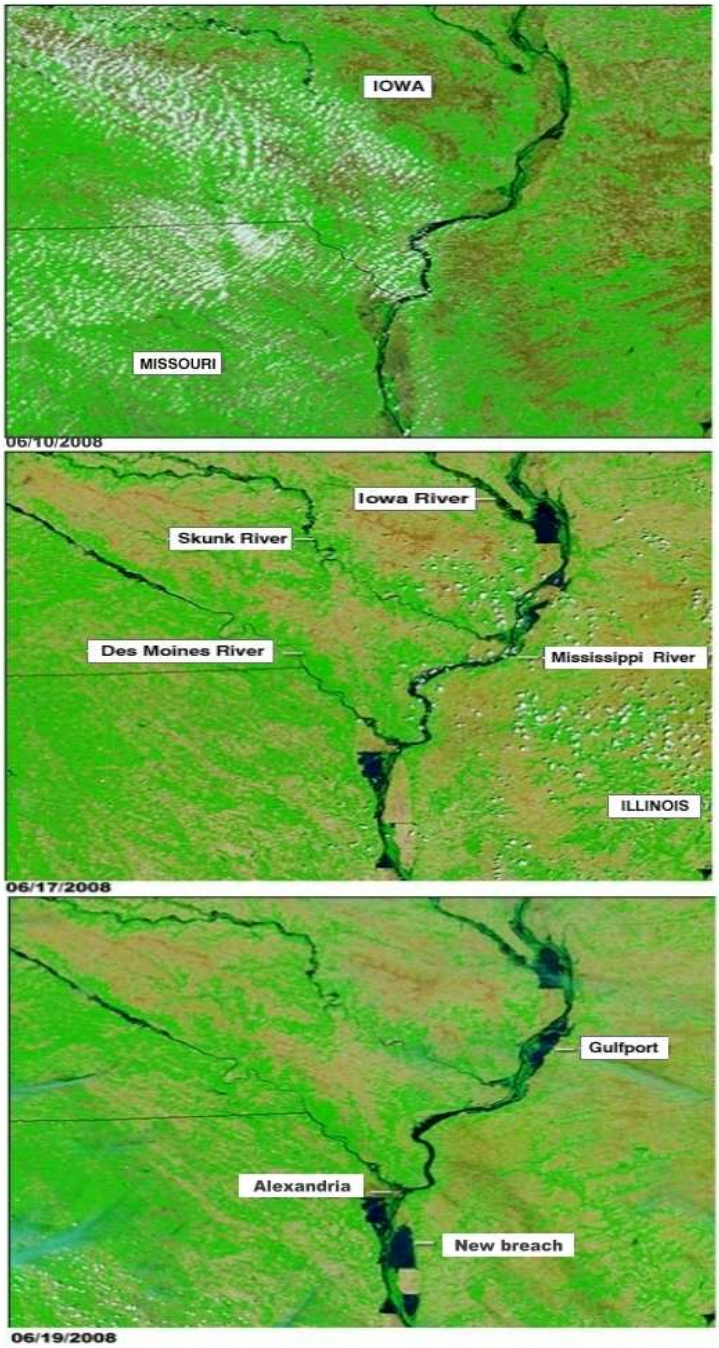

Figure 3.4. MODIS 1km RGB (7, 2, 1) composite images in Midwest on June 10, 2008 before (upper) and on June 17 (middle) and June 19 (lower), 2008 after flooding.

All of the above studies by using spectral features are based on the reflectance difference between water and land. Figure 3.5 shows the histograms in RED, such as the AVHRR/MODIS/SEVIRI CH1, the NIR, such as the AVHRR/MODIS/SEVIRI CH2, (CH2-CH1), CH2/CH1, and NDVI. In our study, we also include shortwave infrared or SWIR

centered at 1.61 µm, for example, the MODIS CH6, and Normalized Difference Water Index (NDWI). The NDWI (Gao, 1996) is a satellite-derived index from the NIR and Short Wave Infrared (SWIR) channels, NDWI= (NIR-SWIR) / (NIR+SWIR). MODIS has 2 bands in the SWIR region: band 5 (1230-1250 nm) and band 6 (1628-1652 nm) while band 2 represents the NIR region. Meanwhile, ABI also has two bands in the SWIR region, band 4 (1.3705 – 1.3855 µm) and band 5 (1.58 – 1.64 µm). According to Gao (1996), NDWI is a good indicator for vegetation liquid water content and is less sensitive to atmospheric scattering effects than NDVI. Since ABI band 4 (1.38 µm) will only have 2 km resolution, so our NDWI will use ABI band 5 (1.61 µm) or MODIS band 6, so our NDWI is more for standing water and may be a little different from that defined by Gao (1996) as (p(0.86 µm) - p(1.24 µm))/(p(0.86 µm + p(1.24 µm)).These histograms further demonstrate that the stratification between Water and Land is pretty clear, and water can be separable from land by using these attributes, especially the CH2 reflectance, the reflectance difference (CH2-CH1), and the reflectance ratio (CH2/CH1).

Perhaps the greatest difficulty with visible/infrared sensors is their inability to image the Earth's surface under cloudy conditions (Rasid and Pramanik, 1993; Melack et al., 1994). For the purpose of determining maximum flood extent, this difficulty is somewhat mitigated by the fact that standing water can be mapped even after flood recession (Rango and Anderson, 1974; Deutsch, 1976). This effect can last from one to two weeks (Hallberg et al., 1973; Rango and Salomonson, 1974; Morrison and White, 1976; Salomonson, 1983).

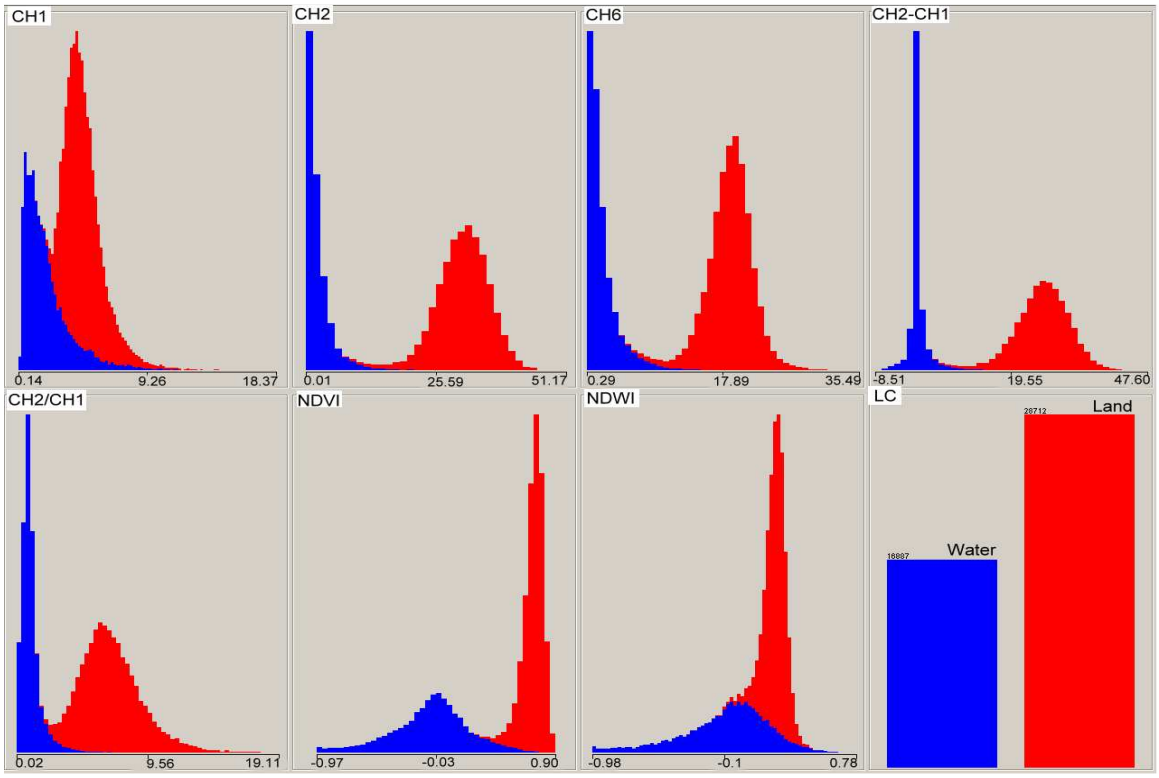

Figure 3.5. Histograms of MODIS surface reflectance at CH1, CH2, and CH6, (CH2-CH1), CH2/CH1, NDVI, NDWI, and Land/Water samples (Sun and Yu, 2010).

3.4.2 Algorithm Selection

Upon simulating the top-of-atmosphere radiances/reflectance with the proxy MODIS data, we then conducted simulation analyses for the algorithm development.

Surface type "water body" is assigned to the 'Water' class; all land surface types are assigned to class 'Land' or 'Non-Water'.

In order to find the best DT approach, several DT algorithms, including the J48graft or J48, which is based on the C4.5, originally proposed by Quinlan (1993), NBTree, which is a Naïve Bayes/Decision Tree hybrid (Kohavi, 1996), Random Tree, Random Forest (Breiman, 2001), REP Tree, BFTree, Decision Stump, FT (final tree), and CART (Classification and Regression Trees) (Breiman et al., 1984), are tested. All of these methods are decision tree techniques which represent supervised machine learning approaches. A total 96,928 around the Mississippi river were input to the algorithms with a split test mode of 50% for training and the remaining 50% for test. A comparison of the test accuracy from different DT algorithms is listed in table 3.8. All the DT algorithms have a capability of discriminating the objects well. The J48graft/J48 or the C4.5 and the CART got the best accuracy of about 97% for water identification. However, if we want to further separate water into different water types, it becomes more difficult. Table 3.9 compares the accuracy for different water types from different DT algorithms. We can see that the accuracies for further separation of different water types are reduced. For water mixed with land, the accuracy is reduced to about 70% from the J48graft/J48 and random forest algorithms, and even to about 60% from other algorithms. For flooded areas, land is inundated by water, so water is usually mixed with vegetation, and/or trees, or even urban. Therefore, the accuracy requirement of 60% for GOES-R yes/no water identification product is reasonable. Since we wish to select an algorithm with the highest accuracy for water identification, as well as the easiest implementation, we selected the J48 or the C4.5 as our baseline algorithm.

Table 3.8. Comparison of DT Accuracy Rate (%) of Classified Instances from different types of Decision Tree Algorithms (Water and Land) (Sun and Yu, 2010)

Trees / Types	ADTree	BFTree	J48graft /J48	Decision Stump	FT	Random Forest	Random Tree	REPTree	CART	NBTree
Water	97.0	97.6	97.8	97.0	97.0	93.7	93.9	97.7	97.7	94.6
Land	99.0	98.9	99.2	99.1	98.0	98.9	98.9	98.9	98.9	99.2
Average	98.7	98.7	98.9	98.7	97.8	98.6	98.1	98.7	98.8	98.5

Table 3.9. Comparison of DT Accuracy Rate (%) of Classified Different Water types from different types of Decision Tree Algorithms (Water and Land)

DT algorithms Water Types	NBTree	J48graft /J48	Random Forest	Random Tree	REPTree	CART	BFTree	FT
Water Mixed with land	69.6%	70.3%	76.8%	66.7%	63.8%	68.1%	66.7%	69.6%
Wetland	62.4%	66.7%	63.1%	56%	58.2%	57.7%	56%	48.9%
Pure River Water	82.1%	82.1%	84.6%	73.2%	82.1%	82.4%	79.7%	80.5%
Pure Lake Water	80.7%	82%	84%	81.3%	79.3%	86.9%	82%	80%
Average	73.4%	73.6%	75.95%	67.5%	71.7%	72%	70.4%	68.5%

Finally, we emphasize that all the results discussed at this point assume perfect cloud detection. That is, all these results are for truly cloud clear pixels. Residual cloud effects in pixels classified as clear may add significant noise to the standing water retrievals.

3.4.3 Mathematical Description of the FSW Algorithm

All of the methods introduced above reflected one aspect of water. We propose to use data mining techniques, such as Decision Trees (DT) technique, to integrate all the above attributes, and at the same time, provide the threshold values and accuracy of algorithm performance.

Compared to traditional statistical models, data mining methods such as DT analysis can help find hidden relationships among multiple attributes/parameters. DT is a classic prediction model to support decision making (Han, 2001) by converting complex data into relatively simple and direct viewing structure. It has proven useful in other research areas, such as land cover/land use classification (Colstoun et al., 2000). Xu *et al*. (2005) employed a decision tree regression approach to determine class proportions within a pixel. The DT was used for wetland composites from the ASTER data (Wei et al., 2008) and flood related land cover classification from multi-temporal RADARSAT-1 SAR data (Parmuchi et al., 2002), in which the wetlands or flood were classified as surface cover types. The M5 model tree has been used to derive flood related discharge (Solomatine et al., 2004). Carroll et al. (2008) have been successful in deriving flood maps for the Midwest flood event with MODIS data using decision tree (DT) approach. These methods improved conventional spectrally based classification.

Methods for analyzing and modeling data can be divided into two groups: “unsupervised learning” and “supervised learning.” Unsupervised learning does not identify a target (dependent) variable, but rather treats all of the variables equally. In this case, the goal is not to predict the value of a variable but rather to look for patterns, groupings or other ways to

characterize the data that may lead to understanding of the way the data interrelates. Cluster analysis, correlation, factor analysis (principal components analysis) and statistical measures are examples of unsupervised learning.

Supervised learning requires input data that has both predictor (independent) variables and a target (dependent) variable whose value is to be estimated. By various means, the process "learns" how to model (predict) the value of the target variable based on the predictor variables. Decision trees, regression analysis and neural networks are examples of supervised learning. If the goal of an analysis is to predict the value of a variable, then supervised learning is the recommended approach. Our goal here is to model (predict) Standing Water and Flood, so we chose the supervised learning approach.

In this study, we will use decision tree algorithm, which is a supervised machine leaning technique, for Flood/Standing Water identification, because unlike the previous studies using piecemeal data, DT algorithms can integrate all the useful predictors. We here take the C4.5 as an example to show how the DT works for classification.

The C4.5 algorithm (Quinlan 1993) is a machine learning method based on decision tree induction (Han et al., 2001). The basic strategy is to select an attribute that will best separate the samples into individual classes by a measurement '*Information Gain Ratio*'. The objective is to produce the most accurate separation with the least amount of information (Han et al., 2001).

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

$$Info(S) = -\sum_{i=1}^{m} \frac{s(C_i)}{s} \log_2(\frac{s(C_i)}{s}) \tag{3.7}$$

Hence, the amount of information needed to partition S into {S_1, S2…S_v} by attribute A (A has v distinct values) is:

$$Info(A|S) = -\sum_{j=1}^{v} \frac{s_j}{s} * Info(S_i) \tag{3.8}$$

The gain is computed as:

$$gainRatio(A|S) = \frac{gain(A|S)}{Info(A|S)} \tag{3.9}$$

where $$gain(A|S) = Info(S) - Info(A|S) \tag{3.10}$$

In this study, five attributes/parameters (CH_2, CH_3, CH_3-CH_2, CH_3/CH_2, and NDVI) and two classes (Water, Land) are integrated for water identification.

3.4.4 The GOES-R ABI Forward Simulations

To test the accuracy and sensitivity of the above DT algorithm for the GOES-R ABI, we performed forward simulations using the latest version of (MODerate resolution atmospheric TRANsmission) (MODTRAN v4.3) to generate a comprehensive simulation dataset.

The MODTRAN atmospheric radiative transfer model (Berk et al., 2000) has been widely used in satellite remote sensing studies for about three decades. It is a moderate spectral resolution model, up to cm^{-1} in frequency. We used MODTRAN version 4, reversion 3, released in 2008. The radiative transfer simulation procedure is illustrated in Figure 3.6.

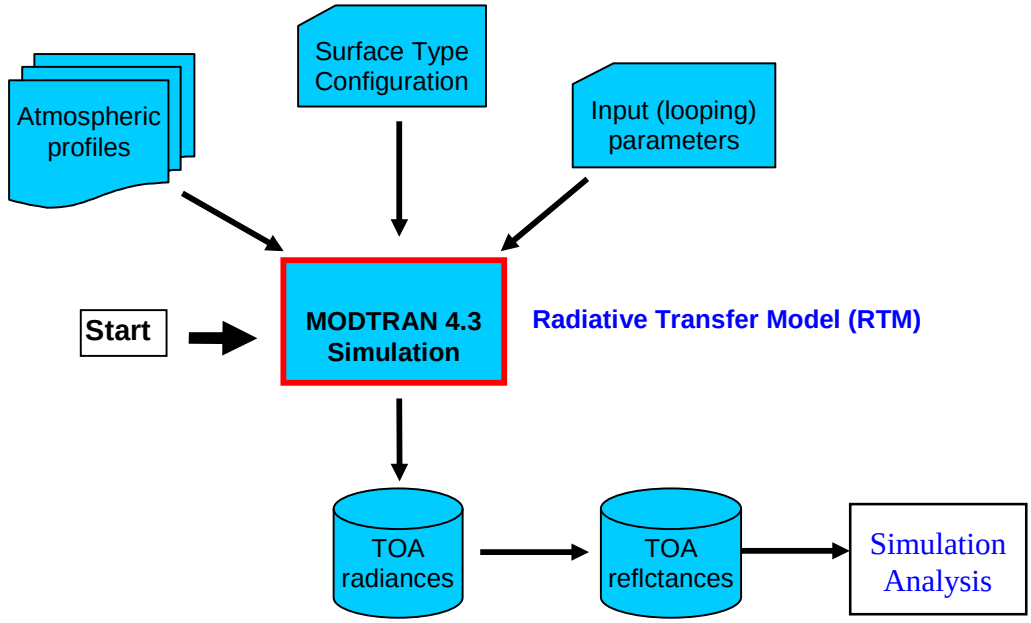

Figure 3.6. Radiative transfer simulation procedure.

An angularly dependent atmospheric radiation model - MODTRAN is used for GOES-R forward simulations (Berk et al., 2000). The spectral range covered is from 0.2 μm (50000 cm^{-1}) to 4 μm (2500 cm^{-1}) with resolution 20 cm^{-1}.

In order to account for the wide range of different atmospheric and surface conditions, a large number of simulations for each season need to be performed with variations in:.

- Geometry of the problem (solar zenith angle, viewing and azimuth angles)
- Atmospheric conditions (profiles of ozone, water vapor, aerosols)
- Surface conditions (spectral characteristics of the surface)
- Characteristics of the instrument (spectral response of the satellite sensors); the latest GOES-R ABI spectral response functions were obtained from the University of Wisconsin (ftp://ftp.ssec.wisc.edu/ABI/SRF/)

In order to represent the variability in solar geometry, the simulations are performed for 10 different solar bins.

Table 3.5. Solar zenithangle bins used for ABI forward simulations.

Solar bin	1	2	3	4	5	6	7	8	9	10
Solar zenith angle (degrees)	12.9	30.8	41.2	48.3	56.5	63.2	69.5	75.5	81.4	87.2

In order to obtain angularly dependent relationships the calculations are performed in 48 viewing bins - 6 zenith angles and 8 azimuth angles.

Table 3.6. Satellite zenith angle bins used for ABI forward simulations.

Zenith Bin	1	2	3	4	5	6
SZA (degrees)	11.45	26.08	40.32	53.75	65.94	76.32

Table 3.7. Azimuth angle bins used for ABI forward simulations.

Azimuth Bin	1	2	3	4	5	6	7	8
Azimuth angle (degrees)	1.91	9.97	24.18	44.02	68.78	97.55	129.31	162.89

Climatological profiles for temperature, water vapor, and ozone grouped into 4 seasons (5 profiles for each season) are used in the simulations. These climatological profiles are derived from the TIGR profiles. For each profile, simulations are performed for 3 different visibilities (18, 23, and 28 km) in order to account for the variability of atmospheric aerosols. These create output files for 4 seasons x 10 solar zenith angles x 5 profiles x 3 visibilities x 3 albedo models (1800 cases total). Fourteen UMD (Univ. of Maryland) IGBP surface types were used plus an additional surface type (Snow/ice) and spectral library from the NASA JPL. Each surface type contains 1800 simulation cases.

The narrowband outgoing radiances at the TOA are obtained by convoluting the spectral radiances with the response function of the specific instrument.

$$I(\mu_0, \mu, \varphi) = \int_{\lambda_1}^{\lambda_2} L(\lambda) I_\lambda(\mu_0, \mu, \varphi) d\lambda \tag{3.13}$$

The radiance is then converted to reflectance. We first determined the mean channel radiance by integrating over the sensor spectral response function (SRF). The latest GOES-R ABI spectral response functions from the University of Wisconsin (ftp://ftp.ssec.wisc.edu/ABI/SRF/) are used.

The channel radiances were then converted into corresponding reflectance.

3.4.5 Algorithm Challenges

3.4.5.1 Cloud shadow effect

Samples of water, vegetation, bare land, and cloud shadow are collected. Scatter points of these samples are shown in Figure 3.7, water is shown in red with vegetation in green, land in blue and cloud shadow in cyan. It can be seen clearly that the spectral signatures of cloud shadow are so close to those of water that cloud shadow may be easily confused and misclassified into water.

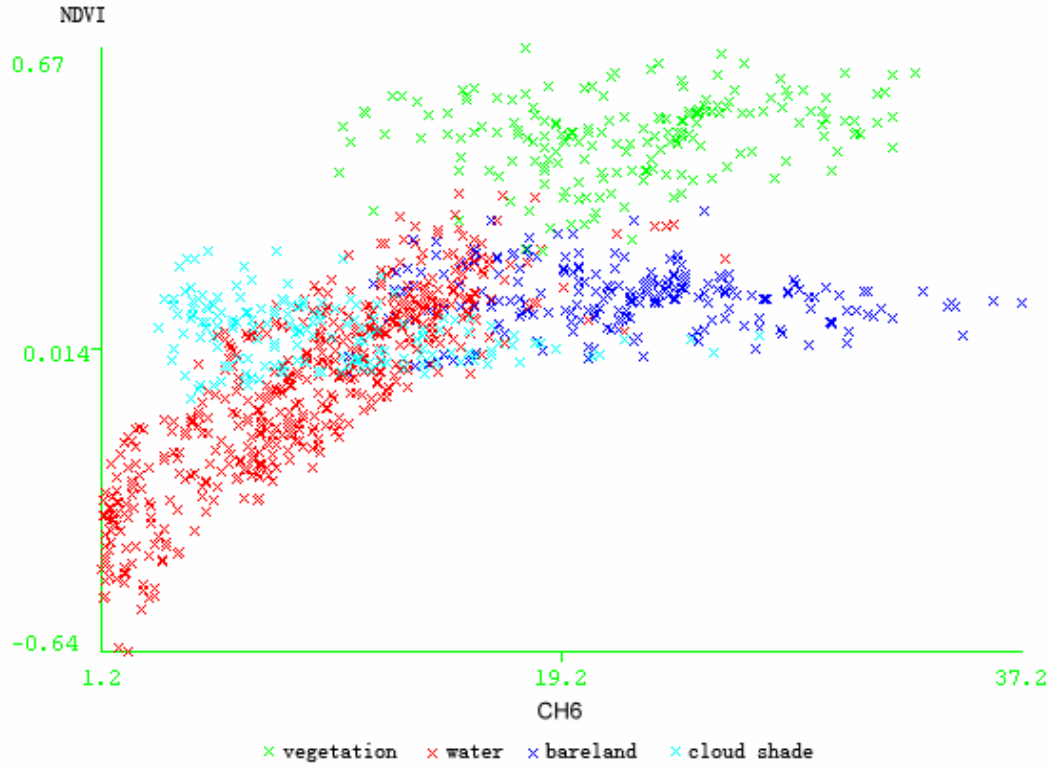

Figure 3.7. Scatter plots of NDVI vs. MODIS CH3 (green: vegetation, red: water, cyan: cloud shadow, and blue: bare land).

- **Geometrical relationship for cloud shadow**

Cloud shadows always exist next to or around the clouds, so the projection of a cloud and its shadow can be assumed as in the same horizontal plane if the earth curvature is not considered. Thus the relationship for cloud shadow in a satellite image can be established based on the geometry relationship. Firstly the coordinates of cloud in an image plane can be calculated by the satellite geometry angles. Then with solar geometry angles, the place of cloud shadow can also be calculated accurately. The coordinates of cloud shadow (X_D, Y_D) in an image plane can be described as the following:

$$X_D = X_c + \frac{h_c}{\tan\theta_v * dx} * \sin\varphi_v + \frac{h_c}{\tan\theta_s * dx} * \sin\varphi_s$$

$$Y_D = Y_c - \frac{h_c}{\tan\theta_v * dy} * \cos\varphi_v + \frac{h_c}{\tan\theta_s * dy} * \cos\varphi_s \quad (3.14)$$

Where (Xc, Yc) is the cloud position in an image, θ_s (solar elevation angle) and θ_v (sensor elevation angle) have a range of 0° to 90°, φ_s (solar azimuth angle) and φ_v (sensor azimuth angle) have a range of 0° to 360°, dx and dy is the spatial resolution in longitude and latitude of a satellite image respectively, and hc is the cloud top height.

- **Some results for cloud shadow detection**

Figure 3.8 shows an example of cloud shadow detection with the MODIS observations (Li et al., 2010).

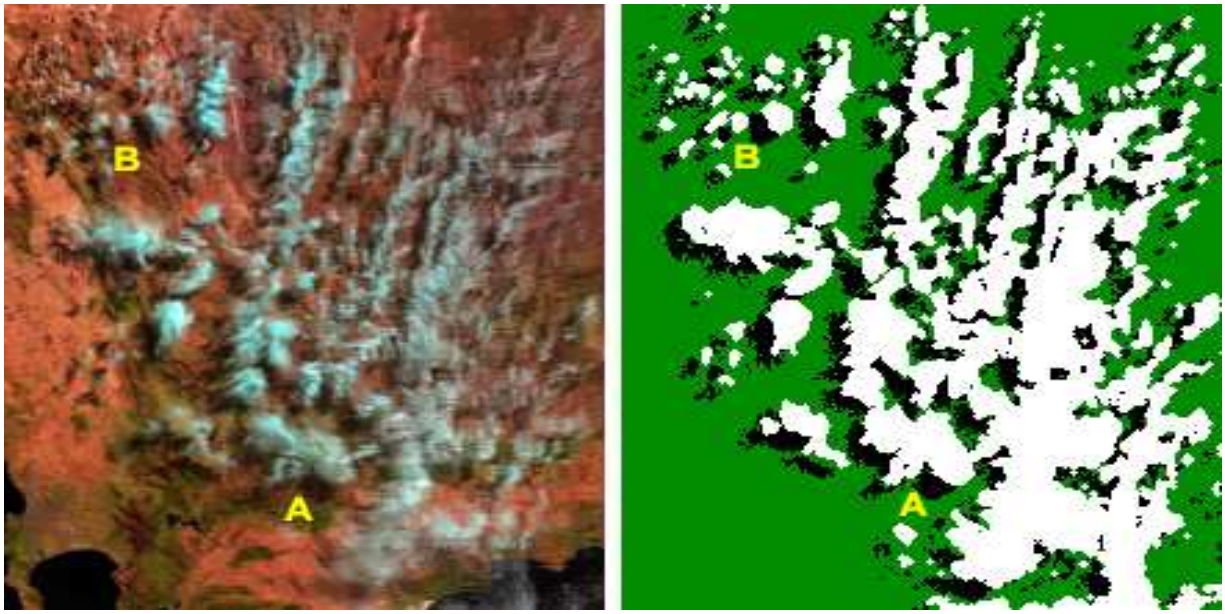

Figure 3.8. MODIS false color (726 RGB composite) image and the detected cloud (in white)/cloud mask (in black) distribution (07:55GMT, 02/04/2008).

3.4.5.2 Anisotropic effect

As shown in Figure 3.9, during early morning, (e.g., 08:00 UTC or 08:45 UTC), and late afternoon, the reflectance over vegetated area is very low, and makes it difficult to separate vegetation from water, or, in other words, the water body is confused with vegetation. If the reflectance of vegetation is as low as that of water, then vegetation may be misclassified into water, as shown in Figure 3.10 (upper, as the area marked with red oval). If ABI surface albedo

product can provide BRDF-corrected nadir surface reflectance, then anisotropic effect won't be a problem to FSW detection.

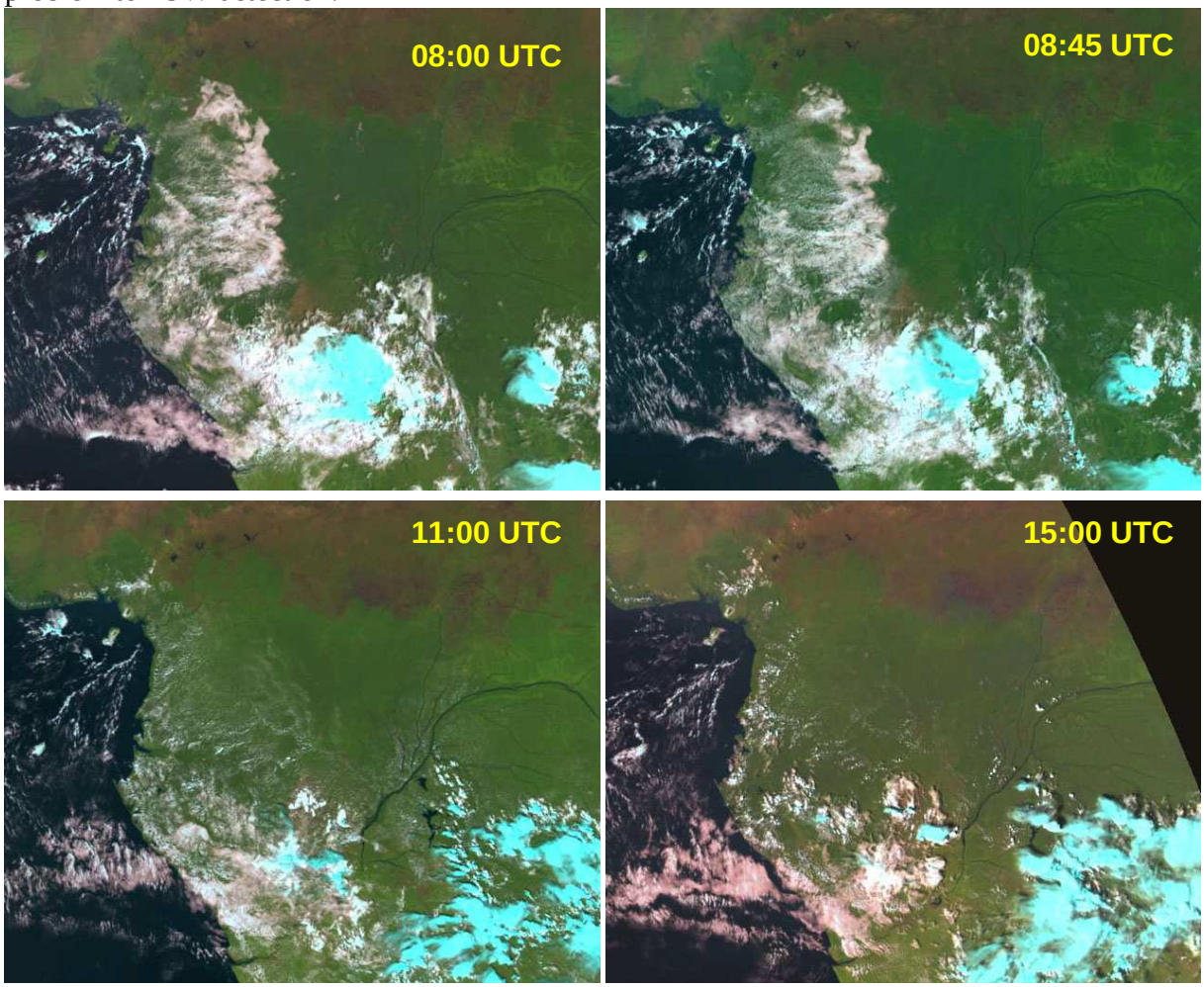

Figure 3.9. Three- channel composite images in the same area at different time on December 26,2007.

3.4.5.3 Sun glint contamination

Sun glint occurs in imagery when the water surface orientation is such that the sun is directly reflected towards the sensor, and hence is a function of sea surface state, sun position and viewing angle (Figure 3.11). A variety of glint correction methods have been developed for open ocean imaging and high resolution coastal applications, since solar glint is usually found over open ocean and coastal region (Kay et al., 2009). While our product is working with inland standing water identification, a land/see mask can help remove open ocean water. For inundated coastal area, it is found that the values of NDVI decreased dramatically in the glitter direction (Figure 3.12), providing an indication of surface water and a method to discriminate water under solar glint conditions (Vanderbilt, 2002).

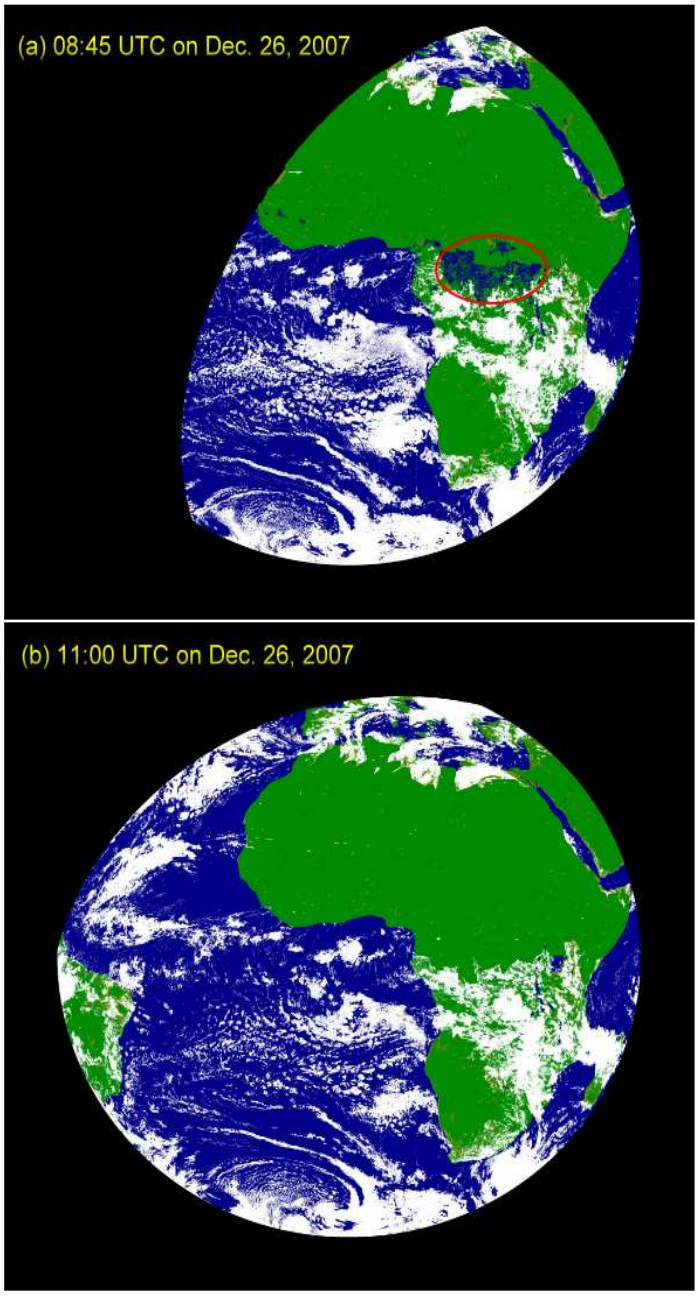

Figure 3.10. Three- channel composite images in the same area at different time on December26, 2007.

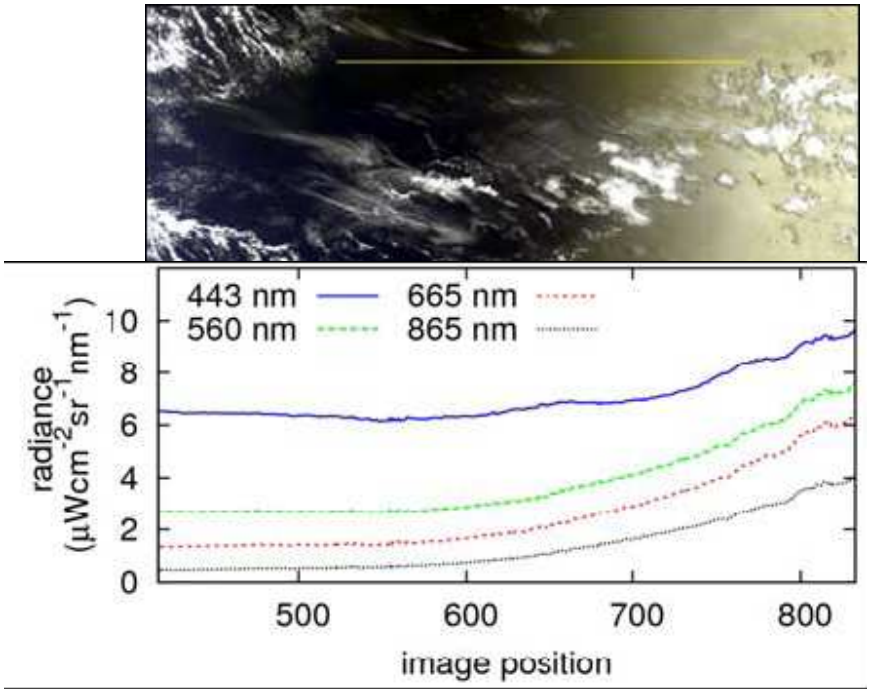

Figure 3.11 Illustration of sun glint in an optical imagery, showing sun glint on the right hand side (upper), and sensor radiance plotted for 4 wavebands along the line marked in (upper) (from Kay et al., 2009)

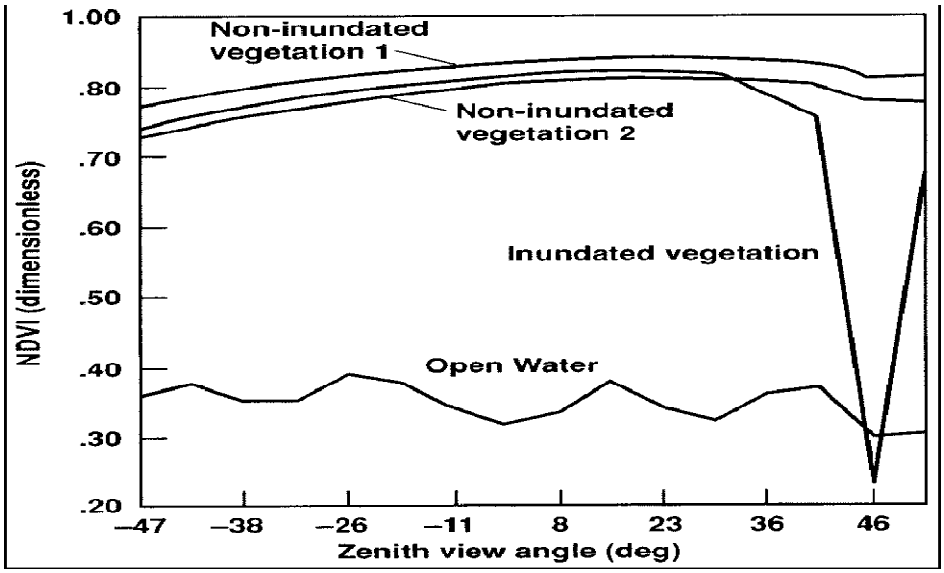

Figure 3.12. The NDVI of various cover types varied with view angle, being less than 0.3 forinundated vegetation viewed in the specular direction. The four classes—noninundatedvegetation 1, noninundated vegetation 2, inundated vegetation, and open water—wereidentified with the aid of a classification performed in spectral space (from Vanderbilt et al., 2002)

3.5 Algorithm Output

Output of the Flood/standing water algorithm mainly contains two data arrays: the Flood/Standing Water identification, plus associated quality control flags, as described in Table 3.8.

Table 3.8 Algorithm output data.

Name	Type	Description	Dimension
Flood/Standing Water class identification	Byte	Determined flood identification for each pixel of the scanning mode as: 1: Flood, 2: Non-Flood, 3: Water body without change, 4: Land without change, 200: Cloud, 100: Cloud Shadow, Snow/Ice: 50, and 255: Missing Value.	grid (xsize, ysize)
QC flags	Byte	Quality control flags for each pixel of the scanning mode: cloudiness, sensor data quality, etc.	grid (xsize, ysize)

Details of the 1-byte product Quality Control (QC) flags are listed in Table 3.9. The value of the entire 1 byte indicates the Quality Flag of the corresponding pixel. QF=0, the pixel of the product is good. QF>0, the pixel of the product is bad.

Table 3.9. Definition of the FSW product Quality Control flags.

Bit	Flag	Definition
0	Availability	0: normal, 1: space or bad or missing pixel
1	Cloud Mask	0: clear, 1: cloudy
2	Cloud Shadow Mask	0: clear or NA, 1: cloud shadow
3	Satellite Angle	0: normal, 1: satellite zenith angle>67º
4	Solar Angle	0: normal, 1: solar zenith angle>67º
5	Snow/Ice Mask	0: clear or NA, 1: snow or ice pixel

In addition, the flood/standing water retrieval processing will also produce some metadata describing processing information (e.g. date/time stamps), as are listed in Table 3.10.

Table 3.10. Metadata defined for the FSW product file.

METEDATA	TYPE	DEFINITOIN
DateTime	common	*Date and time of swath beginning and swath end*
Bounding Box	common	*Product resolution (nominal at nadir), number of rows and number of columns, byte per pixel, data type, byte order information, location of box relative to nadir (pixel space*)
Product Name	common	The ABI FSW

Ancillary Data	common	*Ancillary data name used to produce the product: version number, origin (where it was produced), name*
Satellite	common	GOES-R
Instrument	common	Advanced Baseline Imager
Altitude	common	*Altitude of the satellit*e
Nadir	common	*Pixel in the fixed grid*
Position	common	*Latitude and longitude of the satellite positio*n
Projection	common	*Grid Projectio*n
Mode	common	*Type of Scan mode*
Version	common	*Product version number*
Compression	common	*Data compression type (method) used*
Location	common	*Location where the product is produced*
Contact	common	*Contact information of the producer/scientific supporter*
document	Common	*Citations to documents (i.e., ATBD*)
Product Unit	FSW	NON
Statistics	FSW	*Total number of FSW pixels, total number of Non-FSW pixels,*
Good pixels	FSW	*Total number of "good" pixels (defined by QF)*
Total Pixels	FSW	*Total number of "bad" pixels (defined by QF)*

Note: the definitions in *italic* words are determined at running.

4 TEST DATA SETS AND OUTPUTS

4.1 Simulated and Proxy Input Data Sets

The selected algorithm should be tested using real satellite data. Since the ABI data is not available during the development phase, we used data from other satellite sensors as proxies: the moderate resolution imaging spectroradiometer (MODIS) aboard the NASA Earth Observing System (EOS) Terra and Aqua platforms, and the Spinning Enhanced Visible and Infra-red Imager (SEVIRI) aboard the European Meteosat Second Generation (MSG) satellite. The data from these satellite sensors are considered as good proxies of ABI since they have visible (VIS) and near-infrared (NIR) channels similar to those on the ABI. Table 4.1 lists the sensor spectral specifications of the MODIS and MSG/SEVIRI, as well as ABI.

Table 4.1. Similarity of MODIS and SEVIRI Imagers to ABI in channel spectrum.

Sensor	Channel No.	Wavelength Center (μm)	Bandwidth (μm)	Spatial Resolution
ABI	2	0.64	0.59 – 0.69	0.5 km
	3	0.86	0.8455 – 0.8845	1 km
MODIS	1	0.645	0.620 – 0.670	0.25 km
	2	0.858	0.841 – 0.876	0.25 km
SEVIRI	1 (VIS0.6)	0.635	0.56 ~ 0.71	3 km
	2 (VIS0.8)	0.81	0.74 ~0.88	3 km

4.2. Algorithm Testing with the Simulated Data

The C4.5 classification algorithm (Quinlan, 1993) was applied to the partial simulations as training data to build a decision tree:

- Rule 1:
- CH_3 <= 2.65562%
- -> class Water [99.5%]
- Rule 2:
- CH_3 > 2.65562%
- -> class Land [99.8%]

- Evaluation on training data (10560 items):
- Rule Size Error Used Wrong Advantage
- 1 1 0.5% 2880 0 (0.0%) 2880 (2880|0) Water
- 2 1 0.2% 7680 0 (0.0%) 0 (0|0) Land

- Tested 10560, errors 0 (0.0%) <<
- (a) (b) <-classified as
- ---- ----
- 2880 (a): class Water
- 7680 (b): class Land

The number in brackets under the possibility value (%) in rectangles at each leaf equals the number of training instances which belong to that path in the tree. Meanwhile, each number is followed by the number of classification errors encountered in that particular path of the decision tree. The simulation results show that the near-infrared channel (CH_3 for ABI) reflectance is the most effective attribute to separate water from land. When CH_3 reflectance is less than a threshold value (2.66% here), water can be identified. The total accuracy can reach over 99%. Since in our GOES-R ABI forward simulations, we didn't include calibration error. The simulated data may be an ideal case and quite different from the real situation, the water are also pure water, thus the rules may be too simple, and the water classification accuracy is close to perfect, and may be unrealistic high.

4.3 Output from Input data sets

4.3.1 Test with the Proxy MODIS data

In June 2008, unusually heavy rains from the 17th to the 19th in the upper Midwest triggered flooding throughout the upper Mississippi basin. EOS/MODIS channels 7, 2, and 1 RGB composite images show water bodies before and after flooding (Figure 4.1).

As described in the introduction, several parameters, including MODIS channel 2 reflectance (CH_2) and channel 1 reflectance (CH_1), the difference (CH_2-CH_1) and ratio (CH_2/CH_1) between CH_2 and CH_1, NDVI, brightness temperature at 11 (MODIS channel 31) or 12 µm (MODIS channel 32), and surface temperature, can be used to identify water from land in previous studies. Which parameters or the combination of several parameters are the most effective? Moreover, how can the threshold values can be determined? What is the accuracy? We propose to apply the Decision Tree method to identify Flood/Standing Water from the proxy MODIS measurements, because DT can integrate all the possible candidate predictors, at the same time, it can determine the threshold values, and give accuracy estimates.

4.3.1.1 Data used for training

- **MODIS land cover type products (MCD12Q1)**

  MODIS land cover type 1, which includes 17 International Geosphere-Biosphere Programme (IGBP) types: (0) water, (1) evergreen needleleaf forest, (2) evergreen broadleaf forest, (3) deciduous needleleaf forest, (4) deciduous broadleaf forest, (5) mixed forest, (6) closed shrublands, (7) open shrublands, (8) woody savannas, (9) savannas, (10) grasslands, (11) permanent wetlands, (12) croplands, (13) urban and built-up, (14) cropland/natural vegetation mosaic, (15) permanent snow and ice, (16) barren or sparsely vegetated.
- **MODIS 8-day composite surface reflectance, atmospheric correction products (MYD09A1)**

  The MODIS Surface Reflectance product (MYD09A1) is computed from the MODIS Level 1B land bands 1, 2, 3, 4, 5, 6, 7, G (centered at 0.648 µm, 0.858 µm, 0.470 µm, 0.555 µm, 1.24 µm, 1.64 µm, and 2.13 µm, respectively). The product is an estimate

of the surface reflectance for each band as it would have been measured at ground level after removing the atmospheric scattering and absorption.

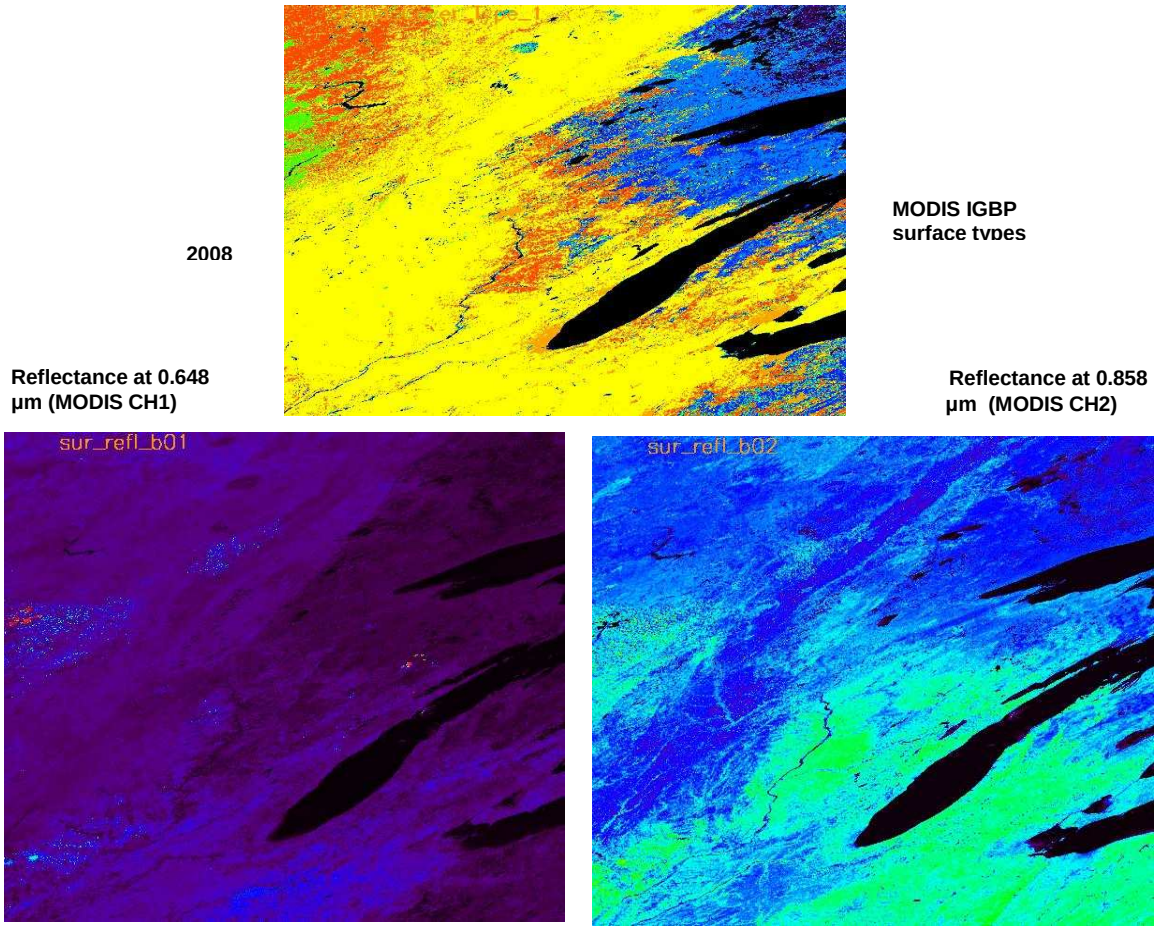

Figure 4.1. MODIS land cover at 1 km resolution and reflectance data at 0.648 µm (MODIS CH1) and 0.858 µm (MODIS CH2) on May 28, 2008 before flooding were used for training.

- MODIS Terra+Aqua Nadir-BRDF (Bidirectional Reflectance Distribution Function) Adjusted Reflectance 16-Day L3 Global 500m SIN Grid V005 (MCD43A4).
- MODIS cloud mask (MOD35) data
  Cloud mask data is used to filter the cloudy conditions.

4.3.1.2 Results from the MODIS training data

The MODIS surface reflectance and land cover data before flooding on May 28, 2008 are used for training (Figure 4.2). From Figure 4.2, we can see clearly that MODIS channel 2 reflectance (CH2) can 'see' water more clearly than channel 1 (CH1), particularly in the case of rivers. In order to perform DT analysis, surface reflectances (%) at 250m are aggregated to the same 1 km resolution as the land cover data. MODIS cloud mask (MOD35) data are used to filter clouds.

Figure 4.4 shows an example of the tree structure derived from the J48 (or the C4.5) algorithm for the discrimination of water from land. The tree employs a case's attribute values to map it to a *leaf*, for designating one of the classes. The number in brackets following each leaf equals the number of training instances that are mapped to this leaf, and the second number after "/" in brackets (if it appears) is the number of instances that are misclassified to this leaf. A non-integral number of cases may arise. This is because when the value of an attribute in the tree is not known, C4.5 splits the case and sends a fraction down each branch. The node in the upper level of the tree has a higher information gain ratio than in the lower level node in the classification. Therefore, as shown in Figure 4.4, attributes/parameters like the CH2-CH1, which appear at the root node of the tree, are more important than those at the lower level, such as the CH2 and NDVI, for identifying water from land.

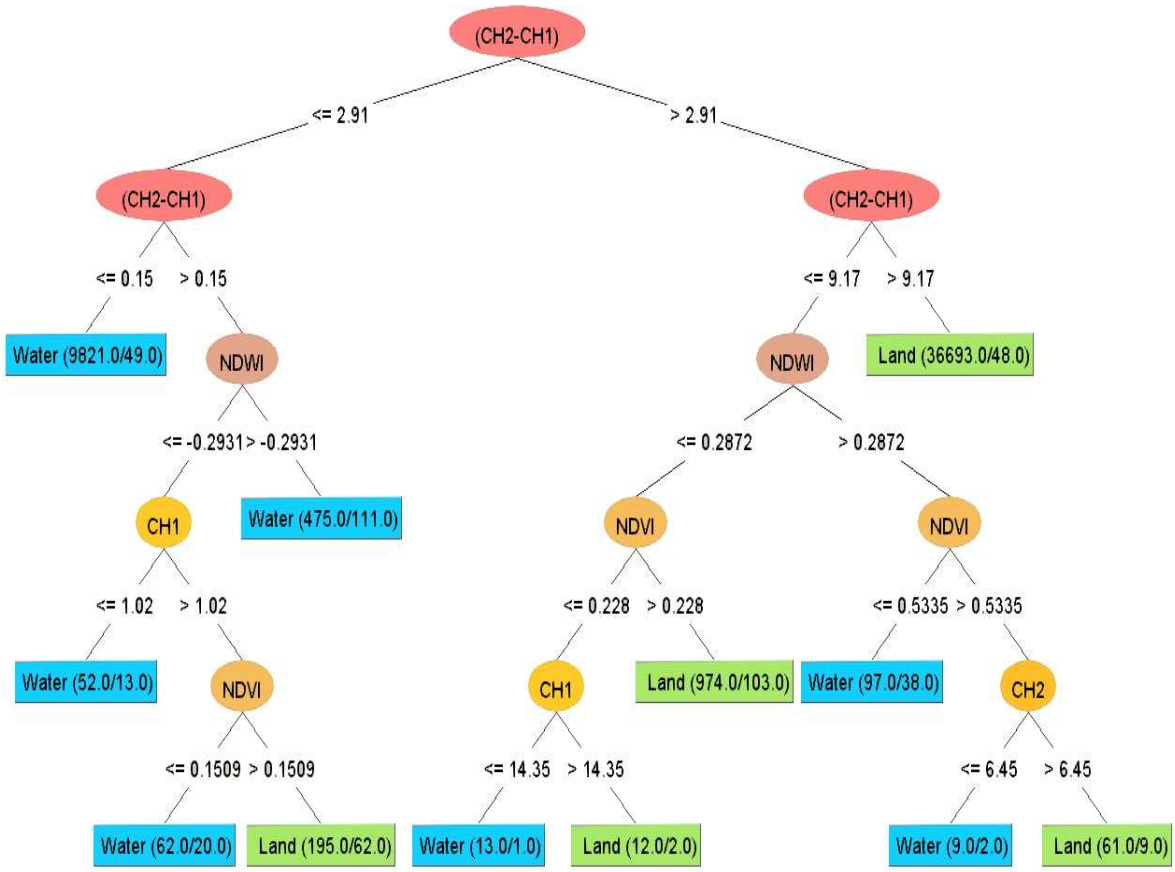

Figure 4.3. An example of decision tree structure derived from the MODIS observations with the C4.5 algorithm.

Rulesets:

Table 4.2. Rule sets generated from the C4.5 algorithm

Rule #	CH1	CH2	(CH2-CH1)	NDVI	NDWI	Class	Accuracy
19			>9.17			Land	99.9%
18		>6.45		>0.5335		Land	99.8%
15			>2.91	>0.228	<=0.2872	Land	99.6%
9	<=1.02		<=2.91			Water	99.2%
10			<=2.91	<0.1509		Water	99.0%
12			<=2.91		>-0.2931	Water	97.7%

Each rule consists of:

- A rule number -- this is quite arbitrary and serves only to identify the rule.
- The set of rules usually consists of at least one rule, which is used to classify unseen instances when no other rule applies.
- Every enumerated rule is composed of attribute-values and a resulting classification, followed by a percentage which represents the accuracy of that rule.

For example, Rule 9: $CH_1 \leq 1.02\%$ and CH2<=2.91%, this rule is accurate 99.2% of the time for water classification.

Rule 19: $CH_2 - CH_1 > 9.17\%$, this rule has accuracy of 99.9% for land classification.

Confusion matrix:

Table 4.3. The confusion matrix from the C4.5 test.

	Classified Water (1)	Classified Non-water(0)
Validation Water (1)	10235	284
Validation Non-water (0)	238	37707

The node in the upper level of the decision tree has a higher information gain ratio than in the lower level node in the classification. Therefore, the reflectance difference (CH2-CH1), which appears at the root node of the decision tree, is more important than those in the lower level, such as the NDVI, NDWI, NDVI, NIR reflectance difference CH_2 and visible reflectance CH_1, or reflectance ratio CH_2/CH_1 between the MODIS CH_2 and CH_1, in identifying the water.

The threshold values from the observation data are different from the simulation results, which show the CH_2 reflectance is the only effective attribute, and the threshold value is 2.65562%. However, for real observation data, the rules are much more complicated: there are several effective attributes with multiple threshold values, and the total error is 1.6%, which equals an accuracy of 98.4%, and is lower than the simulation accuracy of over 99%. Nevertheless, all algorithm performance tests at 1km resolution are well above the required accuracy of 60%. Remember that due to difference in central wavelength and spectral response functions of the visible/near infrared channels, algorithm threshold values applied for different satellite sensors may be different.

Considering that because the surface reflectance may have anisotropic effects, we tested the BRDF-adjusted nadir reflectance data (MCD43). The results show that over the same sample area, the attributes at the top of the output tree structure are the same. Figure 4.4 gives an example for samples around the Mississippi River, where the reflectance difference (CH2-CH1) is the most useful attribute to identify inland water from land. However, it is noted that the threshold values are different. For the BRDF-adjusted reflectance data, the ratio CH2/CH1 appears at a lower level in the tree structure and plays a critical role to separate land from water, but may still be less important than the difference (CH2-CH1), which appeared at the top level of the tree structure (Sun and Yu, 2010).

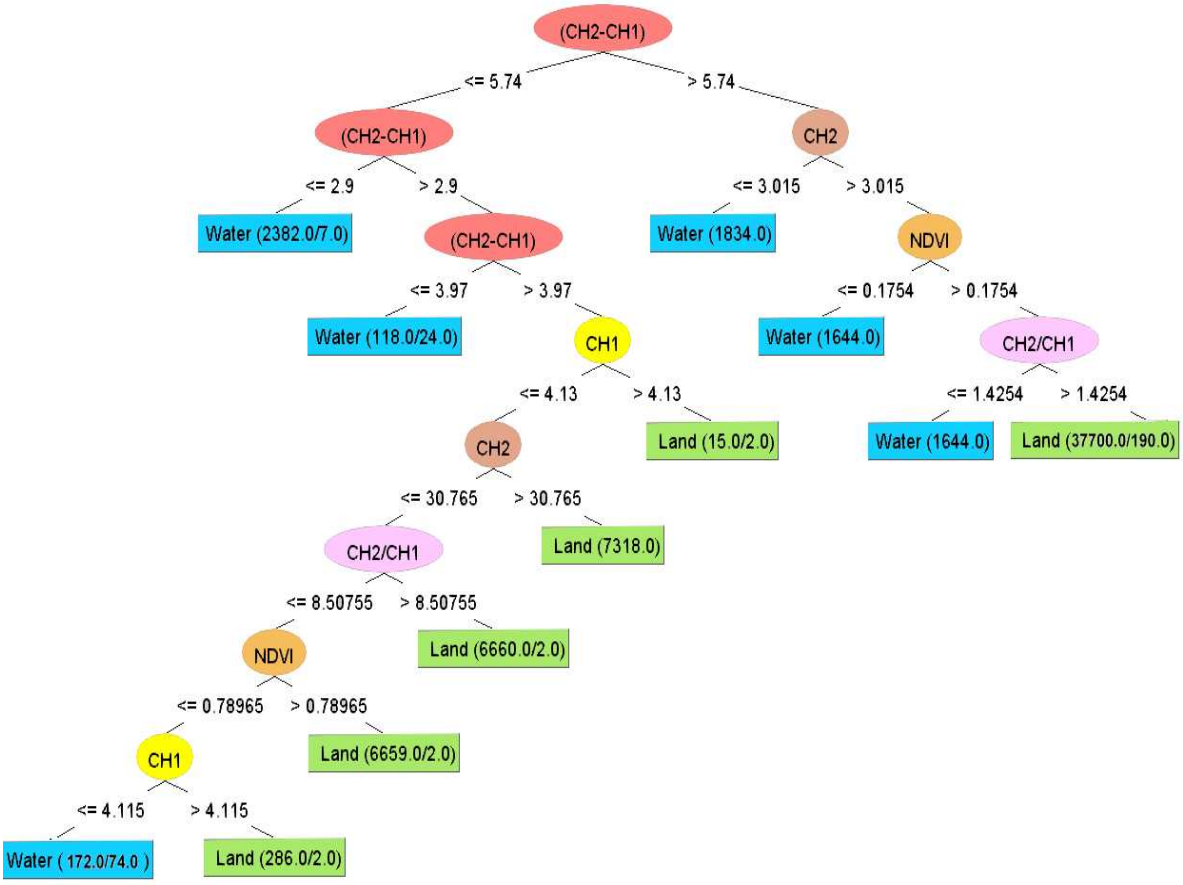

Figure 4.4. An example of Decision Tree structure derived from the BRDF adjusted surface reflectance data with the samples around the Mississippi River using the C4.5 algorithm.

Reflectance in the NIR band (e.g., MODIS CH2) is largely affected by water types. Turbid water usually has a higher reflectance in the NIR band than clean water. When water is contaminated by blue-green alga, the NIR reflectance may increase significantly and become even much higher than the visible reflectance (e.g., MODIS CH1). Therefore, with the increase of water turbidity after flooding, the ratio CH2/CH1 may become more useful than the difference (CH2-CH1). Since we do not have land cover data for after-flooding situation, we tried to manually identify

water and land using the RGB composite images, and then match with the reflectance data. The output tree structure is too complicated to be completely shown in a figure. The results demonstrate that the CH2 reflectance appears at the top of the tree structure; meanwhile, the ratio does appear at a number of leaf nodes in the tree structure, sometimes even at the upper nodes than the difference (CH2-CH1), indicating that the ratio CH2/CH1 may be a more useful parameter to identify turbid water from land than the difference (CH2-CH1).

It may need to note Figures 4.3 and 4.4 just show some examples of DT tree structures. The real DT model may be more complicated, and can't be even shown in one figure.

4.3.1.3 DT training with additional Surface Temperature (ST) data

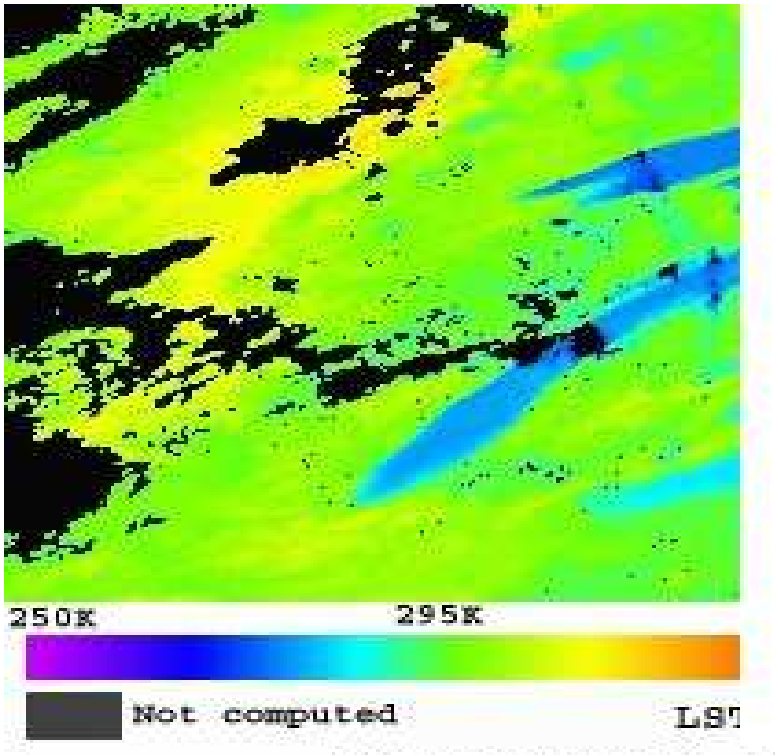

Figure 4.5. MODIS land surface temperature over the training area on sate date (May 28, 2008).

From Figure 4.4, we can see water temperature is usually lower than land surface temperature. If we include surface temperature as an additional attribute for the training, the results are shown below.

***Decision tree*:**

CH_2/CH_1 <= 1.52025 :
| ST <= 289.24 :
| | CH_2/CH_1 <= 1.23318 : Water (2584.8/18.5)

| | CH_2/CH_1 > 1.23318 :
| | | CH_2-CH_1 > 9.79 : Land (6.0/1.2)
| | | CH_2-CH_1 <= 9.79 :
| | | | CH_1 <= 1.65 : Water (24.0/1.3)
| | | | CH_1 > 1.65 :
| | | | | CH_2-CH_1 > 1.16 : Water (47.0/8.3)
| | | | | CH_2-CH_1 <= 1.16 :
| | | | | | ST > 288.64 : Water (2.0/1.0)
| | | | | | ST <= 288.64 :
| | | | | | | NDVI <= 0.141678 : Land (5.0/1.2)
| | | | | | | NDVI > 0.141678 : Water (3.0/2.1)
| ST > 289.24 :
| | ST <= 295.28 :
| | | NDVI <= 0.0836654 : Water (540.0/95.6)
| | | NDVI > 0.0836654 :
| | | | CH_2-CH_1 > 8.29 : Land (12.0/1.3)
| | | | CH_2-CH_1 <= 8.29 :
| | | | | CH_1 <= 1.23 : Water (14.0/1.3)
| | | | | CH_1 > 1.23 :
| | | | | | CH_2-CH_1 <= 0.42 : Land (11.0/2.5)
| | | | | | CH_2-CH_1 > 0.42 : Water (138.0/53.5)
| | ST > 295.28 :
| | | CH_2 <= 2.04 : Water (52.0/18.9)
| | | CH_2 > 2.04 : Land (198.0/71.2)
CH_2/CH_1 > 1.52025 :
| CH_2-CH_1 > 7.63 : Land (110881.0/1558.6)
| CH_2-CH_1 <= 7.63 :
| | CH_2-CH_1 <= 0.52 : Water (79.2/1.4)
| | CH_2-CH_1 > 0.52 :
| | | ST > 293.4 : Land (2206.0/305.7)
| | | ST <= 293.4 :
| | | | CH_1 <= 4.54 : Land (332.0/162.8)
| | | | CH_1 > 4.54 : Water (107.0/36.9)

Evaluation on training data (117242 items):

Before Pruning		After Pruning		
Size	Errors	Size	Errors	Estimate
59	2250 (1.9%)	37	2257 (1.9%)	(2.0%) <<

Rulesets:

Rule 15:

CH_2 <= 2.04%

CH_2/CH_1 <= 1.52025

-> class Water [96.5%]

Rule 13:
CH_2-CH_1 <= 8.29%
CH_2/CH_1 <= 1.52025
ST <= 295.28 K
-> class Water [94.5%]

Rule 20:
CH_2-CH_1 <= 7.63%
ST <= 293.4 K
-> class Water [90.5%]

Rule 22:
CH_2-CH_1 <= 7.63 %
CH_2/CH_1 > 1.52025
ST > 302.82
-> class Land [99.2%]

Rule 25:
CH_2-CH_1 > 7.63 %
NDVI > 0.680956
-> class Land [99.1%]

Rule 17:
CH_2-CH_1 > 5.22
ST > 295.28
-> class Land [98.8%]

Rule 8:
CH_2-CH_1 > 9.79
-> class Land [98.7%]

Rule 14:
CH_2-CH_1 > 8.29
ST > 289.24
-> class Land [98.7%]

Default class: Land

Evaluation on training data (117242 items):

Rule	Size	Error	Used	Wrong	Advantage
15	2	3.5%	2832	91 (3.2%)	20 (36|16) Water
13	3	5.5%	588	100 (17.0%)	26 (57|31) Water
20	2	9.5%	519	209 (40.3%)	101 (310|209) Water

```
 22    3   0.8%  1047     6 (0.6%)       0 (0|0)    Land
 25    2   0.9% 31918    271 (0.8%)      0 (0|0)    Land
 17    2   1.2% 68892    906 (1.3%)      0 (0|0)    Land
  8    1   1.3% 10157    325 (3.2%)      0 (0|0)    Land
 14    2   1.3%   259     62 (23.9%)      0 (0|0)    Land

Tested 117242, errors 2291 (2.0%)   <<
```

Table 4.4. The confusion matrix from the C4.5 test.

	Classified Water (1)	Classified Land (0)
Validation Water (1)	3539	1891
Validation Land (0)	400	111412

Since surface temperature may vary significantly with time, and spatial locations, thus including surface temperature will introduce complexity and instability to the algorithms. Therefore, we will not use surface temperature as an additional predictor or attribute.

Antecedent precipitation should be a good attribute or predictor for Standing Water/Flood detection; however, currently no satellite precipitation product can be available at 1 km resolution.

4.3.1.4 "Future prediction" with the rules obtained from training

Figure 4.6 shows water (values as 1 in blue) and land (values as 0 in green) distributions determined by applying the tree structure (Figure 4.4) obtained from the training data before flooding on May 28, 2008:

The "predicted" water bodies on June 14, 17, and 19, 2008 by applying the rules from the training (Table 2, Figure 4.4) are quite close to the qualitative visual image analysis as shown in Figure 3.4.

Figure 4.7 shows the resulting maximum flood map based on differences between the water maps after flooding on June 17 and 19 and before flooding on June 14, 2008 as shown in Figure 4.6.

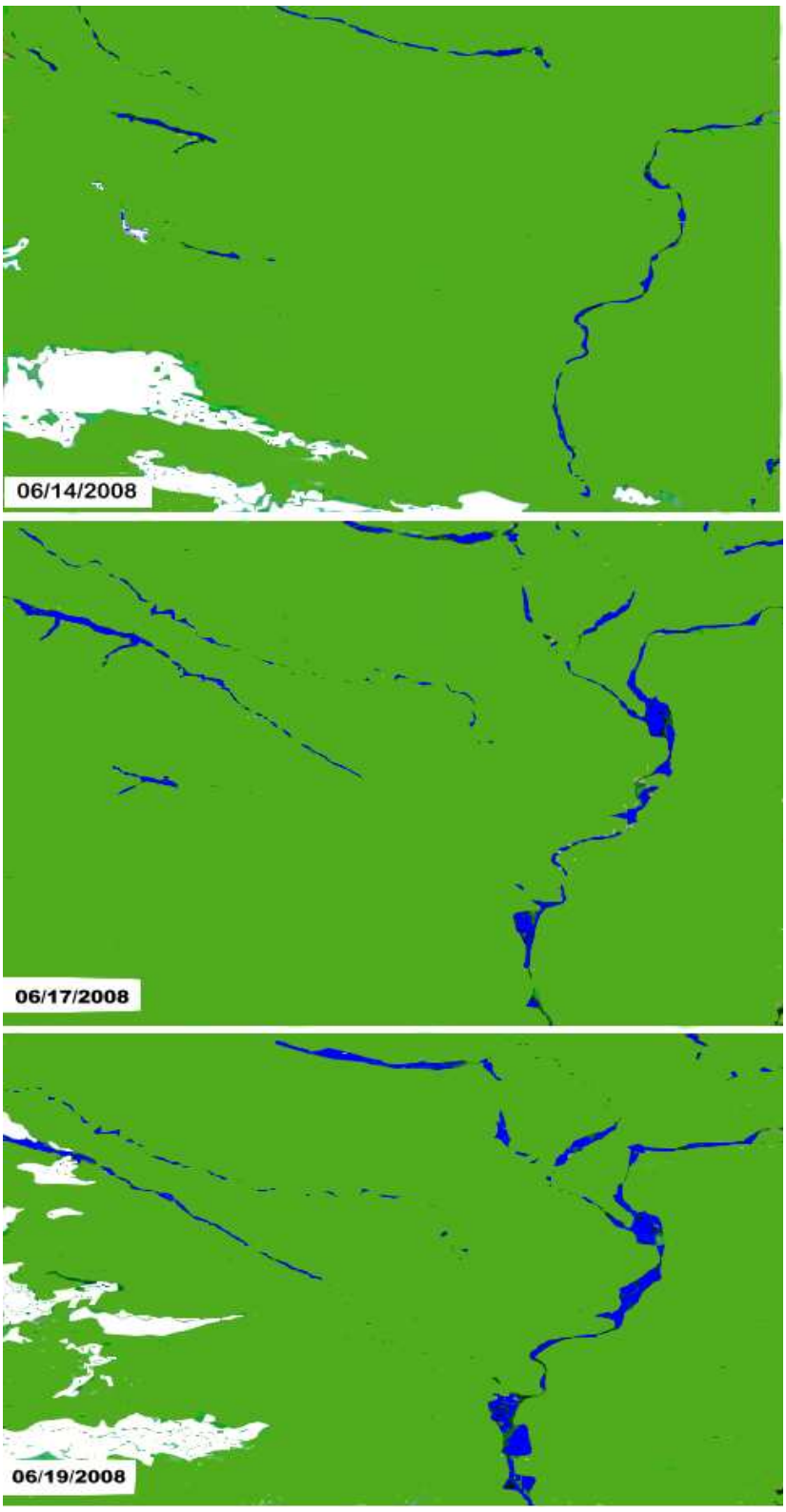

Figure 4.6. Spatial distributions of water (blue)/land (green) identifications over time

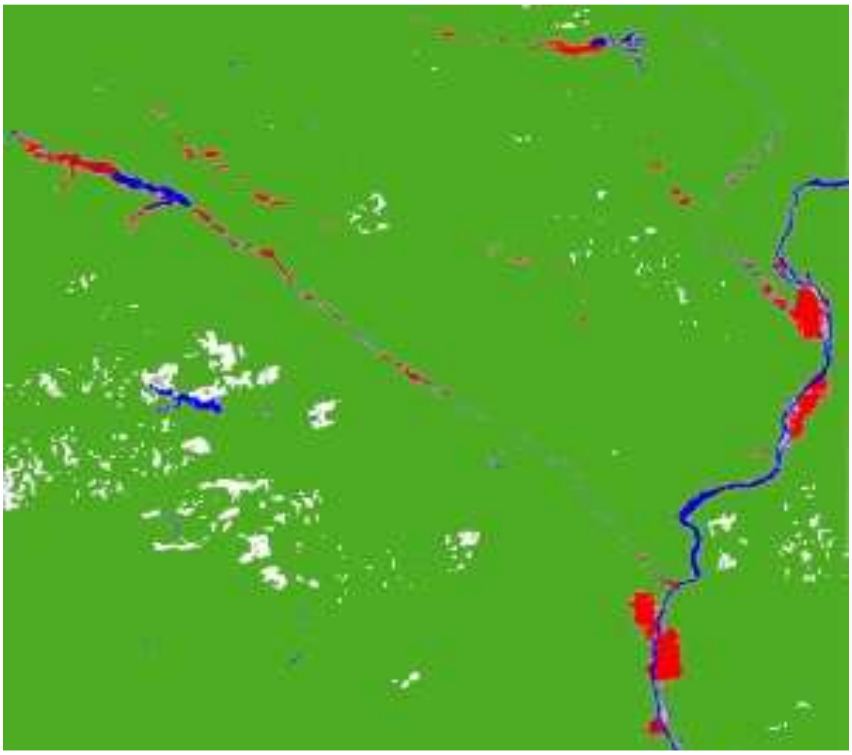

Figure 4.7. The maximum flood extent map during the flood period (original water body is shown in blue, flood is shown in red, and land is shown in green).

We applied these CH_2 reflectance and NDVI predictors and threshold values from the training data at the Midwest of the United States on May 28, 2008 to "re-predict" the New Orleans flooding at the end of August in 2005 due to the landfall of Hurricane Katrina, which caused over 1500 deaths and damage exceeding $50 billion. Figure 4.8 shows the MODIS channels 7, 2, and 1 RGB composite images on August 30 and 31 after flooding, and August 27 before flooding. Figure 4.9 shows the water identification map on these three days calculated by applying the CH_2 reflectance and NDVI predictors and threshold values. From these images, we can clearly detect the flood areas by comparing water fraction maps after flooding with those before flooding. Figure 4.10 depicts the flood map as the difference in water detection values after flooding on August 31 and 30, with those before flooding on August 27. The flooded regions are identified in red colors, the original water bodies are shown in blue color, while clouds are marked in grey color. We can see very clearly that New Orleans and its surrounded areas were inundated on August 30 and 31, 2005 after Hurricane Katrina made landfall on August 29, 2005.

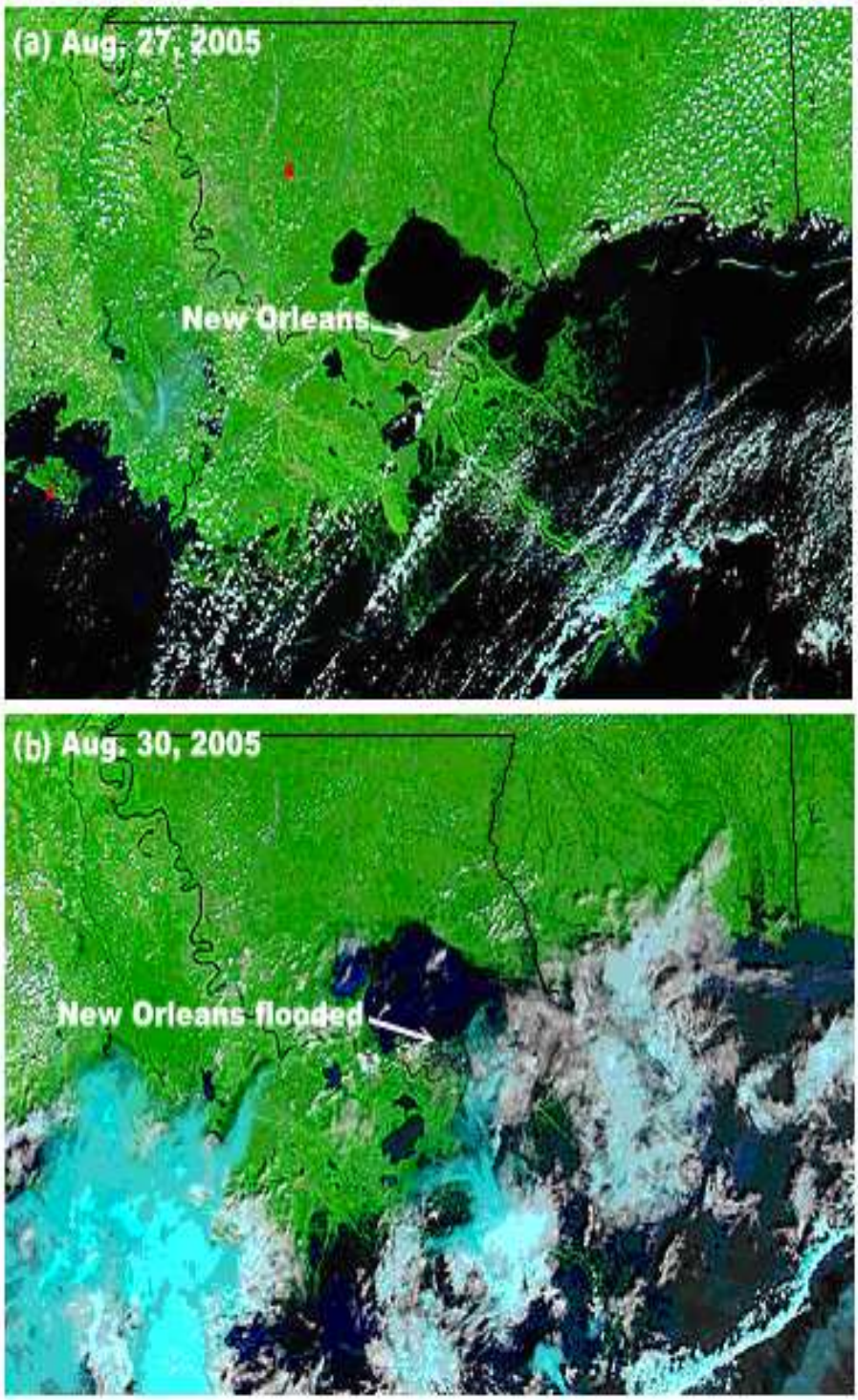

Figure 4.8. MODIS RGB composite images on August 27 (a) and 30 (b), 2008.

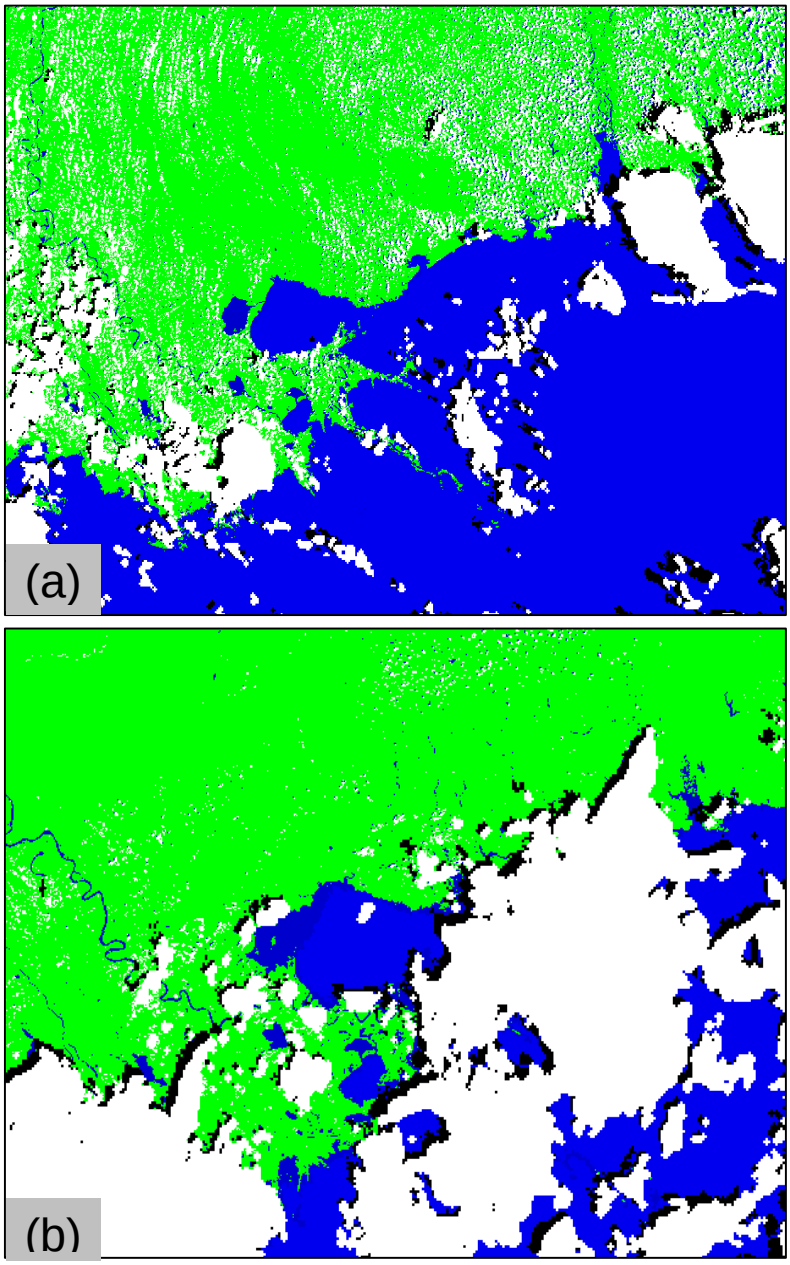

Figure 4.9. Water identification map on August 14 (upper) and 30 (lower), 2005 (water: blue; Green: Land or Non-water; white: clouds; cloud shadow: grey)

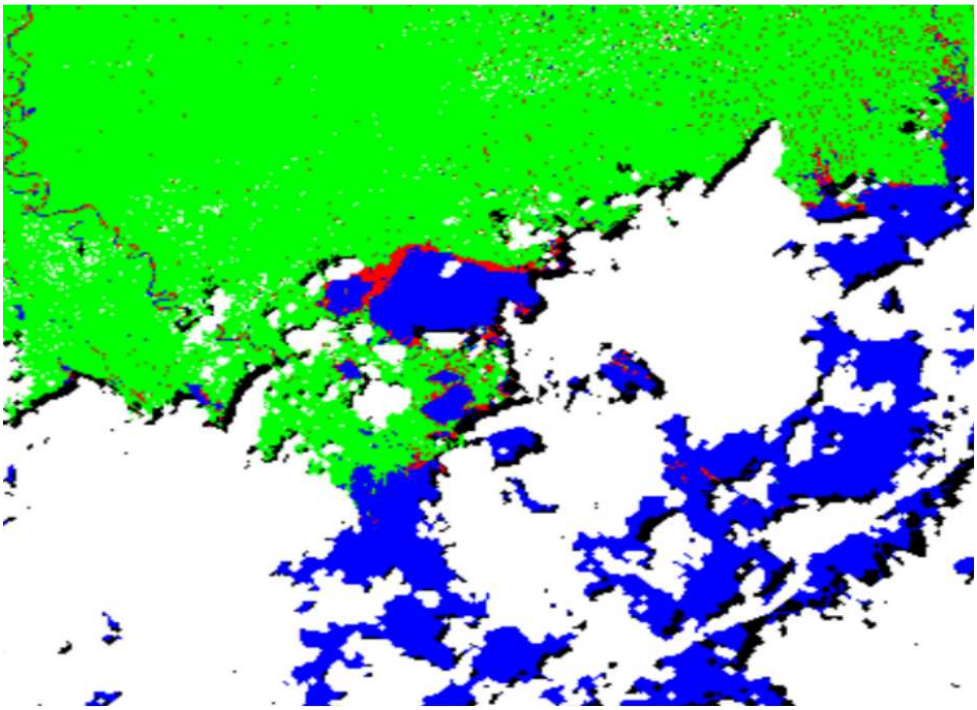

Figure 4.10. Flood map on August 30, 2005 shown as the water difference between August 30 (after flooding) and August 27 (before flooding) (blue: original water body, red: flood, white: clouds, black: cloud shadow, and green: land).

4.3.2 Test with the proxy SEVIRI data

In the SEVIRI data test, several attributes, including SEVIRI channel 1 (CH1), channel 2 (CH2), the difference (CH2-CH1) and ratio (CH2/CH1) between CH2 and CH1, NDVI, are employed as predictors.

4.3.2.1 Data used for training

The reflectance value of CH1 and CH2 are obtained by converting the original digital count to radiances and then to reflectance. The difference, ratio and NDVI values are calculated in the preprocessing stage. Regions of interest are extracted manually according to 18-classes IGBP map, which all non-water classes are combined into land class. The ROIs acquired by different date and time are merged into a large training set with over 110000 records, and the tree structure from the training with the 5 bands SEVIRI data using the C4.5 algorithm (See Figure 4.11). Figure 4.11 shows that NDVI is the most useful predictor to separate water from land, and the ration and the CH1 are also used in the subsequent decision. It is worth noting that the tree structure is not fixed for different datasets. To ensure the discriminative power, the C4.5 algorithm may select different leaf nodes and structure.

4.3.2.2 "Future prediction" with the rules obtained from training

The tree model from the training is applied to real SEVIRI observations. It has been discovered that if the solar zenith angle (SZA) is greater than 67º at 15:15 UTC, as shown in Figure 4.12, then misclassification may happen. The reason for this is the low illumination under large SZA conditions (Figure 4.12 lower). For normal illumination conditions, such as at time of 12:00 UTC (Figure 4.13), when SZA is less than 67º, then no obvious misclassification may happen. Therefore, the FSW product requirement under the SZA to be less than 67º is quite reasonable. In our product, the pixels with large zenith angles as well as large satellite zenith angles are masked out as no-data.

We applied the standing water identification algorithm to a Mozambique flood case occurred on March 8, 2010. The tree structure only includes the difference of the reflectance values of the Ch1 and Ch2 as the predictor in the model, which is obtained from training data of flood cases. Fig 4.14 shows the MODIS composite images on Mar. 8 (Flood) and Mar. 12 (After flood), and Fig. 4.13 demonstrates the Flood/Standing Water identification map, where red pixels are flood. Comparison with the IGBP 18 classes land cover map merged into the reference water/land map is made to derive the flood pixels. Fig 4.15 is a zoomed view of the red box in Fig 4.13 to show the detailed flood coverage. Because the spatial resolution of the SEVIRI is much lower than the resolution of the MODIS, the identified flood coverage is much smaller than the coverage in MODIS composite images.

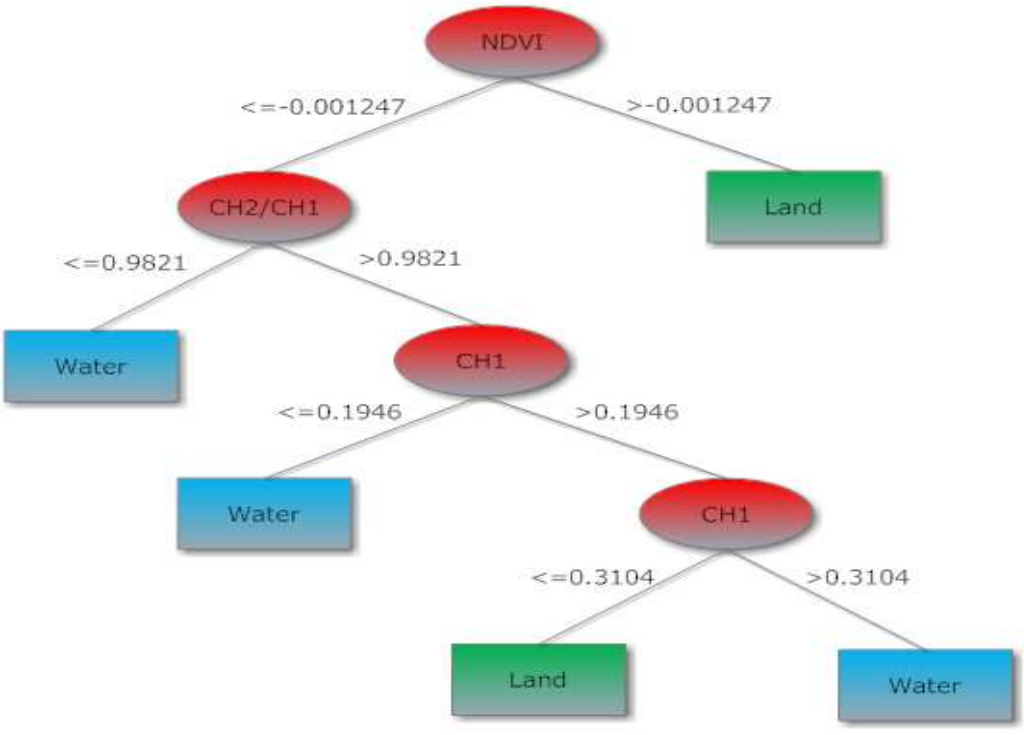

Figure 4.11. An example of decision tree structure derived from the SEVIRI observations with the C4.5 algorithm.

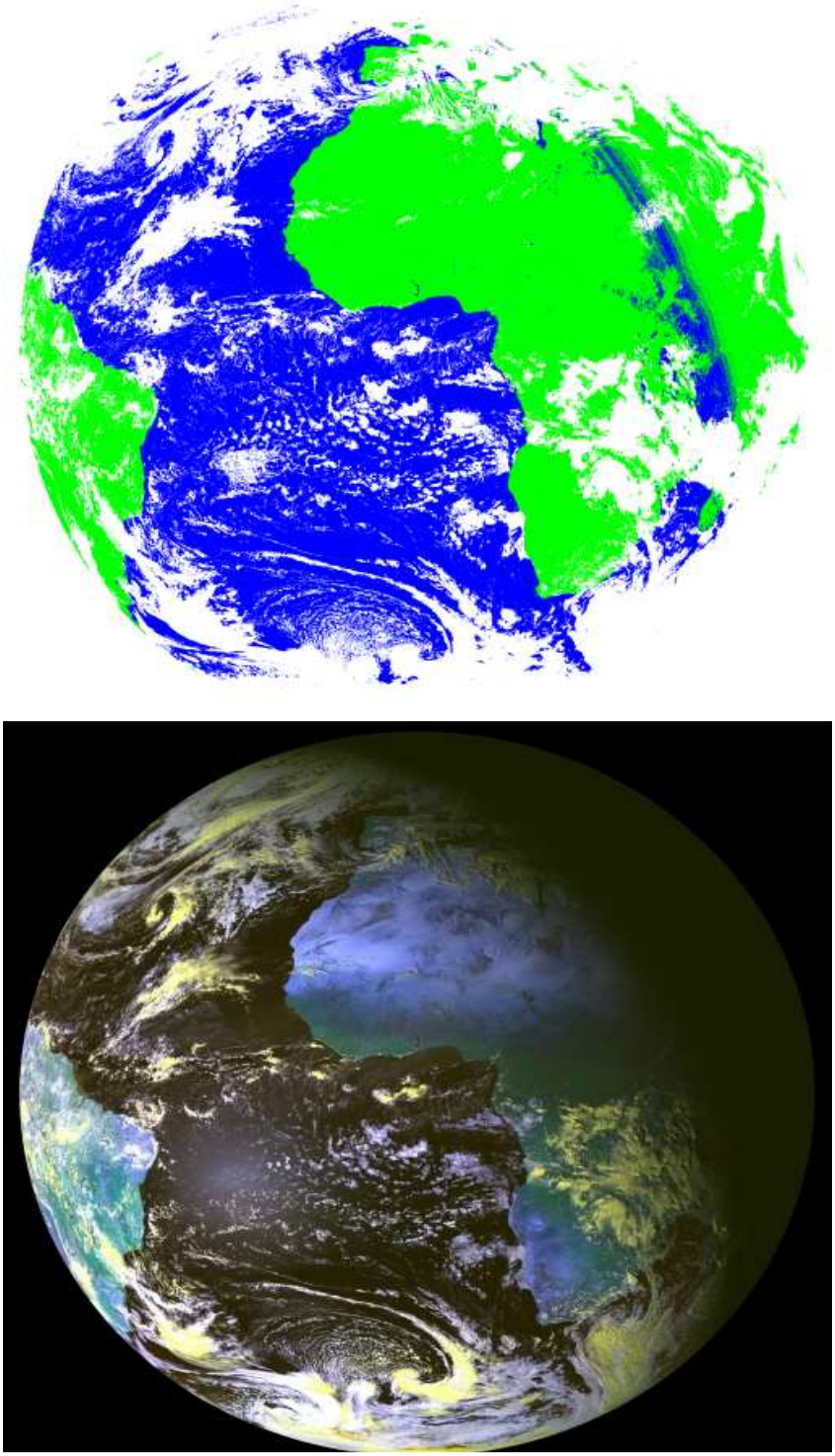

Figure 4.12. Water (blue)/Land (green) derived from SEVIRI observations at 15:15 UTC on12/25/2007.

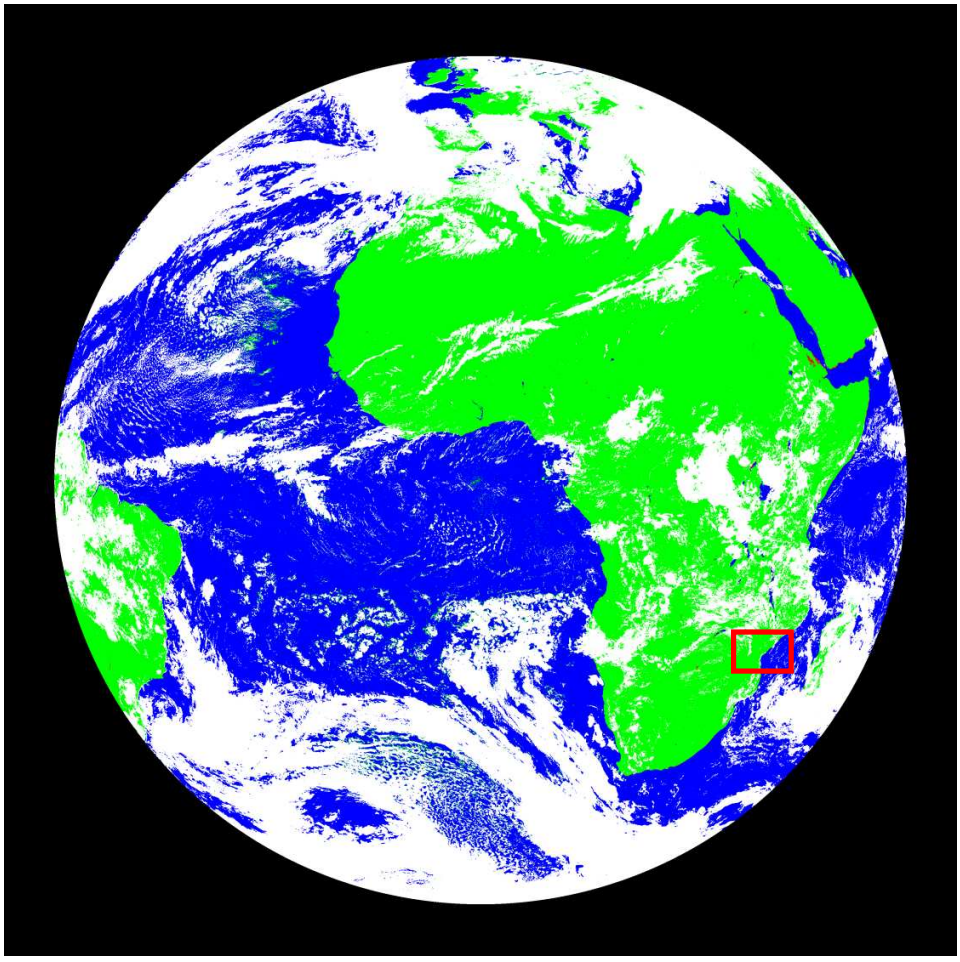

Figure 4.13 Flood (Red), Water (Blue) and Land (Green) derived from SEVIRI observations at 12:00 UTC on 3/8/2010

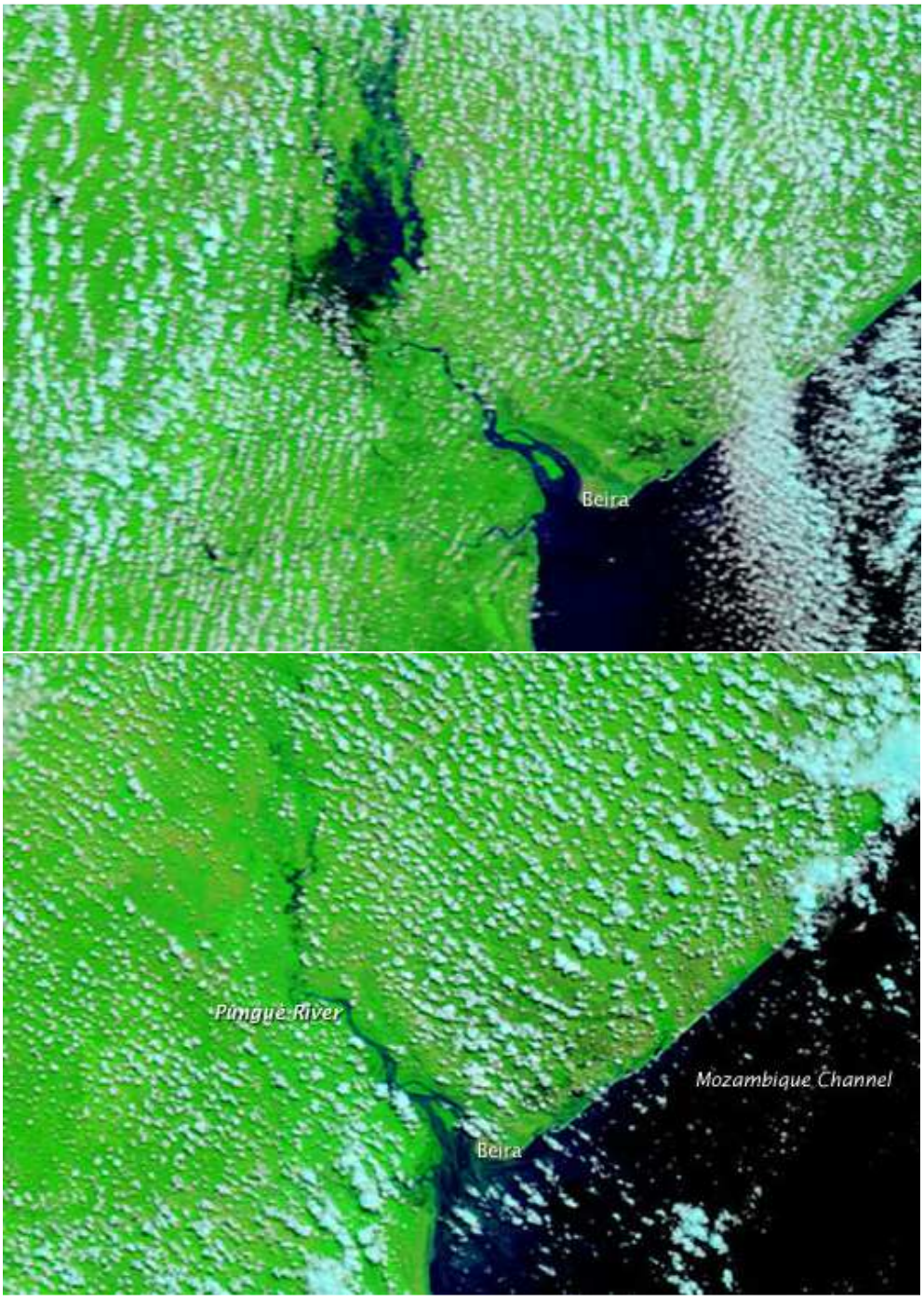

Figure 4.14 MODIS RGB composite images on March 8 (upper) and March 12 (lower)

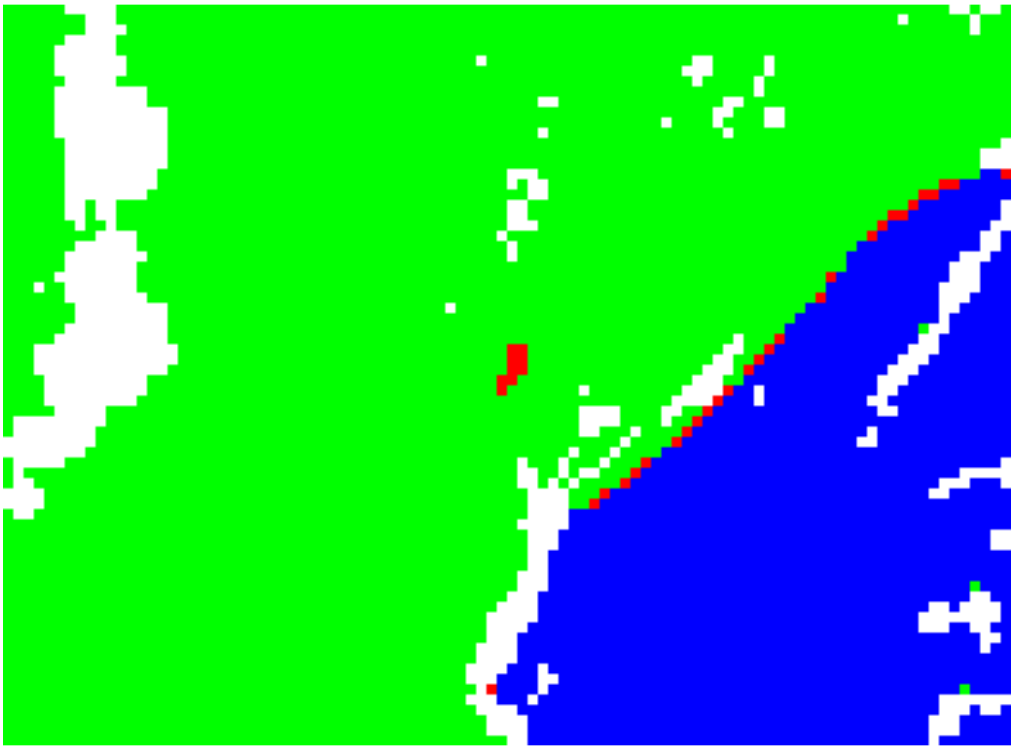

Figure 4.15 Flood map on 12:00 UTC, March 8, 2010. (original water body is shown in blue, flood is shown in red, land is shown in green and white pixels are cloud).

4.4 Precision and Accuracy Estimates

The accuracy of flood identification is determined by the accuracy of water body detection. Since there is no flood map available for MODIS and SEVIRI data now, we can easily think about using ground truth of land cover/land use classification to validate GOES-R FSW preliminary product: yes/no water identification. Figure 4.16 shows the comparison of water/non-water identification map from the MODIS observations with the ground truth of water/land classification map. Some land pixels at the west area of the Great Lakes were misclassified into water, an error known as a commission error. The quantitative validation result is listed in Table 4.5. The producer's accuracy for water identification, which is equal to (100%-omission) is 98.38%, the user's accuracy, which is equal to 100%-commission rate, is 84.91%. The correction classification rate refers to producers' accuracy. The total accuracy for both water and non-water classification is 98.36% with kappa coefficient of 0.90. For testing with the proxy SEVIRI data, we can see that some small water body cannot be detected (Figure 4.17), which may be due to the coarse resolution of the SEVIRI data (3km). The producer's accuracy for water detection is 99.87%, and user's accuracy is 99.88%, which is higher than 84.91% from the MODIS testing. The higher accuracy rate is because less commission error occurred with the SEVIRI observations. The total accuracy for both water and non-water classification is 99.37% with kappa coefficient of 0.99 (Table 4.6).

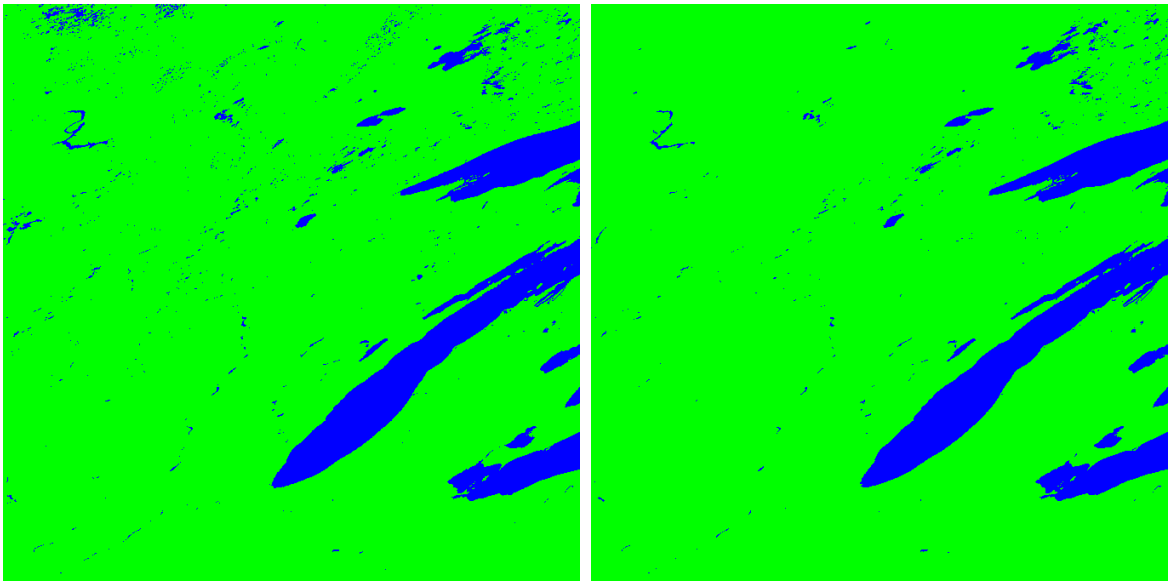

Figure 4.16. Water/land classification from the C4.5 DT algorithm with the MODIS surface reflectance data on 05/20/2008 (left) and ground truth of water/land classification (right)

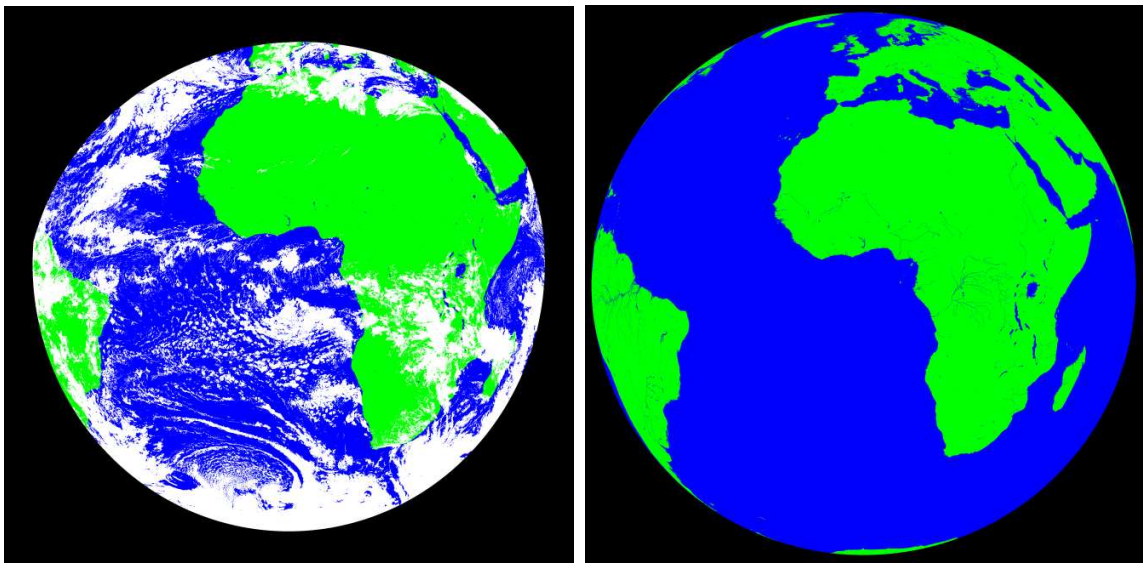

Figure 4.17. Water/land classification from the C4.5 DT algorithm with the SEVIRI data at 11:45 UTC on 12/25/2007 (left) and Ground truth of water/land classification (right)

Table 4.5. Validation results from the proxy MODIS data

Validation Data: MODIS IGBP Land Cover Map (pixels)			
MODIS Classification 05/20/2008		Classified Non-Water (0)	Classified Water (1)
	Validation Non-Water (0)	5185738	9433
	Validation Water (1)	85212	479617
	Total pixels	5270950	489050

Validation Data: MODIS IGBP Land Cover Map (%)				
MODIS Classification 05/20/2008		Classified Non-Water (0)	Classified Water (1)	Producers Accuracy
	Validation Non-Water (0)	98.38%	1.93%	98.38%
	Validation Water (1)	1.62%	98.07%	98.07%
	Users Accuracy	99.82%	84.91%	
			Total Accuracy	98.36%
			Kappa Coefficient	0.90

Table 4.6. Validation results from the proxy SEVIRI data

Validation Data: SEVIRI IGBP Land Cover Map (pixels)			
SEVIRI Classification 12/25/2007		Classified Non-Water (0)	Classified Water (1)
	Validation Non-Water (0)	2445680	29528
	Validation Water (1)	3146	2698957
	Total pixels	2448826	2728485

Validation Data: SEVIRI IGBP Land Cover Map (%)				
SEVIRI Classification 12/25/2007		Classified Non-Water (0)	Classified Water (1)	Producers Accuracy
	Validation Non-Water (0)	99.87%	1.08%	99.87%
	Validation Water (1)	0.13%	98.92%	98.92%
	Users Accuracy	98.81%	99.88%	
			Total Accuracy	99.37%
			Kappa Coefficient	0.99

4.5 Cloud Detection on GOES-R ABI Data

Since we are going to use the GOES-R imager ABI data to detect Standing Water/Floods and the visible and infrared measurements cannot penetrate the clouds, accurate cloud filtering for the Imager data is critical for reliable results. Improved cloud treatment over snow in the shortwave satellite inference scheme has been developed (Li et al., 2007; Pinker et al., 2007) and is ready for use. This algorithm uses four GOES-8 channels, including visible channel 1 (0.67 μm, similar to ABI channel 2), middle infrared channel 2 (3.9 μm, similar to ABI channel 7), infrared channel 4 (10.7 μm, similar to ABI channel 14), and channel 5 (12.0 μm, similar to ABI channel 15), and can be applied to detect cloud during both day and night. We can apply this algorithm for cloud screening.

In general, most clouds cause enhanced reflectance and lower brightness temperatures (except the low level warm clouds) relative to the surface. Snow conditions can be identified from image

sequences, since snow pixels generally tend to be static from one hour to the next hour while clouds move.

Second, if the 11 µm channel brightness temperature is too low (less than 250 K) then most likely the pixel is cloudy. In addition, the absolute brightness temperature difference between 11 µm and 12 µm channels, and 11 µm and 3.9 µm channels, should be in general less than 1.5 K for identifying the pixel as cloud-free.

When the GOES-R ABI cloud mask data is available, we can use this product directly to filter the clouds.

4.6 Error Budget

The algorithm accuracy and product accuracy may be different for our FSW product. The algorithm accuracy will be determined by the accuracy of water classification. The flood is identified by comparison of water classification during flood period with a normal reference water map. So the product accuracy should also primarily be determined by our preliminary product: water classification. The accuracy for algorithm accuracy is compared at Table 4.7.

Table 4.7. Comparison of Algorithm Accuracy Estimates with the Requirement

With Proxy MODIS data	With Proxy SEVIRI data	Requirement
98.07%	98.92%	60%

Since there is no real ground truth for flood/standing water, the river flood forecasting and outlook from the NOAA/NWS and USGS may be used to evaluate our product accuracy for flood detection. But they are some kind of forecasting, not ground truth either. Moreover, they are usually for river discharge and surface elevation data in GIS vector format, how to convert these data to FSW detection is currently under progress. It is expected we can complete these evaluation/validation works at the time of 100% readiness.

5. PRACTICAL CONSIDERATIONS

5.1 Numerical Computation Considerations

We selected the C4.5 DT as our baseline algorithm because it is easily implemented and runs fast. Once the rules are obtained from the training, the Flood/Standing Water algorithm is mathematically simple and requires no complicated mathematical routines. In operations it will be robust and rapid with run time less than 1 minute in terms of the algorithm latency requirement (< 15 minutes, goal) using current computer power. There is no specific numerical computation requirement needed. For storage consideration, 1/0 (FSW/Non-FSW) values should be saved in one-byte integers. Quality flags for each pixel value should be bit-flag definitions, to minimize data storage.

5.2 Programming and Procedural Considerations

The Flood/Standing Water algorithm is a purely pixel-by-pixel algorithm, implemented in sequential mode. Because of the algorithm simplicity, it requires a small amount of code with basic mathematical routines. However, the Flood/Standing Water algorithm requires ancillary datasets such as cloud mask data, pre-trained rules and threshold values for identifying water bodies to the ABI pixel geolocation. The algorithm processing routines should be programmed in block functions to felicitate implementation. Although we selected a well-established and mature decision tree algorithm, since different satellite sensors may have different bandwidth, sensor response function, and calibration, to make the software flexible and applicable to different sensors, we developed our own codes.

5.3 Configuration of Retrieval

The primary adjustable parameters for the Flood/Standing Water retrieval are the rules and threshold values from the training data. These threshold values should also be adjustable in order to optimize the algorithm, if needed, based on the results of post-launch validation. The source of ancillary datasets should be configurable for the best dataset. And finally, it should be noted that metadata used for the product may be modified, reduced and added during late phases of the product generation.

5.4 Quality Assessment and Diagnostics

The Flood/Standing Water detection will be assessed and monitored. First, a set of quality control flags will be generated with the yes/no FSW product for retrieval diagnostics. The quality control flags will indicate the retrieval conditions, including the water/non-water surfaces (i.e., ocean, coast, snow/ice, water etc.), etc. Since reflectance data for the visible channel is needed, Yes/No FSW maps will be generated only during the daytime.

5.5 Exception Handling

The algorithm will handle exceptions through the quality control flags. In identifying the FSW/Non-FSW for each pixel, quality control flags from input datasets will be examined and skipped for bad sensor data (e.g., missing or no sensor data) or cloudy pixels (i.e., “cloud” and “probably cloud”).

6. ASSUMPTIONS AND LIMITATIONS

6.1 Performance

First of all, because the Standing Water algorithm requires ancillary datasets, it is assumed that following data are available before the Water Fraction retrieval is performed:

1) the ABI cloud mask;
2) a high quality dynamic land surface reflectance dataset.

Specifications regarding sensor performance are assumed to be those described in the MRD.

6.2 Assumed Sensor Performance

The algorithm described in this document performs in the visible and near-infrared spectral bands. It is applicable only during daytime on cloudless pixels. The retrieval accuracy may be reduced significantly near the edges of clouds.

6.3 Pre-launch Product Improvements

There are two general areas of prelaunch product improvement: better and more accurate validation and refinement of the FSW algorithm.

6.3.1 Improved Validation Methods

The difficulties of validating satellite retrievals against ground observations are well known and common to a lot of satellite products. These include unknown error characteristics of algorithm and satellite retrievals and calibration uncertainties in the satellite sensor. The only ground truth we can use for GOES-R yes/no FSW detection is the water/non-water classification merged from land cover/land use map, but this map is usually only available during normal conditions. During flooding period, the only available source is the river flooding forecast and outlook maps from the NOAA National Weather Services (NWS), as well as the map of river flooding from the USGS; however, these maps are not direct Flood/Standing Water identification maps, and are not the ground truth of flooding either. Efforts to convert these river flooding maps into FSW identification maps to evaluate flood/standing water detection during flood period will be pursued in the pre-launch period. The method is expected to allow statistically significant error estimates to be made about each source of data, thereby helping specify the error in the satellite FSW.

High resolution ASTER data from the Terra satellite and ETM+ data from the Landsat observations may be used in this effort. A calibration/validation system is being developed and will be implemented first using real time MODIS and Meteosat SEVIRI data as proxies for the ABI. FSW algorithms for the SEVIRI data as the proxy GOES-R ABI type imager will be used. In addition, studies of the GOES-R algorithm as applied to the proxy MODIS data will also be done. Knowledge gained from these studies will be used to improve the algorithm.

6.3.2 Algorithm Improvement

As seen from our training results, the simulation data gave the simplest tree structure with the highest accuracy. This was achieved because the sensor noise and calibration errors were not included. If we applied the tree structure from the training with the simulation data, then it is possible for a large error to occur. Similarly, the tree structure from the training with the proxy MODIS data may not be applied well to the proxy SEVIRI observations, because the calibration and band spectral properties are different between MODIS and SEVIRI. We will provide SEVIRI and MODIS test data simultaneously to AIT in the delivered software package. Instead of using threshold values directly, we use a pattern recognition method; therefore, the detailed tree structure and the specified threshold values in the tree are not very important. Our FSW software will read and parse the model file automatically. There will be no hard code for threshold values in our software. The SEVIRI and MODIS data are just used as the proxy GOES-R ABI data, and our algorithm will process ABI data after launch.

Due to the close spectral features of cloud shadow and water, the cloud shadow may be misclassified into water. We have already worked out an algorithm for cloud shadow detection. A second contributor to FSW uncertainty is the anisotropic effect of surface reflectance as seen from the observing geostationary satellite like the GOES-R. This effect arises due to high solar zenith angle (SZA) during early morning or late afternoon when the shaded surfaces are significantly darker than the sunlit surfaces, creating a low reflectance over land which may cause misclassification. Although the majority of such conditions can be filtered out by working only for pixels with SZA less than 67 º, small number of cases may still be left for SZA greater than 60º. If the GOES-R albedo product can provide BRDF adjusted nadir reflectance data then this would cease to be a cause of concern to our FSW product. For sun glint contamination, which may occur over the coastal area, we can firstly find the glint condition by using geometry method, and then we will be able to use NDVI to identify water under glint conditions. All of these factors should be considered for the FSW algorithm and finding solutions to these problems is planned.

7. REFERENCES

Ali, A., D. A. Quadir, and O. K. Huh, "Study of river flood hydrology in Bangladesh with AVHRR data," *Int. J. Remote Sens.*, 10,1873-1891, 1989.

Barton, I. J., and J. M. Bathols, "Monitoring floods with AVHRR," *Remote Sens. Environ.*, 30, 89-94, 1989.

Berk, A., G. P. Anderson, P. K. Acharya, J. H. Chetwynd, M. L. Hoke, L. S. Bernstein, E.P. Shettle, M.W. Matthew and S.M. Alder-Golden , MODTRAN4 Version 2 Users's Manual, Space Vehicles Directorate, Hanscom AFB, MA 01731-3010, April 2000.

Blasco, F., M. F. Bellan, and M. U. Chaudhury, "Estimating the extent of floods in Bangladesh using SPOT data," *Remote Sens. Environ*., 39,167-178, 1992.

Bonn, F. and R. Dixon, "Monitoring Flood Extent and Forecasting Excess Runoff Risk with RADARSAT-1 Data," *Natural Hazards*, 35, 377-393, 2005.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees," Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1984.

Breiman, L., "Random Forests," *Machine Learning*, vol. 45, no.1, pp. 5-32, doi:10.1023/A:1010933404324, 2004.

Brown de Decolstan, E., L. Di, D. L. Sun, and W. Yang, "NPOESS surface type Algorithm Theretical Basis Document, Version 1," 1998.

Carroll, M. L., C.M. DiMiceli, and J.R.G. Townshend, R.A. Sohlberg, and P. Noojipady, "MODIS Flood Map; 2008 US Midwest Flood, MOD_FLOOD.2008158.Midwest.Geog, version 1.0," University of Maryland, College Park, Maryland, 2008.

Deutsch, M., "Optical processing of ERTS data for determining extent of the 1973 Mississippi River flood," U.S. Geol. Surv. Prof. Pap. 929, ERTS-1, a New Window on Our Planet, pp. 209-222, 1976.

Deutsch, M., and F. Ruggles, „Optical data processing and projected applications of the ERTS-1 imagery covering the 1973 Mississippi River Valley floods," *Wat. Resour. Bull.*, 10,1023-1039.

Deutsch, M., F. H. Ruggles, P. Guss, and E. Yost, "Mapping of the 1973 Mississippi river floods from the Earth ResourcesTechnology Satellite," Proceedings, International Symposium on Remote Sensing and Water Resources Management, American Water Resources Association, Proceedings no. 17, Burlington, Ontario, pp. 39-55, 1973.

Ellrod, G. P., R. V. Achutuni, J. M. Daniels, E. M. Prins, and J. P. Nelson III, "An assessment of GOES-8 Imager data quality," *Bull. Amer. Meteor. Soc.*, 79, 2509-2526, 1998.

France, M. J. and P. D. Hedges, “A hydrological comparison of Landsat TM, Landsat MSS, and black and white aerial photography,” in Damen, Smit, and Verstappen (Eds), Proceedings, 7th International Symposium, ISPRS (Internat. Soc. Photogram and Remote Sens.) Commission VII, Enschede, 25-29 August 1986, pp. 717-720, 1986.

Gale, S. J., and S. Bainbridge, “The floods in eastern Australia,” *Nature*, 345: 767, 1990.

Gao, B.-C., “NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space,” *Remote Sens. Environ.*, 58, 257-266, 1996.

Gupta, R. P., and S. Banerji, “Monitoring of reservoir volume using Landsat data,” *J. Hydrol.*, 77, 159-170, 1985.

Hallberg, G. R., B. E. Hoyer, and A. Rango, “Application of ERTS-1 imagery to flood inundation mapping,” NASA Special Pub. No. 327, Symposium on significant results obtained from the Earth Resources Satellite-1, Vol. 1, Technical presentations, section A, pp. 745-753, 1973.

Hamilton, S. K., S. J. Sipple, and J. M. Melack, “Inundation patterns in the Pantanal wetland of South America determined from passive microwave remote sensing,” *Arch. Hydrobiol.*, 137,115-141, 1996.

Han, J.W., and M. Kamber, ”Data Mining: Concept and Techniques,” Morgan Kaufmann Publishers, San Francisco, CA, 550pp, 2001.

Hess, L. L., and J. M. Melack, “Mapping wetland hydrology and vegetation with synthetic aperture radar,” *Int. J. Ecol. Environ. Sci.*, 20, 197-205, 1994.

Hess, L. L., J. M. Melack, S. Filoso, and Y. Wang, “Delineation of inundated area and vegetation along the Amazon floodplain with the SIR-C synthetic aperture radar,” *IEEE Trans. Geosci. Remote Sens*., 33, 895-903, 1995.

Hillger, D., W. Ellrod, and P. Gary, “Detection of Important Atmospheric and Surface Features by Employing Principal Component Image Transformation of GOES Imagery,” *J. Applied Meteorology*, 42, 611–629, 2003.

Jensen, J. R., M. E. Hodgson, E. Christensen, H. E., Jr. Mackey, L. R. Tinney, and R. Sharitz, “Remote sensing inland wetlands: a multispectral approach,” *Photogramm. Eng. Remote Sens*., 52, 87-100, 1986.

Kay, S., J. D. Hedley, and S. Lavender, “Sun Glint Correction of High and Low Spatial Resolution Images of Aquatic Scenes: a Review of Methods for Visible and Near-Infrared Wavelengths,” *Remote Sens.,* 1(4), 697-730, 2009.

Kohavi, R., “Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid,” Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996.

Li, S., D. L. Sun, Y. Yu, I. Csiszar, “A New Shortwave Infrared (SWIR) Method for Quantitative Water Fraction Derivation with the EOS/MODIS and Landsat/TM”, *IEEE Trans. Geos. Remote Sens.*, in revision, 2010.

Li, S., D. L. Sun, and Y. Yu, “A combined geometrical and spectral method for cloud shadow detection”, *IEEE Trans. Geos. Remote Sens.*, submitted, 2010.

Li, X., R. T. Pinker, M. M. Wonsick, and Y. Ma, “Toward improved satellite estimates of short-wave radiative fluxes-Focus on cloud detection over snow: 1. Methodology,” *J. Geophys. Res*., 112 (D7), D0720810.1029/2005JD006698, 2007.

Lowry, R. T., E. J. Langham, and N. Mudry, “A preliminary analysis of SAR mapping of the Manitoba Flood, May 1979,” Satellite Hydrology, Proceedings, Fifth Ann. William T. Pecora Memorial Symp. on Remote Sens., 1979, Amer. Wat. Resour. Assoc. Tech. Pub. No. TPS81-1, pp. 316-323, 1981.

McGinnis, D. F., and A. Rango, “Earth Resources Satellite systems for flood monitoring,” *Geophys. Res. Let.*, **2** (4), 132-135, 1975.

Melack, J. M., L. L. Hess, and S. Sippel, “Remote sensing of Lakes and Floodplains in the Amazon Basin,” *Remote Sens. Rev.*, 10,127-142, 1994.

Morrison, R. B., and M. E. Cooley, “Assessment of flood damage in Arizona by means of ERTS-1 imagery,” Proceedings, Symposium on Significant Results Obtained from the Earth Resources Satellite-1, Vol. 1, New Carrollton, Maryland, pp. 755-760, 1973.

Morrison, R. B., and P. G. White, “Monitoring flood inundation,” U.S. Geol. Surv. Prof. Pap. 929, ERTS-1, A New Window on Our Planet, pp. 196-208, 1976.

Parmuchi, G. M., H. Karszenbaum, and P. Kandus, “Mapping wetlands using multi-temporal RADARSAT-1 data and a decision-based classifier,” *Canadian J. Remote Sens*., 28 (2), 175-186, 2002.

Pinker, R. T., X. Li, W. Meng, and E. A. Yegorova, “Toward improved satellite estimates of short-wave radiative fluxes-Focus on cloud detection over snow: 2. Results,” *J. Geophys. Res.*, 112, D09204, doi:10.1029/2005JD006699, 2007.

Pope, K. O., J. M. Rey-Benayas, and J. F. Paris, “Radar remote sensing of forest and wetland ecosystems in the Central American tropics,” *Remote Sen. Envion.*, 48, 205-219, 1994.

Pope, K. O., E. J. Sheffner, K. J. Linthicum, C. L. Bailey, T. M. Logan, E. S. Kasichke, K. Birney, A. R. Njogu, and C. R. Roberts, “Identification of Central Kenyan Rift Valley Fever

virus vector habitats with Landsat TM and evaluation of their flooding status with airbourne imaging radar," *Remote Sens. Environ.*, 40,185-196, 1992.

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Francisco, CA, 316pp.

Rango, A., and A. T. Anderson, "Flood hazard studies in the Mississippi River basin using remote sensing," *Wat. Resour. Bull.*, 10,1060-1081, 1974.

Rango, A., and V. V. Salomonson, "Regional flood mapping from space," *Wat. Resour. Res.*, 10, 473-484, 1974.

Rasid, H., and M. A. H. Pramanik, "Areal extent of the 1988 flood in Bangladesh: how much did the satellite imagery show?" *Natural Hazards*, 8, 189-200, 1993.

Salomonson, V. V. (Ed.), "Water resources assessment, in Manual of Remote Sensing, Chap. 29," Amer. Soc. of Photogramm. pp. 1497-1570, 1983.

Schmit T. J., W. F. Feltz, W. P. Menzel, J. Jung, A. P. Noel, J. N. Heil, J. P. Nelson III, G. S. Wade, Validation and use of GOES sounder moisture information, *Wea. Forecasting*, 17,139-154, 2002.

Schmit, T. J., J. Li, M.M. Gunshor, C.C. Schmidt, W.P. Menzel, J. Gurka, and J. Sieglaff, " Study of the Advanced Baseline Imager (ABI) on the GOES-R and beyond," 84th AMS Annual Meeting, Seattle WA, 2004.

Schmit, T. J., W. Paul Menzel, J. Gurka , M. Gunshor, "The ABI on GOES-R," 3rd Annual Symposium Future National Operational Environmental Satellite Systems, San Antonio, January 16, 2007.

Sheng, Y., P. Gong, and Q. Xiao, "Quantitative dynamic flood monitoring with NOAA AVHRR," *Int. J. remote sensing*, 22, 1709–1724, 2001.

Sipple, S. J., S. K. Hamilton, and J. M. Melack, "Inundation area and morphometry of lakes on the Amazon River floodplain," Brazil. *Arch. Hydrobiol.*, 123: 385-400, 1992.

Smith, L. C., "Satellite remote sensing of river inundation area, stage, and discharge: A review," *Hydrolog. Process.*, 11,1427-1439, 1997.

Solomatine, D. P., and Y. Xue, "M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China," *J. Hydrological Engineering*, 9 (6), 491-501, 2004.

Steinbrecht, W., H. Claude, U. Kohler, and K.P. Hoinka, Correlations between tropopause height and total ozone: Implications for long-term changes, *J. Geophys. Res.*, 103, 19183-19192, 1998.

Sun, D. L., and Y. Yu, “Deriving water fraction and flood map with the EOS/MODIS data using decision tree approach”, Journal of Selected Topics in Earth Observations and Remote Sensing, in revision, 2010.

Tucker, C. J., “Red and Photographic Infrared Linear Combinations for Monitoring Vegetation”, *Remote Sensing of Environment*, 8, 127-150, 1979.

Vanderbilt et al., “Inundation Discriminated Using Sun Glint,” *IEEE Trans. Geosci. Remote Sens.*, 40(6): 1279-1287, 2002.

Vila da Silva, J. S. and H. J. H. Kux, “Thematic mapper and GIS data integration to evaluate the flooding dynamics within the Panatal, Mato Grosso do Sul State, Brazil,” Proceedings. *Int. Geosci. Remote Sens. Symp*. (IGARSS '92), pp. 1478-1480, 1992.

Watson, J. P., “A visual interpretation of a Landsat mosaic of the Okavango Delta and surrounding area,” *Remote Sens. Environ*., 35, 1-9, 1991.

Wei, W., X. Zhang, X. Chen, J. Tang, and M. Jiang, “Wetland mapping using subpixel analysis and decision tree classification in the yellow river delta area,” ISPRS Congress Beijing 2008,

Xu, M., P. Watanachaturaporn, P. K. Varshney, and M. K. Arora, “Decision tree regression for soft classification of remote sensing data,” *Remote Sens. Environ.*, vol. 97 (3), 322-336, 2005.