Journal of Environmental Management 201 (2017) 227–240

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Integrating multisensor satellite data merging and image reconstruction in support of machine learning for better water quality management

Ni-Bin Chang [a], Kaixu Bai [b, c, *], Chi-Farn Chen [d]

[a] *Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL, 32816, USA*
[b] *Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, 200241, China*
[c] *School of Geographic Sciences, East China Normal University, Shanghai, 200241, China*
[d] *Center for Space and Remote Sensing Research, National Central University, Taoyuan County, 32001, Taiwan*

ARTICLE INFO

Article history:
Received 6 April 2017
Received in revised form
1 June 2017
Accepted 18 June 2017

Keywords:
Water quality
Remote sensing
Machine learning
Enabling technology
Watershed management

ABSTRACT

Monitoring water quality changes in lakes, reservoirs, estuaries, and coastal waters is critical in response to the needs for sustainable development. This study develops a remote sensing-based multiscale modeling system by integrating multi-sensor satellite data merging and image reconstruction algorithms in support of feature extraction with machine learning leading to automate continuous water quality monitoring in environmentally sensitive regions. This new Earth observation platform, termed "cross-mission data merging and image reconstruction with machine learning" (CDMIM), is capable of merging multiple satellite imageries to provide daily water quality monitoring through a series of image processing, enhancement, reconstruction, and data mining/machine learning techniques. Two existing key algorithms, including Spectral Information Adaptation and Synthesis Scheme (SIASS) and SMart Information Reconstruction (SMIR), are highlighted to support feature extraction and content-based mapping. Whereas SIASS can support various data merging efforts to merge images collected from cross-mission satellite sensors, SMIR can overcome data gaps by reconstructing the information of value-missing pixels due to impacts such as cloud obstruction. Practical implementation of CDMIM was assessed by predicting the water quality over seasons in terms of the concentrations of nutrients and chlorophyll-a, as well as water clarity in Lake Nicaragua, providing synergistic efforts to better monitor the aquatic environment and offer insightful lake watershed management strategies.

1. Introduction

Lake Nicaragua, located at the border line between Costa Rica and Nicaragua, is ranked the largest freshwater lake in Central America. The water quality condition in this lake has been deteriorating over the past decades, mainly due to farming activities, deforestation, and uncontrolled/unregulated disposal of municipal wastewater into the lake. Moreover, the Tipitapa River connects Lake Nicaragua to the highly-polluted Lake Managua upstream, exposing it to the constant discharge of pollutant plumes driven by the natural landscape (World Bank, 2013). In addition to some small towns in the Lake Nicaragua watershed, big urbanized regions, such as Granada, Rivas, and Juigalpa, discharge their domestic sewage and industrial wastewater into Lake Nicaragua directly or through channels (World Bank, 2013). Moreover, fertile soil conditions in the lake watershed have been fostering the agricultural industry since the precolonial age, providing another long-term source of water quality contamination for the lake. For instance, areas dedicated to agricultural cropping, plantation, and cattle farming near the Chontales, Boaco, and Rivas may deliver an enormous amount of fertilizer directly into the lake, thus significantly increasing the eutrophication problem in the lake (World Bank, 2013).

Water quality deterioration has been observed recently around the southern part of the lake, partially caused by the unlimited use of fertilizers on plantations. Another contamination source of the lake water could be due to fish farming of tilapia, which generates large amounts of waste and thus increases the eutrophication potential. The continuous deterioration of water quality in Lake

---

* Corresponding author. Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, 200241, China.
E-mail address: kxbai@geo.ecnu.edu.cn (K. Bai).

http://dx.doi.org/10.1016/j.jenvman.2017.06.045
0301-4797/

Nicaragua is a crucial issue in water quality management that could affect the safe drinking water supply. Frequent monitoring of the water quality status to characterize eutrophication conditions in this lake on a near real-time basis is crucial in such an expansive lake and may be a costly and tedious mission without using satellite remote sensing.

Remote sensing for earth observation is challenging due to many dynamic environmental factors such as aerosols, sun glint, clouds, and others (Gordon and Wang, 1994; Moore et al., 2000; Wang and Bailey, 2001). It is thus noticeable that obtaining a spatial complete, clear image over the study area with one single sensor observation is by no means an easy task. This is especially true in the tropical and subtropical regions where heavy cloud cover emerges frequently in all seasons, making it difficult to collect comprehensive information on a target based on a single instrument's observations. This is exactly the case with the cloud coverage at Lake Nicaragua in Central America, which became an issue for consecutive water quality monitoring with high temporal resolution (Fig. 1). Since the 2000s, the remote sensing community has been experiencing a regime shift to create scientific quality remote sensing data encompassing measurements from multiple satellite missions via data fusion or merging on one hand (Fargion and McClain, 2002), and to conduct intelligent feature extraction with the aid of machine learning or data mining algorithms on the other hand (Chang et al., 2009, 2012, 2013).

Although data fusion with multiple satellite imagery can synergistically advance the understanding of a dynamic environment by improving either spatial, temporal or spectral resolution of image data, possible data gaps may still exist due to either internal (e.g., instrument mechanical failure) or external (e.g., cloud contamination) factors hindering the remote sensing for earth surface observation. The process of blending information from multiple sources into a new dataset is thus commonly referred to as data fusion for better spatial, temporal, and/or spectral resolution. The concept of image or data fusion was initialized in the 1990s (Shen, 1990; Pohl and van Genderen, 1998). In the early 2000s, Waltz (2001) and Luo et al. (2002) formalized the principles and practices of image and spatial data fusion. Data merging research formally began from the data merger activities undertaken by the Sensor Intercomparison and Merger for Biological and Interdisciplinary Studies (SIMBIOS) project in 2002 (NASA, 2002). Later, the normalized water-leaving radiance from Sea-Viewing Wide Field-of-View Sensor (SeaWiFS) and Moderate Resolution Imaging Spectroradiometer (MODIS) was merged to support a semi-analytical bio-optical ocean color model for mapping the ocean chlorophyll concentrations (Maritorena and Siegel, 2005; Maritorena et al., 2010). Very recently, to cope with these gaps caused by cloud coverage, data merging has been used to improve the spatial and temporal coverage, providing effective methods such as the Spectral Information Adaptation and Synthesis Scheme (SIASS) (Bai et al., 2016a). In the first endeavor of implementing the SIASS algorithm, three cross-mission sensors, including the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard Suomi-NPP and MODIS onboard Terra and Aqua, provide a synergistic opportunity to enhance temporal and spatial coverage by merging their complementary observations (Bai et al., 2016a,b). In addition, algorithms like SMart Information Reconstruction (SMIR) can also be used to help image reconstruction via the time–space–spectrum continuum for cloudy pixels that cannot be recovered by data merging (Chang et al., 2015a).

Not only merged images but also fused images may be connected to a wealth of machine learning or data mining algorithms for monitoring challenging water quality constituents of concern such as total organic carbon (TOC), total phosphorus (TP), total nitrogen (TN), and microcystin in fresh water lakes or coastal bays with respect to fused multispectral and hyperspectral remote sensing images (Chang et al., 2012, 2014a, 2014b, 2015a, 2015b). For example, with the aid of genetic programing (GP) methods, Chang et al. (2014a) successfully mapped the distribution of toxic microcystin in Lake Erie by using fused remotely sensed imagery between MODIS and MEdium Resolution Imaging Spectrometer (MERIS). The results suggest that GP models perform better than traditional two-band models in quantifying the concentrations of microcystin in water bodies due to the inclusion of additional spectral reflectance data embedded in hyperspectral images. The unique hyperspectral band helps produce a noticeable increase in the prediction accuracy, especially in the range of low microcystin concentrations. Similarly, Imen et al. (2015, 2016) developed a decision support system integrating data fusion and machine learning tool for monitoring the total suspended sediment concentrations in Lake Mead. These state-of-the-art machine learning techniques applied for processing either fused or merged remote sensing imageries can significantly advance the surveillance of the earth environment (Chang et al., 2014a, 2014b, 2015a, 2015b). Such a regime shift from ‘traditional image processing’ to ‘data/image merging/fusion,’ to ‘machine learning/data mining’ has resulted in significant advancements in certain environmental applications.

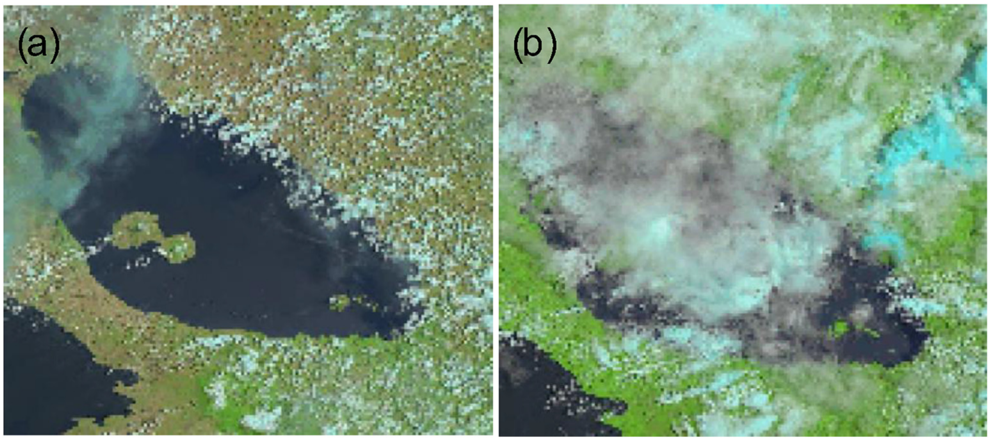

Fig. 1. Landsat 8 operational land imager true color images over Lake Nicaragua on (a) April 12, 2013 (dry season) and (b) August 18, 2013 (wet season). Note that cloud (white area) impacts were severe over the lake in the wet season. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)