

A PROBABILITY-BASED STATISTICAL METHOD TO EXTRACT WATER BODY OF TM IMAGES WITH MISSING INFORMATION

Shizhong Lian^a, Jiangping Chen^{a*}, Minghai Luo^b

^aSchool of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, China - (szlian, chen_jp)@whu.edu.cn

^bWuhan Geomatics Institute, Wuhan, Hubei, China - luominghai@163.com

Commission II, WG II/1

KEY WORDS: Probability statistics, Water body extraction of TM images, Missing information, MNDWI

ABSTRACT:

Water information cannot be accurately extracted using TM images because true information is lost in some images because of blocking clouds and missing data stripes, thereby water information cannot be accurately extracted. Water is continuously distributed in natural conditions; thus, this paper proposed a new method of water body extraction based on probability statistics to improve the accuracy of water information extraction of TM images with missing information. Different disturbing information of clouds and missing data stripes are simulated. Water information is extracted using global histogram matching, local histogram matching, and the probability-based statistical method in the simulated images. Experiments show that smaller Areal Error and higher Boundary Recall can be obtained using this method compared with the conventional methods.

1. INTRODUCTION

Water is a decisive factor to maintain the stability and health of wetland ecosystem (Wang, Lian and Huang, 2012). Using satellite remote sensing image to extract water body information quickly and accurately has become an important approach of wetland investigation, research, and protection (Xu, 2006; Huiping, Hong and Qinghua, 2011; Li, et al, 2013). TM image is an important data source for extracting water body information with high spatial and spectral resolution, high positioning accuracy, and an extremely rich amount of information. However, when multi-period TM images are used to monitor water dynamic process, parts of the images lose the true information because of blocking clouds, cloud shadows, or sensor faults, which made extraction of surface information difficult. The methods with global histogram matching (GHM), local histogram matching (LHM) (Shou, Chen and Ma, 2006), and other common image restorations failed to improve the accuracy of water information well because they used one close-temporal intact image to restore the missing information. These methods successfully improved the classification accuracy of relatively stationary features like houses, roads, vegetation. However, water has the least stability, with the shortest span of two adjacent images in 16 days. Within this short period, the border of water changes relatively more than houses and roads as relatively stationary features. Moreover, accessing qualified images in a 16-day span is difficult because of blocking clouds. In this paper, a new method of water body extraction based on probability statistics is proposed, which improves the accuracy of water information extraction of TM images with missing information.

2. RESEARCH METHODS

2.1 Water extraction methods of TM images without missing information

Numerous scholars have recently conducted research on water body extraction of TM images without missing information (Wang, et al., 2015; Boland, 1976; Jiang, et al., 2014; Hassani, et al., 2015). Jenson extracted water body according to the

threshold, which is decided by the middle-infrared radiation band (MIR), near-infrared radiation band (NIR), and TM5 (Moller, 1990). McFeeter proposed the definition of normalized difference water index (NDWI) to extract water in vegetation areas. NDWI is the ratio of the value results of Green (the green light wave band) and NIR by subtraction and addition (McFeeters, 1996). However, the water body extracted by this type of method is blended with other information, particularly buildings. Considering the weakness that buildings can be easily regarded as water when extracting information with NDWI algorithm, Hanqiu Xu introduced the modified normalized difference water index (MNDWI), which can restrain the vegetation factor and building factor at the greatest extent so as to give prominence to the water body information (Xu, 2005). The radiation value of the water body is high in the green band, which is low in the mid infrared wave band. As a result, the water body information in the MNDWI gray image is highlighted as high value. In this study, MNDWI was used to extract water body of TM images without missing information, the function is shown as follows:

$$\text{MNDWI} = (\text{Green} - \text{MIR}) / (\text{Green} + \text{MIR}) \quad (1)$$

where **Green** is the green light wave band in TM images, corresponding to the second band in Landsat 5 and Landsat 7 and is the third band in Landsat 8; and **MIR** is the middle-infrared radiation band, corresponding to the fifth band in Landsat 5, Landsat 7, and the sixth band in Landsat 8.

The significant step in the process of water information extraction by MNDWI is to determine the threshold of segmentation. The Otsu method is an effective algorithm for image segmentation and is widely used in many fields. In this study, the Otsu method was used to obtain the segmentation threshold. The principle of this method is to divide the original image into two classes: the target and the background; when the variance between the target and background achieves the maximum, the gray value can be the optimal threshold.

2.2 Common image restoration methods

2.2.1 GHM: GHM algorithm aims at the whole filled image and matches its gray histogram to the image to be corresponded. The histogram of the filled image is matched band by band to the gray histogram of the corresponding band of the image to be repaired. Accordingly, the difference of brightness of the two images becomes small. The most commonly used matching method is based on the mean and variance; the function is shown as follows:

$$DN_{ti} = \sigma_t \times \left(\frac{DN_{si} - \overline{DN_s}}{\sigma_s} \right) + \overline{DN_t} \quad (2)$$

where DN_{ti} is the gray value of default location i in the image to be repaired t ; DN_{si} is the gray value of location i in the filled image s ; $\overline{DN_t}$ is the mean of gray value in the image to be repaired t before the repair; $\overline{DN_s}$ is the mean of gray value in the filled image s ; σ_t is the variance of gray value in the image to be repaired t before the repair; and σ_s is the variance of gray value in the filled image.

2.2.2 LHM: Considering the different local brightness at different positions in the image, the LHM algorithm divides the image into some *sub* windows and matches the image to be repaired with these sub parts. The main steps are as follows:

- (1) The sub window size is set to 35×35 in the upper left corner of the filled image. If there are more than 600 image pixels, which have values in both filled image and image to be repaired, then the window size should be extended to 37×37 . The window size increases by 2 each time until $N > 600$.
- (2) The histograms of each band of the corresponding filled image and image to be repaired in a sub window are extracted.
- (3) The two histograms of the filled image and image to be repaired in a sub window are matched according to the GHM method mentioned above.
- (4) The sub window is moved, and the above steps are repeated until the image to be repaired is filled.

2.3 Probability-based statistical method (PSM) to extract water body of TM images

Water is continuously distributed in natural conditions. Water of same water level in one water body exists and disappears simultaneously. As a result, water contour images can be obtained through simulation using probability images of water body distribution. Higher probability indicates deeper water level, and vice versa. Lacking image information causes the failure of water information extraction. The PSM aims to fill the missing information. Specific steps of the algorithm are as follows:

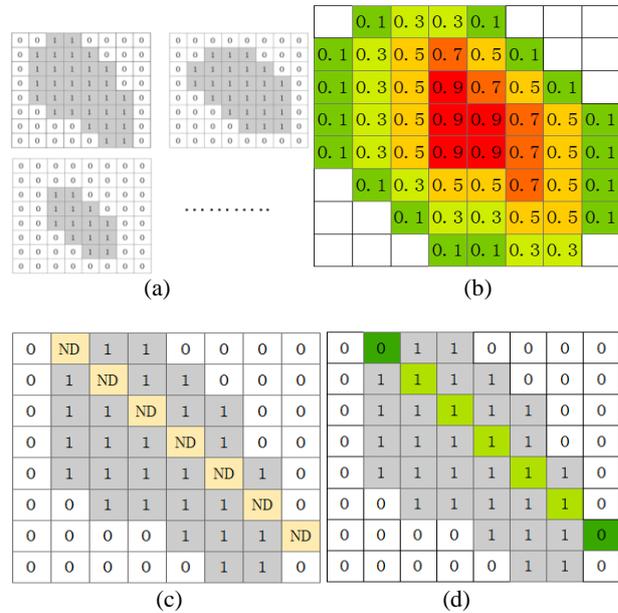


Figure 1. Specific steps of the algorithm
(a) Multi-period water distribution images
(b) Probability images of water body
(c) Water distribution image with missing data
(d) Water distribution image after restoration

(1) MNDWI indexes of multiple-view TM images (without missing data) are calculated, and segmentation threshold of MNDWI greyscale maps is determined using Otsu algorithm. Figure 1(a) shows a multi-period water body distribution image (without missing data), in which 1 represents water body and 0 represents non-water.

(2) Multi-period water body distribution images are overlay analyzed to obtain probability images of water body distribution, as shown in Figure 1(b).

(3) MNDWI indexes of TM images (ND represents missing data) are calculated, and segmentation threshold of MNDWI greyscale maps is determined using Otsu algorithm. Figure 1(c) shows water body distribution image with missing data, in which 1 represents water body and 0 represents others.

(4) Water body distribution images (ND represents missing data) in step (3) and probability images of water body in step (2) are analyzed by overlay. Pixel numbers of water and non-water in water body distribution images (with missing data) are counted in different probability levels. In a certain probability level, missing data are counted as water when water percentage outweighs non-water percentage, and vice versa. Figure 1(d) is a restored water distribution image.

2.4 Accuracy evaluation method

The water information of TM images (with missing data) is extracted using GHM, LHM, and PSM in this paper. The water information (without missing data) is then treated as a reference. The Areal Error and Boundary Recall can be calculated according to Equations (3) and (4).

$$E = \frac{|A_g - A_b|}{A_b} \times 100\% \quad (3)$$

where A_g is the area of the water information (with missing data), which is extracted by different methods; A_b is the area of the water information (without missing data); and E is the value of Areal Error.

$$V = \frac{|L_g|L_g = B_g \cap B_b|}{L_b} \quad (4)$$

where B_g is the boundary of the water (with missing data), which is extracted by different methods; B_b is the boundary of the water (without missing data); L_g is the overlap length of B_g and B_b ; L_b is the length of B_b ; and V is the value of Boundary Recall.

3. WATER DISTRIBUTION OF EXPERIMENTAL DATA

3.1 Experimental data

The remote sensing data used in the experiment are all intact to verify the validity of this method and assess the accuracy of the traditional methods and the method used in this paper in extracting water body information from TM images with missing information. The time span is from January 3, 2015 to October 18, 1992. A total of 40 views of images (122, 039), including 24 views of Landsat 5 images, 7 views of Landsat 7 images, and 9 views of Landsat 8 images, are included.

3.2 Results of the water body extraction from TM images without missing information

The MNDWI index of the 40 views of TM images is calculated by Equation (1), and the corresponding bi-value images can be gained by the use of Otsu method. The water body distribution maps of TM images are shown in Figure 2; only two time points are used as examples. Basing on the contrast between the original images and water body extraction results, we conclude that the MNDWI index can separate water and other features to a great level, and the outline of water body is clear.

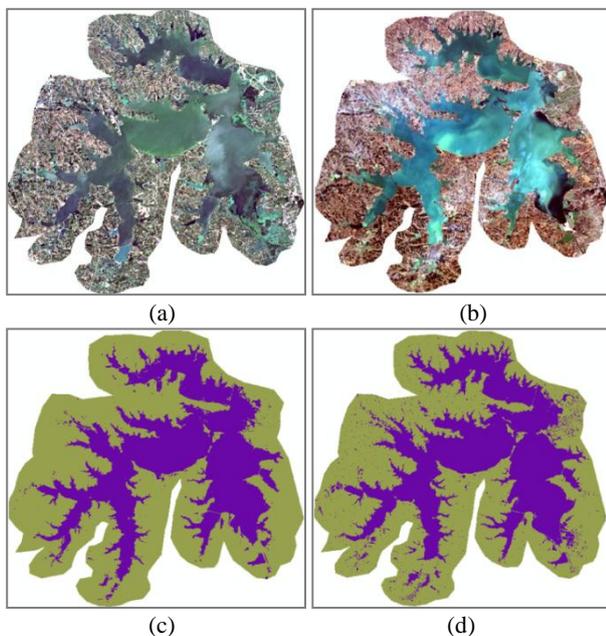


Figure 2. Water extraction results of TM images based on MNDWI

- (a) Original image on 2013.8.9
- (b) Original image on 2013.12.30
- (c) Water body distribution on 2013.8.9
- (d) Water body distribution on 2013.12.30

The probability distribution map of water body can be gained by the superposition analysis of the water body distribution of 40 views of TM images. The probability distribution is shown in Figure 3. The water in different regions has clearly different probabilities. The probability of water distribution significantly changed in the edge area. This result indicates that the water body in the edge water area has a lower water level and tends to evolve into other land types in a short period of time.

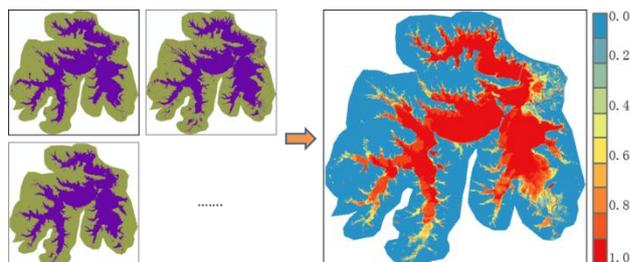


Figure 3. Probability distribution map of water body

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1 Simulating TM image with information missing

The accuracy of the proposed method is verified, that is, to imply quantitative evaluation to the water body extraction results of TM image with information missing by traditional methods and PSM. The TM images to be repaired in the experiment were simulated considering the case of sensor faults and cloud shadows (Figure 4). Figures 4(b) and Figures 4(c) show the simulated image with missing strips and the simulated image covered by clouds, respectively.

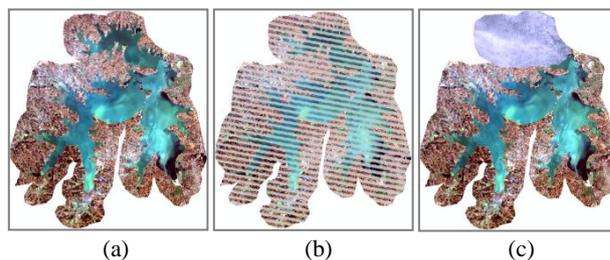


Figure 4. Simulated image with information missing on 2013.12.30

- (a) Original image
- (b) Simulated image with missing strips
- (c) Simulated image covered by clouds

4.2 Results of the water body extraction from TM images with missing information

In this paper, we extracted water body in TM images with missing information by PSM and the traditional GHM and LHM. The comparison of extraction results of various methods is shown below. The results of image with missing strips are presented in Figure 5, and the results of image covered by clouds are shown in Figure 6.

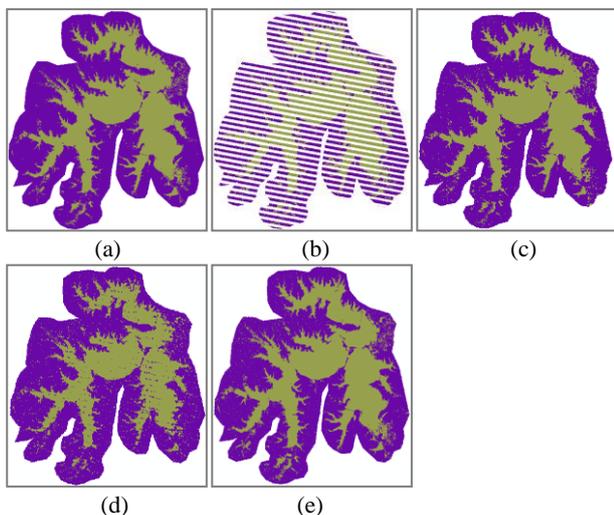


Figure 5. Extraction results of various methods on the image with missing strips

(a) True image (b) Image with missing strips (c) Result of PSM (d) Result of LHM (e) Result of GHM

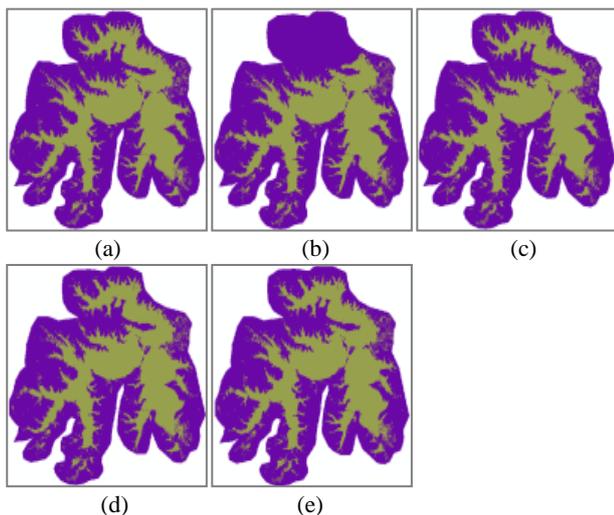


Figure 6. Extraction results of various methods on the image covered by clouds

(a) True image (b) Image covered by clouds (c) Result of PSM (d) Result of LHM (e) Result of GHM

The water extraction effect of the PSM, LHM, and GHM in the two types of damaged image was compared. The result shows that the PSM proposed in this paper is better than the traditional methods, particularly in the area where the changes of water body are more severe, which can be assessed in two highlighted areas in Figure 7. Position 1 shows that the area in the image to be repaired is non-water (Figure 7(a2)), whereas in the reference image is water (Figure 7(b2)). In the same way, position 2 shows that the highlighted area in the image to be repaired is water (Figure 7(a3)), but in the reference image it becomes non-water (Figure 7(b3)). The probability distribution map also indicates that the two highlighted areas have lower probability of water distribution (Figure 7(c)) and tend to convert into non-water area in a short period of time.

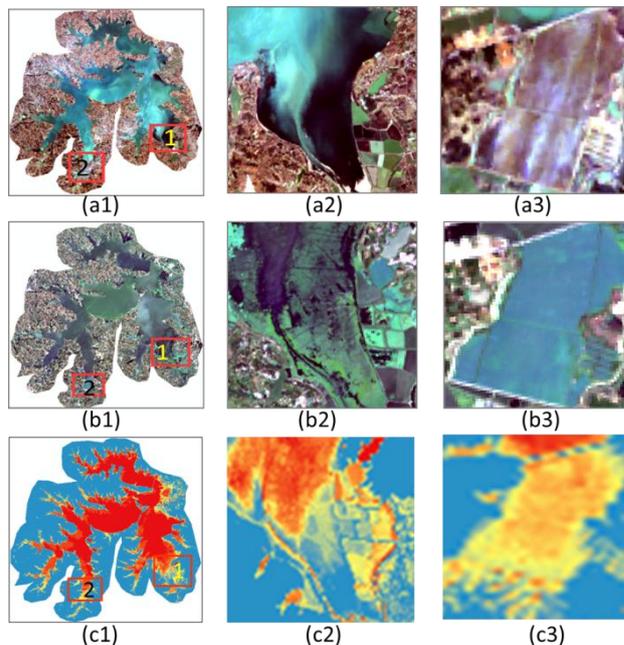
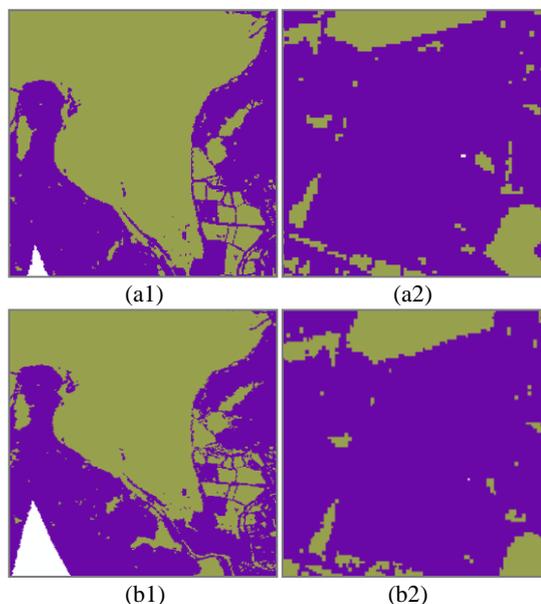


Figure 7. Details of the water body extraction results

(a) TM image to be repaired on 2013.12.30
(b) Reference TM image on 2013.8.9
(c) Probability distribution map of water body

The water extraction results of PSM, GHM and, LHM in the two highlighted areas are investigated, as shown in Figure 7. The comparison result is presented in Figure 8. The PSM can extract the water body information in positions 1 and 2 to a more complete level, whereas the two water areas extracted by GHM and LHM were not consistent with the actual situation. The difference lies in the number of views of TM images considered in the water extraction. The traditional methods only use a period of images to extract water body information, while the PSM is based on the probability distribution map of the water body, which was gained by superposition analysis of multiple periods of water distribution map. In this study, 40 total views of images were absorbed to obtain the probability distribution map.



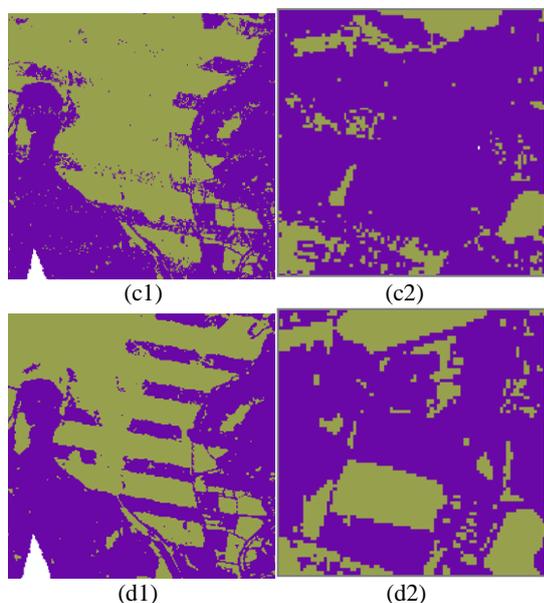


Figure 8. Local effect of the repair on 2013.12.30
(a) True image on positions 1 and 2
(b) Result of PSM on positions 1 and 2
(c) Result of LHM on positions 1 and 2
(d) Result of GHM on positions 1 and 2

4.3 Accuracy evaluation

The water distribution map of 40 views of TM images in chapter 4.1 was set as the true value to evaluate the accuracy of the extraction results based on PSM, GHM, and LHM. The elements considered for accuracy evaluation were Areal Error (Equation (3)) and Boundary Recall (Equation (4)), and the statistical result is shown in Table 1. The visual comparison of PSM, GHM, and LHM is essential to further assess the effect of several methods (Figure 9).

Information deletion type	Accuracy type	PSM	LHM	GHM
Missing strips	Areal Error	2.35%	10.31%	5.36%
	Boundary Recall	96.41%	86.35%	92.11%
Covered by clouds	Areal Error	3.16%	3.56%	3.46%
	Boundary Recall	97.63%	97.51%	97.43%

Table 1. Statistical result of the accuracy evaluation

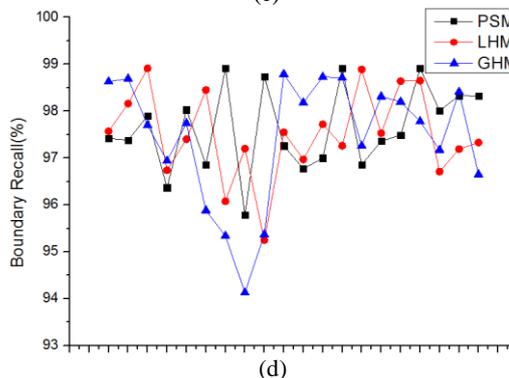
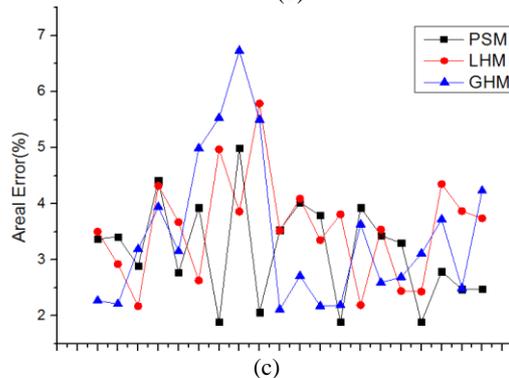
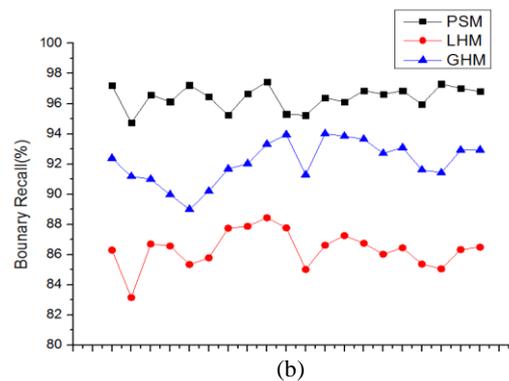
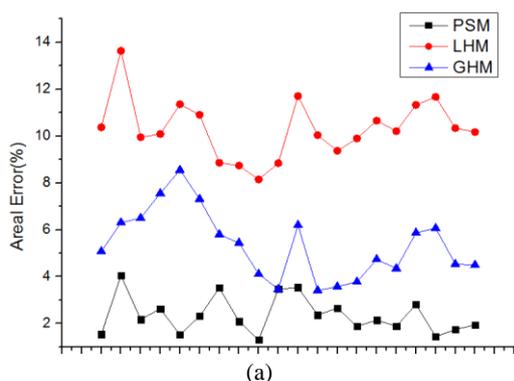


Figure 9. Comparison of extraction accuracy of PSM, GHM, and LHM

- (a) Areal Error of strip repair results
- (b) Boundary Recall of strip repair results
- (c) Areal Error of covered by clouds repair results
- (d) Boundary Recall of covered by clouds repair results

Based on the accuracy evaluation, the PSM proposed in this study is of lower Areal Error and higher Boundary Recall. We conclude that PSM can achieve better water extraction effect than other two methods whether the problem is missing strips or covered by clouds. Therefore, the PSM shows great applicability in water body extraction from TM images with missing information.

CONCLUSION

The accuracy of PSM to extract the water body of TM images with missing information is better than GHM and LHM in both Areal Error and Boundary Recall. The experiment results show that the Areal Error of methods of GHM, LHM, and PSM is generated (5.36%, 10.31%, and 2.35%, respectively) in the case of missing data stripes and (3.46%, 3.56% and 3.16%, respectively) in the case of clouds; the Boundary Recall of GHM, LHM, and PSM is generated (92.11%, 86.35% and 96.41%, respectively) in the case of missing data stripes and

(97.43%, 97.51%, and 97.63%, respectively) in the case of clouds. In conclusion, the PSM can improve water body extraction accuracy of TM images with missing information.

imagery . *International Journal of Remote Sensing*, 27(14), pp. 3025-3033.

ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 41331175. The authors would also like to thank the anonymous reviewers for their comments and suggestions that have greatly improved the work.

REFERENCES

Boland D H P, 1976. Trophic classification of lakes using landsat-1 (ERTS-1) multispectral scanner data. Report EPA-600/3-76-037, Corvallis Environmental Research Laboratory, Corvallis, Oregon, USA

Hassani M, Chabou M C, Hamoudi M, et al, 2015. Index of extraction of water surfaces from Landsat 7 ETM+ images . *Arabian Journal of Geosciences*, 8(6), pp. 3381-3389.

Huiping Z, Hong J, Qinghua H, 2011. Landscape and water quality change detection in urban wetland: A post-classification comparison method with IKONOS data. *Procedia Environmental Sciences*, 10, pp.1726-1731.

Jiang H, Feng M, Zhu Y, et al, 2014. An automated method for extracting rivers and lakes from Landsat imagery . *Remote Sensing*, 6(6), pp. 5067-5089.

Li W, Du Z, Ling F, et al, 2013. A comparison of land surface water mapping using the normalized difference water index from TM, ETM+ and ALI . *Remote Sensing*, 5(11), pp.5530-5549.

McFeeters S K, 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features . *International journal of remote sensing*, 17(7), pp.1425-1432.

Moller-Jensen L, 1990. Knowledge-based classification of an urban area using texture and context information in Landsat-TM imagery . *Photogrammetric Engineering and Remote Sensing*, 56(6), pp.899-904.

Shou J, Chen X, Ma J, 2006. Application and Research on the Recover of Landsat-7 SLC-off images Based on ALR. *Journal of Optoelectronics Laser*, 17(3), pp. 368-371.

Wang S, Baig M H A, Zhang L, et al, 2015. A Simple Enhanced Water Index (EWI) for Percent Surface Water Estimation Using Landsat Data. *Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE Journal of, 8(1), pp. 90-97.

Wang X, Lian Y, Huang C, et al, 2012. Environmental flows and its evaluation of restoration effect based on LEDESS model in Yellow River Delta wetlands . *Mitigation and Adaptation Strategies for Global Change* , 17(4), pp.357-367

Xu H, 2005. A study on information extraction of water body with the modified normalized difference water index (MNDWI) . *Journal of Remote Sensing*, 9(5), pp. 589-595.

Xu H, 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed