Performance analysis of a novel algorithm for large scale water-body surface mapping using elevation and intensity of LiDAR data

Partha P. Acharjee[1], George J. Toscano[1], Collin McCormick[2] and Venkat Devarajan[1]
[1]The University of Texas at Arlington, Texas, USA
[2]Natural Resources Conservation Service (NRCS), Fort Worth, Texas, USA

***Abstract*—Detection of water-body boundary is vital for cartography, environmental study, designing sustainability program etc. LiDAR-based remote sensing algorithms can provide practical solutions of water surface mapping for large areas. Geophysical properties of water and optical properties of laser returns from water surface can simultaneously be exploited for water-body detection and hydro breakline generation using LiDAR technology. Therefore, a new algorithm is proposed utilizing these optical and physical properties such as high absorption rate, specular reflection and local flatness as key characteristics of water surface. It is observed that these key features help successfully detect water-bodies in both rural and complex urban scene. Quantitative analysis shows that for different datasets, the overall accuracy of detection varies from 98.45% to 99.94% when compared to manual detection. Results for a very large-scale data, of around 500 km^2 in size, is also reported. From careful visual inspection on large areas, it is clear that the proposed method correctly detects water-bodies of different sizes and shapes e.g. rivers, small river branches, lakes, ponds, reservoirs etc. Therefore, the proposed method is a fast and reliable tool for hydro-breakline generation from large scale LiDAR dataset.**

***Index Terms*— LiDAR, Water-body, River detection**

I. Introduction

A water-body is defined as any physiographical feature containing water, which includes a pond, lake, reservoir, river sea etc. Mapping water features plays a vital role to plan any sustainability program of aquatic ecosystems and other environmental studies. Traditional water-body boundary extraction approaches involve manual estimation from high resolution imagery and on-site survey, which are usually limited, due practical reasons to only large sized water-bodies for a small selected area. Light detection and ranging (LiDAR) is a top-notch remote sensing technology popularly employed for high resolution earth surface mapping and feature extraction. LiDAR system derives the distance of targeted points from a laser source by emitting laser pulses and analyzing the reflected return. Therefore, the LiDAR system provides highly accurate as well as dense co-registered elevation and intensity information of the targeted surface. However, the size of this high-density point cloud is huge and which makes manual classification difficult and expensive. Automated or at least semi-automated classification or feature extraction algorithms are therefore desired.

Extracting water-body boundaries and hydro break line generation are especially important. Therefore, several results have been reported on automated, semi-automated hydro-breakline generation from LiDAR point cloud for different types of water-body related areas.

Most of the literature covers hydro-breakline generation in coastal areas, where only few results were reported for river and standing water-bodies detection. Early model-based approaches used historic shoreline and examined cross-shore LiDAR elevation profile to detect sudden change of elevation [1, 2]. Later, LiDAR return intensity, data density, historical borders, orthoimages were introduced for water-surface mapping algorithm development [3-6]. In recent years, two new algorithms were proposed for in-land water body detection, which were demonstrated on standing water bodies and still work remained for river and coastal areas [7, 8].

In our proposed algorithm, geometric and optical properties of a water-body were translated to useful features, which decide its expansion or "growth" from a small seed. A new edge detection method called *angular filtering* was also proposed to extract sharp local change of elevation. Seed based methods only analyze potential areas around selected seeds, which make it suitable for large-scale implementation. Quantitative results are shown for two datasets, where ground truth is available. Qualitative performance is shown for very large-scale data as detected water-body polygons overlaid on satellite images. Superior accuracy and significantly low computational time are key factors of the proposed method which enable very large scale practical application. The proposed algorithm does not require orthoimages, historical boundaries etc.

II. Background

A. Angular filtering for edge detection

Local elevation change is a useful property for different feature extraction schemes from rasterized LiDAR data. In general, a water-land interface shows noticeable elevation gradient in comparison to the water surface. Furthermore, elevation gradient is low for natural features such as fields, rolling hills etc. and high for man-made features such as building, tower, dam, bridge etc. Therefore, distinguishing different levels of low elevation change is more useful than distinguishing high elevation changes, when natural features detection is the main concern. In angular filtering [9], elevation changes were non-linearly mapped to an angle using tangent function as shown in equation 01. As shown in Fig.01 (left), lower elevation changes were mapped to larger angle domain in compare to higher elevation change mapping. In short, using this nonlinear transformation, we group less

important higher gradient in a small angular range and provide emphasis on lower gradient values.

For a targeted LiDAR block, n neighboring tiers were selected for angle calculation as shown in Fig.01. Red pixel is the targeted block; yellow and green blocks are first and second tier neighbors respectively, where $n=2$. Maximum angular difference was considered as the angular filter value for the targeted pixel.

$$\theta_i = \tan^{-1}\left\{\frac{Z_i - Z}{d_i}\right\} \qquad [1]$$
$$\theta_{max} = \max\{|\theta_i|\}\ \forall i$$

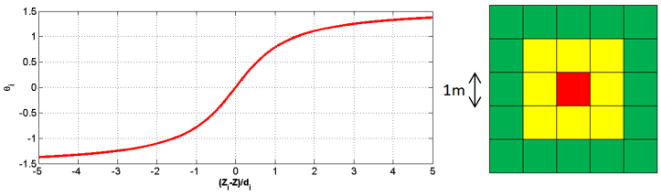

Fig. 01. (Left) Arctangent mapping of elevation change to angle . (Right) The targeted block (red), first and second tier neighbor for angular filtering (yellow and green).

B. *Motivation and hypothesis*

In this section, we perform a heuristic assessment of the LiDAR return characteristics of water bodies in order to develop a set of hypotheses to detect them. Water surface shows different characteristics than other land and man-made features. Specular reflection is prominent from the water surface, which leads to low return density and drop-outs. Drop-out indicates that no LiDAR pulses are returned to the receiver because of total specular reflection or total absorption, and due to smaller number of returns, data density can be lower too. At infrared range, water surfaces show high absorption rate which lead to very low intensity return in general. Thus, it can be safely assumed that at least a small part of the water surface has very low intensity and/or drop-outs. Two examples are shown in Fig.02. LiDAR drop-out areas are shown in orange on RGB image for two sample water-bodies. Therefore, those drop-out regions can be considered as the seed to detect the corresponding water-bodies.

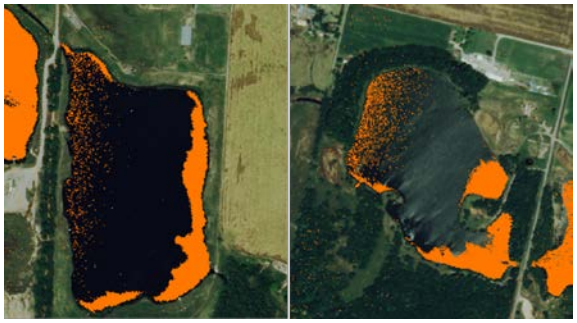

Fig. 02. LiDAR return drop-out areas (orange) are ovalaid on RGB images for two smaple water-bodies. It can be safely assumed that all water-bodies have at least a small very low intensity and/or drop-out region.

Once, a seed is used to get started with a water body, in general very few significant observations were made in LiDAR intensity from water surface. Because of high absorption rate at near infrared range, LiDAR returns are significantly absorbed by water, which leads to very low intensity return. On the other hand, specular reflection may be received by the LiDAR scanner, when scan angle is small. These specular receptions have very high intensity value. Therefore, LiDAR returns from water surfaces are more likely to have very low (absorption) or very high intensity (specular). Intensity image for two sample water-bodies are shown in Fig.03. Blue to red indicates lower to higher intensity. Intensity of drop-out areas are considered are in very low range. It is clear from Fig.03 that water-bodies mainly have very low intensity and only small part may show very high intensity due to specular reflection. Therefore, significant differences in the intensity probability density function (PDF) for water surface returns can be exploited for algorithm development.

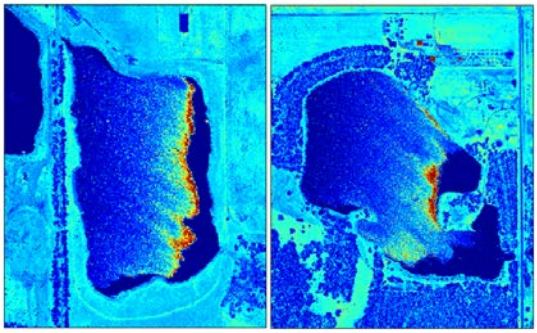

Fig. 03. Intensity map of two sample water-bodies. Blue to red means lower to higher intesity. Water-bodies mainly have very low intensity returns and occatinal very high intensity return due to specular refelction.

From the elevation point of view, water surface is comparably flatter than other topographic features. Furthermore, water to land interface has higher change of local elevation. Therefore, it can be expected that angular filtering value, $|\theta|$ will be higher around water-land interface. Angular filtering mapping are shown in Fig.04 for same sample water-bodies. Red color to blue indicates 0 to 90 degrees respectively. It is clear from Fig.05 that water-land interface has higher absolute angle filtering value.

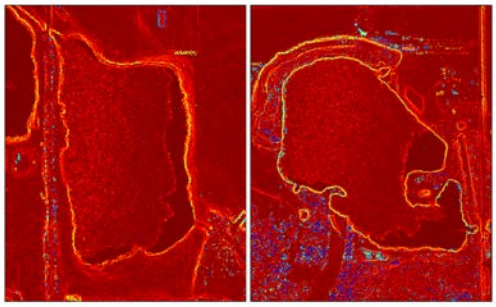

Fig. 04. Angular filtering of two sample water-bodies. Blue is 90 and red is 0. Close to zero angles indicates local flatness. Flatness is significantly changed at water-body boundaries.

III. Proposed Algorithm

An inland water-body boundary detection method for very large scale areas is targeted in this paper. LiDAR return

properties from water surfaces were utilized to develop a computationally fast algorithm. As described in the previous section, three hypotheses were made to design this algorithm. a) At least a small portion of any water-body has LiDAR return drop-outs due to specular reflection. b) LiDAR return intensity is more likely to be very low (high absorption) or very high (specular return). c) Local change of gradient is higher in water-land interface than water surface. These three key characteristics were designed to address inland water-bodies. The whole process flow of the algorithm is shown in Fig.05. Detailed procedure for the each of the working blocks is given below.

A. *Rasterization phase*

For a large scale dataset, working with point cloud costs computational power and time. Rasterization of the point cloud quantizes the irregular point cloud to a regular gridded image-like data. From a point cloud, three different rasters were produced for $2m \times 2m$ blocks. This small block size ensures low spatial quantization errors due to rasterization. Minimum elevation, maximum elevation, return counts and average intensity were used for rasterization. Vegetation height model (VHL) was generated. Finally, a drop-out raster was also produced from drop-out blocks. Vegetation blocks and sharp edge blocks generally have multiple return counts in LiDAR data.

B. *Processing phase*

Both intensity and drop-out rasters were combined to generate seeds. A connected component analysis was performed on intensity and drop-out rasters to extract all connected blocks which have very low intensity or drop-outs. If more than a significant set of connected blocks has no LiDAR return or very low intensity, then that set was defined as a seed. Angular filtering was also performed on the elevation raster to extract local change of elevation profile. Planar surfaces have near zero angle value, whereas rough areas show angle values closer to 90 degrees. A vegetation height model (VHM) was created to distinguish tall trees, using the technique introduced by Antonarakis et al [3].

C. *Iteration Phase*

The iteration phase started with the largest seed generated in processing phase. Seeds were considered as a small portion of a water surface and the area around the seed was included in that particular water-body, if elevation and intensity features support the hypothesis. At first, the angular filtering threshold $\theta_T$was selected as a very small value. Pixels with angular filtering value less than θ_T were added with the seed. Now, the grown area should include more low intensity or drop-out regions if seed was surrounded by water. Therefore, CDF of the new region should be higher or at least follow it.

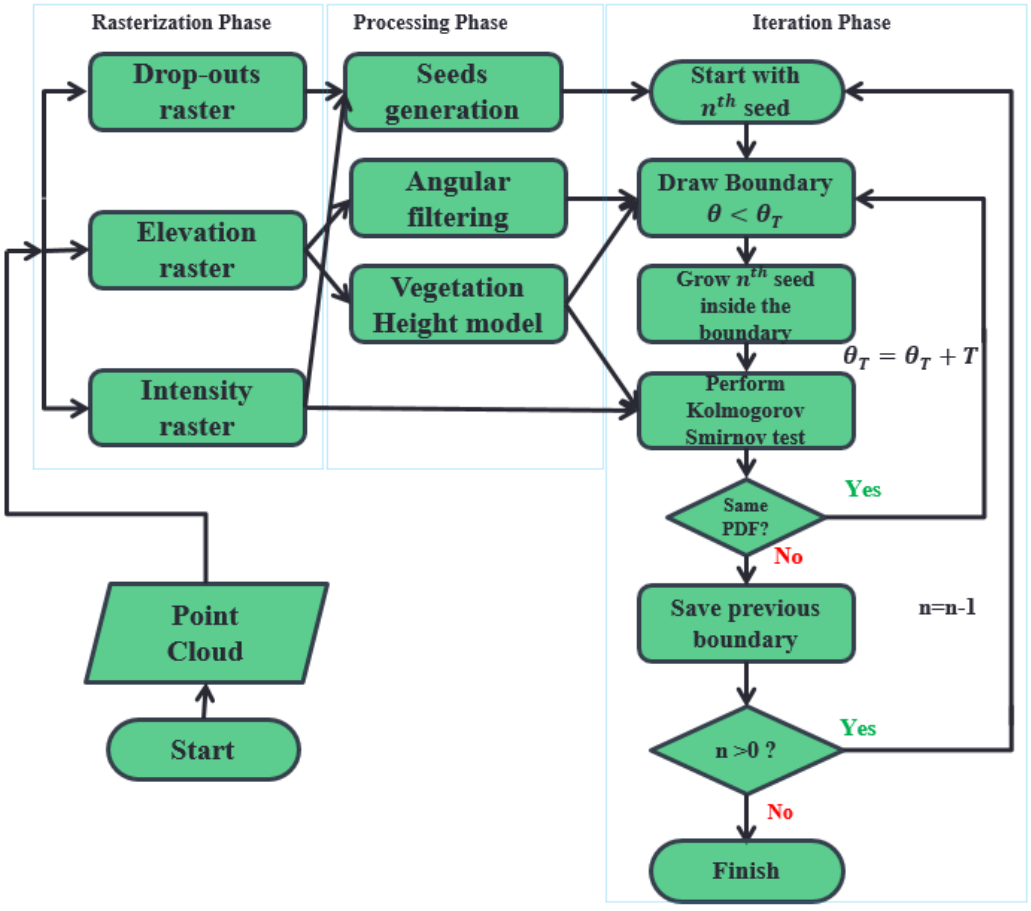

Fig. 05. Algorithomic flow-chart of the proposed water surface detection method.

A Kolmogorov-Smirnov (KS) test was performed to check the similarity of these two PDFs. The KS test was one-sided because if the new CDF was higher that the one from the previous iteration then the added region is also considered as water. If the one sided KS test passed the hypothesis of similar PDF then $\theta_T$was increased for next iteration. When, $\theta_T$was increased beyond the water-land interface value, a large land area would be included with the previous region. This condition is equivalent to virtual "flooding". At "flooding" condition, the CDF will be significantly changed, which would lead to a failed KS test. When the KS test fails for a seed, the previous boundary was restored as resultant break-line for that specific water-body.

The same procedures were followed for the next seed. It may happen that a seed can flood another unprocessed seed or a previously detected boundary. In that case, these regions were merged to represent a single water-body. Only blocks with single return count was useful if low elevation vegetation area around water-bodies were needed to be excluded. This case can be varied according to user needs for any particular project. The whole iteration phase continued until all seeds were processed.

The whole iteration phase for a single seed is illustrated in Fig.06. The detection process started with the largest seed from the left side of the water body, which is indicated as dark red. Angular filtering threshold $\theta_T$was changed iteratively and new regions were added with the seed. Therefore, the CDF was also changing during iteration. Added regions and corresponding CDFs are shown in the same color at Fig.06. It is clear from Fig.06 that the CDF shape was same or similar before flooding. At flooding condition, all blue pixels were added and the CDF drastically changed its shape. Therefore, boundaries of water surfaces were saved just before the flooding condition. It can be seen that water boundary detected after the final iteration was matched with the corresponding satellite image as shown in Fig.02.

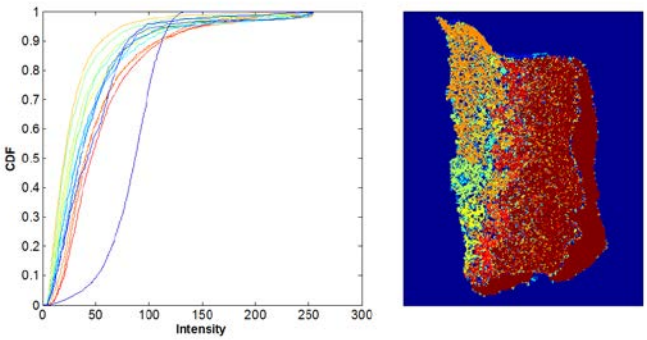

Fig. 06. Change of intensity CDF in every iteration (Left). Added region in each iteration (Right). Added areas shown going from red to blue.

IV. RESULTS

Quantitative results will be presented for two datasets, Nebraska and Oklahoma dataset. The overall accuracy (ACC), Specificity (SPC) and Sensitivity (SEN) were derived as quantitative performance measure from a confusion matrix. Simple coefficient values of performance measure may not fully provide information about the classification performance. Therefore, a detected map was also overlaid on the satellite images to present a qualitative performance measure.

In Fig.07, manually detected border and detected water-body polygons from Nebraska dataset are overlaid on the satellite image. Detected water-body boundaries can be seen to be visually matched with manually detected break-lines, except for a few small mismatches. The small island and a small water-body in that island were also detected correctly. From overlapping areas, it was observed that a 7.02 km^2water surface was detected correctly; a 0.077 km^2 area was missed and a 0.24 km^2area was misclassified as a water-surface in the 20.345 km^2study area. Therefore, it can be noted that the qualitative detection performance is satisfactory with the quantitative overall accuracy at 98.45 %, sensitivity at 96.71 %, and specificity at 99.41 %.

Fig. 07. Detected water-body overlaid on satelite image for Nebraska dataset. Manually detected border (Yellow lines) and Algorithm detected water-body (Blue semi-transparent polygon).

From the Oklahoma dataset, manual break-lines are available for water-bodies larger than 2 acres and rivers wider than 100 feet. In Fig.08, it is seen that the borders of all seven water-bodies are matched with the manual detection. The proposed method is able to detect small water-bodies and narrow rivers. In this case, a minimum size was given as constraint to measure quantitative results because manual detection was available only for large water-bodies. From overlapping areas, it was observed that 0.845km^2 water surface was detected correctly; 0.012 km^2 areas were missed and 0.01 km^2areas were misclassified as water-surfaces in total of 39.06 km^2study area. The numbers for this dataset are: overall accuracy 99.94 %, sensitivity 98.60 %, specificity 99.97 %.

Fig. 08. Seven main water-bodies in Oklahoma dataset. Manually detected border (Red lines) and Algorithm detected water-body (Blue semi-transparent polygon).

Detection performance is shown for a very large scale dataset in Fig.09. The Salisbury dataset contains a few rivers in both rural and urban scenes. Additionally, there are many in-land standing water-bodies too. In Fig.09, detected water-bodies are shown in blue polygons and an overlay satellite image. From visual inspection, it is clear that all major water-bodies were detected correctly. Five sections (A-E) of this large dataset are magnified and shown in Fig.10. From section A, it is observed that a major river and all its curved are detected as expected. Small water-bodies are also extracted correctly as shown in section B. In section C, we see that roads on water-bodies are distinguished from the detected water surfaces. Section D demonstrates detection performance in an urban scene. All bridges were removed from detected water surface. A some small portion of water surface remains undetected because those were separated from main water-bodies by multiple bridge crossings. From section E, it is clear that rivers and its branches can be extracted as expected using the proposed algorithm. Overall, the performance of the proposed algorithm is satisfactory for very large scale dataset contain both rural and urban scenes.

Detection comparisons for a few small water-bodies are shown in Fig.11. It is seen that small water-bodies inside vegetation were detected correctly. Man-made, small square shaped water-bodies, water-bodies with algae and very isolated small water-bodies were detected as well. It is also to be noted that bridges were not detected as water-body part, which is vital for many application. Parts of water-bodies which are divided by a bridge were also detected separately.

V. Conclusions

In this paper, we proposed a novel algorithm for water surface mapping using elevation and intensity of airborne LiDAR data. The quantitative results are shown for two different datasets, where manual hydro break-lines are available. For extensive analysis, qualitative results are shown for a 500 sq.km area. From visual inspection, it was clear that the proposed algorithm is working well as expected.

In future, the tool will be adapted for very large scale application. An ArcGIS toolbox will be developed for state level analysis in a batch mood. An extensive performance analysis will be completed by both UTA and NRCS personnel before final public release of the tool.

References

[1] H. F. Stockdonf, A. H. Sallenger Jr, J. H. List, and R. A. Holman, "*Estimation of shoreline position and change using airborne topographic LiDAR data*," Journal of Coastal Research, vol. 18, no. 3, pp. 502–513, 2002.

[2] S. P. Leatherman, "*Shoreline change mapping and management along the US East Coast*," Journal of Coastal Research, pp. 5–13, 2003.

[3] A. Antonarakis, K. S. Richards, and J. Brasington, "*Object-based land cover classification using airborne LiDAR*," Remote Sensing of Environment, vol. 112, no. 6, pp. 2988–2998, 2008.

[4] A. Brzank, C. Heipke, J. Goepfert, and U. Soergel, "*Aspects of generating precise digital terrain models in the Wadden Sea from lidar–water classification and structure line extraction*," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 63, no. 5, pp. 510–528, 2008..

[5] B. Höfle, M. Vetter, N. Pfeifer, G. Mandlburger, and J. Stötter, "*Water surface mapping from airborne laser scanning using signal intensity and elevation data*," Earth Surface Processes and Landforms, vol. 34, no. 12, pp. 1635–1649, 2009.

[6] J. Smeeckaert, C. Mallet, N. David, N. Chehata, and A. Ferraz, "*Large-scale classification of water areas using airborne topographic lidar data*," Remote Sensing of Environment, vol. 138, pp. 134–148, 2013.

[7] G. Toscano, P. Acharjee, and V. Devarajan, "*Water surface elevation calculation using LiDAR data*," ASPRS Annual conference, Tampa, Florida, 2015.

[8] G. Toscano, and V. Devarajan, "*A LiDAR-based auto hydro break-line generation algorithm for standing water bodies*," PhD dissertation, Uni. of Texas Arlington, 2015.

[9] P. Acharjee, G. Toscano and V. Devarajan, "*A novel angular filter based LiDAR point cloud classification*," ASPRS Annual conference, Tampa, Florida, 2015.

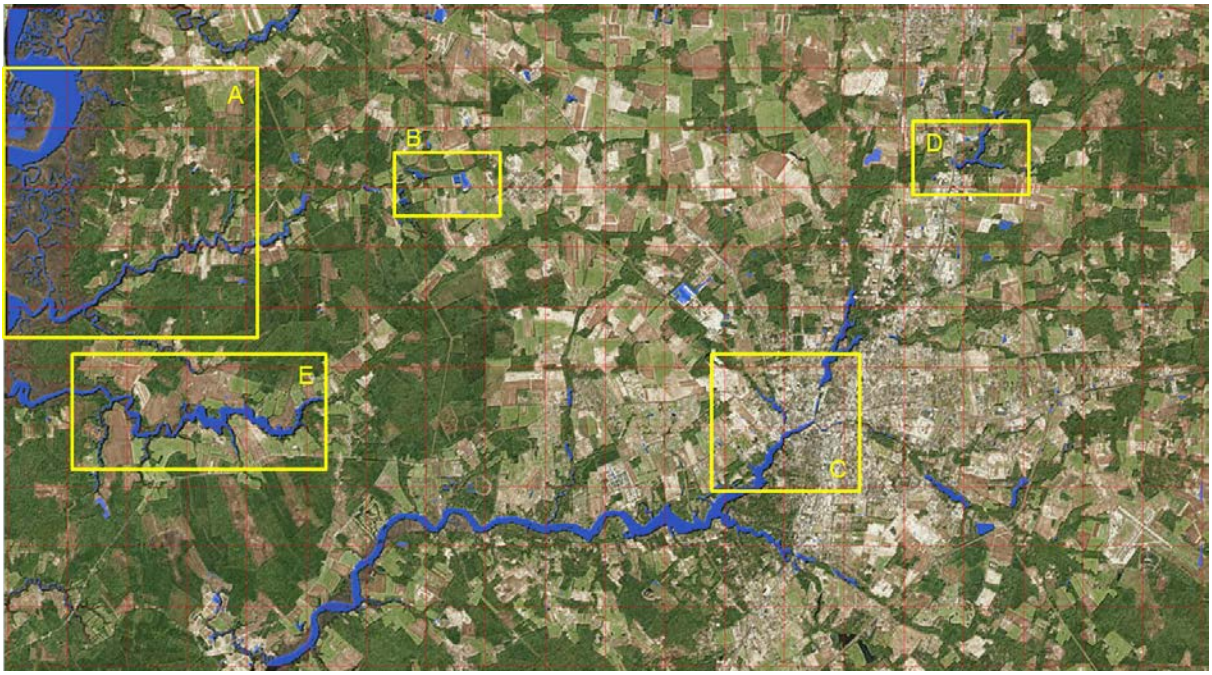

Fig.09. Detected water-body (Blue polygons) overlaid on satelite image for Maryland dataset. Five interesting areas are marked and zoomed in for better visualization. All rivers and small water-bodies were detected according to visual inspection. Grid size is around 2 km^2 and whole area size is around 500 km^2.

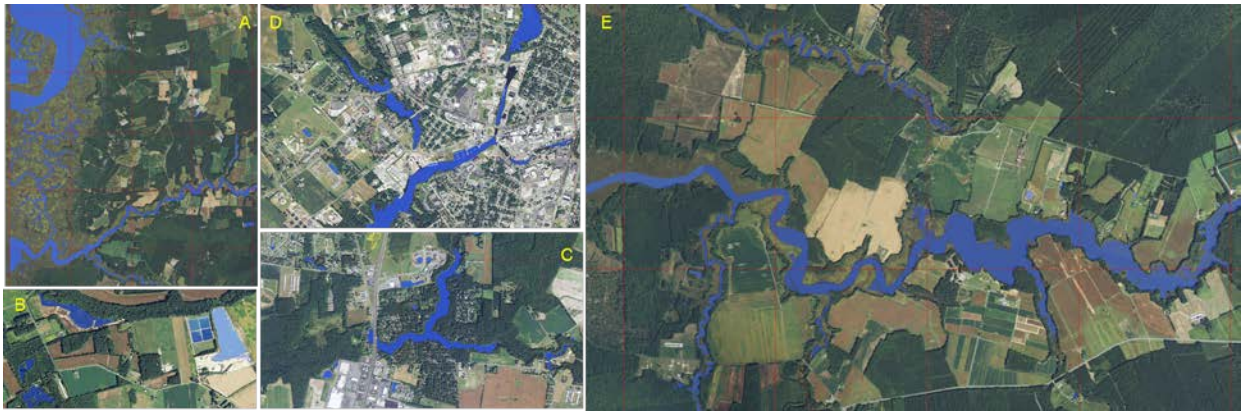

Fig.10. Detected water-body (Blue polygons) overlaid on satelite image for Maryland dataset.

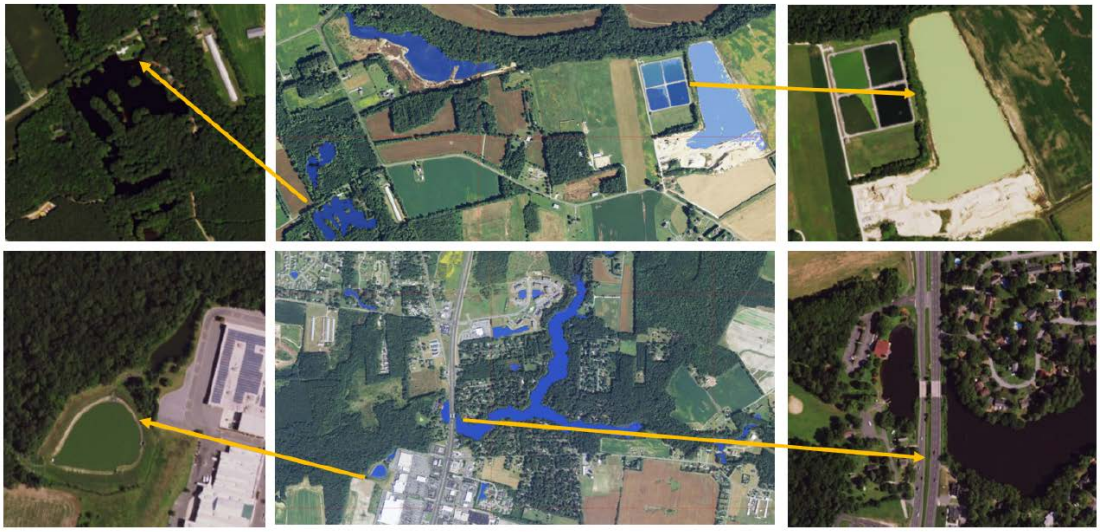

Fig.11. Detected water-body (Blue polygons) overlaid on satelite image are shown for a few small water-bodies