SEMI-AUTOMATED BATHYMETRIC MAPPING PROCEDURE FOR LANDSAT TM

François G. F. Smith
Earth Satellite Corp.
6011 Executive Blvd.
Rockville MD 20852
fsmith@earthsat.com
240-833-8262

Aurélie Shapiro
Special Projects Office
NOS/NOAA
1305 East West Highway, #9653
Silver Spring, Maryland 20910
aurelie.shapiro@noaa.gov
301-713-3000(202)

ABSTRACT

There have been many attempts to calculate bathymetry from the ratio of blue and green wavelengths of satellite images. The most popular attempt is the Lyzenga (1978) method, which implements a linear logarithmic transformation of the remotely sensed data. Many variations have been established since that initial work that have increased the efficacy, and improved on the blue/green ratio approach substantially. One of the most recent attempts is documented by Stumpf et.al. (2003), which introduces a methodology for calculating bathymetry from multispectral satellite data, reducing the number of input variables to four: slope, y-intercept, and blue and green deep-water calibration. The input variables are still manually trained, making this procedure highly dependent on human interaction. This paper describes a method for automating all of the input variables while testing the accuracy of automatically derived products against those derived from human training. The automated method is based on the Stumpf et al. formula but uses a priori sample depths to calculate the most accurate inputs to the variables.

INTRODUCTION

The NOAA Coral Reefs project is a series of tasks in support of the NOAA National Ocean Service (NOS) Special Projects group. One of the tasks in that project was to calculate estimated depth using Landsat TM imagery for specific study areas throughout the Caribbean. The resulting depth layers were used as inputs into models for run-off and change analysis. The estimated depth product was in support of a study seeking to link changes on land with those in the benthic environment.

The purpose of the NOAA Coral Reef Mapping program is to map and monitor U.S. coral reefs in order to focus federal, state and territorial efforts towards mitigating the threats posed on these marine resources. Landsat, along with IKONOS and high-resolution aerial photography are being used to map shallow water coral reefs in the U.S., territories and associated states. With the recent availability of a substantial collection of Landsat imagery acquired for the GeoCover project, NOAA has begun exploring the potential of pairing Landsat scenes from different dates for benthic change detection – identifying areas of sedimentation, seagrass colonization responsible for coral mortality. This project has been expanded to link anthropogenic changes on land with local changes in near-shore coral reefs. This analysis initially requires the ortho-rectification, atmospheric correction of Landsat imagery, as well as estimated depth and water column correction, which was completed by Earth Satellite Corporation.

LANDSAT PROCESSING

Water depth was estimated for each Landsat scene in the shallow shelf, or littoral zone. The values of this layer represent the estimated depth of pixels in feet down to 100 feet. More technically, this layer represents the estimated depth in the photic zone, or area where light penetrates to the bottom (Figure 1). Even more specifically we are mapping a subset of the photic zone, that is, the area in the shallow water environment where bottom reflectance can be measured by the satellite. In the Caribbean this area is typically less than 100 feet. The depth of the photic zone in a given image is a factor of the sediment load, sea conditions, and image quality.

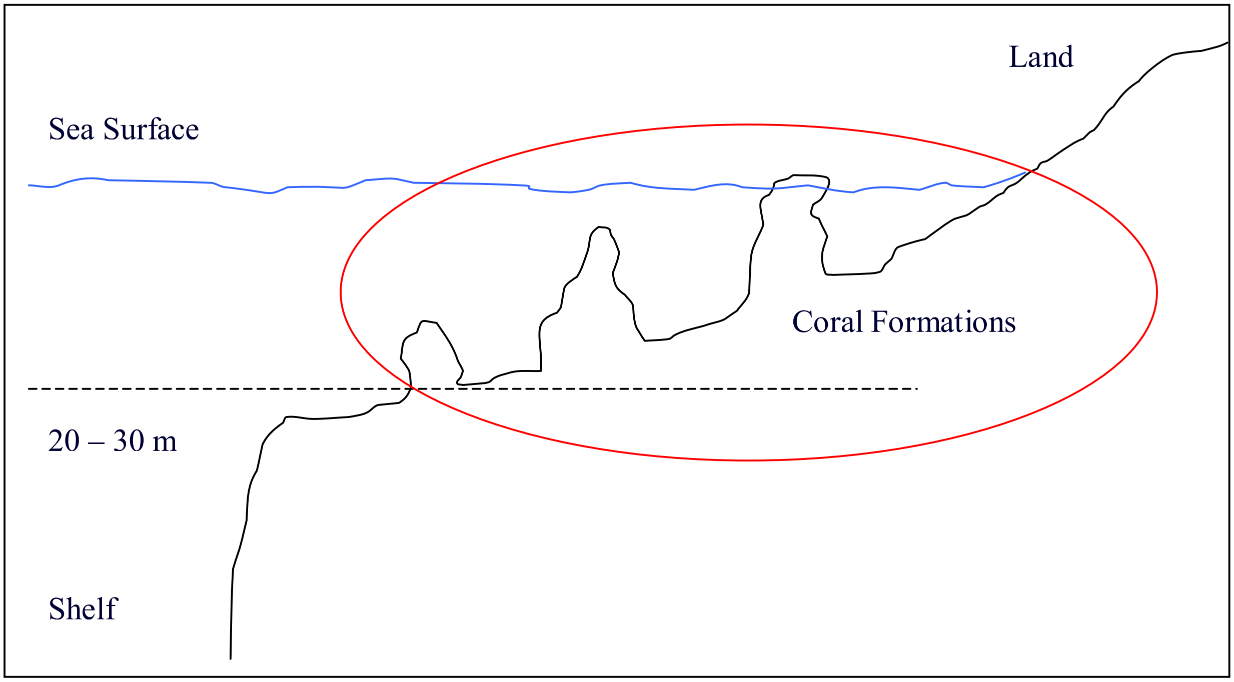

Figure 1. The red circle delineates the photic zone in this example.

The formula that drives this procedure is the Log Ratio Transformation (LRT) as published in Stumpf et al. (2003). First however, the water reflectance is calculated for the image. This reduces sensor and surface effects such as glint that occur similarly across the bands. The water reflectance formula can be represented:

$$\text{Band i} - \text{NIR Band}$$

Where i = visible multispectral Bands 1 –3 (blue, green red). The NIR Band (4) has little or no return in the presence of water, therefore when it does return over water it is probably due to some effect. The results of the water reflectance transformation have little or no land pixels and water surface is less textured.

The Log Ratio formula can be represented as:

$$m_1 * 100 * \text{Log}\ (\text{Band1-rdeepB}_1)/\text{Log}\ (\text{Band2-rdeepB}_2) - m_0 + 0.5$$

where m_1 = slope, $rdeepB_1$ = average value in Band 1 of deep water pixels, $rdeepB_2$ = average value in Band 1 of deep water pixels, and m_0 = shift. In the Stumpf et al., 2003 (Log Ratio Transformation) method, the slope and shift are manually adjusted independently to optimize the best combination for these variables. In the EarthSat approach, these variables are automated. The rdeep variables can also be automated in some scenes but in others it is better to calculate these manually depending on the percentage of deep water pixels compared to shallow water pixels. They are quickly calculated by checking many deep water points and taking one of the lowest measurements.

The first step in the EarthSat estimated depth mapping process is to create the training dataset. The training dataset must be a set of geospatial raster points with values representing depth in feet. In this project the training point layers were created by displaying scanned NOAA nautical charts or British Admiralty charts provided by NOAA

(Figures 2 and 3). Then points were selected at the sounding values. NOAA soundings are in fathoms, while British Admiralty Charts reports depths in meters. When the NOAA Nautical Charts were used, training points were given a value at each sounding (S) of S * 6 to convert the soundings to feet. In the British Admiralty Charts, the points were given a value of S * 3.28 to convert the value to feet from meters posted on the chart. The soundings on the charts are not given at a point but a general area. This makes it difficult to tell exactly what pixel the sounding is representing. Some error in the product undoubtedly results from this fact. No points were collected greater than 100 feet.

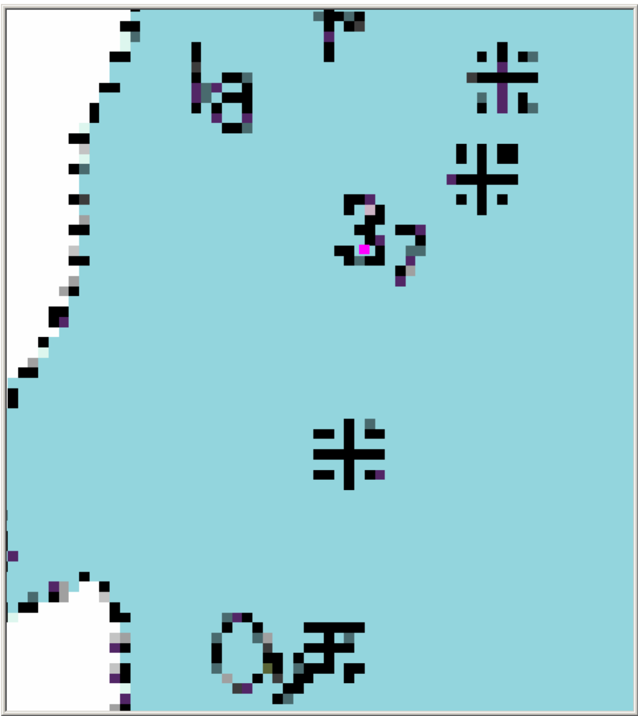

Figure 2. Example of scanned NOAA Nautical Chart. The sounding 3.7 meters is populated with a sample training point (magenta) with a value of 12.

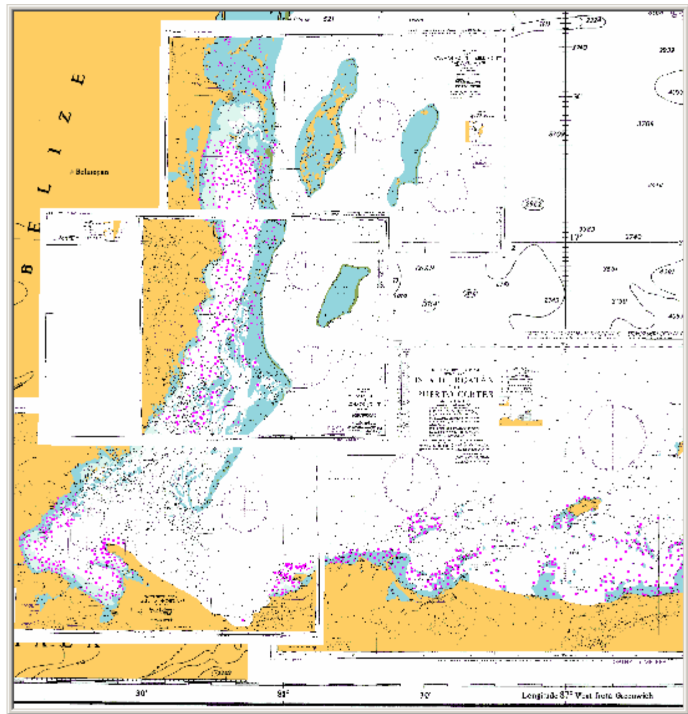

Figure 3. A mosaic of a NOAA Nautical Chart with the Belize sample depth points collected overlayed in magenta.

The Nautical Charts (and British Admiralty Charts) are at different scales and some are decades old. Although this is not the best source to collect training from, it is a consistent dataset for the entire Caribbean and enables the LRT data to be trained for absolute depths.

In some scenes, rdeep was automatically determined by averaging all of the blue band or green band and then dividing by 1.2. This is reasonable in many scenes because only water is present in the water-reflected scenes and the vast majority of most water pixels are deep water pixels. Slope and shift were not optimized in these scenes manually but were calculated by fitting the training data to the initial transformation, with no slope or shift applied, in a linear fashion. This produced the most consistent results, rather than manually adjusting the variables for each scene. Then a 3x3 focal median filter was applied to the estimated depth layer. As depth is a function of its neighboring pixel, and this invariably raised the accuracy of estimated depth for each scene.

The Nautical Chart training data was not used for all scenes. In an area such as Belize where many final depth layers were mosaicked, it was necessary to train the data to calculate the slope and shift based on the overlap of adjoining complete depth layers. First one scene was calculated using the training data, and then the subsequent scene areas were calculated using the overlap as training to create a chain of scenes. The rdeep variables were still independently determined. This chain method clearly reduced mosaic lines and made for more useable and consistent data over large areas, but probably the accuracy in the overlap-trained scenes was not increased.

For areas where more than one scene was available, all the data were used to produce a final estimated depth for that area. This allowed for continuous data with few if any “holes” in the data. There were still some holes in the data when only one or two scenes were used because sometimes clouds or other contamination overlapped between two dates. For additional scenes the above processes were performed and then the median value among all scenes were calculated. In the case where there were only two scenes for the same footprint, the average was returned at each pixel. Finally, the estimated depth layers were individually masked by a shelf or photic zone delineation to remove all of the noise in the deep-water pixels where the formula is not relevant.

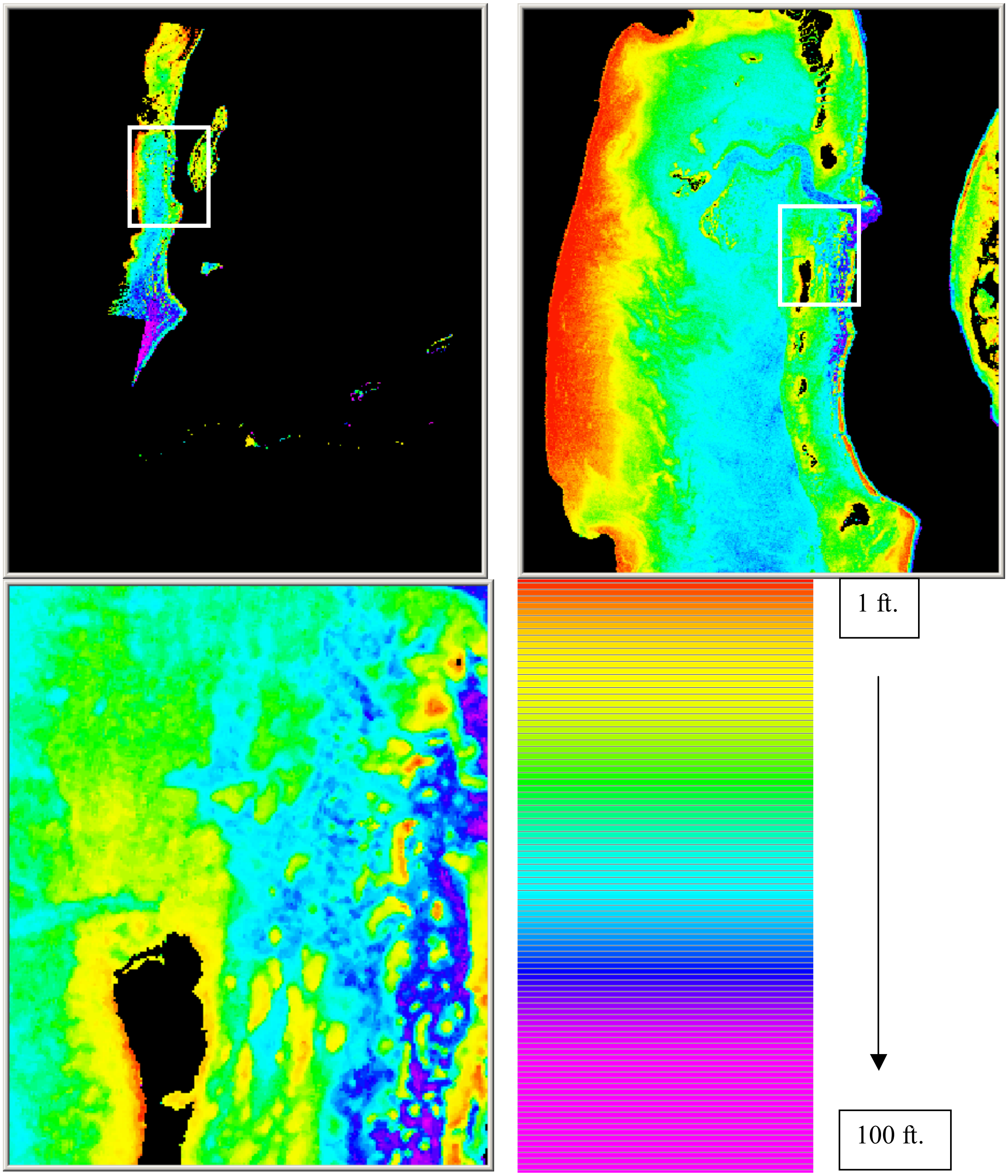

Figure 4. Depths derived from EarthSat's estimated depth procedure. Example is from Belize Landsat imagery. Two dates of imagery were used per scene footprint to make this 8 TM scene footprint mosaic.

There exists some unavoidable error in this procedure, mostly due to suspended sediment, which acts much like cloud cover. It is impossible using this technique to obtain reliable depth information in areas of sustained sedimentation. The algorithm tends to estimate these areas to be shallower than they are. Cloud contamination remains a problem even with two scenes per scene footprint. Using more than two scenes generally minimizes the effect of clouds. Often the cloudy pixels influence the digital numbers of the surrounding pixels in a large buffer. Another important factor is the quality and quantity of the training data. This affects the average error of the final depth product significantly.

ACCURACY ASSESSMENT

A formal accuracy assessment was not completed for this project, however informal tests of the accuracy were completed. The Belize study area was tested during the course of the estimated depth production. Buck Island, off of Saint Croix in the U.S. Virgin Islands was also tested although it was not part of this study area. NOAA was able to provide EarthSat with large amounts of field collected shallow water soundings for the Buck Island area to test the accuracy of the depth estimation procedure. In the tables below, the results are reported for several areas in terms of correlation coefficient and average error. A high correlation is a good indication of the potential of the algorithm for depth classification; it reports the relative accuracy of the results. The average error indicates the absolute error in the data. It is the average difference between the points returned in the results and the validation data.

Many different variations of estimated depth mapping procedures were attempted in the validation process of the depth estimation layers. The product as delivered to NOAA utilized a linear fit of the log ratio transformation to determine the slope and shift. Another method tested for fitting the data was a weighted least squares (WLS) approach. This allowed the fit to be weighted based on depth. Therefore where the LRT is more accurate, generally in the lower depths, the influence of the fit to the regression line was greater. Where the correlation is lower, and thus the LRT has less correlation to depth, the data is not as much influenced by the line of fit. Another approach attempted was to apply At-Satellite-Reflectance (ASR) to the imagery prior to analysis. This correction creates at-satellite radiance first by calibrating the imagery based on the gains and biases reported in the image header. A formula is then applied to take into account the sun angle, Earth-Sun distance, and mean solar exo-atmospheric irradiance. For more on this procedure, consult the USGS website: http://landcover.usgs.gov/pdf/image_preprocessing.pdf
Principal Components Analysis (PCA) was also tested for correlation with depth in an uncorrected form. Finally, the manual method that is generally used at NOAA was tested for accuracy compared to the automated process.

The study area for this test was a two-scene mosaic of the shelf off the coast of Belize. Training for the automated processes was based on digitizing points off of the British Admiralty charts discussed earlier. The validation points were collected at the same time from the same source. The initial ground truth layer was subset into training and validation points by grouping the points into clusters of 10. Odd numbered ground truth points were used to build the training point layer, and even points were set aside for use as validation points. The clumps of 10 points were used to alleviate spatial auto-correlation inherent in continuous data such as bathymetry or elevation. Training data included 255 points used in classification while 256 validation points were withheld.

For Belize, several attempts were tested including: a linear fit (same as delivered product), a WLS weighted by 1, and an ASR corrected linear fit. The accuracies of the depth layers were reported in terms of Pearson's Correlation Coefficient (PCC), overall average error, and average error < 20m deep. Table 1 provides the results of this test.

Table 1. Results of the estimated depth tested in Belize.

Layer	Pearson's CC	Average Error	Avg Error < 20m
A=not at-sat, linear	0.67065	17.42073	9.941748
B=WLS- 1, not at-sat	0.648791	27.09146	6.044444
C=at-sat, linear	0.601741	19.08537	11.10417
D=Ground Truth	1	0	0

These results suggest that the non-ASR corrected linear fit version has the highest correlation with validation depth measurements (67%), and the lowest average error overall. Not surprisingly, the average error below 20 meters for the WLS approach is lower than the other methods. It is weighted to be more influenced by the data of lower depth because the LRT is better correlated at the lower depths. It would be logical to assume that an atmospheric correction algorithm applied before the transform would aid in depth mapping but in this case that was found to be false.

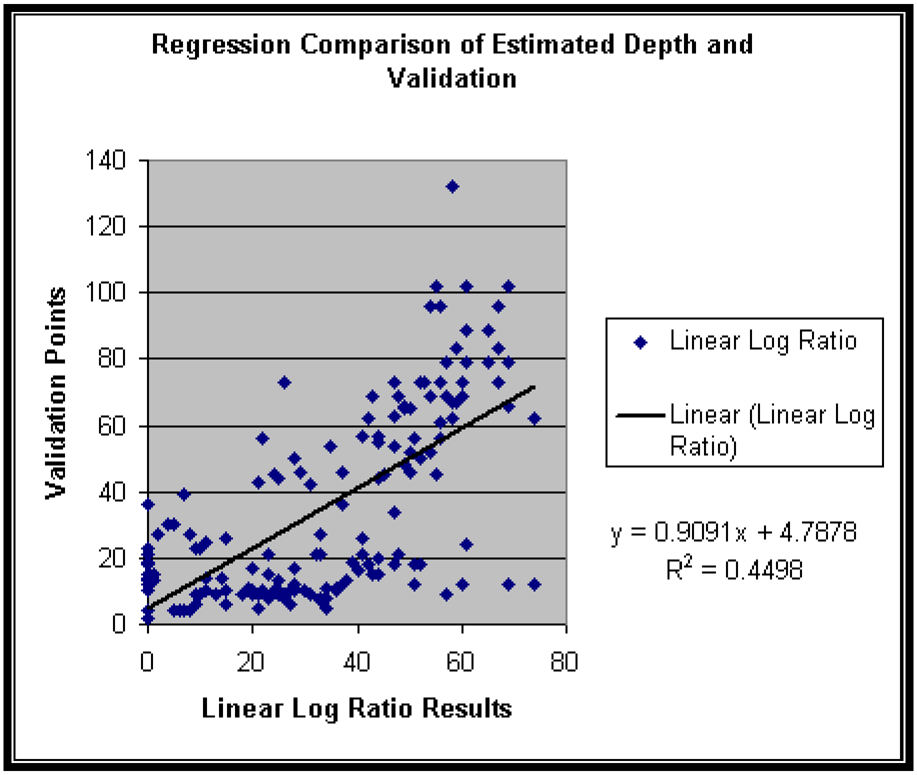

Figure 5. Regression calculated for the linear fit attempt and the validation points for the Belize data.

The area around Buck Island, Saint Croix was the other site in the Caribbean used to test the estimated depth procedure. Though Buck Island was not part of the study area, testing of the procedure was performed on Landsat data of this area for two reasons. First, unlike Puerto Rico and some other areas nearby, the sea state and lack of sediment around Buck Island created an ideal situation for testing of the algorithm itself without those factors to confound the results. Buck Island also contains a large shelf or littoral zone. Much of this zone is situated in a protected channel. There was little run-off at least at the time of image acquisition. Secondly, this is an area in which NOAA has collected extensive depth soundings on the shallow water shelf.

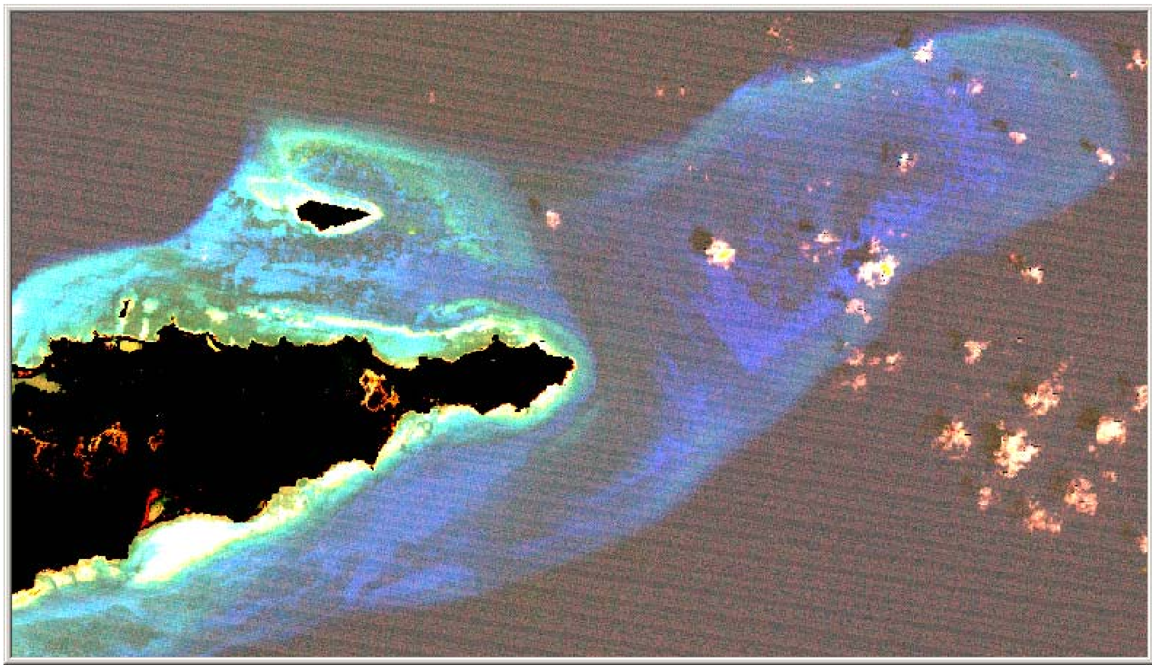

Figure 6. Buck Island, water reflectance transform.

NOAA provided 20,041 depth points after the data were rasterized and reprojected. The points were subset into training and validation points. After subset there were 9,857 training points and 10,184 validation points. These data were delivered to EarthSat in depth units of meters.

Many variations of the log ratio transformation technique were applied to the Buck Island data. They included running the weighted least squares (WLS) approach four different ways including setting the weighting to 1, 3, 5 and 1 with a 3x3 median filter applied. A PCA transformation that was fit to the training data was tested as well as the manual NOAA approach, and the automated linear fit. The results of these attempts are listed in Table 2.

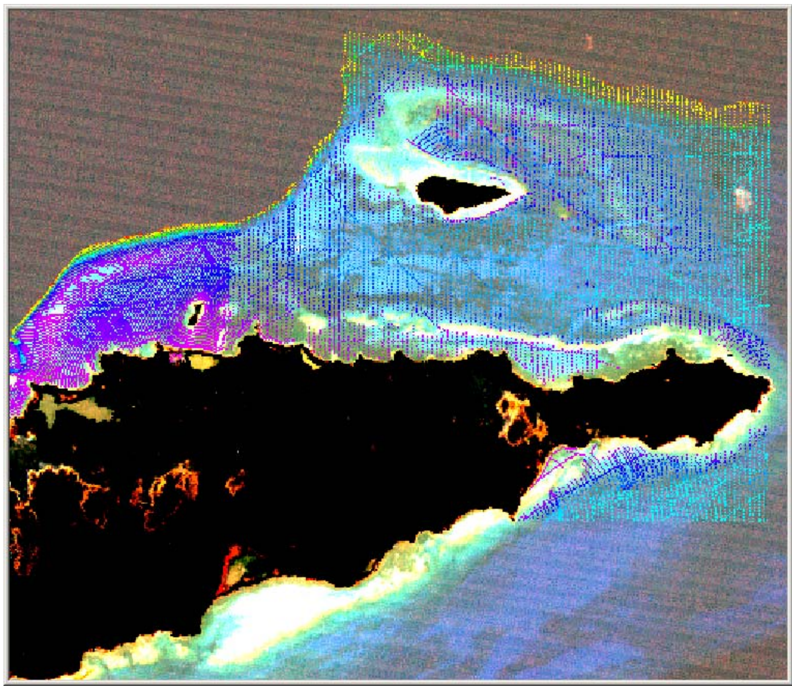

Figure 7. Buck Island water reflectance transform with field-collected depth soundings overlaid.

Table 2. Accuracies of estimated depth attempts.

Data Layer	Pearson's CC	Average Error	Average Error < 20
A=Weighted Least Means approach, weighting of 1	0.646868024	4.377693156	2.885200553
B=WLS, 3	0.645305501	4.129872733	2.702516191
C=WLS, 5	0.651260551	3.992784848	2.622655505
D=PCA 4th transform	0.561640544	69.78645155	No Data
E=Manual NOAA approach	0.697072907	5.680929953	5.018129079
F=WLS,1,3x3 median filter	0.677267027	4.30604269	2.800574651
G=Log Linear fit	0.65400906	4.290109229	3.406181476
H=Field-collected ground truth validation	1	0	0

The table clearly shows the manual NOAA approach to have the strongest correlation with the data. It is sometimes the case that the manual approach has higher correlation however, the correlation varies by scene. This technique is much less consistent than the automated approaches but carries a higher potential for relative accuracy. This accuracy is probably related to the rdeep variable. It is theoretically probable that if the rdeep variables are determined manually and the slope and shift are determined automatically that the highest and yet most consistent results will be attained. This was the theory applied for the final Belize study area estimated depth product, which was not tested. However, the absolute accuracy as represented by the average error is lower in the automated technique. This is related to the slope and shift variables in the formula. These variables are particularly difficult to optimize in the NOAA manual technique. In this test, the linear fit data is more accurate overall than the WLS fit data.

The table shows the overall average error for the estimated depth layers tested as well as the average error under 20 meters depth. This is important to know because depending on the application it may be necessary to achieve greater accuracy at certain depths.

The PCA attempt was not as successful. The theory behind trying this approach was that the fourth principal component in a PCA is naturally well correlated with depth. It was expected that the average error would be high and even off the scale as no values were less than 20 meters, but it was also expected that the correlation to depth would be stronger. Some studies have shown that transforming the data using component analysis is one way to map depth. But this has only been done successfully when components are manually altered in transformed space to find the optimal settings for depth correlation. This manual component shift was attempted in Belize but no results were produced or tested. This appears to be an unstable method and inaccurate if no component shift is applied.

A spatial filter with a kernel size of 3 x 3 pixels was applied to the WLS product with a weighting of 1. It was applied to test how a spatial filter can increase the correlation of the results. In this case, it increased the correlation from 65% to 68%. This is because depth is spatially auto-correlated. Each pixel has a higher likelihood of being similar to its neighbors. Therefore if a spatial filter is performed, it smoothes out error in the results and creates a more realistic surface to represent bathymetry. Applying a spatial filter had no significant impact on the average error however. Average error remained at about 4.3 meters overall and 2.8 meters under 20 meter depth.

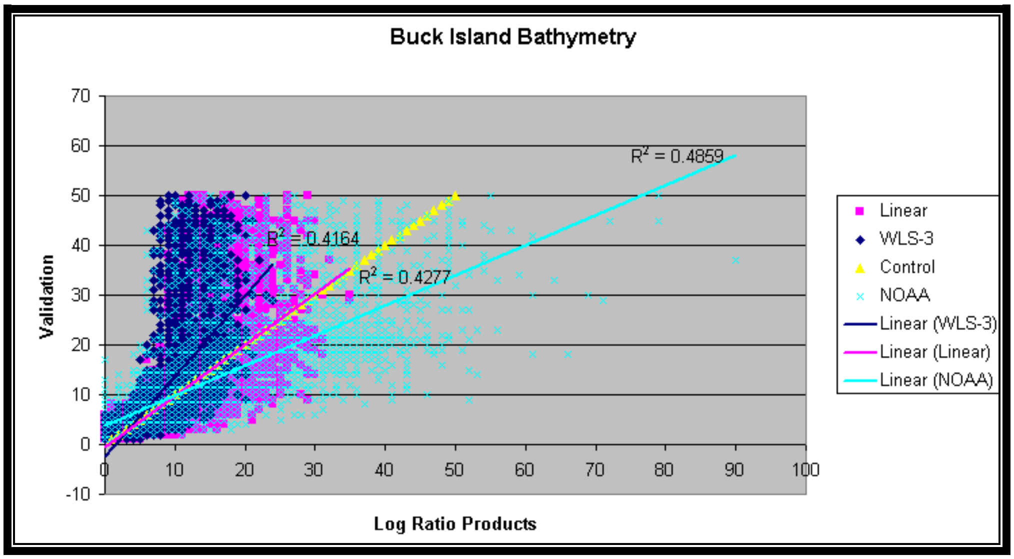

Figure 8. Regression calculated for several depth products and the validation points for the Buck Island data.

As expected, the WLS approaches are much more accurate in the shallower depths, but beyond 15-20 meters the accuracies fall off dramatically. Over 20 meters there is no correlation. In the linear fit approach, the shallower depths are less accurate than the WLS approach. After 20 meters, the accuracy falls off but there is still correlation. However, after 30 meters there is no longer substantial correlation with the ground truth soundings. The manual approach (in Figure 8 labeled "NOAA"), which has not been transformed to fit the data, returns depths that are still somewhat correlated as deep as 80 meters. But the correlation in the shallower depths are much weaker than in the other two methods. There is still a logarithmic relationship in the manually factored data as well (see Figure 9).

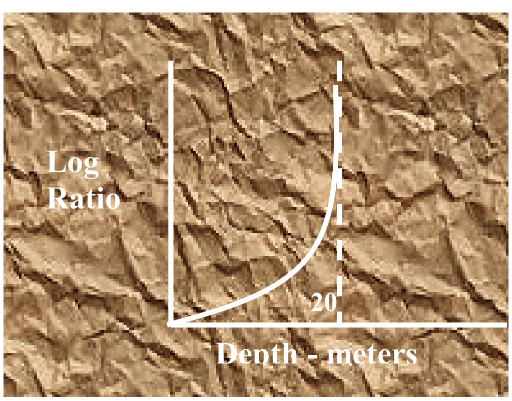

Figure 9. Graph of the relationship of LRT products to depth. Due to the logarithmic behavior of the estimated depth results, the accuracy of the points is inversely related to the depth.

DISCUSSION

Manual and automated attempts of estimated depth were compared in several validation tests and it was found that accuracies for the manual depth estimations (where the four variables present in the LRT formula were manually derived as per NOAA's recommended technique), were highly variable. Theoretically, a user who is very familiar with a particular area and has a priori knowledge of the depths in the photic/littoral zone would be able to optimize these variables such that the maximum overall accuracy would be higher than if the variables were automatically derived. However, with the manual NOAA approach, if the area is not well known, the accuracy returned may potentially be much lower than when using the automated means.

The manual technique is also more time-consuming. Whereas the limiting factor time-wise in applying the automated technique is processing the training data and the computing time for the models, the manual technique is limited by human interpretation. In more complex areas, where there is a large variability in terms of depth and sea state, it often takes one or more days to complete one scene. This same area can be mapped with the automated technique within two hours including pre- and post-processing. Therefore, the most appropriate method depends on the project, available resources, and the study area.

The automated technique is faster, more objective, more consistent, and ultimately necessary for mosaicking scenes. If the study area is large and comprises many scenes that need to be mosaicked, and the imagery is not of the best quality, the automated method is most appropriate. In many projects, the ideal method may begin with one scene that is well known and therefore transformed with manually derived variables and the adjacent scenes generated by automatically deriving variables based on the overlap area. Although estimated depths produced by the manual method have the potential for higher overall accuracy, they more often are not. Given recent experience, higher accuracy is obtained when the correlation is higher. Often, the overall average error is not as high as the automated method even with a greater correlation, and under any assumed circumstances, the average error < 20 meters would be higher (worse) in the manually derived product.

Based on preliminary findings in this report, a good estimate of automatically derived depth products have a correlation to depth at around 65-70% and the average error depends mostly on the training dataset used. If nautical charts are used, the overall average error will be around the 6-meter range (20 ft.). If a ground-truth layer consisting of many soundings is used, then it will be in the 4-meter range (10-15 ft.). The quality of training data can nearly cut the average error in half though the correlation coefficient will remain largely unaffected.

The average error in the linear and WLS transformed approaches is correlated to the depth returned in the product. In other words, the deeper the estimated depth, the higher the potential error. Based on preliminary findings in this study, the intensity of that correlation is increased with the order of magnitude of linear adjustment. This means that the error in the manual method is less due to depth, and more due to depth in the linear fit, and the error is mostly corresponding to depth in the WLS product.

Issues that likely limit the accuracy of log transformed depth estimations are: 1 – training data, 2 – image quality, and 3 – the formula itself. The effect of the training data was discussed in the last paragraph. The effect of the image quality is tremendous if sediment is present, waves are particularly intense, algal blooms are visible, or cloud and cloud shadows are present. Sensor noise and speckle can also contribute to inconsistent results. Generally, the more recent Landsat 7 data is less noisy and better for depth estimation than earlier Landsat 4 or 5. The formula itself is limited due to the small range of blue values in Landsat imagery (8-bit), which is scattered intensely by the atmosphere. A high percentage of the blue band values are atmospheric noise. It does not appear to increase accuracy of the depth product to correct for this atmospheric contamination. In addition, the formula is both limited and made possible by the attenuation of the blue and green bands through water.

CONCLUSIONS

During the estimated depth study, EarthSat found that it was possible to automate the estimation of depth from satellite imagery that correlate with ground truth to 65-70%, with an average error of 10 to 20 feet depending on the training data quality. EarthSat determined that applying a 3 x 3 spatial filter to a depth surface increases the correlation to training data by several percent. EarthSat provided evidence to suggest that pre-processing the data for ASR correction does not effectively increase the accuracy of the depth results. EarthSat demonstrated the strength of the depth estimation procedure based on efficiency, accuracy and most importantly consistency, allowing for large-area analysis. This study also helped determine the relationship between error and the procedure. This includes the relationship between training data quality and average error, the relationship between spatial filtering and correlation, and the relationship between depth and average error.

This study opens the discussion for global depth estimation. This procedure has the potential to map depths for all the photic zone area of the Earth due to its efficiency, consistency in results regardless of the interpreter, and consistency from image to image if overlap is used as training.

REFERENCES

Lyzenga, D.R. (1978). Passive Remote Sensing Techniques for Mapping Water Depth and Bottom Features. *Applied Optics*, 17:379-383.

Stumpf, R P., K. Holderied (2003). Determination of Water Depth with High-Resolution Satellite Imagery Over Variable Bottom Types. *Liminology and Oceanography*, 48(1):547-556.