Error Analysis of Bathymetric Data derived from IKONOS Imagery
Location: Tinian Island, Commonwealth of the Northern Marianas Islands

Produced by: Kyle R. Hogrefe: Hogrefe Natural Systems Analysis in affiliation with Davey Jones Locker GIS Laboratory, Oregon State University

Produced for: Pacific Islands Benthic Habitat Mapping Center (PIBHMC) / NOAA Fisheries' Coral Reef Ecosystem Division (CRED)

Analysis Overview

Bathymetric data were derived from IKONOS multispectral satellite imagery provided by the National Center for Coastal Monitoring and Assessment (NCCMA). The original imagery, purchased from Space Imaging, Inc. (now Geoeye, Inc.), was orthorectified to correct for detected geographic offsets. Three images, acquired on different dates, were analyzed to extend the spatial coverage of the final derived bathymetry product by combining data from cloud free areas. The image file names were "tinian_233_msi.pix", "tinian_109297_msi_000_rat.img" and "tinian_109297_msi_001_rat", but they will be referred to as Tin233, Tin297-00 and Tin297-01, respectively, in this analysis.

Processing steps were based on methods originating in Lyzenga 1985 with refinement as described in Hogrefe et al. 2008 and Hogrefe 2008 (http://oregonstate.edu/~hogrefek/Cookbook/).
An overview of the processing steps as follows:

1) Conversion of file type to view images
2) Data conversion from digital number to radiance values
3) Correction for atmosphere and water surface reflection
4) Linearization of spectral decay of as function of depth
5) Masking of data not applicable to depth derivation
6) Georectification of Images
7) Extraction of linearized spectral values and depth data
8) Perform multiple linear regression to determine formula variables for depth derivation (and derivation of depth)
9) Integration of derived bathymetry with multibeam sonar bathymetry

This analysis focuses on the statistical accuracy several products that result from step 8 to determine the most accurate data for integration with multibeam sonar bathymetry collected by PIBHMC/CRED. Once the multiple linear regression was performed in step 8 (above), the resulting variables were plugged into the multivariate slope intercept formula (below) to derive bathymetry. These variables can be adjusted to increase the accuracy and coverage of the product. The two basic changes in derived bathymetry that can be accomplished by adjusting the original multiple linear regression (MLR) variables are:

1) Depths can be changed equally across the entire image by adjusting the Y intercept. Depths are increased when the Y intercept is decreased and depths are decreased when the Y intercept is increased.

2) The slope of the regression line (of the derived depths against sonar bathymetry) in the error analysis can be changed by adjusting the slope of the linearized blue and green spectral values. Thus, increasing shallow depths more than deeper depths.

These adjustments to the MLR variables allow for greater depth range and spatial coverage in the derived bathymetry. Depths derived in areas of very shallow water often have positive values which are then lost when the product is "trimmed" to include only depth (i.e. negative) values. I hypothesize that this effect is due to an inversion of the spectral relationship between the blue and green bands in these very shallow areas. Where depths are greater than ~3 meters, blue radiance values are always greater than green radiance values, however, where depths are less than ~3 meters the inverse is often true. Because the multibeam bathymetry is seldom shallower than 10 meters in the Tinian data, this "shallow inversion" is not captured in the values extracted for the multiple linear regression that determines the variables for depth derivation. While this problem is reduced by adjusting the MLR variables, the adjustments also impact the statistical accuracy of the product.

This error analysis validates the choice of which product(s) to integrate by establishing the statistical accuracy of derived bathymetry from each image using the original variables and then documents the statistical differences as the variables are adjusted.

The formula used to derive bathymetry is a multivariate slope intercept formula as follows:

Depth = Y_{int} + $(m_{blue})(x_{blue})$ + $(m_{green})(x_{green})$

Where:

Y_{int} = Y intercept

m = slope

x = linearized spectral value

Part 1: Analysis of Derived Bathymetry using Extraction Points.

In processing step 7, ArcGIS point features are created to extract sonar depth and linearized spectral values for use in the step 8 MLR analysis. Over 150 points are chosen per satellite image where pixels with clear spectral signal are concurrent with depths between the shallowest available and 25 meters. These same point features are used to extract derived depth values for comparison with sonar depth values in the following liner regression analyses. For image Tin233, 504 points were used while 384 points were used for image Tin297-00 and 435 points were used for image Tin297-01.

Image Tin233

As shown in Figure 1, the R^2 value for the bathymetry derived using the original MLR values is 0.4872 while the slope of the regression line is also 0.4872. The moderate R^2 value (approaching 1) represents a fairly tight grouping of the derived depth scatter plot around its regression line while the moderate slope value indicates a good correlation between derived and sonar depth, represented by the (red) plot. The resulting raster grid provides realistic bathymetric data, but its shortcoming is the aforementioned phenomenon of positive values in very shallow areas.

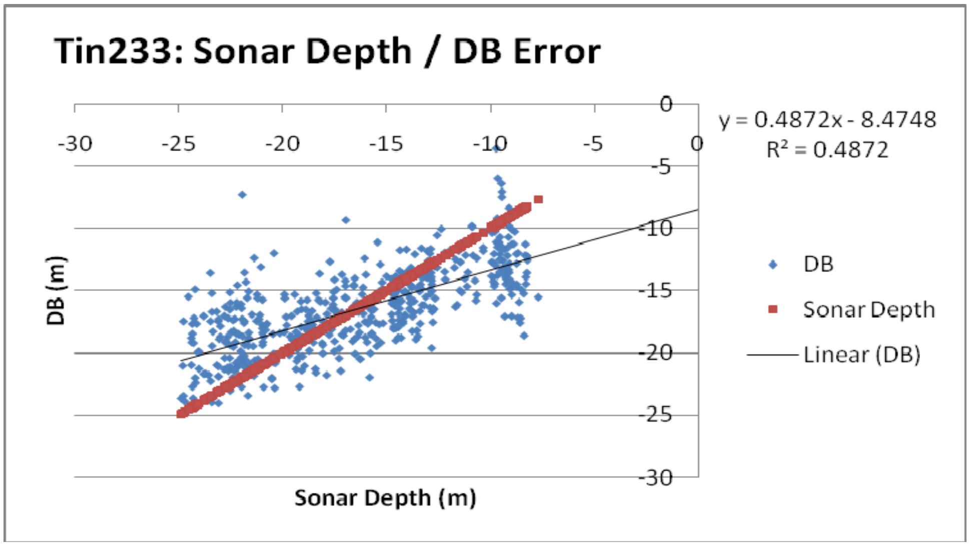

Figure 1. Error analysis of derived bathymetry from Tin233 using original MLR variables**.** Formula applied: Depth = 17.9857 - 12.0532 * X_{blue} + 31.5961 * X_{green}

In order to increase shallow coverage (0-5 m), new bathymetry was derived by both decreasing the Y-intercept (Figure 2) and then by simultaneaiusly decreasing the slope values derived for the linearized blue and green spectral data (Figure 3). The decrease in Y-intercept increased depths “across the board” while resulting in more accuurate derived values at greater depths. However, mid-range and shallow depths are made less accurate. The decrease of the slope of the linearized radiance values of both the blue and green bands, causes an increase in shallow depths with minimal changes to derived depth values from 15 to 25 meters. Both of these adjustments to the MLR variables result in an approximately 4 m increase in shallow depths (Note the simailar Y-intercept values in Figures 2 and 3) and comparable increases to the spatial coverage of the derived product (detected by visual assessment).

In order to take advantage of the increased coverage in the 0-5 meter range achieved by adjusting the MLR variables while maintaining the accuracy in the original product, each of the two adjusted products were integrated with the original into a mosaic. During the mosaic, the data from the original product was prioritized over that from those with changed variables so that the more accurate data, in the 5-15 m range, was retained while the shallow water coverage was extended.

The statistical analyses of the Tin233-DB/DB1 and Tin233-DB/DB5 mosaiced products are provided in Part 2 (Figures 13 and 14).

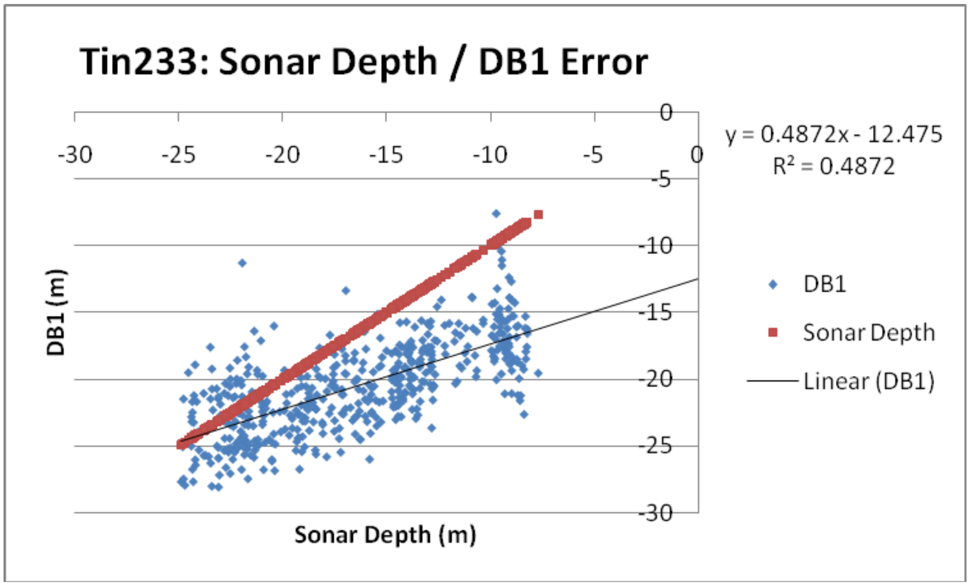

Figure 2. Error analysis of derived bathymetry from Tin233 using a decreased Y-intercept value**.** Formula applied: Depth = 13.9857 - 12.0532 * X_{blue} + 31.5961 * X_{green}

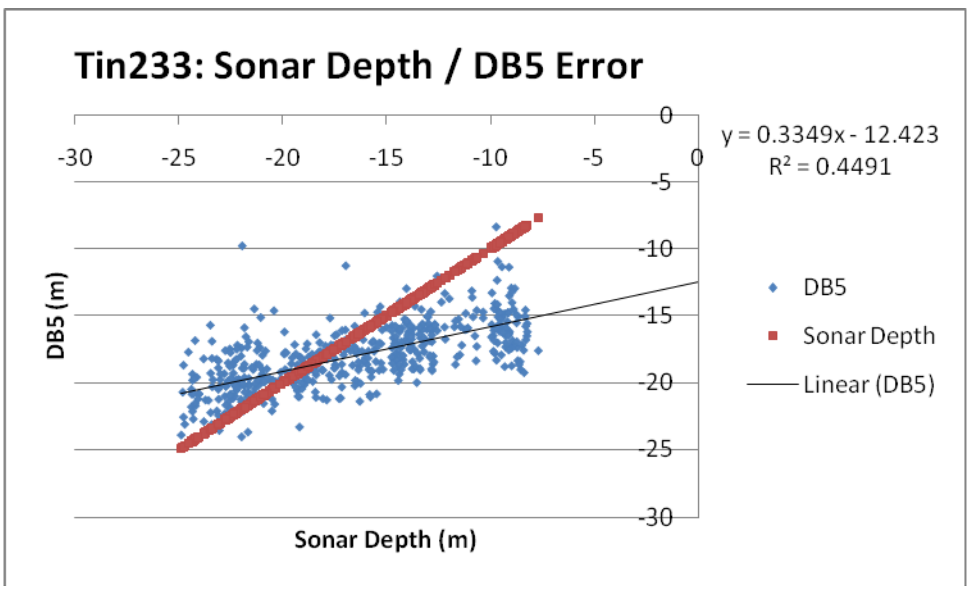

Figure 3. Error analysis of derived bathymetry from Tin233 using decreased blue and green slope values**.** Formula applied: Depth = 17.9857 - 7.5532 * X_{blue} + 27.0961 * X_{green}

Image Tin297-00

As shown in Figure 4, the R^2 value for the bathymetry derived using the original MLR values is 0.3375 while the slope of the regression line is also 0.3375. This R^2 value represents a statistically significant grouping of the derived depth scatter plot around its regression line while the slope value indicates a correlation between derived and sonar depth, represented by the (red) plot. The resulting raster grid provides realistic bathymetric data, but its shortcoming is the aforementioned phenomenon of positive values in very shallow areas.

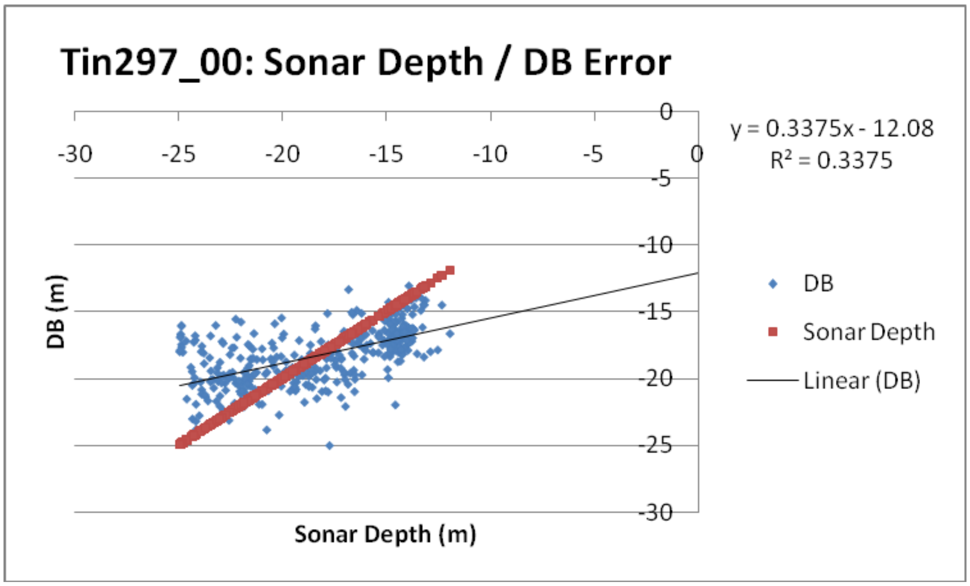

Figure 4. Error analysis of derived bathymetry from Tin297-00 using original MLR variables**.** Formula applied: Depth = 12.1661 - 4.1476 * X_{blue} + 19.2248 * X_{green}

In order to increase shallow coverage (0-5 m), new bathymetry was derived by both decreasing the Y-intercept (Figure 5) and then by simultaneaiusly decreasing the slope values derived for the linearized blue and green spectral data (Figure 6). Though specific values vary, the effects of the changed variables exhibit the same patterns as described in the discussion of the Tin233 products (above).

Again, each of the two adjusted products was integrated with the original into a mosaic with the data from the original product prioritized over that from those with changed variables to increase coverage in shallow terrain without completely sacrificing the better statistical accuracy of the original.

The statistical analyses of the Tin29700-DB/DB1 and Tin29700-DB/DB3 mosaiced products are provided in Part 2 (Figures 18 and 19).

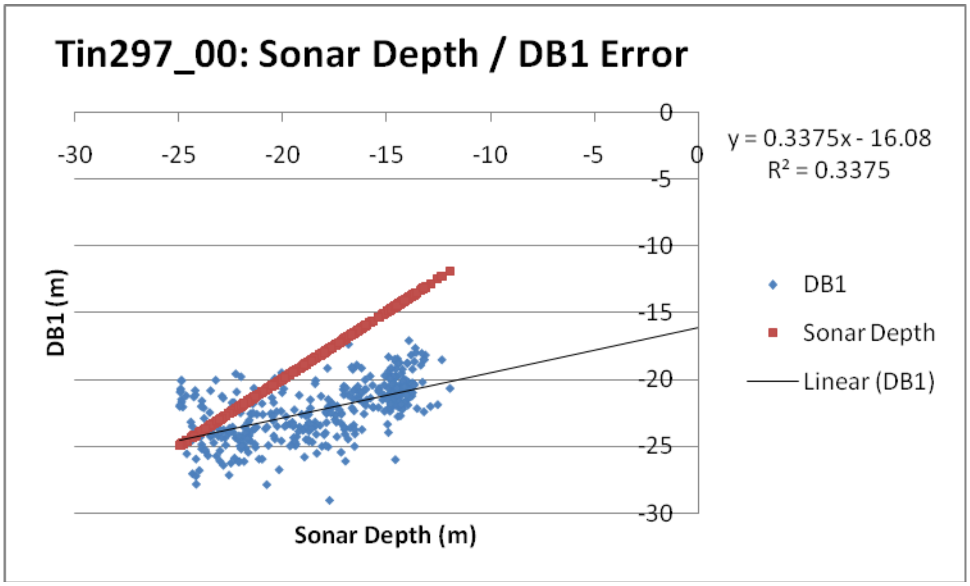

Figure 5. Error analysis of derived bathymetry from Tin297-00 using decreased Y-intercept value**.** Formula applied: Depth = 8.1661 - 4.1476 * X_{blue} + 19.2248 * X_{green}

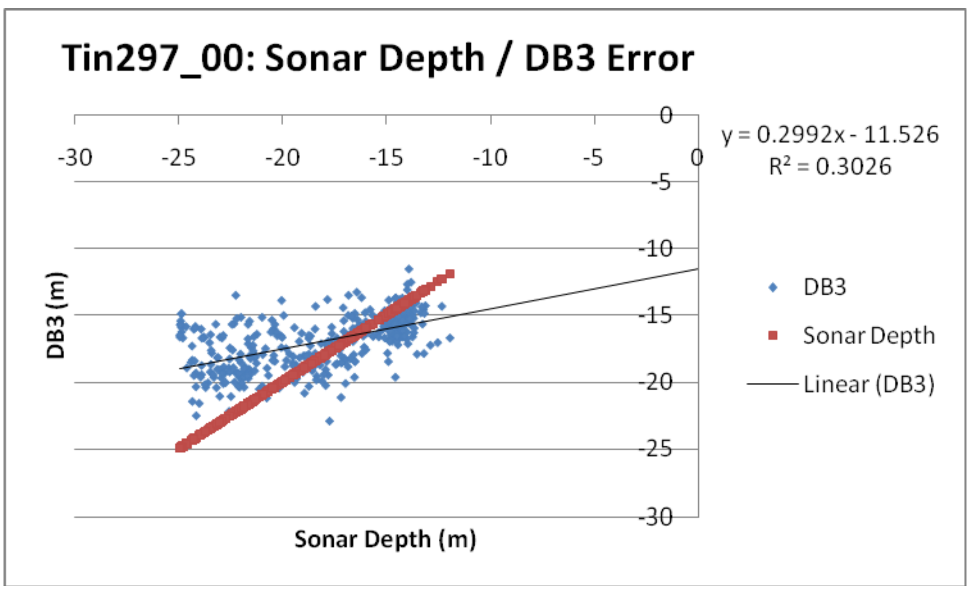

Figure 6. Error analysis of derived bathymetry from Tin297-00 using decreased blue and green slope values**.** Formula applied: Depth = 12.1661 - 0.1476 * X_{blue} + 15.2248 * X_{green}

Image Tin297-01

As shown in Figure 7, the R^2 value for the bathymetry derived using the original MLR values is 0.5512 while the slope of the regression line is also 0.5512. This moderate R^2 value represents a statistically significant grouping of the derived depth scatter plot around its regression line while the slope value indicates a good correlation between derived and sonar depth, represented by the (red) plot. The resulting raster grid provides quite realistic bathymetric data, but its shortcoming is the aforementioned phenomenon of positive values in very shallow areas.

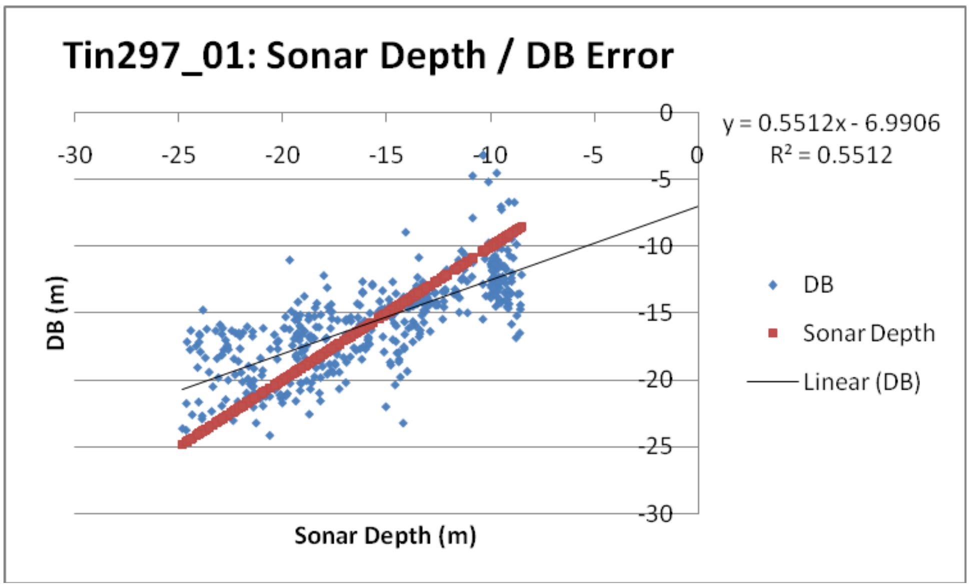

Figure 7. Error analysis of derived bathymetry from Tin297-01 using original MLR variables**.** Formula applied: Depth = 5.1266 - 15.8184 * X_{blue} + 25.5864 * X_{green}

In order to increase shallow coverage (0-5 m), new bathymetry was derived by both decreasing the Y-intercept (Figure 8) and then by simultaneaiusly decreasing the slope values derived for the linearized blue and green spectral data (Figure 9). Though specific values vary, the effects of the changed variables exhibit the same patterns as described in the discussion of the Tin233 products (above).

Once again, each of the two adjusted products was integrated with the original into a mosaic with the data from the original product prioritized over that from those with changed variables to increase coverage in shallow terrain without completely sacrificing the better statistical accuracy of the original.

The statistical analyses of the Tin29701-DB/DB1 and Tin29701-DB/DB4 mosaiced products are provided in Part 2 (Figures 23 and 24).

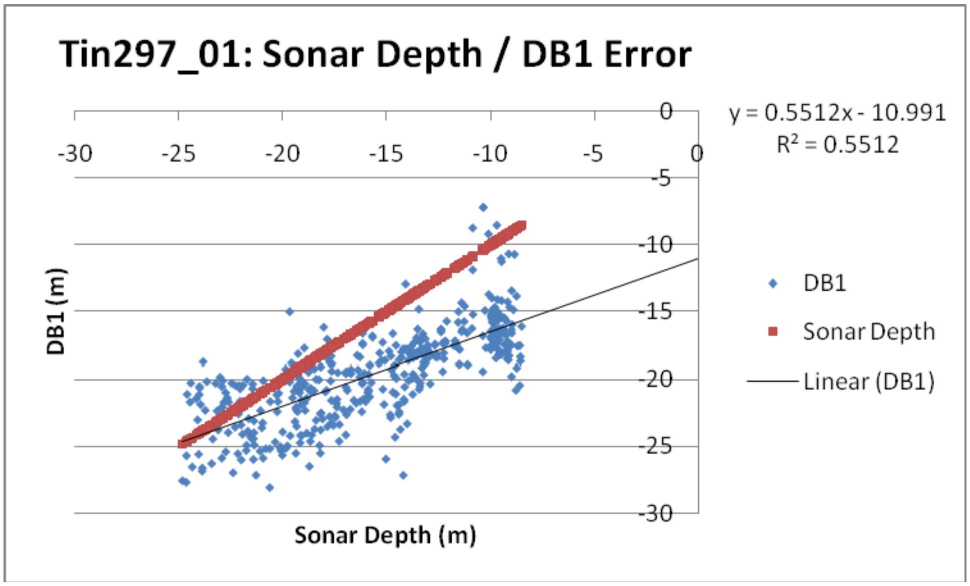

Figure 8. Error analysis of derived bathymetry from Tin297-01 using decreased Y-intercept values**.** Formula applied: Depth = 1.1266 - 15.8184 * X_{blue} + 25.5864 * X_{green}

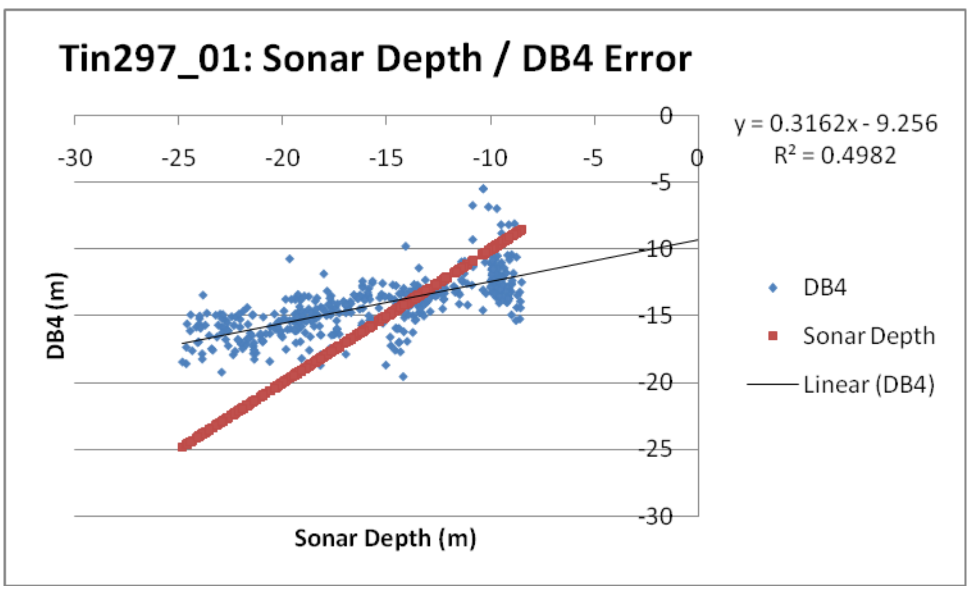

Figure 9. Error analysis of derived bathymetry from Tin297-01 using decreased blue and green slope values**.** Formula applied: Depth = 5.1266 - 7.8184 * X_{blue} + 17.5864 * X_{green}

Part 2: Analysis of Derived Bathymetry where Derived Depth Concurs with Multibeam Sonar Depths of less than 15 m.

The following analyses are a comprehensive statistical review of three derived bathymetry grids from each image as they are integrated to form an integrated derived bathymetry mosaic, Tin_DBall_mos. An analysis of this final derived bathymetry grid, which is subsequently integrated with the multibeam bathymetry into the product Tin_DBMB_mos, is also included.

After the Part 1 analyses were conducted, the derived bathymetry raster grids were prepared for integration with multibeam bathymetry by applying masks to exclude:

1) Derived depths of greater than 25 meters or less than 0 meters
2) Values derived from island areas and areas of cloud cover
3) Depths derived in areas deeper that 25 meters as indicated by the multibeam sonar bathymetry
4) Values derived inside the Tinian coastline as indicated by a shapefile provided by the NCCMA with the IKONOS imagery.

All remaining values were considered to be potentially valid derived depths, however, derived data deeper than 15 m are seldom used because the multibeam data usually reaches shallower depths. Therefore, the following error analyses utilize the derived and sonar depth values of each grid cell where derived data overlaps with multibeam sonar data of 15 m or less. Microsoft Excel was used for these analyses.

Because the Tinian multibeam data is deeper than 20 m for much of the island's perimeter, Tin_DBMB_mos includes derived depths up to the derivation methodology's limit of 20-25 meters. The large number of datapoints when all derived depths to 20 m are considered (over 32,000) overloads Excel's capacity so that S-Plus was used for further analyses of this depth range for the final derived bathy mosaic.

Helpful notes:

Figures 10, 11 and 12 analyze the same product as Figures 1, 2 and 3: bathymetry derived from image Tin233 using original MLR variables, reduced Y-intercept and reduced linearized blue and green band slope values, respectively.

Figures 15, 16 and 17 analyze the same product as Figures 4, 5 and 6: bathymetry derived from image Tin297-00 using original MLR variables, reduced Y-intercept and reduced linearized blue and green band slope values, respectively.

Figures 20, 21 and 22 analyze the same product as Figures 7, 8 and 9: bathymetry derived from image Tin297-01 using original MLR variables, reduced Y-intercept and reduced linearized blue and green band slope values, respectively.

Image Tin233

For the derived bathymetry products from the image Tin233 presented in Figures 10, 11 and 12, the R^2 values for the derived products are reduced significantly (from Part 1) due to a greater variability in departure from the mean value when all derived depth less than 15 m are considered. This is due to the variable sea state and cloud cover in this image adding to the inherent variability in the derived depth data when considering over 5,000 data points. However, notice that the slope value increases in each case when more points are considered. This indicates that the slope of the scatterplot trend line more closely matches the 1:1 relationship of the sonar depth plotted against itself, a better reflection of reality.

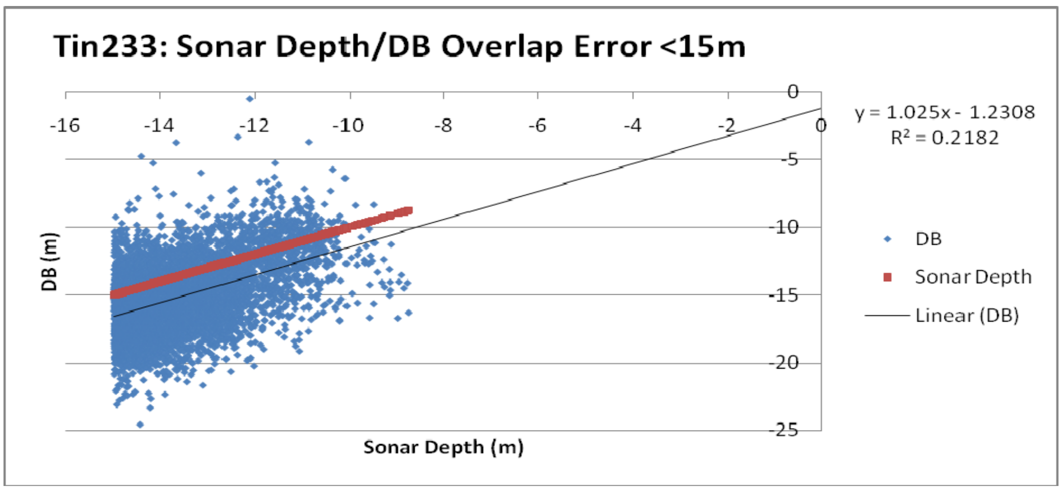

Figure 10. Comprehensive error analysis of derived bathymetry from Tin233 using original MLR variables**.** Formula applied: Depth = 17.9857 - 12.0532 * X_{blue} + 31.5961 * X_{green}

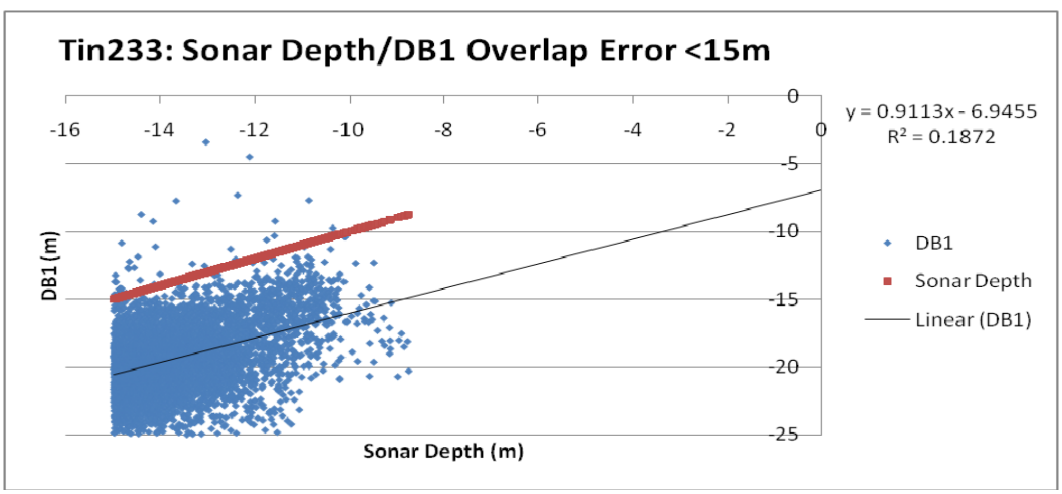

Figure 11. Comprehensive error analysis of derived bathymetry from Tin233 using decreased Y-intercept value**.** Formula applied: Depth = 13.9857 - 12.0532 * X_{blue} + 31.5961 * X_{green}

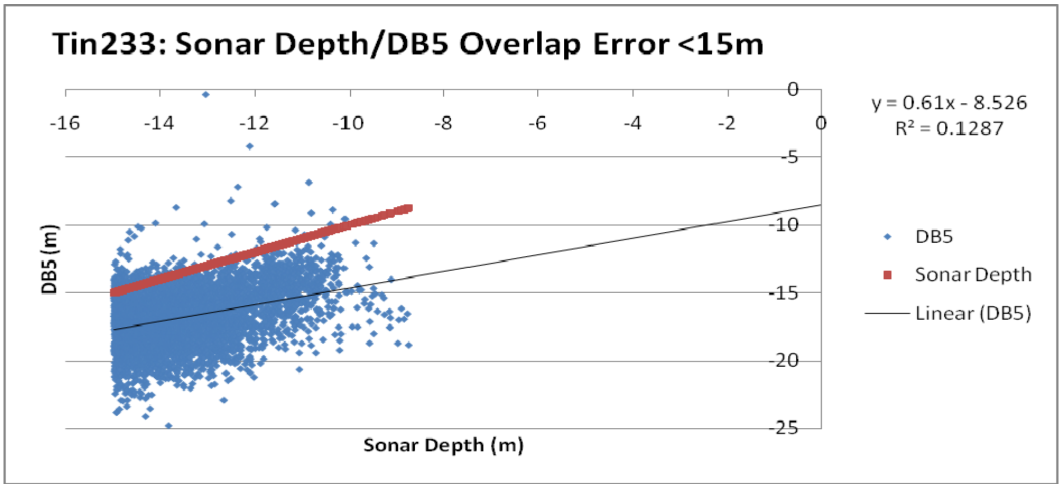

Figure 12. Comprehensive error analysis of derived bathymetry from Tin233 using decreased blue and green slope values**.** Formula applied: Depth = 17.9857 - 7.5532 * X_{blue} + 27.0961 * X_{green}

Both products that resulted from adjusted variables, DB1 and DB5, were mosiaced with the original product, DB, to expand spatial coverage in shallow areas.

Considering the derived bathymetry mosaic Tin233_DBDB1_mos (Figure 13), the statistical accuracy of the integrated product decreases from that of the original product (Figure 10) with decreased R^2 and Y-intercept values. In contrast, the statistical accuracy of the derived bathymetry mosaic Tin233_DBDB5_mos (Figure 14) actually increases over that of the original product with increased R^2 and Y-intercept values.

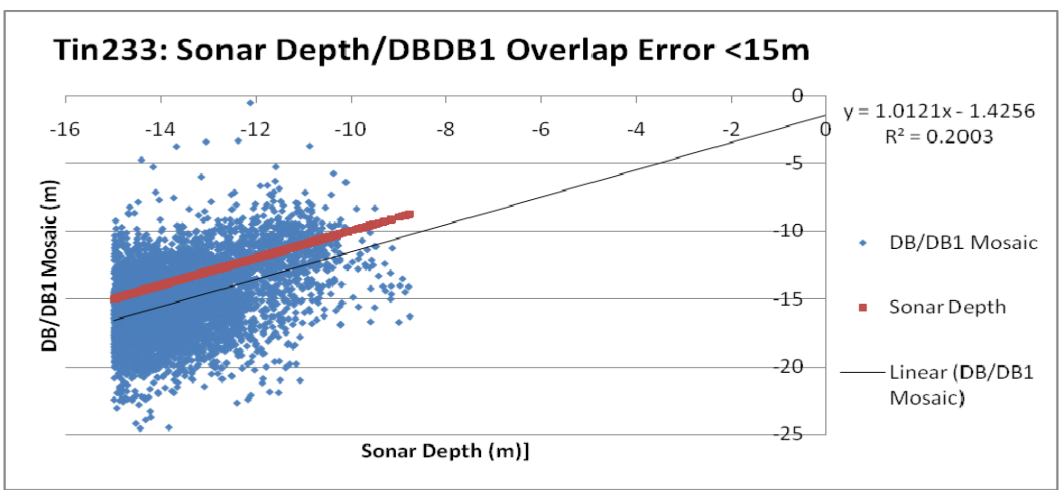

Figure 13. Comprehensive error analysis: mosaic of derived bathymetry products DB and DB1.

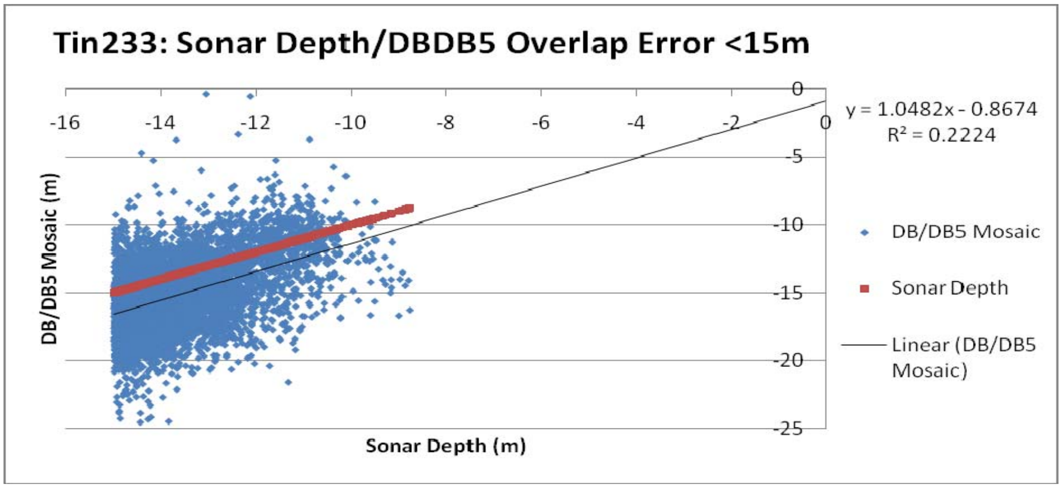

Figure 14. Comprehensive error analysis: mosaic of derived bathymetry products DB and DB5.

Given the results of these statistical analyses, the mosaic of derived bathymetry products DB and DB5 from image Tin233 is considered to be the best product for eventual integration with multibeam bathymetry. However, large areas of shallow water terrain were obscured by cloud cover in this image. Images Tin297-00 and Tin297-01 also have significant cloud cover, but many of areas obscured in Tin233 were cloud free. Bathymetry was also derived from these images to extend spatial coverage of the final product.

Image Tin297-00

For the derived bathymetry products from the image Tin297-00 presented in Figures 15, 16 and 17, the R^2 values for the derived products are reduced significantly (from Part 1) due to a greater variability in departure from the mean value when all derived depth less than 15 m are considered. This is due to the moderate sea state and cloud cover in this image adding to the inherent variability in the derived depth data when considering more points. In addition to these factors, the poor statistical results for these products are the result of very few shallow multibeam readings (none < 10m) in cloud free areas of the image. Not only are there few data points available to establish the spectral decay relationship within this depth range, but the resulting derived depths are not able to be tested. The LiDAR data recently made available will directly address both of these issues in future versions of this product.

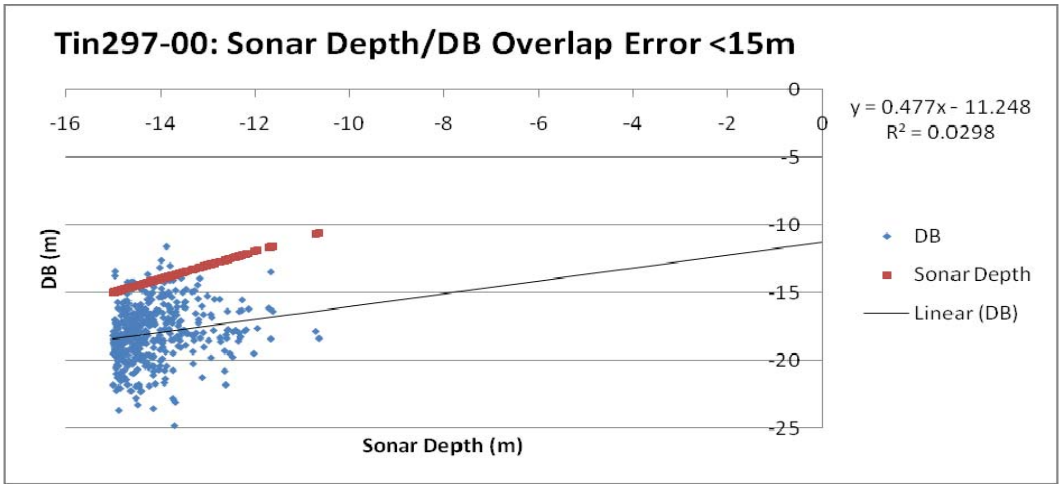

Figure 15. Comprehensive error analysis of derived bathymetry from Tin297-00 using original MLR variables**.** Formula applied: Depth = 12.1661 - 4.1476 * X_{blue} + 19.2248 * X_{green}

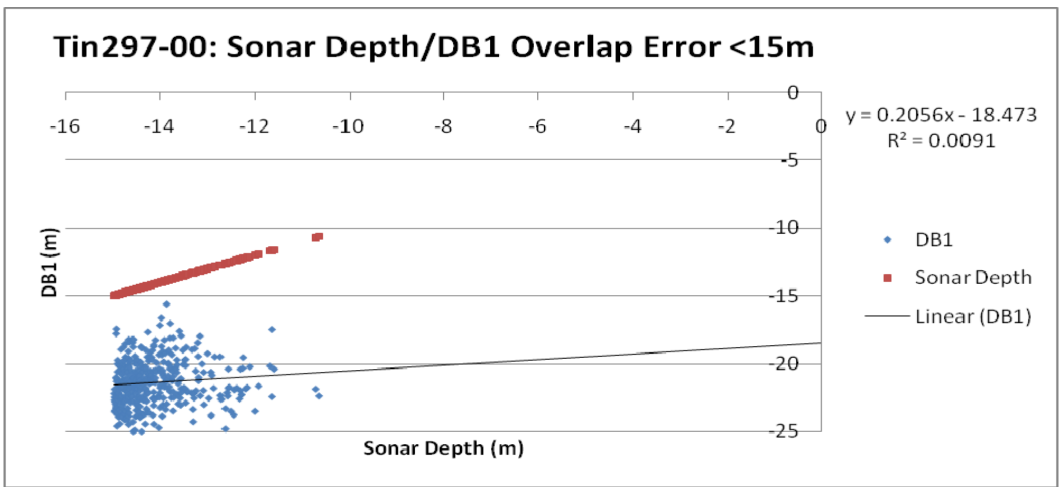

Figure 16. Comprehensive error analysis of derived bathymetry from Tin297-00 using decreased Y-intercept value**.** Formula applied: Depth = 8.1661 - 4.1476 * X_{blue} + 19.2248 * X_{green}

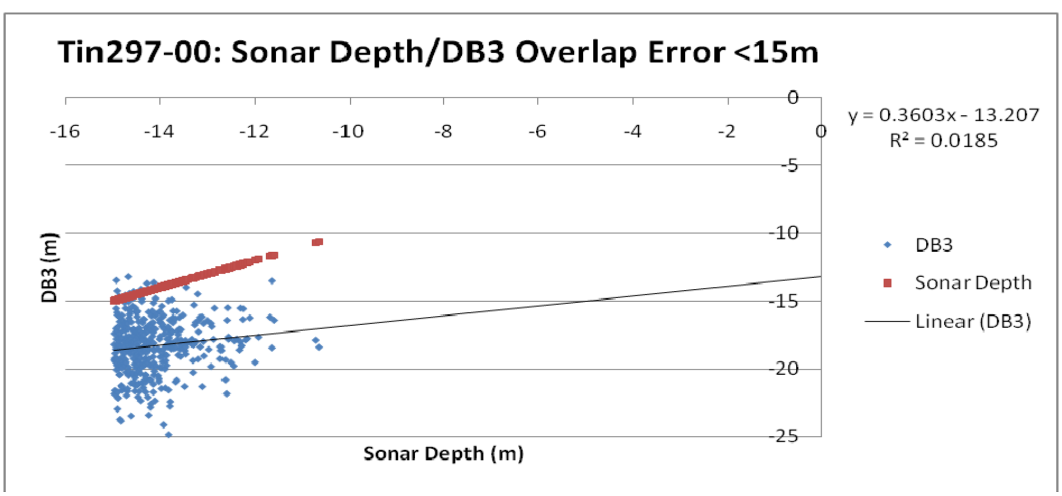

Figure 17. Error analysis of derived bathymetry from Tin297-00 using decreased blue and green slope values**.** Formula applied: Depth = 12.1661 - 0.1476 * X_{blue} + 15.2248 * X_{gree}

Both products that resulted from adjusted variables, DB1 and DB3, were mosiaced with the original product, DB, to expand spatial coverage in shallow areas.

Considering the derived bathymetry mosaic Tin29700_DBDB1_mos (Figure 18), the statistical accuracy of the integrated product decreases from that of the original product (Figure 15) with significantly decreased R^2 and Y-intercept values. However, the statistical accuracy of the derived bathymetry mosaic Tin29700_DBDB3_mos (Figure 19) only decreases slightly in both measures.

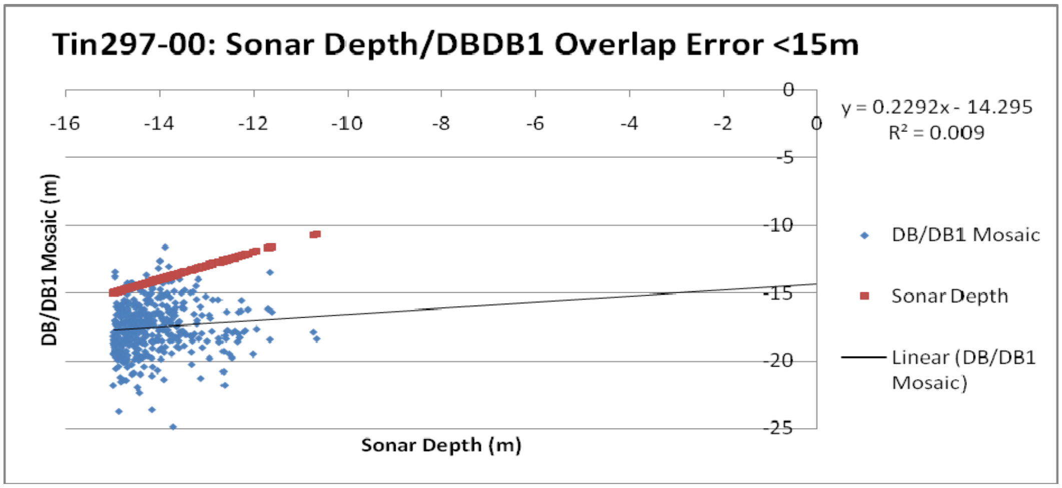

Figure 18. Comprehensive error analysis: mosaic of derived bathymetry products DB and DB1.

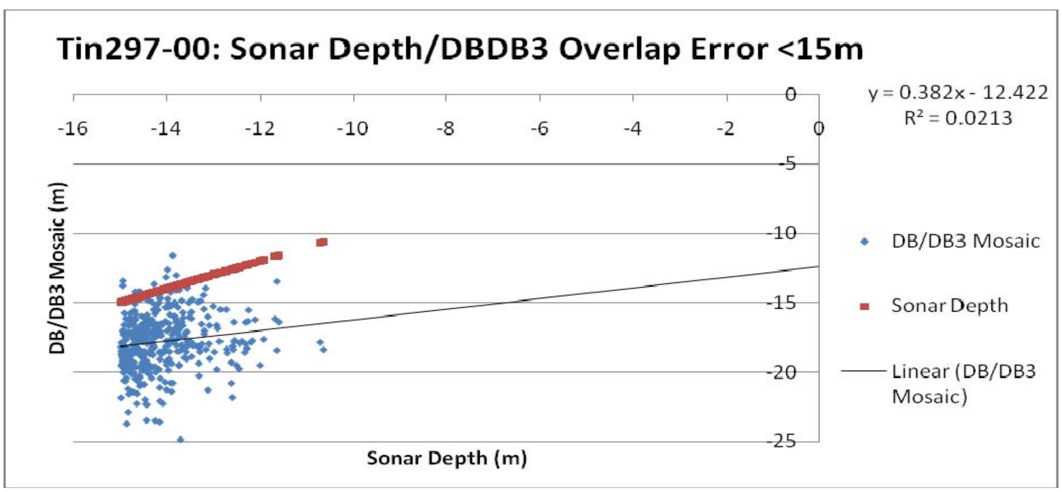

Figure 19. Comprehensive error analysis: mosaic of derived bathymetry products DB and DB3.

Given the results of these statistical analyses, the mosaic of derived bathymetry products DB and DB3 from image Tin297-00 is considered to be the best product for eventual integration with multibeam bathymetry. Though the statistical results of this product are less than satisfying, visual inspection of the product shows reasonable bathymetric figures so the product was included in the final mosaic. Its spatial coverage is very limited so that there is little impact to the accuracy of the final product.

Image Tin297-01

For the derived bathymetry products from the image Tin297-01 presented in Figures 20, 21 and 22, the R^2 values for the derived products are reduced significantly (from Part 1) due to a greater variability in departure from the mean value when all derived depth less than 15 m are considered. This is due to the variable sea state and cloud cover in this image adding to the inherent variability in the derived depth data when considering over 7,000 data points. However, notice that the slope value increases in each case when more points are considered. This indicates that the slope of the scatterplot trend line more closely matches the 1:1 relationship of the sonar depth plotted against itself, a better reflection of reality.

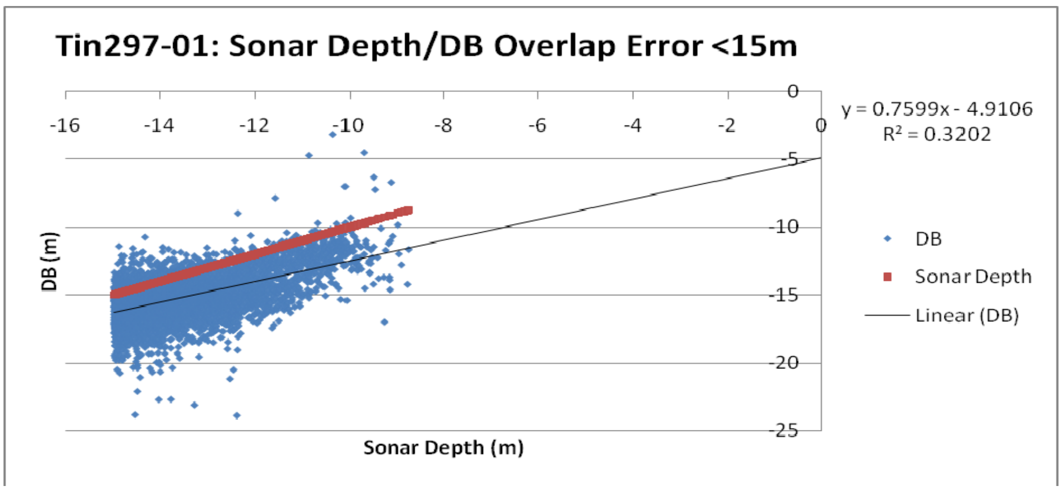

Figure 20. Comprehensive error analysis of derived bathymetry from Tin297-01 using original MLR variables**.** Formula applied: Depth = 5.1266 - 15.8184 * X_{blue} + 25.5864 * X_{green}

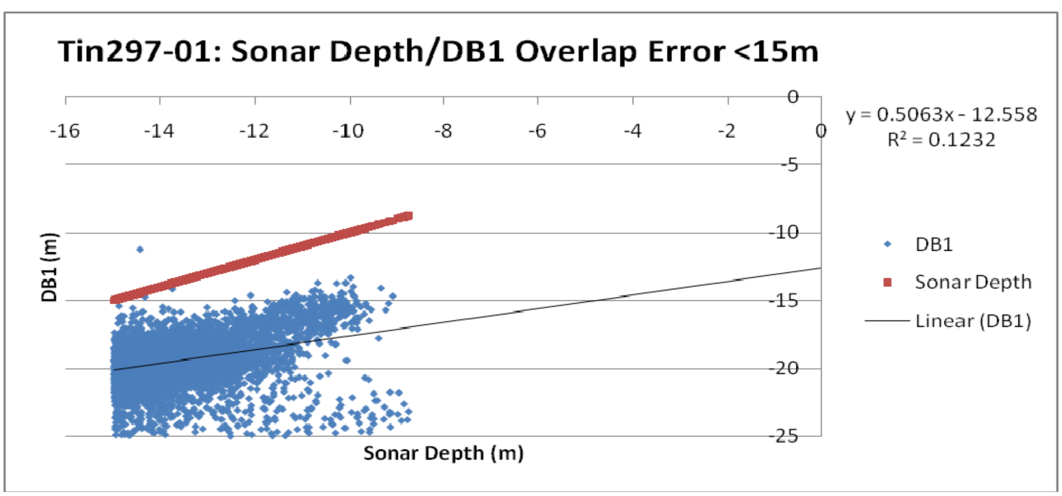

Figure 21. Comprehensive error analysis of derived bathymetry from Tin297-01 using decreased Y-intercept values**.** Formula applied: Depth = 1.1266 - 15.8184 * X_{blue} + 25.5864 * X_{green}

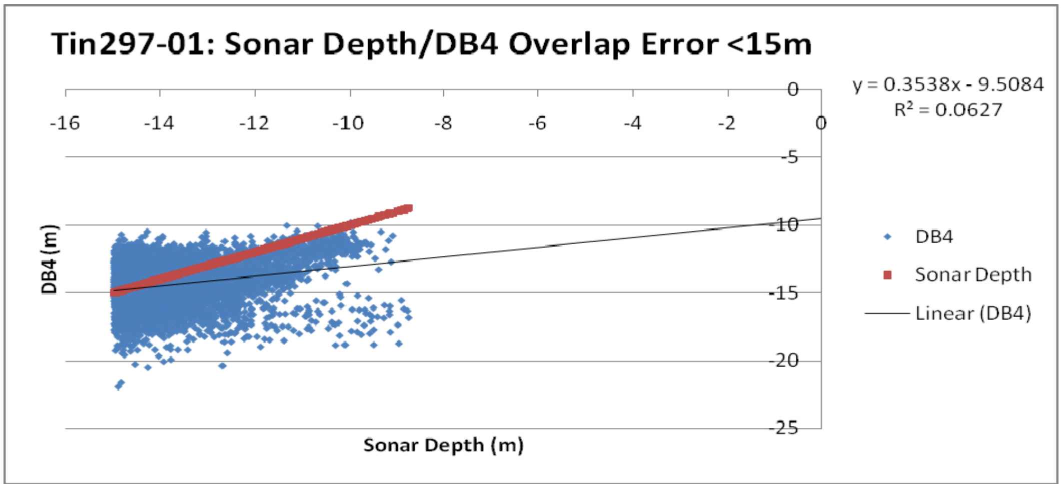

Figure 22. Comprehensive error analysis of derived bathymetry from Tin297-01 using decreased blue and green slope values**.** Formula applied: Depth = 5.1266 - 7.8184 * X_{blue} + 17.5864 * X_{green}

Both products that resulted from adjusted variables, DB1 and DB4, were mosiaced with the original product, DB, to expand spatial coverage in shallow areas.

Considering the derived bathymetry mosaic Tin297_DBDB1_mos (Figure 23), the statistical accuracy of the integrated product decreases from that of the original product (Figure 10) with decreased R^2 and Y-intercept values. However, the statistical accuracy of the derived bathymetry mosaic Tin297_DBDB4_mos (Figure 24) actually decreases to a greater degree.

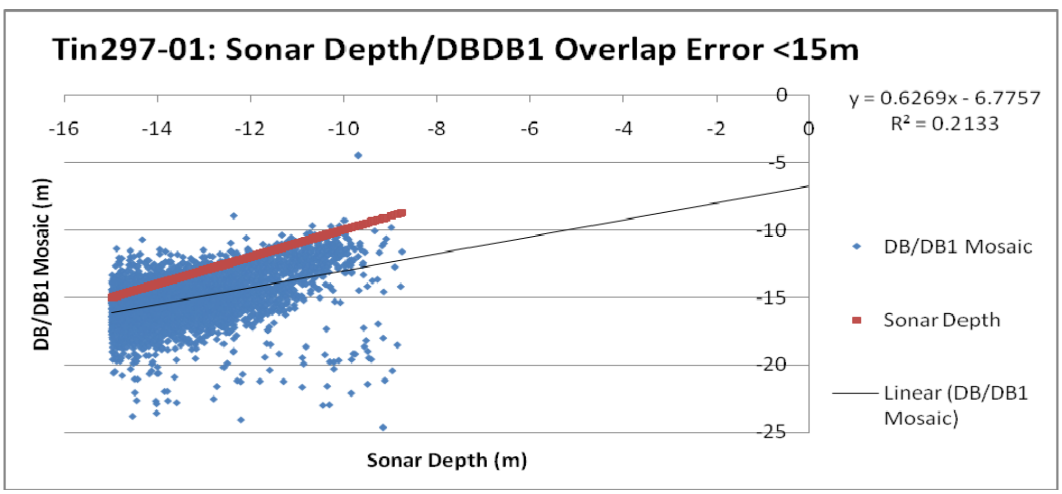

Figure 23. Comprehensive error analysis: mosaic of derived bathymetry products DB and DB1.

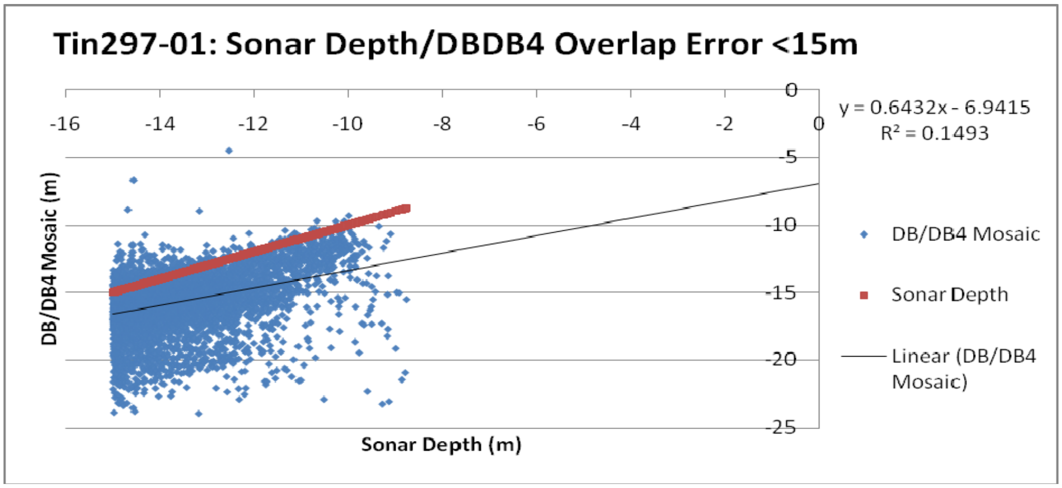

Figure 24. Comprehensive error analysis: mosaic of derived bathymetry products DB and DB4.

Given the results of these statistical analyses, the mosaic of derived bathymetry products DB and DB4 from image Tin233 is considered to be the best product for eventual integration with multibeam bathymetry.

Final Mosaics of Derived Bathymetry from all Images

In order to achieve the greatest spatial coverage possible, the derived bathymetry mosaics from each image were in turn mosaiced for the final integration. During integration first priority was given to the product data from image Tin233 (tin233dbdb5_mos) since it was the most statistically accurate and provided the most extensive coverage around the island. Second priority was given to the data product from Tin297-01 (tin29701dbdb1_mos) and third priority to the data product from Tin297-00 (tin29700dbdb3_mos).

The statistical analysis (Figure 25) and visual inspection of this integrated product demonstrates a significant degree of statistical validity and excellent representation of near shore terrain. Note that the statistics for this final product (tin_DBall_mos) indicate superior accuracy to any of the interim products that were derived from either original or adjusted MLR variables. This also holds true for any of the image specific mosaics.

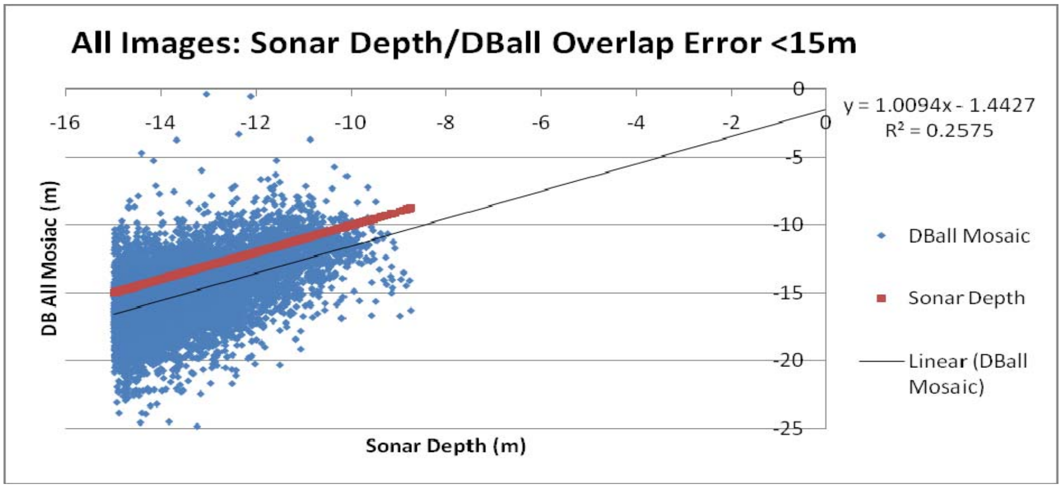

Figure 25. Comprehensive error analysis: integration of derived bathymetry mosaics from all three images.

Because the Tinian multibeam data did not cover shallows very well well for most of the island's perimeter, the final multibeam/derived bathymetry mosaic includes derived depths up to 25 m. Thus, the accuracy of derived depth up to 20 m and 25 m were assessed.

The final S-Plus statistics presented below include data first from all pixels where sonar bathymetry < 20 m and derived bathymetry concur (47,372 data points) and then where sonar bathymetry < 25 m and derived bathymetry concur (102,680 data points). An Excel plot is not included because the program cannot handle that much data.

S-Plus Statistics for Depths < 20 m

```
Residuals:
    Min     1Q  Median    3Q   Max
 -7.331 -1.471 -0.1837 1.301 8.088

Coefficients:
                 Value Std. Error   t value  Pr(>|t|)
 (Intercept)  -11.3516    0.0607  -187.0994    0.0000
DBall.Mosaic    0.3368    0.0036    94.4757    0.0000

Residual standard error: 1.923 on 47370 degrees of freedom
Multiple R-Squared: 0.1585
F-statistic: 8926 on 1 and 47370 degrees of freedom, the p-value is 0
```

S-Plus Statistics for Depths < 25 m

```
Residuals:
    Min     1Q  Median    3Q   Max
 -11.44 -2.327 -0.1619 2.259 10.59

Coefficients:
                Value Std. Error   t value  Pr(>|t|)
(Intercept)  -11.8491    0.0618  -191.8015    0.0000
      Depth    0.4598    0.0034   133.5626    0.0000

Residual standard error: 3.02 on 102678 degrees of freedom
Multiple R-Squared: 0.148
F-statistic: 17840 on 1 and 102678 degrees of freedom, the p-value is 0
```