Intermap Technologies
STAR-3i
Accuracies & Evaluation Strategies

Ian K. Isaacs, Director NEXTMap Programs
Intermap Technologies
Ottawa, Ontario Canada
iisaacs@intermaptechnologies.com

ABSTRACT

The underlying technology, instrumentation, and software that comprise Intermap Technologies proprietary interferometric synthetic aperture radar (IFSAR), known as STAR-3*i* ®, were upgraded in late 2001. The upgraded system became fully operational in 2002 and has since been actively employed in the United Kingdom, United States of America, South East Asia and Indonesia. The two chief reasons for the system upgrade were to achieve:

- higher absolute vertical accuracies,
- smaller image pixels (higher resolution)

Two accuracy evaluations, prior to the STAR-3*i* upgrade, are presented to provide a baseline for accuracy improvements achieved by Intermap Technologies. In brief these two independent studies analyze data that was acquired by the STAR-3*i* sensor at a nominal 20,000 feet above mean ground level. Two additional evaluations undertaken by two different agencies, from the public and private sectors, following the STAR-3*i* upgrade are further described and summarized. The first documents a horizontal accuracy evaluation, and the second a more comprehensive horizontal and vertical accuracy assessment. In both cases the data was acquired by the STAR-3*i* sensor at a nominal 30,000 feet above mean ground level.

This paper focuses on accuracy evaluation strategies and techniques employed both by Intermap clients and third party agents of these clients. Techniques are proposed for future IFSAR data accuracy evaluation, drawn from the experience of Intermap clients and third party agencies. Finally a brief summary of on-going Intermap validation of elevation data acquired under the NEXTMap USA Program is presented.

Introduction

The STAR-3*i* system upgrade was accomplished in approximately four months culminating in December 2001. Subsequent improvements to software and operating procedures occurred over a three-month period in the spring of 2002. Optimization of data processing software was carried out over a six-month period between March and August of 2002. Continual upgrades to data processing hardware are permitting ever-shorter data processing cycles; resulting in dramatic reductions in the total length of time between data acquisition and client receipt of both elevation and image data.

The following table provides a summary of the stated accuracies of digital elevation and radar image data derived from Intermap Technologies STAR-3*i* system.

Flight Elevation	Swath Width	Vertical Accuracies of 5m posted data	
		Prior to Upgrade	Post Upgrade
30,000 above ground	10 Km	2.0 m RMS	1.0 m RMS
20,000 above ground	6 Km	1.0 m RMS	0.5 m RMS
		Image Pixel Size	
Resolution altitude independent		2.5 m	1.25 m
Horizontal accuracy alt. independent		2.0 m RMS	2.0 m RMS

Table 1 STAR-3*i* Accuracies

This paper focuses on several third party evaluations that show that Intermap Technologies, has previously met, and is meeting these stated accuracies with respect to the provision of digital elevation models and ortho-rectified radar imagery. Finally evaluation strategies are presented to provide a guide for the meaningful horizontal and vertical evaluation of IFSAR data

Evaluation Projects pre sensor up-grade

The evaluation undertaken by two different public sector agencies are described and summarized. These summaries are presented to provide a baseline for accuracy improvements achieved by Intermap Technologies as the sensor has been upgraded. In brief the following two independent studies describe data that was acquired by the STAR-3i sensor at a nominal 20,000 feet above mean ground level.

Fort Knox Mine Site, Alaska

In October 1998 Intermap Technologies acquired approximately 260 km2 data in the vicinity of Fort. Knox Alaska. The Alaska Science Foundation contracted Intermap to supply IFSAR data to demonstrate to the business community in general in Alaska that IFSAR was a viable technology for the creation of elevation models and image maps in Alaska [Gansen, 1999]. Alaska has a short summer season with challenging weather conditions un-favorable for traditional photogrammetric mapping projects. IFSAR was presumed to offer both an innovative and economical solutions to these real problems. The purpose of this small project was to determine if IFSAR mapping in Alaska could be carried out to meet the general stated accuracies of Intermap Technologies.

Following the data acquisition and the supply of the digital elevation models and imagery, the client contracted for the survey of 34 ground control points in mixed terrain, ranging from valley bottoms to flatlands to hilltops, in conditions that contained a good deal of scrub brush or light scrub. Aeromap US of Anchorage Alaska was retained by the Alaska Science Foundation to carry out the ground surveys and report on the accuracies obtained. The following table summarizes the vertical accuracy of the IFSAR data as compared to high order GPS static surveys.

GPS Points	Mean error	Root Mean Square
34	0.5 meters	0.92 meters

Table 2 Fort Knox Vertical Accuracy

Red River, North Dakota

In September 1998 Intermap Technologies acquired approximately 900 km2 data in the flood zone of the Red River adjacent to the Canada USA international border. The purpose of this project was to determine how IFSAR and LiDAR data could be combined to better facilitate the mapping and resulting understanding of large areas of land prone to flooding. This work was undertaken following a catastrophic flood in 1997 that destroyed residential, commercial and agricultural properties over a large region of North Dakota and Minnesota USA, and Manitoba Canada [Damron, James J. et al 2000]. The International Joint Commission of the US and Canadian governments asked the U.S. Army Corps of Engineers for assistance with technologies that may better prepare flood management and mitigation agencies to both understand and delineate flood prone areas, and indicate where floods may affect current habitation.

Following the data acquisition and the supply of the digital elevation models and imagery, the Topographic Engineering Center, of the US Army Corps of Engineers surveyed approximately 1,800 ground control points, using a combination of kinematic and static GPS survey techniques. The following table summarizes the vertical accuracy of the IFSAR data as compared to higher order GPS kinematic and static surveys:

GPS Points	Mean error	Root Mean Square
1,800	0.09 meters	0.55 meters

Table 3 Red River Vertical Accuracy

Evaluation Projects post Up-grade

The evaluations undertaken by two different agencies, from the public and private sectors, are described and summarized. The first documents a horizontal accuracy evaluation, and the second a more comprehensive horizontal and vertical accuracy assessment. In both cases the data was acquired by the STAR-3i sensor at a nominal 30,000 feet above mean ground level. This means that the sensor was nominally 10,000 feet higher than the acquisition elevation for the first two studies presented.

Panhandle region, Texas

In April 2002 Intermap Technologies acquired data covering an area of approximately 15,900 square kilometers encompassing six rural counties in the ‘panhandle’ region of northern Texas. The primary purpose of this project was to create topologically structured vector files extracted from the radar imagery [List, John E. 2003]. The end products delivered were analyzed by the US Geological Survey (USGS) under funding by and in cooperation with the US Census Bureau (USCB) to determine the suitability of enhanced IFSAR data for feature extraction.

The U. S. Census Bureau requested that the USGS Rocky Mountain Mapping Center establish 110 GPS test points across a six-county area in the Texas Panhandle. The counties were Parmer, Castro, Swisher, Briscoe, Hall, and Childress. Census provided specific, detailed criteria regarding the distribution of the points and the desired survey accuracy of the test points.

The Census Bureau provided a digital file listing over 16,000 possible locations within the project area from which the USGS was to select the 110 specific test points. The Census Bureau divided the six-county area into 63 equal-size cells, each given a unique letter-number designation, (e.g., D3). At most a cell could have three points if extenuating circumstances arose. The USGS did not have access to the radar imagery or to the vector files extracted from this imagery prior to surveying the location of the 110 specific test points. However, the USGS actually surveyed the location of 113 evaluation points. The USGS reported that the final horizontal absolute accuracy of each of the 113 test points was better than 10cm. The key location factors that the USGS used to determine where to survey a test point were:

1. Point was to be located at the center of intersection of two well traveled roads,
2. Roads were to intersect at 90o. The intersection shape should form either a robust “T” or an equally robust “+”.
3. Each road should be named in the Census database and found on existing mapping.

Project area

The project is located in northwestern Texas where the terrain is generally undulating with sparse vegetation coverage.

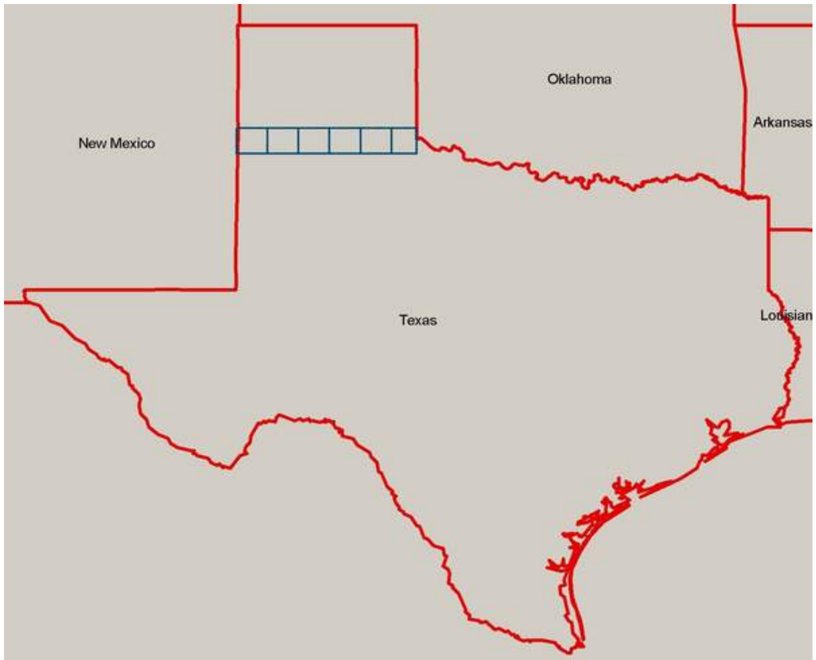


Figure 1 Texas Panhandle – 6 Counties

Vector extraction

The project deliverables were limited to two dimension Arcinfo formatted vector files and did not include the ortho-rectified radar image or the digital elevation model. Therefore, the horizontal accuracy assessment undertaken by the USGS was designed to prove the absolute accuracy of map worthy features, as extracted, and not the inherent absolute accuracy of a well-defined feature as observed in the imagery. There is a notable difference in this objective in that the final accuracy will also include some measure of interpretation of two un-defined points. The first undefined location is the 'center' of the intersection that is identified by the person surveying that general location and the second is that same general 'center' as defined by the person undertaking the map feature extraction. However, the final accuracy as represented by this process is very useful to anyone whom is using the radar image for traditional vector mapping. This accuracy will represent the typical horizontal location than can be expected for digital mapping using the radar image as the visual backdrop for feature extraction.

Horizontal Positional Accuracy

The USGS surveyed points, representing the center of road intersections, were known to be of higher accuracy than the coordinates extracted from the vector files derived from the radar imagery. These were overlaid graphically on the vector files. The node that was formed at the intersection of the two roads closest to this survey point was then saved to a text file and given the same identifier as the surveyed point. The following tables summarize the results of the comparison of the horizontal coordinates of the appropriate pairs of points. "Delta-x" is the difference between the "X" values of each pair of coordinates and “Delta-y” is the difference between the “Y” values. “Delta-r” is the linear distance from the surveyed road intersection to the mapped road intersection.

	Delta -x	Delta -y	Delta -r
Accuracy 95%			4.75
RMSE	1.89	1.99	2.74
Minimum	-5.77	-2.71	0.14
Maximum	5.91	6.90	7.78
Median	0.90	0.73	1.99
Average	0.80	0.84	2.33
Stdev	1.72	1.81	1.45
Count	111	111	111

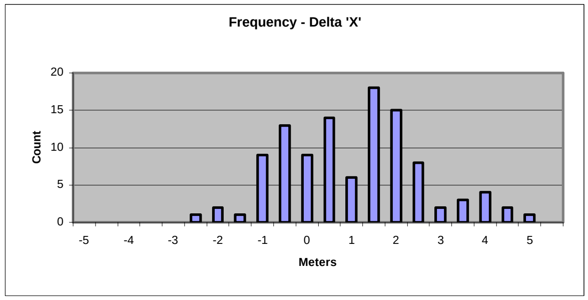
Table 4 Texas Horizontal Accuracy (i)

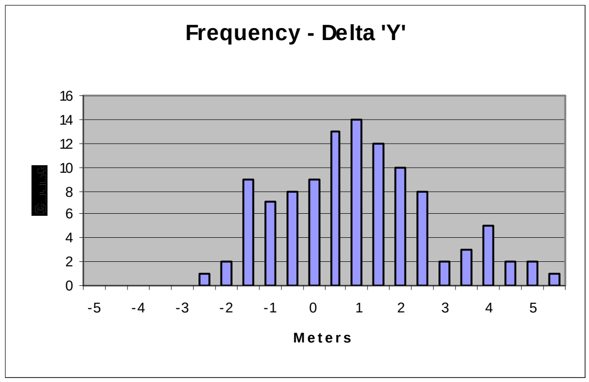
The reader will note that 111 of the 113 surveyed points have been used in the above assessment. Two points were removed because these two intersections were not represented in the final digital files due to the quality of the roads. The second table represents the final accuracy as three blunders were removed as the linear difference between the surveyed point and the mapped intersection were each greater than 3 times the standard deviation. This type of blunder removal is possible given the room for human interpretation in the position of the virtual center of the roads in question.

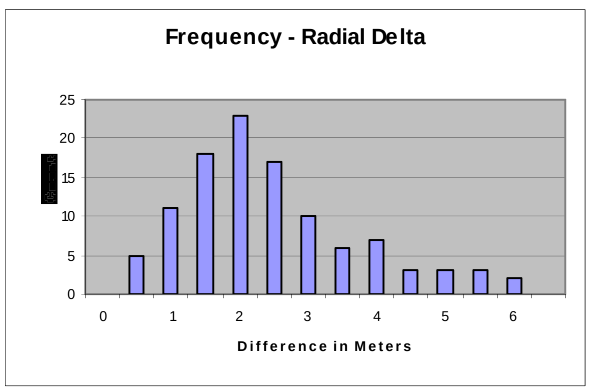
	Delta -x	Delta -y	Delta –r
Accuracy 95%			4.08
RMSE	1.71	1.88	2.54
Minimum	-2.99	-2.71	0.14
Maximum	4.74	5.60	5.83
Median	0.86	0.72	1.97
Average	0.79	0.77	2.21
Stdev	1.52	1.73	1.26
Count	108	108	108

Table 5 Final Texas Horizontal Accuracy

The RMSE error of 2.54 indicates that Intermap Technologies has exceeded the stated horizontal absolute accuracy 2.50 meter for the imagery and elevation model; as the RMSE above includes the sum of interpretations errors made by both the surveyor and map compilation staff. The following tables also show that the errors are distributed normally within the data. “Delta X”, “Delta Y” and “Delta Radial”, are as explained before.







NEXTMap Britain

In early 2002 Intermap Technologies Inc. commenced the acquisition of about 210,000 square kilometers of data covering England, Wales and southern Scotland. Data acquisition, of this area, was completed in the March of 2002. Subsequently the remaining data in Scotland was acquired in 2003. Data was acquired at both 30,000 and 20,000 feet above mean ground level. Approximately 160,000 square kilometers were acquired at 30,000 feet above mean ground elevation, with about 50,000 additional square kilometers being acquired at 20,000 feet above mean ground elevation. Digital data delivery commenced in the summer of 2002 and is now complete. The initial client, Norwich Union Insurance, engaged the University College of London (UCL) to undertake an evaluation of the data, as delivered, to determine if Intermap Technologies met the specifications. Subsequent evaluations by the Environment Agency of England and Wales are also underway as this Agency has also purchased a license to NEXTMap Britain data. A summary of both of these evaluations is presented herein.

Figure 2 NEXTMap Britain

University College of London (UCL)

The evaluation on the data acquired at 30,000 feet has been completed and the Executive Summary of the report states that horizontally … "*It was found that the IfSAR DSM and the ORRI fit well to Ordnance Survey Landline Plus data and to GPS check points. The data sets are free from blunders and inconsistencies. The positional accuracy of the IfSAR data was +/- 1.5m when using photogrammetric check points as a reference*." [Dowman, Ian, et al 2003] The Ordnance Survey Landline Plus data is horizontally accurate to 0.5 meters RMSE that is sufficient quality to evaluate the horizontal displacement of the IFSAR data, which is stated to be ±2.5 meters RMSE. Similarly the vertical evaluation was summarized as … "*The best accuracy of the IfSAR is obtained over an open field which is interpreted as bare earth, where a mean difference between the IfSAR DTM and aerial photography was -0.001m and the RMSE was ±0.172m (IfSAR higher). The IfSAR and Lidar surfaces are in good agreement. Over a cropped area the IfSAR DSM had a mean difference from the Lidar DSM of –0.608m and a RMSE of 0.770m. A comparison of the 2 DTMs gave corresponding figures of –0.376m and ±0.480m. Over the whole area the IfSAR DTM was –0.223m difference from the ALS (Lidar) with a RMSE of ±1.013m*." The LiDAR data was stated to be ±0.25m RMSE. The authors of this report did not provide raw measurements that could be summarized but clearly the report provides confidence that Intermap Technologies has met the stated requirements for the Type II product; that is absolute vertical accuracy of ±1.0m RMSE and absolute horizontal accuracy of ±2.5m RMSE.

The Environment Agency of England and Wales

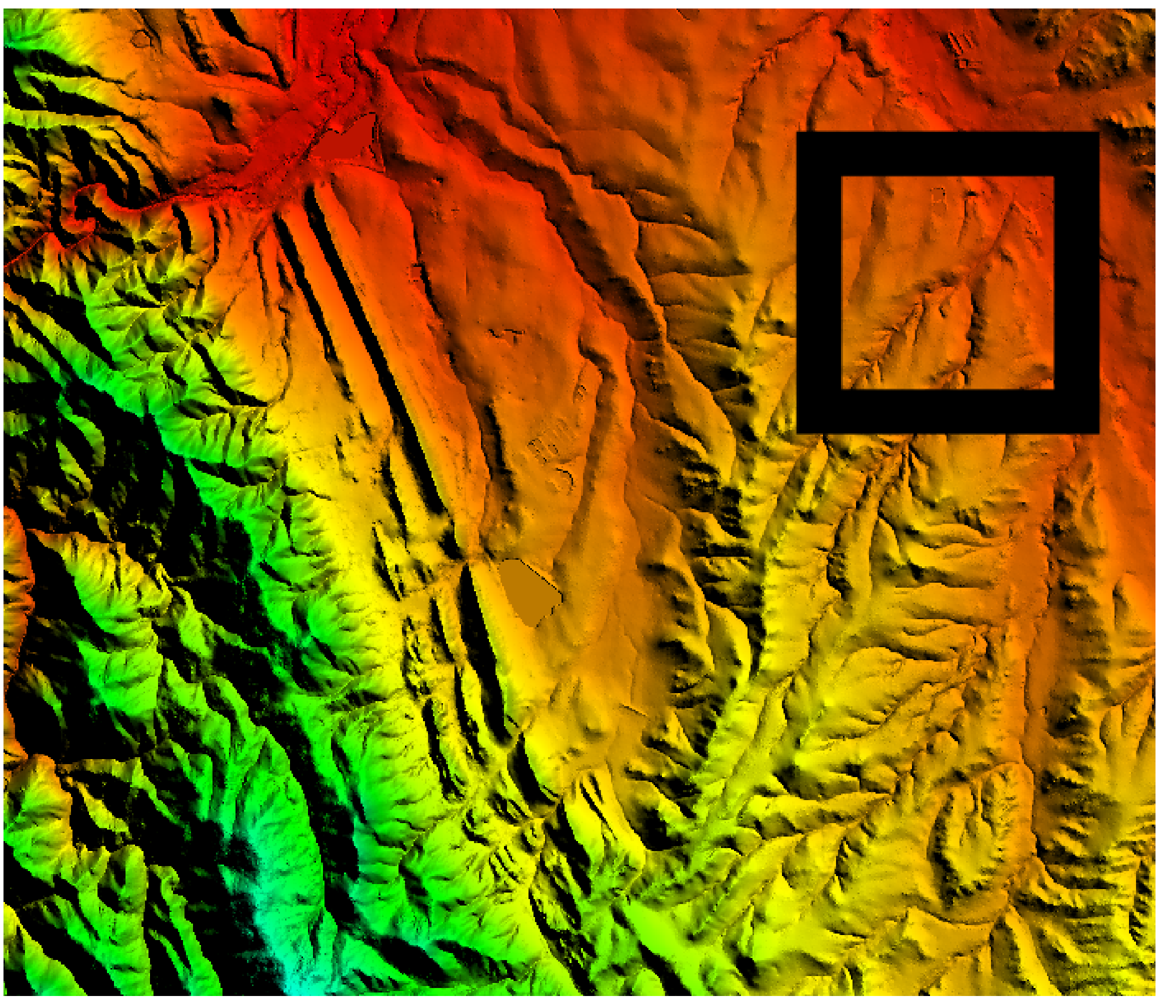
The evaluation of the vertical and horizontal accuracy is now virtually complete. This government agency has a considerable catalog of LiDAR and other survey data spread throughout England and Wales. An initial vertical accuracy assessment based upon 205 high accuracy (GPS) surveyed points over a wide area in central England indicated the following:

1 meter DEM	1st Surface DSM	Bare Earth DTM
Bias	**35 cm**	**2 cm**
RMS	**100 cm**	**37 cm**

Further evaluations relying on LiDAR are summarized below:

Evaluation method	Points	LiDAR Area	RMS
LiDAR	> 100,000,000	> 2,500 km2	90cm
Note: The bias was not provided. The ground area evaluated to date represents more than ? of the total NEXTMap Britain dataset.			

The approach taken by the Environment Agency is shown in the diagram below:



The NEXTMap elevation data set represented by the shaded relief is (10 x 10)km in extent while the black rectangle represents the LIDAR dataset used to evaluate the tile in question. The LIDAR dataset is approximately (2x2)km and can occur anywhere within the NEXTMap tile. The Environment Agency computes the elevation difference between the two intersecting areas and then reports the RMS statistic for the NEXTMap Britain tile in question. This process has been repeated on more than 630 NEXTMap Britain tiles and the total RMS has been reported as 90cm nationwide.

This evaluation by the Environment Agency of England and Wales is probably the most expansive undertaken by any client of Intermap Technologies. The authors have not provided raw measurements that could be summarized; however, this is the second detailed study that has once again confirmed that Intermap Technologies has met the stated requirements for the TYPE II product.

Intermap internal NEXTMap USA validation

Intermap commenced the acquisition of data under the recently announced NEXTMap USA Program in late 2003. The company has acquired more than 120,000 km2
of Type II data in disparate blocks in Californian, West Virginia and Mississippi. All of this data has been validated against high accuracy survey data extracted from the National Geodetic Surveys database. The following table summaries this validation:

State	Block area km2	NGS Pts.	Mean meters	RMS m
California	24,000		-0.29	**+/- 0.91**
West Virginia	39,000		0.07	**+/- 0.68**
Mississippi	40,000		0.22	**+/- 0.72**

The approach used in the above validation was to extract all NGS that fall in the respective areas, except those noted as being located upon abutments, bridge walls and so forth. The validation team has not culled any other points from the test set nor have they determined if some of the points fall in ground areas normally obscured from the view of the sensor.

Evaluation strategies

The mapping industry has always struggled with the quantification of spatially distributed error. The problem has two chief components:

1. How much sampling is required to infer adherence to a specification
2. How should the samples be spatially distributed.

The smaller the sample size and the more concentrated the spatial distribution of the sample points to the data being evaluated, the lower would be the confidence in the result. There are no specific rules that govern the sample size and distribution. The USGS states that a minimum of 24 “well distributed” points are required to infer statistical significance with respect to the evaluation of spatial data. The Census Bureau uses a minimum of 110 “well distributed” points for the same purpose. To date the most comprehensive testing in North America has been done by the US Army Corps of Engineers Topographic Engineering Center. They used a test range composed of six million photogrammetrically derived points. The nature of spatial sampling implies that the statistical measure references an area of sampling and assumes that within that area the error distribution is homogeneous under the particular rules of sampling. When mapping products are delivered over large areas the sampling area should be comparable to the delivery area.

The Root Mean Square Error calculation assumes a normal distribution of error and no bias. It simply states that ~67% (one standard deviation) of the evaluation points when subtracted from the target data points will yield a root mean square difference of one meter or better for Intermap Technologies TYPE II data. Conversely, ~33% of the samples will test outside of this envelope. When this computation is done on very large sample sizes the impact of outliers caused by the radar characteristics and blunders in evaluation/sample point measurement are both minimized and the computation will yield a value below one meter (for TYPE II). When testing is done on small sub-sets of the data one must be very careful to ensure that the test points are valid, free of blunders and are positioned to reflect the understanding of radar characteristics.

Radar Characteristics vis-à-vis Evaluation Points

A Digital Surface Model (DSM) represents the scattering surface that is observed by the radar. This may include buildings, and other structures as well as vegetation and bare ground. The Digital Terrain Model (DTM) however is 'extracted' from the DSM. Basically, an automated process samples what the algorithm perceives to be as bare ground and an interpolative surface can then fit to these points to create a regular raster representing the 'bare-earth'. An interactive editing process modifies the points thus collected so that the final DTM that is delivered has 'blunders' removed to the best judgment of the operator.

There are several pertinent issues that affect the DSM and thus the derived DTM. These are described in the Intermap Technologies Product Handbook. A few are noted here:

1. The radar integrates over a footprint that is approximately 5 meters square. Therefore, the DSM sample at that point will contain the effects of all the scattering objects within it. For example if it contains bare ground and a raised object such as a structure or tree, both will contribute to the sample elevation. Similarly if the sample is at the edge of a road, it may also contain the ditch at the side of the road. If the DSM sample is being compared with a GCP somewhere in the footprint, it may give an over-estimate or under-estimate of the elevation. Therefore it is important that the GCP be in a situation of unobstructed, modest and constant slope such as an open field or park.
2. The radar views to the side of the aircraft with local incident angles (in flat terrain) of about (45° +/- 10°). Therefore in the direction perpendicular to the flight path there are shadow effects behind tall structures and layover effects in front. This has two consequences in urban areas:

    i. These areas are sometimes void of data, unless filled with second look, and are interpolated. These interpolations may have associated errors.
    ii. In areas with narrow streets parallel to the flight line, the buildings may obscure the streets, so there may be little sampling of the bare earth in the street itself.

Evaluation Point Accuracy

The validation of an elevation model requires the identification and use of other higher accuracy elevation data that could be referred to as the "bench-mark" or "truth data set". Having obtained this "bench mark" it is imperative to know that the "bench mark" data is:

1. Un-equivocally higher in both vertical and horizontal positional accuracy
2. Identical in geo-location projection
3. Identical in vertical datum
4. Identical in elevation model posting spacing in the case of higher order grid elevation data

Once these four fundamental facts are ascertained then the comparison process will require the subtraction of the IFSAR elevation data from the "bench mark" data. In the case of grid data the grid subtraction will be valid for the ground areas that are modeled by both grids. In the case of high order point data, such as high accuracy survey points, then the subtraction will be valid for each discreet point. There are generally three ways to create vertical accuracy validation statistics for radar-derived DTMs. Each has advantages and disadvantages:

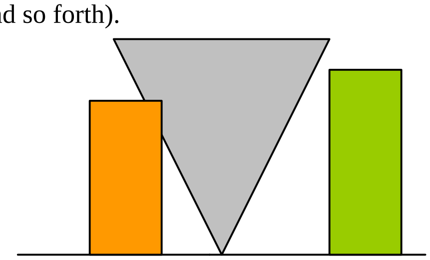
1. Individual Ground Control Points (GCP):
   The advantage is that these GCP are usually high precision points (accurate to a few cm) in (x,y,z), tied in to High Accuracy Reference Networks or other high-order benchmarks. The disadvantage is that they are often relatively few in number and may not represent the spatial variability of the subject elevation model over a range of conditions. For a large area with variable terrain and coverage, relatively large numbers of, well distributed points are required to ensure statistical validity. Moreover, as previously noted, GCP site selection is important and will be addressed later.

2. Transects:
   It is possible to attach a GPS antenna to a vehicle, and using kinematic differential processing to sample the roads over large areas quickly and relatively inexpensively. However the results are usually less accurate than point measurements and are subject to intermittent GPS outages due to obscuration from bridge overpasses, overhanging trees, etc. More importantly, if the vehicle is constrained to roads, there will be an interpretation problem related to roadside ditches, and radar-obscuration by adjacent buildings and trees.

3. High Accuracy DTM:
   If a LIDAR or other high accuracy DTM is available, with high sampling density and accuracy at the sub meter RMSE level, the comparative accuracy over larger, continuous areas can be obtained. The disadvantage is the expense. However it may be possible to obtain smaller sub-sets and use them to characterize the performance of the radar DTM in a variety of conditions, as was the approach taken by the Environment Agency of England and Wales. Several of the validation exercises conducted by Intermap have used LIDAR-derived DTMs and their associated point sets as comparative 'truth'. Of course these systems have their own errors and anomalies so care must still be exercised.

Evaluation Point Location

The main rule is that each GCP site should consist of unobstructed, level, bare terrain. A rough guide would be that the site should be clear of objects within a circle of radius dependant upon the height of features surrounding the desired GCP location. The target within this circle should be flat or of uniform slope (less than 20°). An additional constraint is that large aggregations of buildings or wooded areas in the vicinity (i.e. where the radar is unable to sample the ground) will create edge effects and should be avoided. A useful 'rule of thumb' is … a suitable GCP should be located within the center of a clear circle where the horizontal distance from the GCP to the vertical obstacle is at least twice the height of the obstacle (tree, tower, building, cliff and so forth).



In the above diagram the "V" shape actually represent a cone between two buildings. The edge of this cone represents the nominal depression angle of the radar. Therefore any building or vertical obstruction that intersects this cone edge will result in perturbations to the digital elevation adjacent to this feature. Any point that is placed in partially obscured areas or is placed in sloping terrain, with slopes greater than 20o, could yield elevation differences of several meters thus rendering that point useless towards the overall evaluation of the data. The choice of location must be consistent with the local conditions at the time of the radar data collection. For example, if the area being evaluated contains rotational crops it is imperative that there be no appreciable crop growth during the time of the radar data acquisition. Transient features can also affect the elevation model cars or trucks on roads often change the elevation value of the road at that point. Therefore while roads are often seen as good physical locations due to the lack of obscuring features these could have been present during data acquisition and the use of these areas to support GCP's need to be evaluated carefully. Care must be taken to avoid grouping these ground control points in any one area thereby inferring wide area accuracies that are only relevant to that one area of interest.

Given that the IFSAR is a side looking sensor one must also be careful to avoid testing in areas of foreshortening or shadow since these areas are interpolated within the DSM and the DTM is derived from this interpolated data. Similarly, in areas of very high or very low signal return, (areas of water and areas of saturation due to concentrations of buildings acting as corner reflectors) the DSM can be interpolated.

Conclusions

Intermap Technologies has consistently provided digital elevation models and ortho-rectified radar imagery that meets or exceeds the specifications that describe the respective products. Several notable government and research agencies have rigorously evaluated this data. These evaluations have confirmed that Intermap Technologies data has often exceeded the specifications. In all four of these evaluations, both prior to and following the radar sensor upgrade, the final data has been shown to meet the specifications quoted by Intermap Technologies.

REFERENCES

Gansen, Henry [1999] "STAR-3i Data Acquisition & Data Processing Final Report Fort Knox Mine, Site Alaska." Intermap Technologies

Damron, James J. and Daniel Carlton [2000] "Evaluating IFSAR and LIDAR Technologies Using ArcInfo: Red River Pilot Study." US Army Topographical Engineering Center.

List, John E. [2003] "Texas 6-County Census Bureau Project Parmer, Castro, Swisher, Briscoe, Hall, and Childress Counties Final Field Report". USGS Rocky Mountain Center Denver Co.

Dowman, Ian and Fischer, Peggy [2003] "Validation of IfSAR Digital Elevation Models from Intermap Technologies" University College of London.