Modern Applied Science
www.ccsenet.org/journal.html
Vol. 3, No. 1
January 2009

Investigation of Differences of Topographical Map and GIS-derived Spatial Map with Actual Ground Data in Peninsular Malaysia

Mohd Hasmadi Ismail (Corresponding author)
Survey and Engineering Laboratory, Department of Forest Production
Faculty of Forestry, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
Tel: 60-3-8946-7220, Fax: 603-89432514
E-mail: mhasmadi@putra.upm.edu.my

Taylor, J.C.
Cranfield Centre for Geographical Information Management
National Soil Resources Institute, Cranfield University, Silsoe, Bedford
MK45 4DT, United Kingdom
Tel: 44-1525 863064
E-mail: j.c.taylor@cranfield.ac.uk

Abstract

In a geographical information system (GIS), digital maps usually used to show multiple views of geographical objects either through two-dimensional or three-dimensional, which topographical parameters are digitally generated. Digital maps are often used in extensively environmental application without quantifying the effect of their errors. This study was carried out to investigate the difference of elevation and slope of topographical map and GIS-derived spatial map with actual ground data. The analyses of differences were quantified from interpolation process, sampling and measurement in the field. The RMSE of the DEM creation for the test site was 0.62. The result was based on the 10 m DEM resolutions and 20 m contour interval. From the analysis of differences (elevation and slope) of topographical map and actual ground data, it's showed that the difference is only about 2 % and 28%, respectively. The great differences on slope may be due to error during data collection by different enumerators and also inconsistent reading of slope measurement and target. Despite the difficulty occurs during ground data collection, estimation method was applied and this relatively simple procedure but appears acceptable in regard to sufficient data sets at nominal map scale 1:50000.

Keywords: Peninsular Malaysia, DEM analysis, Validation, Elevation and slope differences

1. Introduction

A fundamental for natural resources survey whether at local, regional or global scales is availability of map with high quality, reliable and up to date information. An additional consideration is the increasing use of satellite remote sensing data combine with global positioning system (GPS) in natural resources surveys Kardoulas et al. (1996). Topographical maps at any scale are beneficial to be used as a baseline data for designing ground survey (Thomas et al., 2000). These maps are usually used to determine the range of an area, shows a comprehensive information and logistic about certain area. It is also represents the earth surface as digital elevation data (DEM) that has been specified mathematically and is illustrated upon identified scale. Most usually, a map is a two-dimensional, geometrically accurate representation of a three-dimensional space. Previously, people used the topographical maps to know the place by looking at the geographic feature in the paper map. With the advancement of technologies, the topographical map was digitized into GIS system and they called GIS map or digital map. In the beginning era of digital maps there were only exclusive use of digital data, and the data were assumed to ready precisely. However, all maps are always inherent errors that

constituent uncertainty. According to Helen (2003) uncertainty need not be excised as a flaw, but needs to managed and accepted as an intrinsic part of complex knowledge.

Digital maps data are often used in analyses without quantifying the effects of these errors. In addition, elevation data in a particular grid cell is often based on sample elevation points. If the sampling scheme is inefficient, the resulting grid may be biased. It is a practical impossibility to obtain information on the exact source and amount of error in a particular digital topographical map. The underlying assumption of this research is that specific error within digital topographical map cannot be known and therefore the map is remaining question with uncertainty. Many digital topographical maps are generated from paper topographic maps. The grid location on the ground may vary and most likely will not correspond directly with the co-located value in a digital map. However, issues on the data errors or uncertainty are become a critical in GIS especially when users are taking this thing as an important key in decision making. According to Blakemore (1985) the main concern is being expressed within the geographic information science to effectively deal with uncertainty and manage the quality of information outputs. The basic scientific requirement of being able to describe how close their information is to the truth it represents. Meanwhile Stoms (1987) discussed knowledge-based approaches which employ various methods of reasoning under uncertainty for specific applications. Many studies has been carried out for calculating topographical parameter related to DEM, and each method were produced with different results (Skidmore, 1989; Ryder and Voyadgis, 1996; Suzanne and Chales, 2006; Qiming et al., 2006). Thus, more research needs to be done in this area to verify the uncertainty of topographical used in many environmental and resources analysis. Therefore, the intent of this study is to investigate the differences of digital topographical map in geographical information system (GIS) with reference to elevation and slope compared to actual ground survey data.

2. Methodology

2.1 Quantifying of DEM accuracy

The topographical map was acquired from the Department of Survey and Mapping Malaysia. The map is scaled 1: 50000, based on aerial photograph/photogrammetric restitution (Figure 1). The selected test site is at the northeast Jerantut region in Pahang State, Peninsular Malaysia. The topographical isolines contours are 20 m intervals and were digitised into GIS. The contour lines were assigned an attribute value according to their height in meters above sea level. The resulting dataset was then used to produce a DEM using Arc View software with the 3D extension analyst. Height value were added to the existing contour line used previously in generating the DEM. Adding the height information to the contour lines was the most time consuming stage of the process to generate a DEM.

The methodology to quantify DEM accuracy is the Root Mean Square Error (RMSE). It measures the dispersion of the frequency distribution of deviations between the known or measured elevation and the interpolated elevation (DEM surface). Based on the study by Gao (1997), the accuracy of a raster DEM is related to the contour density and the DEM resolution was derived as follows:

$$\text{DEM accuracy [RMSE (m)]} = \frac{(7.274 + 1.666\ S)\ D}{1000 + \varepsilon}$$

where, S stand for resolution in meters; D stands for contour density expressed as km km^{-2}; ε is an error term related to D. Contour density was calculated by dividing the total length of contour by the size of the study area. The accuracy of terrain representation was evaluated against root-mean-square-error (RMSE) of elevation residual at 50 selected check points, mathematically expressed as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Zi - zi)^2}$$

where n is number of check points; *zi* stands for the interpolated elevation; *Zi* is the elevation read directly from measured elevation at the same position. In particular, the RMSE for the study area was based on the 10 DEM resolutions and 20 m contour interval (m). The contour density in this area is 25.82km km^{-2} , thus the RMSE of the DEM creation for the study area is 0.62.

2.2 Validation of the slope and elevation

Ground surfaces potentially have an infinite number of points that can be measured. However, it is impossible to record every point; consequently a sampling method must be used to extract representative points to build an elevation model that approximates the actual surface. The choice of data collection strategy and techniques is critical for the quality of the results. Field data should represent adequate information in the modelling. Since field data tend to be very important, the survey of terrain surface characteristics is adapted to provide accurate information about the slope and elevation. However, as this particular collection technique is relatively time consuming, the limited small sampling area is used as a representation of a whole study area.

Slope and elevation samples were collected from two different approaches, (i) On road tracking using GPS and (ii) Off road sampling. On selected site, a five linear transect was designed in the forest with the distance interval of 50 m for each transect. In the second method by GPS tracking, two existing forest roads were taken as a sample track. To test the "fitness" of the two data sets in representing the ground surface a regression analysis was performed. A linear regression technique was used to regress both ground slope and elevation data against the corresponding slope and elevation and from the contour spatial map. Correlation and regression analyses will be performed in order to correlate the “fitness” of slope and elevation obtained from digital elevation model and field data for the entire test site. Measured and estimated data values were considered as x and y data points respectively. If there was 1:1 correspondence between x and y values, the regression coefficient and the intercept of regression line in y-axis would have the following equation:

$y = a + bx$

where,

y = estimated data (from map)

x = measured data in field

A test of significance can be made by comparing the ideal data of a= 0.0 and b=1.0 with the best-fit and chosen significance limit values of a and b from the fitted linear regression. The letter r is the correlation coefficient that when it approaches value 1.0, it gives an indication of close association between two data points. A total of 108 sample measurements were collected for calibration and validation. The calibration data set was used to establish the coefficient in the regression equation. The predictive ability of the calibration equation was assessed using the coefficient of multiple determinations (r^2).

2.3 On road tracking using GPS

The GPS was placed on the vehicle while the tracking was carried out. On the road where a slope break or ‘breakline’ is observed (breaklines are lines in the topography where grade changes exist, such as tops and toes of slopes), measurements were taken to estimate the ‘original’ of the slope and elevation due to cut and fill work during the road construction. Figure 2 illustrated the example of road cross section, cut and fill and slope breakline profile, while Figure 3 illustrated the sample track.

Geo Explorer 3 is functioning as rover was used to collect road data. The data base was created by Data Dictionary Editor was opened in Geo Explorer 3. After all data were collected, the data was first transferred into a computer. Several pieces of equipment were required such as cable from GeoExplorer 3 to desktop PC and platform to hold the GPS. The transferred data were then opened and displayed in the Pathfinder Office software. This software provides a step-by-step guide to the data transferred. The data has been saved as a Pathfinder Office format (*.ssf) format. The differential correction was implemented to correct the data. The correction was performed using the base reading, which was obtained from the reference map. The coordinated system was set up and the value of latitude and longitude were recorded. The extension of a base format created in .dat. format. The procedure to transfer the data from GPS into computer was the same as transferring the field data into computer. The final file of the road maps was converted to the .shp format to allow this data to be displayed in the Arc View GIS.

2.4 Off-road sampling

Line transect sampling describes a class of methods and estimations which have been developed (and mostly applied) in environmental surveys to estimate population size. The best known line transect method is known as distance sampling. The methods are specifically intended to deal with the problem of undercount or incomplete detection, which is widespread in environmental surveys but also occurs in other surveys. The method used for this survey is the line-transect sampling method: Five lines transect 600 m long, carried out in a deep forest and the elevation and slope measured within a constant interval of 60 meter. The point sample of slope, elevation and distance were recorded at every interval. Layout of the slope and elevation sampling is shown in Figure 4. Altogether were five transects or 3000 meters.

3. Results and discussions

3.1 Evaluation of DEM and slope

Most of the elevation data in this region of study range from 400 m to 800 m. There is very little high area that is more than 900 meters. In fact, a DEM model that generated at the 20-meter contour interval revealed that a ground surface from this contour interval is suitable for road planning purposes. Evaluation of elevation and slope was made by the regression analysis of acquired data from field and GIS spatial maps. The good result and success of the evaluation depended upon the size, location of sample collected and consistency in data recorded by enumerators. From the data collected for off road and on road sampling, the coefficients of determination for regression analysis are presented in the next section.

3.2 Evaluation on off-road sampling.

The graphs show that r^2 for slope and elevation range from 0.5047 to 0.9871(slope) and 0.0327 to 0.9816(elevation). From the total of the samples (Transect 1-5) in Figure 5, it reveals that elevation and slope data that are created in the GIS as a spatial map are more accurate compared to slope maps, whereas r^2 of elevation is 0.9547 and slope is only 0.6813. Therefore, it is important to note that the final results of spatial map produced have their limitations or random errors. Random error can result from mistakes such as inaccurate validation survey or improper recording of slope and elevation data will remain.

3.3 Evaluation of on road tracking by GPS

Slope and elevation measurements which were taken from existing forest roads with the aid of GPS revealed improvement in the regression analysis (Figure 6). The r^2 for slope data were 0.6296 and 0.8486, while r^2 for elevation data was 0.997 and 0.9889, respectively. Meanwhile, the regression analysis of all data (Road 1 and 2) revealed a higher result of r^2 with 0.7462(slope) and 0.9928 (elevation). The result of linear regression showed that the elevation and slope created in the GIS from topographical map and data recorded from the field from off-road and on road were correlated with the topographical map.

Figure 7a and Figure 7b presents a plot of this regression analysis from ground data and estimated data from the spatial map of both methods. This graph representing regression analysis shows the relationship between slope and elevation in the ground as represented on the X-axis, and slope and elevation that derives from spatial information map in GIS on the Y-axis. Over 220 sample locations were used and measured in the analysis that contained the length of 9139 m. The coefficient of determination for all collected data from both methods in this analysis is 0.7207(slope) and 0.9889(elevation), significant at the .05 levels. This result implies that approximately 72% of the appraisal data of the slope and 98% of the elevation data were attribute to variations on the ground surfaces and spatial data that are present in spatial information.

The scattergram showed that most of the slope measurements were distributed and far away from the regression line. Those patterns are probably due to the error of data collection by different enumerators and also a different shoot of slope target during the field survey. It can be stated that the regression is only efficient if the coefficient of determination is high, which suggests that small errors occurred during the data collection and the regression relationship must be linear with the equation slope around one.

4. Conclusions

This paper addressed the DEM accuracy and differences of digital topographical map derived from GIS with actual ground data with emphasised on elevation and slope parameters. As a result, the contour density in this area is 25.82km km^{-2} , thus the RMSE of the DEM accuracy creation for the test site is 0.62. Greater difference are exist in slope than elevation but less or similar in the flat surface. The differences of both methods are less than 30 % for slope ($r^2 = 0.72$); and less than 5 % for elevation ($r^2 = 0.98$). The great differences on slope were due to error of data collection by different enumerators and also inconsistent reading of slope measurement and target during field survey. It can be stated that the regression is only efficient if the coefficient of determination is high, which suggests that small errors occurred during the data collection. It is recommended that, future studies should be carried out using difference sampling design and statistical approaches. Comparison studies on difference map scale (e.g. 1:10000, 1:20000 and 1:50000) are also recommended in order to reveal the uncertainty and its effect on topographical parameters.

5. Acknowledgements

This work was carried out under collaboration project between Faculty of Forestry University Putra Malaysia (UPM) and Pahang State Forestry Department. This research article is a part of the Ph.D thesis completed in the Cranfield University, Silsoe at the United Kingdom. To Dr. Graham Thomas and Mr. Tim Brewer at Cranfield University thank for giving an invaluable discussion in this work. The authors are also grateful for the comments provided by the referees that helped to enhance this article. Above all, I wish to thank all Forest Rangers and Foresters who are involves in this work, especially during field work in the forest.

References

Blakemore M. (1985). High or low resolution? conflicts of accuracy, cost, quality and application in computer mapping. *Computers & Geosciences,* 11. 345-348.

Gao J. (1997). Resolution and accuracy of terrain representation by grid DEM at a microscale. *International Journal of Geographic Information Systems,* 11. 199-212.

Helen C. (2003). The certainty of uncertainty: GIS and the limits of geographic knowledge. *Transaction GIS,* 7. 165-175.

Kardoulas N G Bird A C and Lowan A L. (1996). Geometric correction of Spot and Landsat imagery: A comparison of map and GPS derived control points. *Photogrammetry Engineering & Remote Sensing,* 62. 1173-77.

Qiming Z Xuejun L and Yizhong S. (2006). Terrain complexity and uncertainties in grid-based digital terrain analysis. *International Journal of GIS,* 20. 1137-1147.

Ryder W and Voyadgis D. (1996). Measuring the performance of algorithms for generating ground slope *Proceeding at the 2nd International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences,* 21-23, May, 1996. Fort Collins Colorado U S A pp 208-216.

Skidmore A. (1989). A comparison of technique for calculating gradient and aspect from a gridded digital elevation model. *International Journal of GIS,* 3. 323-334.

Stoms D M Davis F W and Cogan C B, (1992). Sensitivity of wildlife habitat models to uncertainties in GIS data. *Photogrammetric Engineering & Remote Sensing,* 58. 843-850.

Suzanne P W and Chales N K. (2006). Quantifying DEM uncertainty and its effect on topographic parameter. *Photogrammetric Engineering & Remote Sensing,* 72. 1081-1090.

Thomas G Cristophe S and Taylor J C. (2000). Mapping systems and GIS: A case study using the Ghana national grid. *The Geographical Journal,* 166. 306-311.

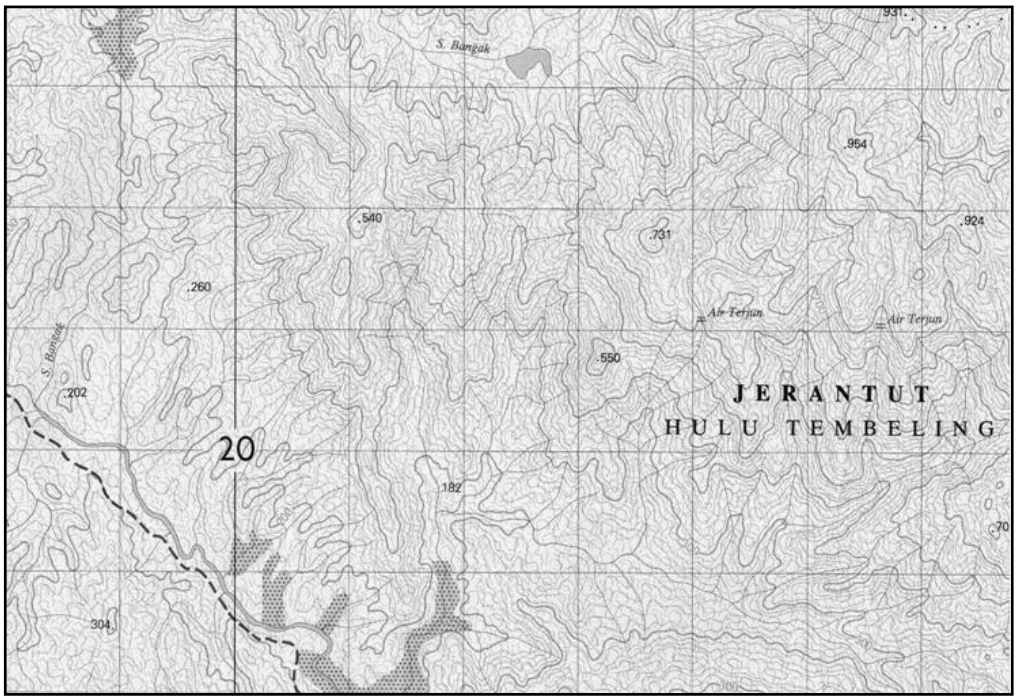


Figure 1. A 1:50000 scale topographical map of the study area

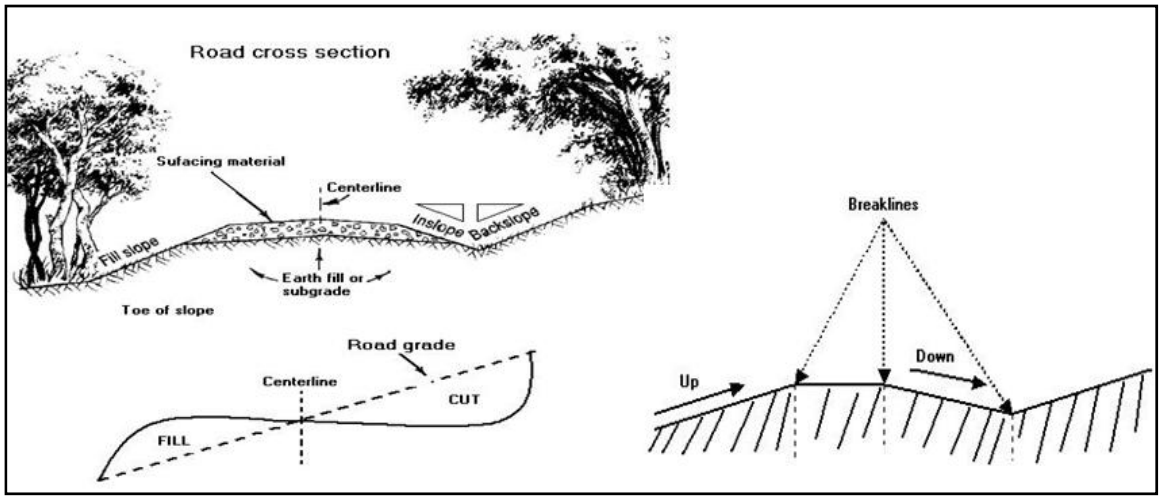


Figure 2. Illustration of road cross section, cut and fill and slope breakline profile.

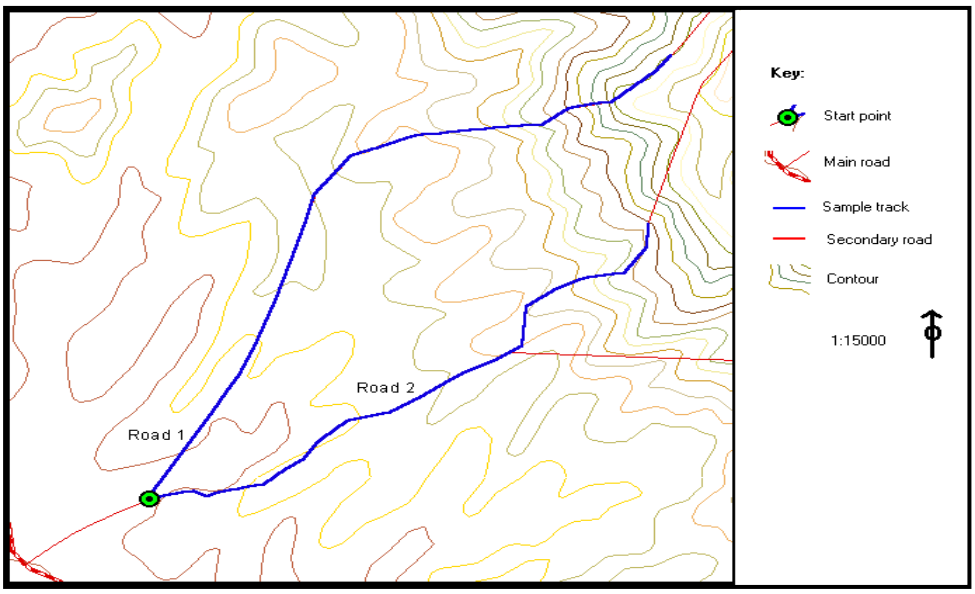


Figure 3. The selected sample tracks in the study area

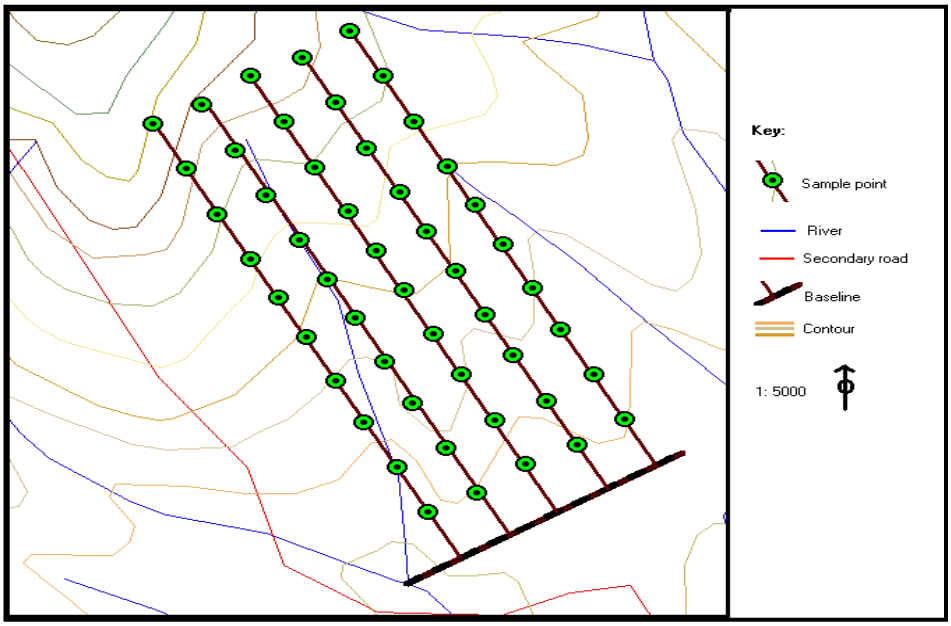


Figure 4. Layout sampling site for off road survey (elevation and slope)

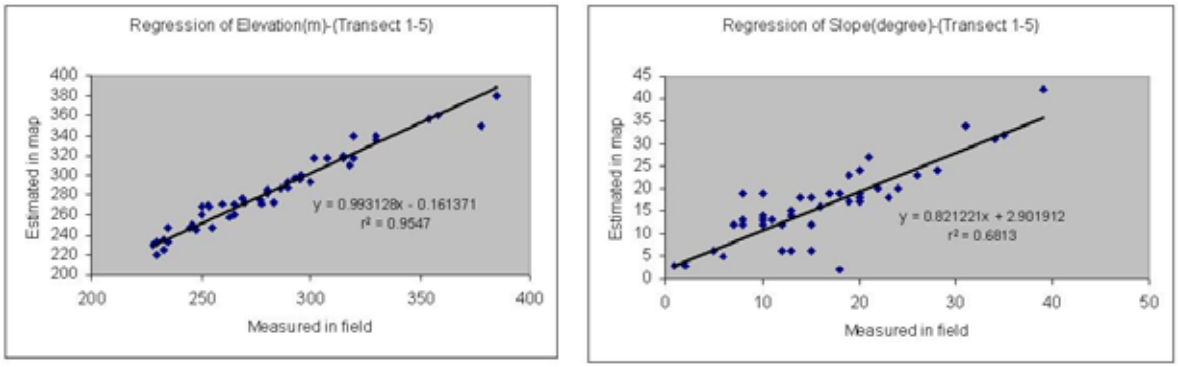


Figure 5. Relationship of measured and estimated elevation and slope from all transect data.

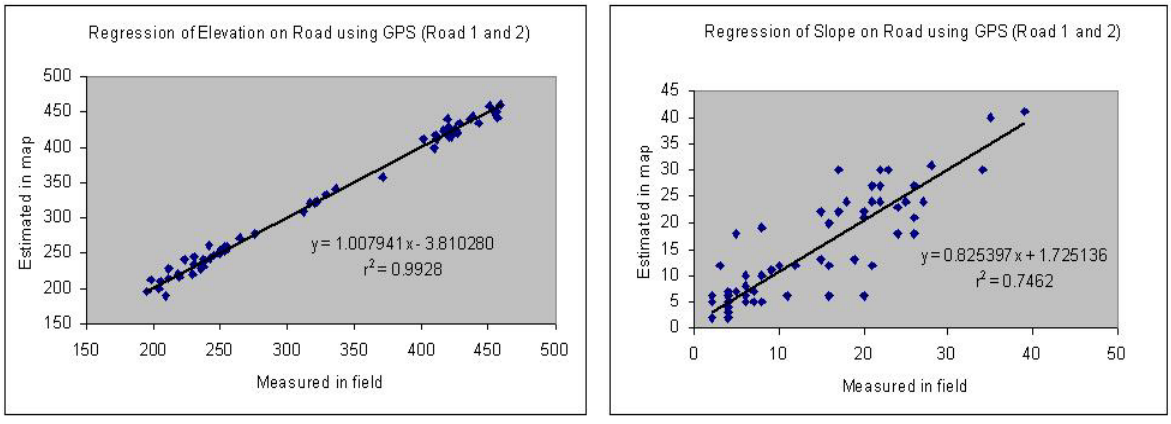


Figure 6. Relationship of measured and estimated elevation and slope from all sampling data.

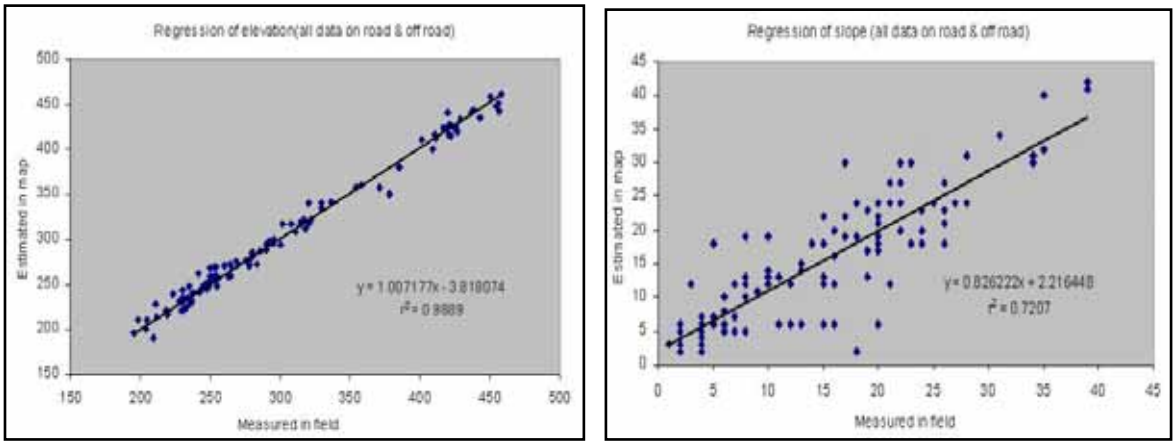


Figure 7. A regression plot derived from the slope analysis of both methods.