DATA INTEGRATION AND ACCURACY ISSUES IN DIGITAL TOPOGRAPHIC DATABASES

Dr. R. Siva Kumar
Survey of India
Siva_kumar@vsnl.com

Working Group IV/III.2

KEY WORDS: Accuracy, Error propagation, Feature extraction, GIS, Integration, Photogrammetry

ABSTRACT

There are certain problems of matching the data when the database is obtained from many sources. One of the major issues is the geodetic datum, which is different depending upon the source. In addition, maps on different projections necessitate conversion of the data. In the absence of accurate transformation parameters, the integration of databases does suffer. The absolute accuracy is poorer though the relative accuracy of objects extracted from satellite images was comparatively higher. Building up of a system of data bases, which may be integrated vertically and accessed by many customers, has accuracy implications. The vertical integration of land information assumes that the function, scale, projection and spatial referencing systems used by all data bases may be integrated. As it is not an ideal world, spatial data bases are neither static nor derived from the source material of the same scale. Relationships between all these factors must be determined for successful integration of spatial databases. Accuracy is also temporal in nature. It should be possible to query the data base and find out the quality of each and every object.

1 INTRODUCTION

The basic input data for the National Topographic Data Base for most of the developing, and even developed nations is the topographical map cover. In countries like India the maps are based on Everest Ellipsoid whereas the modern era satellite imagery is available on the World Geodetic System 1984 (WGS-84). Another associated problem is that all the topographical maps of India are on the polyconic projection whereas the satellite imagery is on Universal Transverse Mercator (UTM) projection. Conversion programs do exist but are not accurate enough as the transformation parameters needed are not available till the national datum is redefined. Data matching is a problem when the data base is derived from many sources. Thus the integration of databases does suffer on this account besides leading to other issues such as seamless contiguity. Thus the absolute accuracy is poorer though the relative accuracy of objects extracted from satellite images was comparatively higher.

A multi-source database is a multi-layered view of the world which is easily accessible for management and decision making. Building up of a system of data bases, which may be integrated vertically and accessed by many customers, involves the accuracy implications. The vertical integration of land information assumes that the function, scale, projection and spatial referencing systems used by all data bases may be integrated. As it is not an ideal world, spatial data bases are neither static nor derived from the source material of the same scale. These are neither on the same projection, nor based on the same geodetic datum; and also are not all setup for same function. Relationships between all these factors must be determined for successful integration of spatial databases.

1.1 Indian Geodetic Datum

The local geodetic datums are the reference ellipsoids adopted by a country to fit a particular region of the earth in the best possible manner, and hence, these are generally not geocentric. Such local geodetic datums are established through an integrated adjustment of the geodetic network in the country and their relationship with geocentric datums is established through a seven parameter transformation of three translations for shift of origin, three rotations about the three axes of the rectangular Cartesian coordinate system and one scale factor. To these datums the space geodetic data such as Global Positioning System (GPS) data is referred. According to Kulkarni (1996), the importance of establishing a local geodetic datum to the desired level of accuracy is obvious from the fact that the accuracy of the entire geodetic data depends directly on the accuracy of the datum to which the data refers.

One of the major problems encountered is the geodetic datum of data. The Indian Geodetic Datum- Everest Ellipsoid- was established in 1880 through an adjustment of the Geodetic network, computed using the models, techniques and methods then existing in practice at that time. The geodetic network in India was divided into five parts, four quadrilaterals and the Southern Trigon.

1.2 Accuracy Issues

The accuracy of data in a GIS is prone to errors from many sources. Right from the stage of aerial photography till the data is stored it is burdened with creeping errors. Considering that the maps are derivatives of photogrammetry, the final accuracy of a map may be expressed by law of propagation of variances;

$$\sigma_{Map} = (\sigma^2_{Control} + \sigma^2_{Aerial\ Triangulation} + \sigma^2_{Orientation} + \sigma^2_{Plotting} + \sigma^2_{Cartography} + \sigma^2_{Printing})^{1/2} \quad (1)$$

The basic input for the digital topographic data base is map and hence the accuracy of digital data by the law of propagation of errors will be;

$$\sigma_{Digital\ Data} = (\sigma^2_{Map} + \sigma^2_{Digitisation})^{1/2} \quad (2)$$

σ_{Map} **= 12.5** Meters on ground in the case of a 1:50,000 scale map. As we take the film positives the error in digitisation is minimised compared to that of digitisation from the paper map. This is due to the fact that the film based material is relatively more stable dimensionally than paper.

2 INVESTIGATIONS

Investigations were carried out for different operations performed. It is found that the accuracy levels attained were not only different for different operations, but also dependent upon the terrain features and the way they are depicted on the maps.

2.1 Paper Maps

The errors in the digital data due to the source material such as maps printed on paper were investigated. Paper maps were scanned and the raster data was rubbersheeted to theoretical dimensions. A number of points were chosen and the variations in the positions when compared to the original positions. The three basic entities such as point, line and area were investigated. For points, established points such as triangulation stations, intersected points and surveyed positions of the temples were chosen. For each type of feature 10 objects were chosen and the mean differences are tabulated. Where ever the coordinates were available from previous surveys such as Triangulation Stations and Intersected Points, they were used for comparison. At other places, the coordinates read from the film positives were assumed to be accurate and used for comparison. For some points GPS coordinates were also recorded in differential mode. The raster and vector coordinates were recorded from the displayed coordinates on the Microstation design files.

It is observed that the errors in coordinates sometimes get reduced due to digitisation as is the case in Triangulation Stations. The errors get increased due to symbology such as huts, temples etc. Fine points such as fine stream (nalla) junctions still retain good accuracy. Similar study was conducted for line and area features also. For line features the maximum deviation from the true alignment was taken for each features, 10 such objects were investigated. For linear features, the centre line was taken as the true position. The map coordinates were read through coordinatograph.

It was revealed that that the linear feature's accuracy remains almost the same as that of the map and it improves in vector data. A similar investigation was carried out in the case of the areas, the results of which show that the accuracy of the data slightly improves when vectorised. This is due to the fact that very fine line may be drawn under magnification (Zoom mode) with resultant improvement in accuracy and the centre line is taken for measurement in the case of area and line features.

2.2 Accuracy From Stereo Photogrammetry

When the data is derived from a stereo model directly, the accuracy achieved is much higher. Aerial photography on scale 1:30,000 was used for stereo restitution, and the data set for undulating terrain was chosen for collecting Digital Elevation Model by progressive sampling with the grid interval of 250 metres. The DEM data was changed into a Triangulated Irregular Network (TIN) file which containing the vertices of slope triangles as X, Y, Z triplets. The TIN file was converted to a GRID file and then graphically the terrain is obtained. The DEM generated data was compared with the heights of some prominent points . In this case RMSE in Position was 17.2 m and for Elevation it was 11.9 m.

The RMSE value for height is about 12m which is rather high. The accuracy could be further improved by reducing the sampling grid interval. However, the accuracy achieved with the interval of 250m is sufficient only to produce contours at 50m vertical interval (VI) as 1/5 th of the VI is the accuracy specification for Indian topographic maps.

3 SATELLITE IMAGERY FROM IRS 1C

India is in a unique situation amongst the developing countries as it has its own satellite which is capable of giving stereo imagery with a resolution of 5 - 6 metres. IRS 1C has also multispectral data bands besides the panchromatic image band. The stereo suffers from the inherent problems of across the track stereo derived by changing the angle of imaging in off nadir viewing mode. It is apparent that sun angles and shadows play havoc. Similarly the spectral signatures change drastically with the time on the same day due to many factors including the moisture content etc., Best stereo imagery for precision topographical feature extraction would be along the track stereo imaged simultaneously, which is likely to be available in near future.

3.1 Investigations with IRS 1C

The investigations carried out for updating digital cartographic data base using IRS 1C imagery are not encouraging. The data base was derived from 1:50,000 scale topographic maps and the insertions and deletions were plotted from the IRS 1C imagery and then field verified. The results shown in Table 1 speak for them selves.

Feature	% Error Commission	% Error Omission
Roads (Major)	0	0
River/Streams	22	0
Residence/Bldg	12	3
Tracks/Minor Rd	51	37
Habitation limit	14	0
Minor canals	0	83

Table 1. Data updating with IRS 1C imagery

3.2 Integration issues

It is evident from the above that the data derived from stereo images of even 5-6 metres resolution, can not be relied upon for features which are of smaller width/size even though they are more than 2 pixels in size. This is due to problems enunciated above, source being across stereo imagery. Its reliability could be improved by musing a combination of multispectral data with panchromatic data. The deterioration in accuracy in identifying features could be attributed to the lack of ground knowledge of the operators. This leads to an important deduction that field surveyors are best suited for this task. It has been observed earlier that good field surveyors proved to be better photogrammetry operators during the analogue/analytical era.

4 ACCURACY IMPLICATION DUE TO DATUM CHANGE

As the uneven and rough physical surface of the Earth is not suitable for mathematical formulation and computations of geodetic data reduction, a hypothetical geometrical reference surface, *Geodetic Datum,* must be defined. The geocentric datums are earth-centred, earth-fixed reference (ECEF) ellipsoids, normally established through space geodetic observations, fitting the entire Earth in the best possible manner. The local geodetic datums are the reference ellipsoids adopted by a country to fit a particular region of the earth in the best possible manner, and hence, these are generally not geocentric. Such local geodetic datums are established through an integrated adjustment of the geodetic network in the country and their relationship with geocentric datums is established through a seven parameter transformation of three translations for shift of origin, three rotations about the three axes of the rectangular Cartesian coordinate system and one scale factor. To these datums the space geodetic data such as Global Positioning System (GPS) data is referred. According to Kulkarni (1996), the importance of establishing a local geodetic datum to the desired level of accuracy is obvious from the fact that the accuracy of the entire geodetic data depends directly on the accuracy of the datum to which the data refers.

One of the major problems encountered is the geodetic datum of data. The Indian Geodetic Datum- Everest Ellipsoid- was established in 1880 through an adjustment of the Geodetic network, computed using the models, techniques and methods then existing in practice at that time. The geodetic network in India was divided into five parts, four quadrilaterals and the Southern Trigon. Due to the computational limitations, these sectors could not be adjusted together and also several corrections were not applied. In 1937-38, an attempt was made to readjust this network which also suffered from the same limitations. Now this gigantic task is being re-attempted and is likely to be completed by the turn of the century due to the fact that modern fast computers are available (Kulkarni, 1996). According to Rajal et al., (1996) adjustment of the North-West and South-East quadrilaterals was completed and the work is progressing in the South - West quadrilateral. The results of the North-Western quadrilateral have now been computed and the differences between the old and new coordinates is shown in Table 2.

Sl No	No of Stations	Maximum difference in Latitude	Maximum Difference in Longitude	Minimum Difference
1	**767**	**10.302 m**	**10.47m**	**0.0**

Table 2. Maximum and minimum differences in coordinates

To illustrate further a number of triangulation stations and Intersected Points were chosen and re-projected onto UTM on the WGS-84 using the software available with approximate parameters for conversion from Everest to WGS-84. The results are not uniform in all the three types of the terrain viz., Desert, Plains and the Snowclad. The results are tabulated in Tables 3 to 5. It may be seen that the errors do not have any pattern and could not be modeled.

Sl No	Type	No of Stations	Maximum Difference in Latitude	Maximum Difference in Longitude	Minimum Difference in Longitude	Minimum Difference in Latitude
1	Triangulation	**50**	**85.210 m**	**103.523m**	**22.765**	**10.600**
2	Interpolated Points	**50**	**98.780m**	**112.230m**	**26.430m**	**36.800**

Table 3. Differences in the coordinates in normal area in the central part of India

Sl No	Type	No of Stations	Maximum Difference in Latitude	Maximum Difference in Longitude	Minimum Difference in Longitude	Minimum Difference in Latitude
1	Triangulation	**50**	**86.010 m**	**90.350m**	**12.660m**	**11.260m**
2	Interpolated Points	**50**	**119.170m**	**102.520m**	**16.640m**	**26.430m**

Table 4. Differences in coordinates in desert area

Sl No	Type	No of Stations	Maximum Difference in Latitude	Maximum Difference in Longitude	Minimum Difference in Longitude	Minimum Difference in Latitude
1	Triangulation	**50**	**115.150 m**	**111.710m**	**9.890m**	**23.835m**
2	Interpolated Points	**50**	**129.000m**	**120.120m**	**6.760m**	**12.460m**

Table 5. Differences in coordinates in high hills

From the results obtained above, it is not possible to model the data unless all the quadrilaterals and the Trigon are adjusted together in 3D.

5 QUALITY INDICATORS

International Standards Organisation (ISO) published 5 standards under ISO 9000 series. They are ISO 9000, ISO 9001, ISO 9002, ISO 9003 and ISO 8402. One may say that these standards are functional because they describe guidelines and not actual systems and procedures. They are intended to be general in the sense that they apply not only to the development of information systems, but to all types of products and services. For each situation a particular implementation has to be designed and made operational. The meaning of quality may be inferred from the statement: a product or service has quality when it meets requirements and/or expectations. If a product or service meets more than requirements or expectations, one may say that its quality is better than required or expected.

The quality of digital spatial data depends upon the procedures employed for assuring the quality. Digital mapping is quite complicated process and the digital data is liable to get corrupted, if adequate quality control procedures are not adopted and there are inconsistencies in data, of any form, would render the efforts to create data base futile. Hence rigorous quality control procedures are essential to maintain the level of accuracy as well as to ascertain the accuracy of digital data outputs. To lay down a proper quality control procedure, the entire gamut of digitisation, and creation of DTDB was studied and the error prone areas are noted.

5.1 Quality Stamp

Quality Stamp is also added as another additional attribute at the end of the list of attributes. The quality indicator is designed as two digit code in the scale of 01 to 09 depending on the source of data. The highest indicator, 01 is given to Ground Survey based on modern survey techniques. The indicators are given in the Table 6. The quality of each element of data individually or as a group may be queried with simple logical operators such as > or <.

Sl No	Data Collection Source	Quality Indicator
1	Ground Survey	01
2	Aerial Photographs > Scale 1: 10,000	02
3	Aerial Photographs on Scale < than 1:10,000	03
4	Satellite Imagery of Resolution > 10m	04
5	Satellite imagery of Resolution < 10m	05
6	Satellite imagery of Resolution < 1m	06
7	Published govt. reports	07
8	Published non govt. reports	08
9	Hearsay	09

Table 6. Quality Indicators

6 TEMPORAL DATA

Temporality of topographic data is quite natural and it also helps the planners and developers in seeking solutions. In all surveyed maps, we find the temporality from the year of survey and field surveys are conducted during a particular time of the year, in India during the months October to March, one could handle temporality of the data. In the digital domain, this will give additional advantage while analysing the data top handle the same feature at different periods of time. The problem at hand how it could be incorporated in the data bases. A simple method was devised which was found to be of immense use.

6.1 Date Stamp

In working towards multidimensional data bases, an additional attribute was added as Date Stamp. The Date Stamp designed contained 6 digit numeric filed with year in 4 digits and the next two digits indicating the month. For example, the data pertaining to a new road inserted from the satellite imagery of November 1994, will be shown as '199411'. It is so designed to enable queries based on the year as well as month as an operator such as < or > which yields the desired answer.

7 CONCLUSION

Uncertainty is particularly a significant problem in GIS because spatial data tend to be used for purposes for which they were never intended, and because the accuracy problem in GIS requires consideration of both object oriented and field oriented views of geographic variation (Goodchild, 1991). Map accuracy is relatively a minor issue in cartography, and map users are rarely aware of the problem. Maps use a very constrained technology of pen and paper to communicate a view of the world to their users. Cartographers feel little need to communicate information on accuracy, except indirectly through map quality statements or in detailed legends. But when the same map is digitised and input to a GIS, the mode of use changes. The new uses extend well beyond the domain for which the original map was intended and designed.

The quantitative analysis of maps is not new to GIS, but GIS brings the accuracy issues into focus. Moreover the machines used to make measurements in GIS (digital computers) are inherently far more precise than the machines of conventional map analysis. The bitter truth is that all geographical data are inherently inaccurate, and that these inaccuracies will propagate through GIS operations in ways that are difficult to predict.

The quality indicator given as additional attribute in the scale of 0 to 9 provided a good opportunity to retrieve the quality information. This is a primitive attempt in classifying the quality based on source of data and the operations that go through the process that will yield end accuracy and not the source accuracy alone. However, as a concept of *quality stamping,* this worked well. To whatever features the quality tag was appended, the same could be retrieved with simple logical operators in the query. Of course, the ground survey, was accorded highest quality as all surveyors take pride in their work and the ground truth of seeing is believing adds to quality. This may not necessarily be true in all cases. It is possible that information extracted from a very high resolution data received from satellites of 1metre resolution or better is likely to yield similar if not better accuracy if viewed in stereo mode. However, these are still not available.

The five indicators of accuracy given in US Spatial Data Standards (Clarke, 1992) viz., *Lineage, Positional Accuracy, Attribute Accuracy, Logical Consistency and Completeness* were tried though lineage was the basic factor considered for giving quality indicators. Positional Accuracy and Completeness were tested by superimposing the raster and vector data in the digitisation stage itself. The attribute accuracy was also taken care while giving the quality indicator. The Logical Consistency was checked on the data base generated performing the topological tests that were possible in the MGE environment. The aim was to obtain 'Topologically Clean' data by ensuring that all the chains intersect at nodes, cycles of chains and nodes are consistent around polygons and Inner rings embed consistently in enclosing polygons. This is very rigorous testing and it took quite sometime to ensure to get the data topologically clean even for small areas. It will indeed be a daunting task when we talk of a many tera bytes of data covering large countries such as India. The basic step in quality control was to generate a check plot on a vector plotter for comparison with the input data to ensure completeness and integrity of the data.

REFERENCES

1. ASPRS Interim Accuracy Standards for Large - Scale Maps, 1988, The PE&RS Journal, 1988, 1079 - 1081.

2. Barker, T.B., 1985, Quality by Experimental Design, *Dekker*, 384.

3. Bolstad, P., 1990, Uncertainty in Manually Digitized Mapped Data, *International Journal of Geographical Information Systems*,4(4),399 - 412.

4. Clarke, Andrew L., 1992, Data Quality Reporting in the Proposed Australian Spatial Data Transfer Standard, *Proceedings of the Symposium on Spatial Database Accuracy, Melbourne, Australia, June 19-20, 1991,* 256-259.

5. Doyle, J. F., 1983, Surveying and Mapping with Space Data, *A Paper presented at Corten Retirement Symposium, ITC, The Netherlands.* 1 - 15.

6. Goodchild, M.F., and Gopal, S., 1989. Accuracy of Spatial Databases, Taylor and Francis, New York.

7. Goodchild, M.F., 1991, Keynote Address at Symposium on Spatial Database Accuracy, *Proceedings of the symposium on spatial database accuracy, Melbourne, Australia,(ed) Hunter, Gary, J.Hunter, University of Melbourne, Australia* . 1-16.

8. Iyer, R.P., and Siva Kumar, R., 1996, Softcopy Photogrammetry - Challenges and Potential, *Proceedings of Carto Vision 2001 at New Delhi,* 57 -59.

9. Jankowski, P., 1995, Integrating Geographical Information Systems and Multiple Criteria Decision-Making Methods, *International Journal of Geographical Information Systems,* 9 (3), 251 - 274.

10. Joshi, M.D., and Siva Kumar, R., 1995, Maintenance of Topographic Database in India; Photogrammetric Solutions, *South African Journal of Surveying and Mapping,* 23 - 2 (138), 85 - 89.

11. Kulkarni, M.N., 1996, GPS for Mapping, Proceedings of "Carto Vision 2001", 6-7.

12. Masters, E., 1991, Defining Spatial Accuracy, *Proceedings of the symposium on spatial database accuracy, Melbourne, Australia,(ed) Hunter, Gary, J.Hunter, University of Melbourne, Australia,* 215 - 224.

13. Nagar, V.K., and Siva Kumar, R., 1987, Revision of 1:250,000 Scale Maps using Thematic Mapper Data Products, *Indian Cartographer,* 7, 93 - 98.

14. Rajal, B.S, and Kulkarni, M.N., 1996, Redefinition of the Indian Geodetic Datum, Proceedings of the 'Carto Vision 2001', 23 - 25.

15. Siva Kumar, R., 1987, SPOT Imagery - Relevance for Topographical Mapping in India, *Indian Cartographer,7,* 120 - 128.

16.. Siva Kumar, R., 1989, A Case for Survey Net, *Indian Cartographer,* 9, 348 - 361.

17.. Siva Kumar, R., 1994, Accuracy Standards for Ground Control, *Indian Surveyor,* 42 (1), 53 - 62.

18.. Siva Kumar, R., 1996, Surveyors at Cross Roads, *Indian Surveyor,* 48 (1), 51 -59.

19. Siva Kumar, R., 1996, Challenges in Creating a Digital Topographic Data Base in a Developing Country, Proceedings of Carto Vision 2001 held at New Delhi, 62-64.

20. Siva Kumar, R., and Vachher, P.K., 1996, GIS - A Dream or Reality, Remote-Sensing for Environmental Management, (ed) Khan, I.H., and Ahmad, S., Vishal Publishing House, New Delhi, 123 - 131.

21.. Siva Kumar, R., 1997 and Joshi, M.D., (1997), Potential use of Remote Sensing and GIS in monitoring water distribution in the command areas in the 21st Century, *Indian Surveyor,* 48 (2), 8 - 11.

22. Siva Kumar, R., and Joshi, M.D., (1997). Sectoral GIS- The only Alternative for Sustainable Land Development in Developing Countries, *International Soil Society conference at Enschede, The Netherlands, 1997.* (on CD-ROM).

23. Siva Kumar, R, (1998). An Integrated Multi-Source Digital Topographic Data Base for India, *Ph.D Thesis submitted to Indian Institute of Technology, Delhi, India.*

24. Siva Kumar, R, (2000). Some aspects of accuracy in GIS, *GIS at Development,* 4 (2), 45 - 49.

25. Terry R., 1991, Spatial Accuracy as a Quality Indicator, *Proceedings of the symposium on spatial database accuracy, Melbourne, Australia,(ed) Hunter, Gary, J.Hunter, University of Melbourne, Australia,* 162 - 174.