THE EFFECTS OF GIS ARCHITECTURE, DATA MODELS AND DATA SOURCES ON THE ACCURACY OF DIGITAL MAPS: AN EXAMPLE OF DIGITAL TERRAIN MODEL OF IBADAN REGION.

Fabiyi O.Oluseyi
Department of Geography
University of Ibadan
Ibadan, Nigeria.
E-mail seyifabiyi@yahoo.com
Tel. +234 –02-8102900, 234-02-8108090
MobileTel +234 08023271512

ABSTRACT

Maps have been widely accepted as a veritable instrument in national development. They are graphic representations that facilitate a spatial understanding of things, concepts, conditions, processes, or events in the human world. Map making; which is an important means of graphically communicating information has been greatly influenced by modern technological developments including digital technology and development in geographic information systems

However different type of maps exist and at differing scales. The usefulness of any map is heavily dependent on the level of accuracy of the map. The issue of map accuracy becomes prominent in developing countries where substantial parts of the countries are only partially surveyed. With the advent of GIS (Geographic Information Systems) the procedures for production of basic maps and derived maps have been revolutionized accordingly. While the conventional methods for data capture in the production of analogue maps utilized intuition and experience of the cartographer, the automated mapping functions of GIS depend on data algorithms, software techniques and architecture for carrying out data conversion, transformation and spatial analysis in map making.

The paper discuses different types of errors in digital map making and their It identifies the unique problems of analogue-Digital map conversion in Nigeria, using the specific example of l production of digital terrain model of Ibadan region. The paper compares different transformation errors and the accuracy levels of digital maps. It also highlights different sources of errors in digital map-making and suggests techniques of minimizing errors in digital maps.

INTRODUCTION

Maps have been widely accepted as a veritable instrument in national development. Since the first world map compilation by Idrisi and his team, the issues of mapping and accuracy levels of maps have been of serious consideration in Geography , cartography, surveying and many other land based disciplines. However different type of maps exist and at differing scales. The usefulness of any map is heavily dependent on the level of accuracy of the map. The issue of map accuracy becomes important in developing countries where substantial parts of the countries are unmapped or mapped with error prone methods. With the advent of GIS (Geographic Information Systems) and cartographic packages the procedures for production of basic maps and derived maps have been revolutionized accordingly. While the conventional methods for data capture for producing analogue maps (often called map generalization) utilized intuition and experience of the cartographer, the automated mapping functions of GIS depends of data algorithms, software techniques and architecture for carrying out data conversion, classification, transformation and spatial analysis.

The contribution of GIS to map making will not be of significant value if the procedure does not remove or rather ameliorate the numerous problems associated with precision and accuracy in conventional map making. The improvements of digital map-making over the conventional method is documented in Fabiyi (1996). GIS makes use of data from different platform and sources for the purpose of producing digital maps. Each of these data sets needs to be transformed to a common platform for the purpose of digital map making. The procedures for digital map-making adopted by different software relates to the architecture of the package. The software architecture relates to the structure of data storage, analytical procedures and spatial analysis algorithms.

CONCEPT OF ACCURACY AND PRECISION IN DIGITAL MAP MAKING.

Digital map-making is the process of utilizing computer technology to produce map. A map is a two-dimensional scale model of a part of the surface of the earth. This model is a systematic description or representation of the part of the earth, generally using symbols to represent certain objects and phenomena. Maps are effective ways of presenting a great deal of information about objects and the spatial relationship of objects.

Maps are of different categories for example, topographic maps, planimetric maps/ cadastral maps, thematic maps, cartograms among others. Most earth related disciplines benefits immensely from cartography and revolution in the procedures and techniques of cartography is obviously affecting the processes of utilization of cartographic products as well as the revolution taking place in deferent disciplines based on the discoveries of digital/ automated cartography.

With the advent of electronic computers and the development of graphic cards in modern computers, efforts were directed towards the use of computer to do some graphic jobs that were manually handled.

Map Accuracy

Map accuracy have been variously observed and defined. Some have differentiated accuracy from precision by indicating that accuracy is relative while precision is absolute. In this case accuracy is the relative correctness of the position of objects on a map from the real position of the object on the earth surface, Precision is the level of exactness of the object position. Aronof (1989) gave a broader consideration of Accuracy. The direction of research in Digital map making is in the area of absolute precision of data capture and map presentation.

The accuracy of a digital map is not dependent on the map's scale. Instead, it depends on the accuracy of the original data used to compile the map, how accurately is this source data has been transferred onto the map, the algorithms of transformations, and the resolution at which the map is printed or displayed. In the case of analogue-digital map conversion the digital map cannot be more accurate than the original sources (paper map). However some software allow minor corrections through the input of ground measurements (ground truth) For example digital map in Arc/info format can be transformed using the ground co-ordinates (at least 3 point) obtained through GPs, this will adjust the digital map in accordance with the in put co-ordinates. In this case the resulting map is more accurate and has better topology than the source map.

To create the spatial database for the purpose of digital map making, existing maps or manuscripts may have been digitized or scanned, and other original data, such as survey reports, aerial photographs and images, and data from third parties may also have been used. The final map will therefore reflect the accuracy of these original sources.

Types and sources of error in Digital map making

There is no error free digital map, the purpose of mapping would determine the standard error level that can be tolerated. Digital map-making involves four basic steps which are:

(a) Data collection and data capture
(b) Data transformation and spatial analysis
(c) Map embellishments and annotation
(d) Map presentation and output.

Error can be introduced at any stage of digital map making, and the errors can be extrapolated as the process continues. Some of these errors can be enumerated as follows

(a) Data collection error

Errors can occur in field data collection, In the existing map used during analogue digital conversion, Errors from the base map materials (paper shrinks in wet weather and expands in dry season, this can result in inaccurate data.

(b) Data capture error

Error in digitizing caused by operator and equipment inaccuracies inherent in the geographic features. Error due to imperfections of users, during line-following digitizing process. This can cause Slaver (sliver) polygons or spaghetti, undershoots and overshoots. Raster data has inherent error based on the spatial resolution of the device for data capture.

(c) Data Storage and spatial analysis errors

Error due to insufficient numerical precision, inefficient spatial algorithms and insufficient spatial precision.

(d) Map embellishment and output error.

Error caused by inaccuracy of the output device and instability of the output medium especially the printouts.

These sources of error can be classified into three types based on the cause and effects on map accuracy. These are data sources, data model and GIS software architecture.

Four types of accuracy issues in digital map making

In digital map making there are essentially four types of accuracy issues.

- ***Positional Accuracy***

Positional accuracy is the deviation in the geographic location of an object on a map from its true ground position. There are two components of positional accuracy; the ***bias and precision:*** the bias refers to systematic discrepancies between the represented and true position. Bias is commonly measured by the mean or average positional error of the sample points.

Precision on the other hand refers to the dispersion of the positional errors of the data elements. Precision is commonly estimated by calculating the standard deviation of the selected test points. A low standard deviation indicates that the dispersion of the positional errors is narrow. i.e. the error tends to be relatively small. The higher the precision of the measurement, the greater the confidence in using data. A measure of positional accuracy commonly used digital map is the root mean square error (RMS). It is calculated by determining the positional error of the test points. Squaring the individual deviations and taking the square root of their sum. This measure does not distinguish between the bias and precision components of accuracy. Usually a more accurate base map is used to test the sampled points of digital maps and the deviation from normal is estimated to establish the level of positional accuracy. In Nigeria however the available maps are dated and many of them were produced through less efficient methods. It becomes difficult to establish which map has precision accuracy problems. If GPs points are used the cost expended on establishing the level of precision becomes too high to justify the benefits accruable from accurate digital map.

It is noteworthy that there are no hundred percents accurate digital maps, it becomes a professional issue on the level of accuracy that must be attained by digital map before the map can be acceptable for use.

- ***Classification or Attribute Accuracy***

Attribute accuracy relates to the error in classification during map generalization. Attributes may be discrete or continuous variables. Discrete variables can take on only a finite number of values whereas continuous variables can take on any number of values. Categories like land use class, Vegetation type, or administrative area are discrete variables. Ratings are also discrete variables.

The method of assessing accuracy for continuous variables is similar to that discussed for positional accuracy of discrete variables is the domain of classification accuracy assessment. The assessment of the classification accuracy is a complex and somewhat controversial procedure. It has received considerable attention in the remote sensing literature. The difficulties in assessing classification accuracy arise because accuracy measurement is significantly affected by such factors as the number of classes, the shape and size of individual areas, the way test points are selected, and classes that are confused with each other.

- ***Logical Consistency***

Logical consistency refers to how well logical relations among data elements are maintained. For example, it would not be consistent to map some forest stand boundaries to the center of adjacent roads, and some of the road edge. They are normally all mapped to the road edge. Political and administrative boundaries defined by physical features should precisely overlay those features. The edge of a property that borders a lake should coincide with the lake boundary.

Consistency problem is often escalated when dealing with features that have fluctuating boundaries e.g. a lake, the boundaries of such lake is a function of the season which measurement is carried out. Consequently data obtained at different season may not overlay properly and may result in slaver polygons As a result; the lake boundaries may be accurately delineated, but logically inconsistent among data layers.

Some GIS software are able to accommodate this type of positional discrepancy by assigning a band of uncertainty which would be treated as if they were coincident. (A boundary that is treated as a band of uncertainly is often termed a fuzzy boundary).

Completeness

The completeness of coverage is the proportion of data available for the area of interest. A data set may not provide complete aerial coverage of the area of interest or attribute data may not be available for some portion of the data set. Ideally, a data set would provide 100% coverage. However, in some cases extrapolation or interpolation may be carried out to account for areas that lack data. The procedure is based on certain assumption which have some level of errors.

DATA MODELS AND SOURCES ON MAP ACCURACY.

The earth is a complex system consisting of several elements interacting with one another. Each element may also be defined by several parameters. In geographic information systems each of the different elements of the environment is isolated and each of its different parameters identified constitutes a theme. Data on each theme are recorded and stored in separate data layers or data planes. There are different ways of organizing the data on each layer in the computer for spatial analysis in a GIS framework. The different ways in which data are organized within information systems are referred to as data models or data structures. The database therefore actually presents a model of that segment of the earth's surface

There are two basic types of data structures that is two ways in which data are referenced and stored in an information system: these are raster or cellular organization of spatial data and vector data structure.

Vector Data

Vector data records spatial information as x, y coordinates in a rectangular (planar) coordinate system. Point features are recorded as single x, y locations. Line features, including the outlines of polygons are recorded as an ordered series of x, y coordinates. Vector data is well suited to recording the location of discrete geographic features with precise locations like streets, parcel boundaries, streams and telephone poles.

Vector data can accurately record the actual ground location of features. However, vector data is highly dependent on the number of x, y coordinate points that are chosen to represent features, especially natural features like streams and coastlines. A typical vector data is made up of series of arcs, nodes and vertexes. Features that are captured as points are represented by nodes that have no direction but with specific coordinates. x, y, and z . Line features are captured as series of nodes and vertexes connected by arcs. Area features are captured as series of nodes connected by arcs and the starting node or the first node share the same location with the last or the ending nodes. In order word a line that closes represents an area feature.

Raster Data

Raster data records spatial information in a regular grid or matrix organized as a set of rows and columns. Each cell within this grid contains a number representing a particular geographic feature, such as soil type, elevation, land use and slope. Raster data are commonly, but not exclusively used to store information about geographic features that vary continuously over a surface, such as elevation, reflectance and groundwater depths. Image data is a form of raster data in which each cell or pixel stores a value recorded by an optical or electronic device.

Data model error

The types of data structure adopted during map compilation and digital map making will inevitably influence the level of accuracy obtainable from such sources. Vector data is continuous representation of data therefore discrete features or data collected at discrete aggregation units can only be represented with some level of error. The permissible level of error however will depend on the purpose of map making.

In a raster-based model, the error level is determined by the pixel size of data model. The pixel size is also dependent on the spatial resolution of the data collection device.

Another major source of error is during raster vector conversion and vice versa. The algorithms of converting from vector to raster have different assumptions and generalization that either information is lost or added to the data boundaries. The implication is that when there is any form of transformation in data structure the level of accuracy reduces. It is noteworthy that it is practically difficult to complete process of digital map-making in one software platform. Some software are vector based while some are raster based. When data is converted from one software platform to another .The data may be exploded (enlarged) or contracted based on the architecture of the two software types and the data linkage modules available in each of the software. These inevitably affect the transformation error. If the user would present the map in a vector environment all raster data must be converted to vector, and vice versa if the data output would be in raster environment. Raster-vector conversion and vector –raster conversion are with extensive positional and classification error. The errors are magnified when many layers of data need to be converted from one format to another. See figure 1.

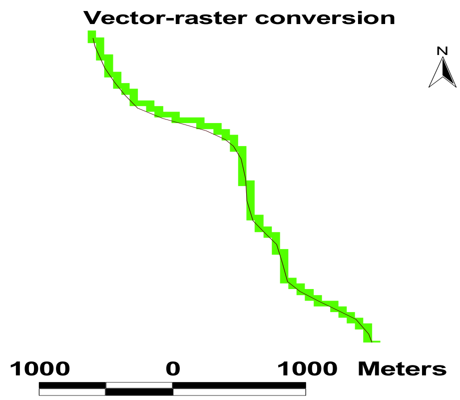

Figure 1. Vector Raster conversion of a poly line.

GIS SOFTWARE ARCHITECTURE AND MAP ACCURACY

In geographic context, all human activities and the natural environment where the activities are taking place can be captured as data. GIS is data driven, it makes use of data in space.

Geographic data have topology, which represent the mathematical relationship of objects in space. In GIS topology generally means the spatial relationships between adjacent or neighbouring features. It is expected that the spatial data stored in the computer should bear the same topology with the real life objects that they represent. This is a function of directions or, bearing and distance. Spatial relationships between object and among features can be important although these considerations are not always apparent to the eye. Distance from one feature to another is available through simple measurement; also proximity of features such as connections or nearness may be determined through various measures adopted by the software design to determine 'neighbourhood' characteristics.

A GIS does not store map, rather the data are stored in the computer memory with their unique 'IDs' that serve as pointers or identifiers. Although the eye can estimate some of the basic properties of spatial data, a GIS must provide highly accurate measures and readings of their properties. All objects on earth or phenomena correspond to spatial data and can be captured as thus. However our interest may not be on the spatial component of the object or phenomenon.

Spatial data characteristics of GIS and its implications on error occurrence in digital map making

Locational Specificity
Location refers to the position of the data in space, usually in form of a set of x, y, co-ordinates. The objects could be distributed at points, along lines or within given areas, location is the major starting point of measurement, normally using a known geographic co-ordinate system, such as latitude-longitude. Most cartographic packages stores spatial data in different file while the location characteristics are stored as attributes. The effectiveness of the link between these files is directly or indirectly influence the speed and the efficiency of spatial operations.

Area and Perimeters
These are other properties of spatial data and are easily calculated in most GIS software. Furthermore, the eye quickly appreciates the shape of features but computers need special programming to describe the

spatial forms of object in space. To calculate area, or perimeter of an object in a digital map, the location of centroid of the features must be established. The location of a centroid is software dependent as different approaches are utilized by different software vendors.

Location Of Centroids And Its Implication On Digital Map Making.
GIS automatically calculate the centroid of objects and this is maintained in the database environment. the position is based on the generalization assumptions adopted during software design. For example; For ***a region*** the centroid is its mathematical centre but most GIS packages adjust the centroid to fall within the region, if the centroid lies outside the borders of the region. If the centroid of is different from one software to another then maps from different software platforms can con be compared. The centroid of a line is the calculated mid point along the entire line it is the mid point of the longest segment in the line. The centroid of a point is the co-ordinates of the point.

Topology

Mathematical topology assumes that geographic features occur on a two- dimensional plane. Through planar enforcement, spatial features can be represented through nodes (0-demensional cells) edges, sometimes called arcs (one-dimensional cells); or polygons (two –dimensional cells). In most GIS packages topology is implemented through data structure for example in ArcInfo, the coverage is a topological data structure. The data architecture is designed in such a way that topological relationships are stored in the arc attribute tables called AAT by storing the adjacent polygon IDS in the Lpoly and Rpoly fields. Adjacent lines are connected through nodes and this information is stored in the arc-nodes table.

Spatial relationships between objects and among features can be important although these considerations are not always apparent to the eye. Distance from one feature to another is available through simple measurement; also proximity of features such as connections or nearness may be determined through various measures to determine ‘neighbourhood’ characteristics.

Topology becomes important when the area covered by the map is large. Whichever the types of projection types adopted topology must be maintained. However the focus of recent researches in software design is to de-emphasise topology in digital map-making. (ArcNews June 2001).

Advanced spatial analyses is enabled by the topological relationships of the software Analyses such as adjacency, connectivity, and containment utilise topological relationship for accurate results and presentation of results.

The principles adopted by a software to determine the structure is determined by the architecture which dictates the GIS database structure. There are three types of data architecture in GIS environment. These are hierarchical structure, Network structure and relational data structures. These types of data structure determine the procedures of relating the data attributes with the pointers and identifiers.

The architecture of GIS software also relate to the structure of data storage and methods of compacting or compressing data. For example in Raster Data environment there are four types of data architecture which are: (a) chain coding; (b) run-length encoding; (c) block coding; and (d) quadtree. These show different level of data configuration and package to encourage efficiency. On the other hand the vector data environment can also have different types of data architecture such as: (i) whole polygon structure; (ii) Dual independent Map Encoding (DINE) file structure; (iii) arc-node structure; {iv) relational structure; and (v) digital line graph.

The structure of database organisations can also contribute to the level of accuracy obtainable from digital map-making. The major types of database organisation in GIS can be grouped into three. These are; hierarchical data structure, network and relational data structure.

When spatial calculation of any kind is carried out such as spatial statistics, buffering, proximity analysis among others the adjusted centroid is used see illustration. The implication is that since the algorithms to establish the adjusted centroid differ from one software to another it is difficult to spatial analyses carried out on data captured through different software will inevitably differ and the result cannot be precisely compared with other results. Products from such spatial analysis have some level of precision error on the site.

The data was collected from a range of sources such as spot height and contour information from topographic map

Methods
The paper examines the effects systems architecture on the accuracy of digital elevation model of Ibadan region. Ibadan region is the largest traditional urban area in the sub Saharan Africa. See figure 2. Spot heights were obtained from existing records in map by Federal surveys. The spot heights were captured in ArcViewGIS environmental and a Digital elevation model was produced.

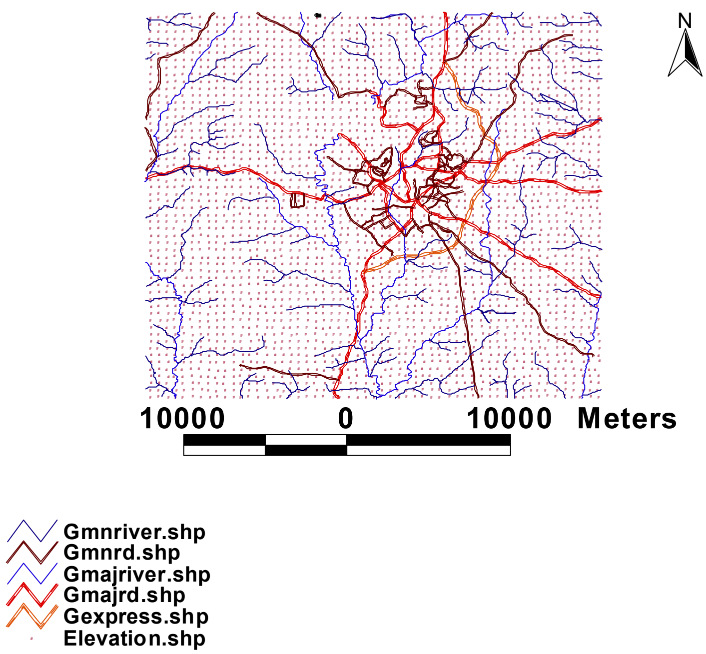

Fig.2. Map of Ibadan region

Digital elevation model of Ibadan metropolis was produced through the Spatial Analyst extension of ArcView GIS a product of ESRI (Environmental System Research Institute) The spot heights were captured and used as Z values for surface interpolation. The resulting map was super imposed on drainage and road map see Figure 3.

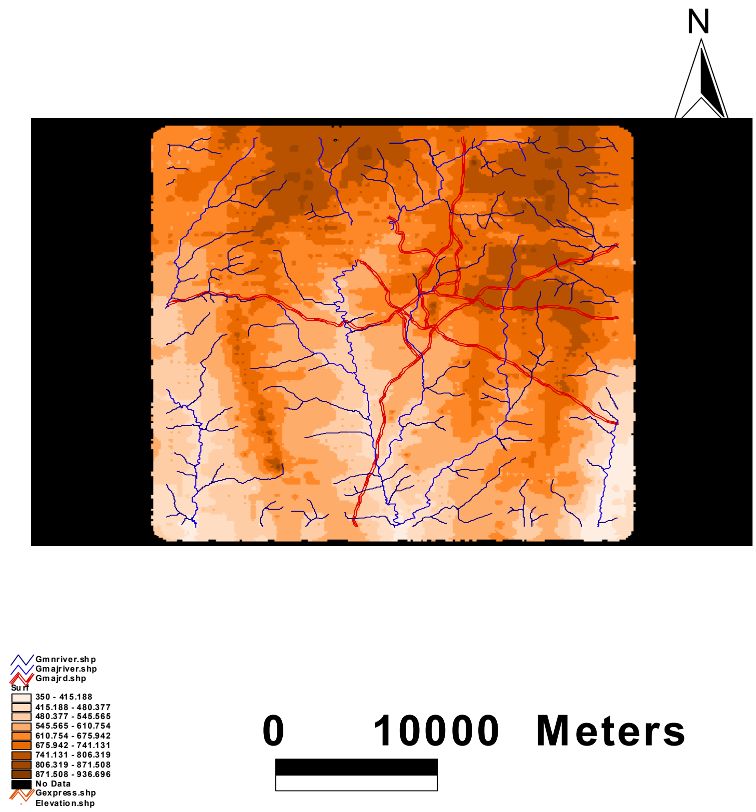

Fig 3. DIgital Terrain model of Ibadan region

A hand held GPS was used to control areas where spot heights are not available.

Discussion

The following were observed.

1. The drainage and the DEM map reveal some discrepancies e.g. rivers cross a contour twice, and stream flowing up hill in some parts of the region.

2. Different methods of surface interpolation yielded different accuracy level. See figure 4.

3. There is high deviation between the interpolated

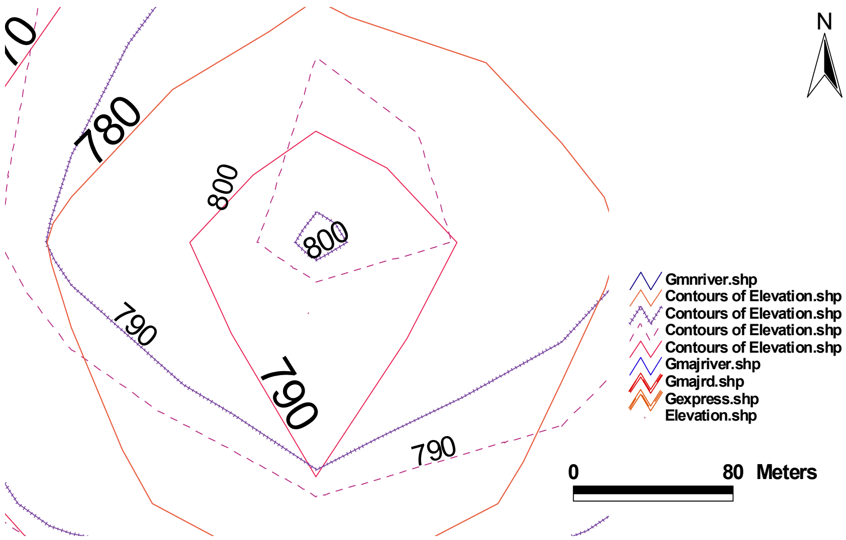

Figure 4: different error levels from Digital map- making

IMPLICATIONS FOR ENVIRONMENTAL RESEARCHES.

The advancement brought about by the integration of digital technology to the conventional map-making introduce another dimension to the techniques and the paradigms of map making. More people who are untrained in the conventional cartography are heavily involved in digital map making. Though the importance of accuracy in digital map is related to the purpose of the map but some map have diverse purposes.

The need for specifications in the algorithms and the database design of digital map-making software become of uttermost importance. Maps have security implications and when error level becomes large despite the level of advanced technology it becomes a legal questions what level of precision error is allowable in the law court especially at the level of cadastral maps. The level of accuracy of spatial analysis result to make it scientifically valid should also be resolved especially a test for estimating the significant level of accuracy of spatial analysis in GIS environment should be the focus of GIS software vendors.

However, while digital map making is still in the elementary forms the level of accuracy achievable in the process will increase with time or maturity.

CONCLUSION

The computer age has brought a new dimension to traditional cartography. It has expanded the circles of those who can produce and use maps. The landscape is still very slippery with some attendant problems. The error which hinge on the usefulness of the map information must be addressed before we can fully integrate automated map-making to all land related disciplines. Many spatial analyses in GIS are still error prone.

REFERENCES

ArcNews June 2001:.Environmental Systems Research Institute Inc.

Fabiyi, O.O. (1996). Application of Satellite Remote Sensing to the Monitoring and Surveillance of Master Plan Implementation; A case of Abuja Federal Capital City. *Proceedings of Nigerian Institute of Town Planners.* Pp 48-63.

Aronoff S. (1992). *GIS: A management Perspective WDL* Publications.