SS476

International Association for Official Statistics

International Statistical Institute

United Nations Economic Commission for Africa

African Statistical Association

JOINT IAOS/AFSA CONFERENCE

Addis Ababa, 22-24 May 1995

IAOS/AFSA/IS.4/05

Managing Development in the 1990s and Beyond: New Trends in Statistics

INVITED SESSION 4:

DATA FOR DIVERSE AREAS; THE GEOGRAPHIC DIMENSION IN DATA

AN OVERVIEW OF COMPUTERIZED GIS

AND SOME OF ITS USES IN STATISTICAL OFFICES

By

Yacob ZEWOLDI

UNECA
Statistics Division

ADDIS ABABA
MAY 1995

AN OVERVIEW OF COMPUTERIZED GIS AND SOME OF ITS USES IN STATISTICAL OFFICES

By

Yacob Zewoldi
Statistics Division, ECA

Paper presented to the Joint Conference of
International Association for Official Statistics
and African Statistical Association

22 - 24 May, 1995

Addis Ababa, Ethiopia

. *The views expressed in this paper are those of the author and do not necessarily reflect those of the Economic Commission for Africa.*

ABSTRACT

The objective of this paper is to introduce Geographic Information System (GIS) to the novice and to point out some of its many possible applications in statistical offices and research institutes. GISs are computer systems that integrate spatial and attribute data. Sources of spatial data for a GIS are paper maps, aerial photographs and satellite images and that of attribute data are censuses, surveys and administrative records.

A number of maps can be stored and analyzed through overlaying in GIS to find answers to complex questions based on common geographic occurrences. Most GISs are designed to store spatial data and link attributes to the stored features; to analyze these data on those attributes; model future scenarios, make predictions, and display spatially referenced data for solving complex planning and management problems.

GIS can also be powerful tool for statistical offices in data collection, storage and update; and in integrating, analyzing and visualizing information. Statistical maps and graphs make it easier for the general user to understand the geographic distributions and patterns inherent in statistical tables. Microcomputer GIS offer extensive possibilities for the statistician to produce his own maps and charts.

RESUME

L'objectif de ce document est d'introduire le Système d'Information Géographique (SIG) au novice et de montrer quelques-unes de ses nombreuses applications possibles dans les services statistiques et les instituts de recherche. Les SIG sont des systèmes informatiques intégrant des données spatiales et individuelles. Les sources de données spatiales pour un SIG sont les cartes géographiques, les photographies aériennes et les images par satellite; celles des données individuelles sont les recensements, les enquêtes et les registres administratifs.

Certaines cartes peuvent être stockées et analysées en les superposant dans le SIG pour trouver des réponses à des questions complexes en relation directe avec des événements géographiques courants. La plupart des SIG sont destinés à stocker des données spatiales et relier des variables aux caractéristiques stockées; analyser les données spatiales concernant ces variables; modeler les scénarios futurs, faire des extrapolations, et mettre en exergue des données de référence spatiales pouvant aider à résoudre des problèmes complexes de planification et de gestion.

Le SIG peut être aussi un puissant outil pour les services statistiques dans le cadre de leurs activités de collecte, de stockage et de mise à jour des données, ainsi que pour l'intégration, l'analyse et la visualisation des informations. Grâce au SIG, il est plus facile à l'utilisateur habituel de graphiques et de cartes statistiques de comprendre la distribution et les caractéristiques géographiques des phénomènes telles qu'elles ressortent des tableaux statistiques. Le SIG sur micro-ordinateur offre de grandes possibilités au statisticien dans la production de ses propres cartes et graphiques.

1. INTRODUCTION

There is an increasing recognition that spatial and statistical data should be integrated for sound socio-economic development planning and implementation. This has given impetus to the collection, analysis and publication of geo-referenced data. Consequently, significant technological developments have been made in recent years to integrate spatial and statistical databases.

The term used to characterize the computer systems that integrate spatial data and statistical information is Geographic Information Systems, GIS for short. The term GIS appears to have a wide variety of definitions (Carter, 1989; Rhind, 1989; Wellar, 1993). A GIS is a computer assisted system for the acquisition, storage, analysis and display of geographic data (Eastman, 1992). According to Rhind (1989) a comprehensive definition of GIS is "A system of hardware, software, and procedures designed to support the capture, management, manipulation, analysis, modeling, and display of spatially referenced data for solving complex planning and management problems." For the new-comer to GIS, he suggests a simpler definition of GIS: "A computer system that can hold and use data describing places on the Earth's surface."

In some cases users substitute another word or words for 'Geographic' to give such expressions as the Urban Information System or the Natural Resource Information System (Carter, 1989). Some of the micro-based GIS software packages on the market are ARC/INFO, ATLAS*GIS, GEO/SQL, GeoScope, GEOSYS, GisPlus, M.A.P., MapInfo, PAMAP GIS, SPANS and Idrisi.

The objective of this paper is to introduce GIS to the novice. It is also a modest attempt to point out some of its many possible applications in statistical offices and research institutes.

2. DATA AND DATABASES IN GIS

2.1 Data in GIS

The data stored and manipulated within a GIS are of two kinds. The first is spatial or geographic data such as longitude and latitude defining the location of objects in the form of points, lines and polygons. Spatial data identify where objects of interest are located, their distributions and extent, and their adjacency, proximity, or connectivity in relation to other phenomena for which a geographic location can be assigned (Wellar, 1993).

The second kind of data consists of the attributes or characteristics of the spatial objects under consideration. For example, the location of a lake can be expressed by a pair of
geographic coordinates such as longitude and latitude. Non-spatial or attribute data can be its depth, width, aquatic life in it, water quality, etc.

The cartographic data can be vector (contour lines or boundary lines) or raster (cell or grid format). Attribute data can be nominal, ordinal, interval or ratio.

2.2 Databases

A GIS links spatial data of geographic features with defined location on the earth's surface, represented by geographic coordinates, with information about these features on a map. Spatial data are stored in a geographic database. Attribute data of the geographically represented feature are stored in another database. Thus, for example, we might have a road network defined in the spatial database and qualities such as its width or type stored in another database logically connected to the spatial data. The logical connection enables to manipulate the non-spatial data along with the spatial data and the databases can be queried to display the roads according to the type of information that needs to be shown.

The database concept is central to a GIS and is the main difference between a GIS and a simple drafting or computer mapping system which can only produce graphic output (Eastman, 1992; Rhind, 1989). One of the many powers of a GIS lies in the link between the spatial and the attribute data. All contemporary GISs incorporate a database management system.

2.3 Data sources

Spatial data for a GIS can be obtained from paper maps, aerial photographs and satellite images. Devices such as digitizers and scanners are used for encoding geographic data into a computer from existing paper maps. Most GIS incorporate software packages which permit spatial data capture in digital form and conversion of vector format to raster and vice versa.

Processed remotely-sensed images and aerial photos are another source of machine-readable cartographic data. Image processing software allows one to take raw remotely sensed imagery (such as LANDSAT or SPOT satellite imagery) and convert it into interpreted map data according to various classification procedures (Eastman, 1992). This approach is becoming a major means of map data acquisition, particularly in the developing countries, because of its savings in cost and time. The converted map can then be manipulated in a GIS for interpretation and analysis. Land-use and land-cover mapping are two of the most common applications for satellite remote sensing.

The sources of attribute data are censuses, surveys and administrative records. Censuses and surveys could be of the population, housing, agriculture, industry or business type. Some of the data that can be collected from these sources from which statistical maps can be presented include total populations by subareas; population density; other population characteristics like age, sex, language, fertility, mortality, migration, educational attainment, employment, income, number of housing units, type of housing structure and material of construction, tenure, number of rooms, sanitary facilities, sources of light and heat, percent of population engaged in agriculture, size of holdings, amount of land in crops by type of crop, crop production, size and type of livestock and poultry, irrigation, etc.

Preparing maps in computers and entering large volumes of attributes in a database is a time-consuming and costly process for many organizations. One advantage of a GIS is that it allows for sharing of data among users. GIS makes possible the integration of spatial data from different sources. It is, therefore, feasible to exchange information between organizations and to combine spatial data and their attributes from different sources, provided the data sets are reasonably accurate for the intended purpose.

Attribute data linked to some geographic component are available in computer readable form free of charge or can be obtained at reasonable prices from national statistical offices, universities and research institutes, and international agencies such as UNECA, FAO, UNIDO, and UNESCO.

In recent years, national mapping agencies of developing countries have began automated mapping and computerized data collection. The development of modern computer techniques have also facilitated direct conversion of aerial photos and satellite images into readable information for GIS. Computerized spatial data sets can be, therefore, obtained from national mapping agencies or international agencies such as UNEP. Furthermore, present developments and future improvements in networks could make remote access to databases at other institutions possible.

3. SPATIAL DATA ANALYSIS AND MODELING

A GIS is a powerful analytical tool (Eastman, 1992; Rhind, 1989). Its powerful database management provides the means of storing a wide range of data and updating them. A GIS can use the stored data to answer questions and compute new information about map features such as the length of a particular road, or generate buffer zones on either side of a road, or determine the total area of a particular soil type.

3.1 - GIS query

A GIS can be used to query what is to be found at a given location from the stored information in the databases. This important tool provides retrieval of data that meet certain conditions.

Two types of queries are answered by a GIS. The first type of query refers to questions regarding the absolute or relative locational properties of the spatial entities. For example, given a network of lines representing railways, what is the shortest route to get from point A to point B? It can identify the area of the region lying within 10 kms of a country's coastal line, or show all areas that are adjacent or within 10 kms of floodplains in an administrative region.

The second type of query refers to the attributes of the spatial features. The database can be, for instance, queried to indicate how many of the countries in Africa are covered by at least 20 percent vegetation or the population densities of the administrative regions of a country. A GIS allows one to quickly search the database for items eligible for output, and then locate and display the relevant areas on the map.

Complex queries with several conditions can also be undertaken. One may inquire of the system to find all property parcels bordering a certain length of a river, where households own cattle and to produce a map of the result.

3.2 Spatial analysis

One of the many important capabilities of a GIS is the ability to store spatial data and link attributes to the stored features; to analyze these data on those attributes; and to map out the result. Many software which are not strictly GIS also have these capabilities. These software, however, do not have the ability to analyze data based on truly spatial characteristics (Eastman, 1992).

In a GIS both kinds of information and the relationships between them are used to analyze the data, model future scenarios, make predictions, and reach conclusions (Paulsson, 1992). It can be used to produce new information through calculation. For example, to find soil erosion risk prone areas one can calculate the size of areas of all occurrences of land on steep slopes that are under agriculture.

A number of maps can be stored and analyzed through overlaying to find answers to complex questions such as which land parcels are at least two hectares large, in a commercial zone, vacant for sale, not subject to flooding, not a kilometre from a heavy-duty road, and have no slope over 10 percent. It is this overlaying capability to compare different entities based on their common geographic occurrence that gives GIS its true identity (Eastman, 1992). The overlaying operations include applying mathematical expressions and Boolean logic, performing proximity searches and analyzing topography, clustering, and aggregating (Paulsson, 1992). These kinds of analyses require fully-fledged GIS systems.

3.3 Modeling

Modeling allows scientists and planners to build working scenarios of the systems which they are studying to see whether decisions result in the best solutions possible. What happens, for example, to traffic flow or price of adjacent land in the future if a new road is added to an existing network? Answering this type of question requires both geographical and attribute information. A flexible GIS enables users to simulate different scenarios to reach an optimal decision by manipulating these kinds of information.

Modeling, which is a fairly new activity in GIS, helps one to understand processes that act in space. According to Eastman (1992, p.31), modeling

> "... is based on the notion that in GIS, our database doesn't simply represent an environment, *it is an environment!* It is a surrogate environment, capable of being measured, manipulated and acted upon by geographic and temporal process. Our database thus acts as a laboratory for the exploration of processes in the full complexity of true environment. Traditionally, in science, we have had to remove that complexity in order to understand process in isolation."

4. OUTPUTS IN GIS

The final outputs from a GIS include maps, statistical tables, and graphs or charts related to geographical areas of interest. For most users, the important GIS capability is to present data on a map (Landis, 1993). Maps enliven presentations and communicate complex information faster and with more impact than any other graphical tool (see attached Figures). Presentation maps can be used to show viewers spatial patterns of information. Thematic maps, i.e. maps that consist of information on attributes, are essential tools for business, government, research, and education, both for analyzing data and making effective presentations.

Some typical applications of interpretive maps in business activities are to identify trade sites, target marketing areas, design sales territories, monitor sales performance and analyze traffic patterns. City planners can use maps to assess future water consumption and sanitary infrastructure, plan future transport networks, locate social services, or monitor slums and new residential areas. Research groups use them to analyze outbreaks and spread of diseases, water management or to study environmental changes. Schools and universities use maps to train students in geography, engineering, sociology, architecture, city planning, etc.

Most GIS take selected elements of the database and produce map output on the screen or produce hardcopy printouts on a printer or plotter. Many GIS software systems provide only very basic cartographic output, and rely upon the use of other high quality software systems for the production of final maps for publications or computer display.

5. SOME OF THE USES OF A GIS IN STATISTICAL OFFICES

The use of GISs has grown dramatically in the last decade in business, universities, research institutes and governments. GIS can also be powerful tools for statistical offices in data storage and update, and in integrating, analyzing and visualizing geo-referenced information.

Statistical maps and graphs make it easier for the general user to understand the geographic distributions and patterns inherent in statistical tables (see Figures 1 and 2). A GIS can produce statistical maps to show quantitative relationships that are not readily apparent from tabular data. It highlights the important relationships or changes in the data and raises the interest of the reader in examining the tables for further detail. It can be used to show relationships between geographical areas and to indicate the high and low incidence of a variable, such as the incidence of poverty or the literacy rate.

GIS provides the opportunity to locate and show the boundaries of administrative areas for which data are reported in statistical publications. It is possible to present statistics by small areas within major subdivisions, statistical or metropolitan areas. It also allows spatial aggregation. Many GIS software packages can present a thematic map and its associate table in one presentation. Some are able to display thematic maps, statistical graphs, and the output of statistical analysis in one integrated presentation.

Microcomputer GIS offer unlimited possibilities for the statistician to produce his own maps and charts and not be totally dependent on a graphic artist or a cartographer. It allows users to experiment with the data and related maps and to select the most appropriate display of the results of their analyses. It, therefore, reduces cost, and enhances timeliness and efficiency by enabling statistical offices to produce relevant maps at low cost and in time.

Simple trend and pattern analysis can be done using a GIS. It enables the analyst to explore a complex database in search of patterns and relationships that might be linked to geographic processes. Government offices which conduct censuses want to use census data to make comparisons of time periods to study what has changed within an area over time, or to see what spatial pattern exists in certain variables, or to know the anomalies that do not fit the pattern and where they are located. For example, GIS applications can be used by planners in national and local governments to get insight into the socio-economic characteristics of the population and their spatial variation by administrative hierarchy. Features such as population density, total population of a defined area, fertility and mortality rates, illiteracy rate, and income per capita can be geo-referenced in different ways, depending on the aggregation level used in collecting socio-economic features. Migration patterns can be studied in a GIS by linking population movements to the earth's surface, for instance with the place of origin and destination.

Data management functions permit updating the information when necessary and retrieving only the relevant information when required. The capability of taking the output of an analytic process and placing it back into the geographic database for future analysis is extremely important.

The capability to use the geographic database for future analysis is another very important use of GIS. A GIS can help to improve the updating and use of spatial data at all levels of an organization. Maps of enumeration areas (EAs) can be digitized and kept in the database for updating and for use in future censuses or surveys. Geographic data sets for a country or sub divisions can be brought together from the EAs using a GIS. Computerized individual EA maps or their aggregation at different levels of geographical divisions or statistical areas and the base map on which EA's were delineated can be used for planning and management of data collection or to monitor work progress in censuses and surveys.

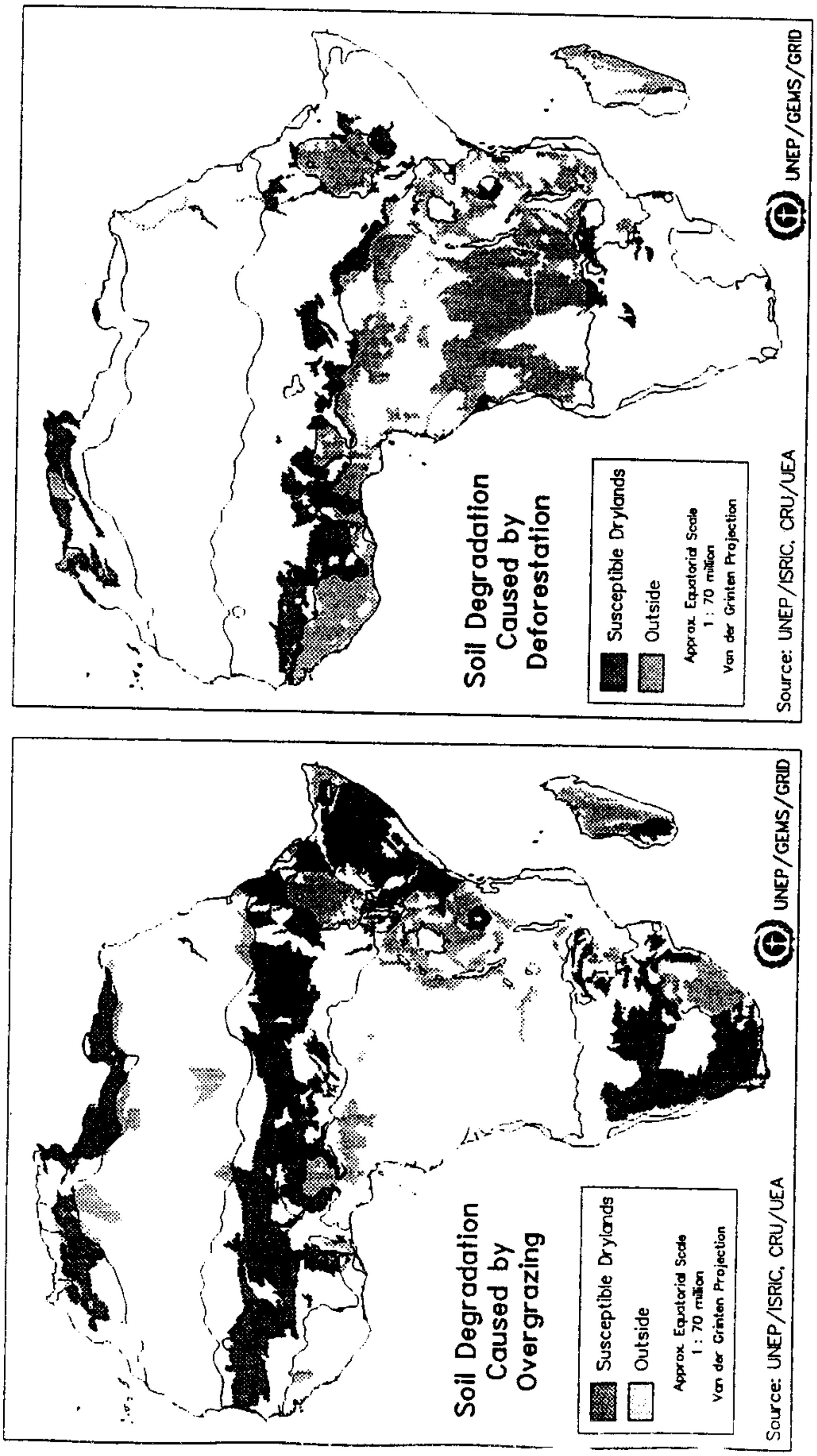
GIS also plays a useful role in the data collection phase of a statistical production. An area sampling methodology for socio-demographic surveys of urban areas was developed by Dureau et al. (1989) where high-resolution satellite imagery was used to optimize the area sampling plan. For stratification, a relationship between field-observed density of built-up areas and reflectance data was established and used for classification. The methodology was first applied on a migration study in Quito, Ecuador, and a detailed handbook was published. The advantages of this approach to institutes and individuals dealing with surveys are lower costs and an accuracy at least equivalent to simple random sampling, with no reliance on any other sampling frame than the satellite data. There is also a thinking that fast population estimates could be made using satellite remote sensing and GIS with some information from field sampling.

6. CONCLUSION

The paper has attempted to introduce GIS to the new comer and to show its use in enhancing spatial information in planning and management. It was also attempted to indicate some of the utilities of GIS to potential users such as statistical offices and research institutes.

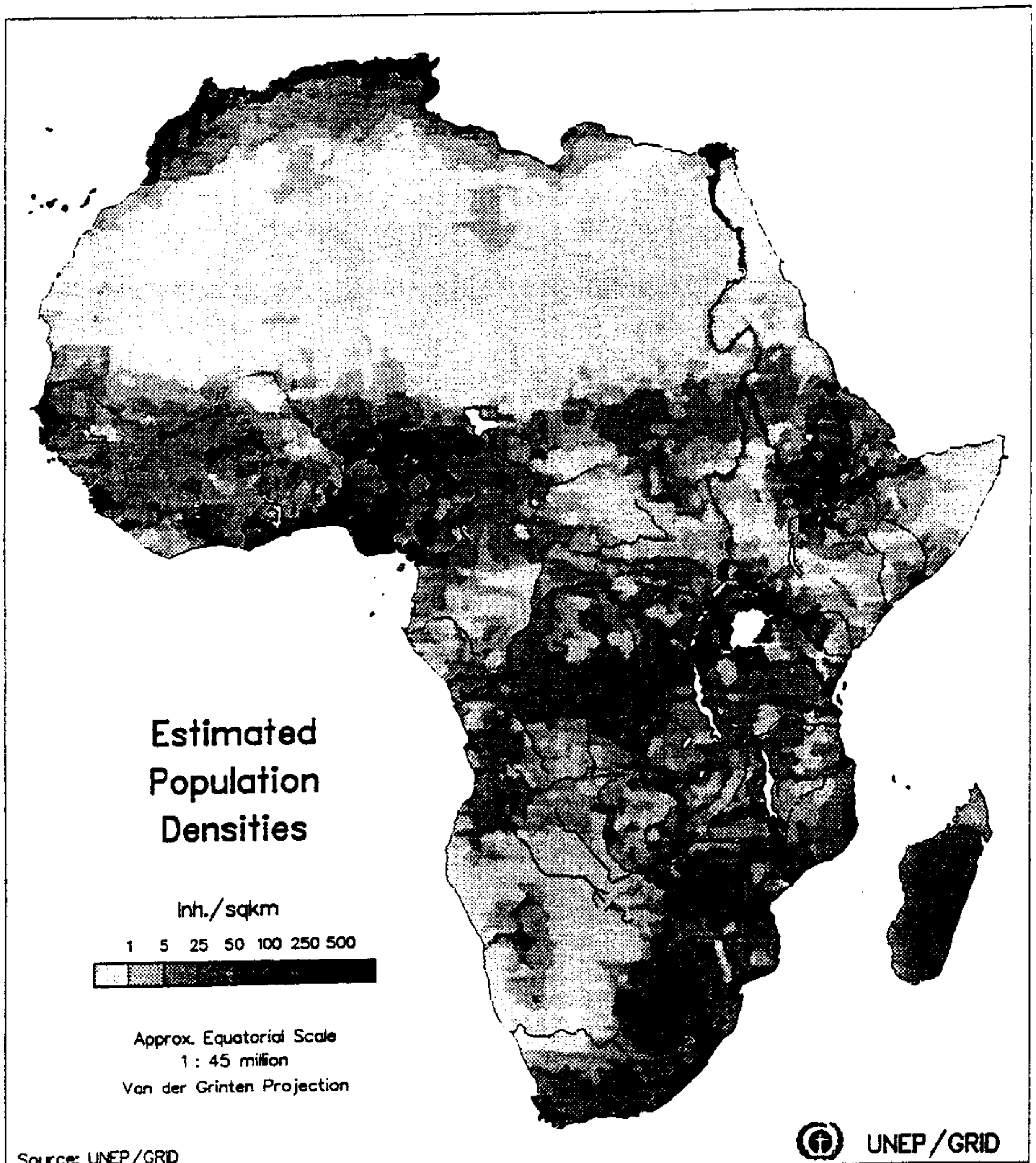
The rapid advancement in computer technology is making micro-computers and related peripherals cheaper and powerful. The decline in prices of computers and the development of less costly, user-friendly and powerful GIS software packages are expected to increase its role in statistical applications. Attempts are already underway to integrate GIS and statistical software packages. GIS is, therefore, expected to become an important tool to statisticians and others interested in socio-economic research in the future.

Figure 1 : Soil degradation causes derived from the GLASOD African dataset



Taken from Deichmann and Eklundh, 1991, pp. 21

Figure 2 : Estimated African population densities



Taken from Deichmann and Eklundh, 1991, pp. 58

REFERENCES

Carter, James R., 1989, 'On defining the GIS' *Fundamentals of GIS: A Compendium*, W.J. Ripple (ed.), American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, Bethesda, pp. 3-7.

Deichmann, U. and L. Eklundh, 1991, *Global Digital Datasets for Land Degradation Studies: A GIS Approach*, United Nations Environment Programme, Global Resources Information Database, Case Study Series No. 4, Nairobi.

Dureau, F. O. Barbary, A. Michel, and B. Lortic, 1989. *Area sampling from satellite image for socio-demographic surveys in urban environments: Training handbook*, Editions de l'ORSTOM. Bondy: France.

Eastman, J. Ronald, 1992, *Idrisi*, Version 4.0 rev.1,Clark University, Graduate School of Geography, Masssachusettes, USA, March.

Kam, Tin-Seong, 1993, 'Visualization: Linking computer graphics with geographical analysis system for resource management and spatial planning' in *Sustainable Natural Resource Management and Spatial Planning in Developing Countries: Geo-Information Technology Perspective*, J.S. Edralin and Kam Tin-Seong (eds.), United Nations Centre for Regional Development, pp. 271-290.

Landis, John D., 1993, 'Gis Capability, Uses and Organizational issues', *Profiting from a Geographic Information System*, Castle, Gilbert H., III (ed), GIS World, Inc. USA, pp. 23-53.

Paulsson, Bengt, 1992, *Urban Applications of Satellite Remote Sensing and GIS Analysis*, Urban Management Programme Discussion Paper, The World Bank, Washington D.C.

Rhind, David, 1989, 'Why GIS' *ARC News*, Reprint, Summer 1989 Issue, Vol. 11, No. 3, Environmental Systems Research Institute, Inc.

Wellar, Barry, 1993, 'Gis fundamentals', *Profiting from a Geographic Information System*, Castle, Gilbert H., III (ed), GIS World, Inc. USA, pp. 3-21.