The Measurements of Data Accuracy and Error Detection in DEM using GRASS and Arc/Info

Sung-Min Cho[1]

GRASS와 Arc/Info를 이용한 DEM 데이터의 정확도와 에러 측정

조 성 민[1]

ABSTRACT

The issue of data accuracy brings a different perspective to the issue of GIS modeling, calls into a question the usefulness of data models such as DEM. Accuracy can be determined by randomly checking positional and attribute accuracy within a GIS data layer. With the increasing availability of DEM and the software capable of processing them, it is worthwhile to call attention for data accuracy and error analysis as GIS application depends on the priori established spatial data. The purpose of this paper was to investigate methods for data accuracy measurement and error detection methodology with two types of DEM's: 1 to 24,000 and 1 to 250,000 DEM released by U.S. Geological Survey. Another emphasis was given to the development of methodology for processing DEM's to create Arc/Info and GRASS layers. Data accuracy analysis with DEM was applied to a 250 sq.km area and an error was detected at a scale of 1:24,000 DEM. There were two possible reasons for this error: gross errors and blunders.

KEYWORDS : DEM, GIS, Data Accuracy, Error Analysis,Arc/Info, GRASS

요 약

GIS 데이타의 정확도 문제는 DEM과 같은 데이터의 유용성과 적용에 대한 서로 다른 견해를 불러일으킨다. 데이터의 정확성은 좌표의 정확한 위치와 속성정보를 무작위적으로 검색하여 결정할 수 있다. DEM은 과거 보다는 손쉽게 취득할수 있고 이를 처리할수 있는 소프트웨어도 다양해 졌으나 GIS의 응용은 이미 만들어진 데이터에 따라 그 결과가 달라질수 있으므로 데이터의 정확도와 에러에 대한 주의를 기울일 필요가 있다. 본 연구의 목적은 1:24,000과 1:250,000 DEM 데이터를 이용하여 DEM의 정확도를 검색하고 데이터가 지닌 에러를 찾아내는 방법을 모색하는데 있다. GRASS와 Arc/Info를 이용하여 DEM을 레이어로 만들어내는 과정 또한 연구 되었다. 연구지역은 250 km^2의 면적을 지녔으며 연구 결과 1:250,000 DEM에서는 실제 등고값이 정상적으로 처리 되었으나 1:24,000 DEM에서는 실제의 등고값이 아닌 0으로 표현된 에러가 발견되었다.

주요어 : 정확도, DEM, GIS, GRASS, Arc/Info, 에러

1998년 7월 15일 접수 Received on July 15, 1998
1 Honam University 호남대학교 도시조경학부 조경학과 (mincho@honam.honam.ac.kr)

INTRODUCTION

Geographic Information Systems(GIS) data are typically acquired in a variety of formats, including graphic and attribute data(i.e. nominal data, descriptive or attribute data) from both printed and digital files, such as LANDSAT or DEM(Digital Elevation Model). The variety of data and formats pose serious problems in comparing and relating spatial data to well-defined specifications and standards for modeling(Feezor et al., 1989). That is, variations in spatial data types, formats, scales, and projections result in data inconsistencies. These inconsistencies can affect the spatial and attribute characteristics of data for GIS application.

The issue of data accuracy brings a different perspective to the issue of GIS modeling, calls into question the usefulness of data models such as DEM. Users must question, outline, and identify the accuracy of location, size, shape, and attribute of the spatial data developed and utilized for GIS. Accuracy can be determined by randomly checking positional and attribute accuracy within a GIS data layer. Although most GIS users are unfamiliar with accuracy measurements it is crucial that the user understand the origin of positional and attribute inaccuracies before employing procedural checks. Most geographical data is uncertain, and it is important to devise ways of describing uncertainty and determining its errors in spatial data. Therefore, with the increasing availability of DEM and the software capable of processing them, it is worthwhile to call attention for data accuracy and error analysis.

OBJECTIVES

DEM is a numerical representation of surface elevations over a region of terrain. Two types of DEMs can be distinguished: gridded DEM and feature DEM. Gridded DEM records a surface elevation for every intersection in a two-dimensional coordinate grid covering the region under investigation. Feature DEM records only a random distribution of elevations across the region, specially those elevations along the boundaries of the Earth's surface features, such as roads and rivers(Swann et al., 1988).

The purpose of this paper was to investigate methods for data accuracy measurement with two types of DEM's: 1 to 24,000 and 1 to 250,000 DEM released by U.S. Geological Survey. Another emphasis was given to the development of methodology for processing DEM to create Arc/Info and GRASS(Geographic Resource Analysis Support System) layers.

MATERIALS AND METHODS

This study was accomplished by using DEM, Arc/Info(ESRI, 1992),and GRASS. In addition, Arc Macro Language(AML) and AWK were used for writing several programs. USGS DEM at a scale of 24,000 was purchased from USGS and DEM at a scale of 1:250,000 was obtained through an Internet anonymous File Transfer Protocol(FTP) account as ANSI-standard ASCII characters in fixed length blocked record format. These DEM's later were imported to Arc/Info and changed to UTM Zone 18 coordinate system from State Plane coordinate system. As DEM's were acquired in ASCII format, it was necessary to generate spatial coverage in Arc/Info and then convert them into a GRASS layer. This process was done to compare two different tools for DEM generation. To permit conversion of Arc/Info coverages to GRASS, several alternative methods were used. DEM's were converted into Arc/Info grid by using ASCIIGRID command. After creating a single DEM, LATTICECLIP in Arc/Info was used to delineate the DEM covering only the watershed area. LATTICECLIP creates a lattice defined by the overlap between a grid and a polygon coverage. An existing watershed boundary was used as a clipping coverage and the outgrid was used as the grid to be clipped. GRIDASCII command in Arc was used to create a simple ASCII text file from a grid. Table 1 shows the ASCII file of DEM generated by GRIDASCII. The file consisted of header information containing key words and values, followed by cell values in run-length order. Header information reorganized by GRIDASCII must be modified to an appropriate format followed by a stream of grid-cell values. For this purpose, an awk program, fixgridascii.awk, was written to reformat the header file (Table 2). The fixgridascii.awk program modified the header file so that R.IN.ASCII command in GRASS could recognize its header format to generate a GRASS layer. R.IN.ASCII converted an ASCII raster text file created by GRIDASCII into a binary raster map layer. To fulfill this process the input file must have a header section which describes the location and size of the data. The header information consists of the origin, the number of rows and columns, and the values used to represent NODATA mesh points(Table 2). This information was followed by a space-delimited list of grid-cell values. The north, south, east, and west field values entered are the coordinates of the edges of the geographic region. The rows and columns field values describe the dimensions of the matrix of data to follow. Figure 1 shows the Broadhead watershed DEM of 1:250,000 generated for GRASS with R.IN.ASCII.

TABLE 1. ASCII header file and digital elevation values generated by GRIDASCII in Arc/Info.

nclos	332
nrows	349
xllcorner	464592.18115985
yllcorner	4535474.0035082
cellsize	92.9200241022554
NODATA-value	-9999
-9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999	
-9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999	
-9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999	
-9999 -9999 -9999 -9999 -9999 -9999 -9999 -9999···	

TABLE 2. ASCII header file and grid-cell values which were modified by using fixgridascii.awk program.

north	4546703.091919999
south	4535474.00350820
east	495441.62916190
west	464592.18115985
rows	349
cols	332
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
548 548 548 548 · · · · · · · · · · · · · · · · · · ·	

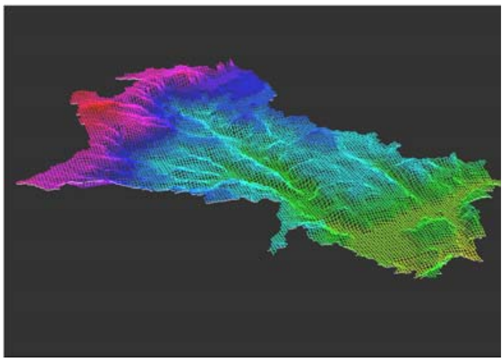

FIGURE 1. 3-D image of 1:250,000 DEM generated for GRASS with R.IN.ASCII.

RESULTS AND DISCUSSION

An error is a difference between the true value and the observed value of a quantity caused by the imperfection of equipment, by environmental effects, or due to imperfections of the observer.

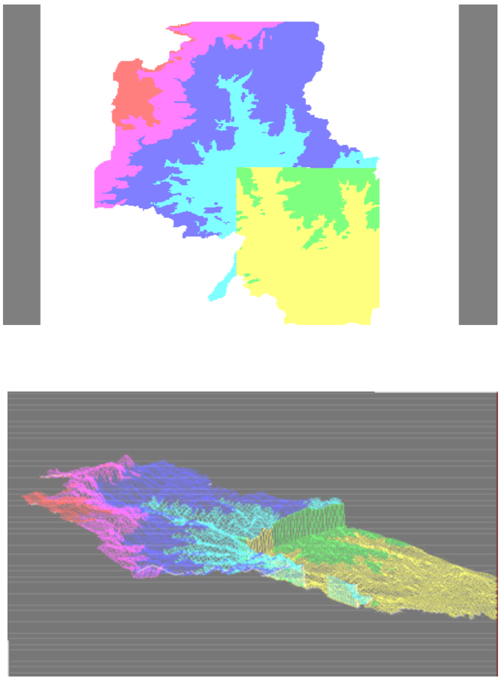

FIGURE 2. These two images show that 1:24,000 DEM did not contain correct elevation values. Light gray color in rectangular box in top image is the area where an error was found. Bottom image shows the same erroneous area resulted in distorted 3-D image.

While processing the collected datasets purchased from USGS into necessary formats, an error was detected in USGS DEM at a scale of 1:24,000. Figure 2 shows that it did not have correct elevations compared to neighboring DEM's. There are two possible reasons for this error: (1) gross errors and blunders, and (2) systematic errors. Gross errors are caused by carelessness or inattention of the observer in using equipment, reading scales or dials or in recording observation(Mikhali et al., 1981, Thapa, 1992).

For example, the observer may bisect the wrong target in angle observations, or may record observations by transposing numbers, e.g., by writing 82.23 instead of 28.23. Gross errors may also be caused by failure of equipment. According to Mikhail and Gracie (1981), some of the techniques used in detecting and eliminating gross errors include taking multiple readings on scales and repeating the whole measurement and checking for consistency.

Systematic errors or blunders occur in accordance with some deterministic system which, if known, may be represented by some functional relationship(Thapa, 1992; Moore et al., 1991). In surveying and photogrammetry, systematic errors occur because of environmental efforts, instrumental imperfections, and human limitations. Some of the environmental effects are humidity, temperature, and pressure changes. These factors affect distance measurements, angle measurements, and GPS(Global Positioning System) satellite observations. Instrumental errors include lack of proper calibration and adjustment of instruments. Systematic errors must be detected and observations must be corrected for systematic errors or they must be corrected by some mathematical model. In a statistical sense, systematic errors include bias in the observations. Unlike gross errors, they cannot be detected or eliminated by repeated observations. Therefore, if systematic errors are present, the measurements may be precise but they will not be accurate.

CONCLUSION

One current trend in GIS is towards the analysis of data accuracy from diverse sources and in different formats(Piwowar et al., 1990). Most systems available today provide a simple solution to this problem by supplying the user with a few external conversion programs to accommodate these data. While these conversion routines are able to bring together some disparate forms of data, they tend to be an extra barrier to be crossed prior to performing any analytical operations such as hydrological feature generation with DEM's.

Through this research, pre-processing and post-processing procedures of DEM's were developed to simplify the use of integrated GIS. Attention has been focused on data accuracy analysis because it is the first, and often insurmountable, hurdle to be overcome when beginning GIS application. Data accuracy measurement with DEM's was applied to a 250 sq.km watershed in USA and an error was detected at a scale of 1:24,000 DEM. It did not have correct elevations showing only 0 values compared to neighboring values. There were two possible reasons for this error: gross errors and blunders.

In conclusion, it was revealed that corrective actions must be taken to minimize the errors in DEM; DEM must be viewed and edited before GIS application. KAGIS

REFERENCES

ESRI. Environmental Systems Research Institute. 1992. Understanding GIS: The Arc/Info method. ESRI, Inc. Redlands, CA.

Feezor, D. R., M. Hirschi and B. J. Lesikar. 1989. Effect of cell size on AGNPS prediction. ASAE paper No.89-2662. St. Joseph, MI.

GRASS. Geographic Resource Analysis Support System, Version 41. 1994. U.S. Army Corps of Engineers, Construction Engineering Research Lab. Champaign, Illinois.

Mikhail, E. M., and G. Gracie. 1981. Analysis and adjustment of survey measurements. Van Nostrand Reinhold Company, New York. pp.340.

Moore, I. D., Grayson, R. B., and Ladson, A. R. 1991. Digital terrain modeling: A review of hydrological, geomorphological and biological application. Int. J. of Hydrological Process 5(1):3-30.

Piwowar, J. and LeDrew, E. F. 1990. Integrating spatial data: A user's perspective. PE&RS 56(11):1497-1502.

Swann, R., D. Hawkins., A. Westwell-Roper, and W. Johnstone. 1988. The potential for automated mapping from geocoded digital image data. PE&RS 54(2):187-193.

Thapa, K. 1992. Accuracy of spatial data used in Geographic Information Systems. PE&RS 58(6):835-841 KAGIS