REMOTE SENSING AND GEOGRAPHIC INFORMATION SYSTEMS DATA QUALITY PERSPECTIVES

Hilda Maria Clauzet Ferraz de Mello
Neusa Dias

Faculdade de Tecnologia de São Paulo
Depto. de Edifícios
Pça. Cel. Fernando Prestes 30
São Paulo, SP - CEP 01124
Pho: +55-011-229-5847
Fax: +55-011-228-0123
BRAZIL
Commission III

ABSTRACT

Remote Sensing data have been used increasingly as GIS source data.

Decisions based on inaccurate and imprecise data are likely to implement uncorrect lines of action. Ass soon as Remote Sensing, GIS and connected acessories will be in current use to manage resources and solve problems, the utmost problem will be a capability to describe and prescribe products to be used by managers.

This paper aims at a general study of the sources of errors connected to data acquisition, conversion and presentation of the final product when Remote Sensing is integrated directly to GIS. The present methods are used to quantify this errors as the data are processed and are studied so as to overcome the existing problems.

KEY WORDS: data base, data quality, GIS--LIS, remote sensing

INTRODUCTION

The total error contained in a GIS analysis can be minimized by the experience of errors related to the products source and to the errors generated by the data handling function.

Generally, the Geographic Information Systems user is no conscious of his basic products errors source, neither of the inadequacy of the creation and acquisition of integrated data in a GIS. These users shall have a stronger experience of that errors possibilities in the product sources and of how these errors multiply through these data analysis, so that they can appraise the inadequacy of the obtained data and accept them or not.

Due to the wide amplitude of its applications, the remote sensing is nowadays an important space information acquisition technique used in Geographic Information Systems.

REMOTE SENSING AND GEOGRAPHIC INFORMATION SYSTEM DATA INTEGRATION

For the data obtained by remote sensing to be used as a direct input for a GIS the space grade and the resolution must be appropriate to the GIS and its application. This question has been debated for years and, in the future, with the new orbital systems, a satisfatory solution may be found to attend to the GIS necessities.

The largest existent problems with data derived from remote sensing, specially the orbital ones, are the accuracy of the classification scheme used, which normally remains under the desired level for the GIS data and the location accuracy of the remote sensing information, which frequently is lower than the required for a GIS data. In many cases this error combinations lead remote sensing data to be rejected for operational use in a GIS.

REMOTE SENSING AND GEOGRAPHIC INFORMATION SYSTEM DATA INTEGRATION

The remote sensing data acquisition process in a Geographic Information System consists of the procedures of acquisition, error ratio determination and final product presentation. The error may accumulate in the process in an additive or multiplicative way and isolated error processes may be hidden by larger errors.

Theoretically, the errors input in each stage can be estimated; although this is made only at the end, as a whole. Usually the decision making is based in graphic final products, statistic data or modeling results, which may or not contain indications related to the accuracy of the generated information, what limits the accuracy in the decision implementation. It is necessary a large hability to quantify the errors connected to the data and monitor these errors through the applications of a Geographic Information System.

Acquisition

It is important to have an acquaintance with the kind and quantity of possible errors in all the data sources and, if possible, its control.

Some data acquisition errors are common to any data collecting way, as in the local obtained data, so as in the ones obtained by distance measures. Others, like geometry and radiometry errors, can be controlled. But the largest difficulty consists on the error evaluation and quantification caused by the human subjectivity during the data analysis and interpretation.

Concerning to the geometry and radiometry errors, some aspects must be considered so that we can keep track of them.

The processing of several data levels in a Geographic Information System database is based on an adjusted and accurate space register between data levels. Therefore, it is essential that all the remote sensing data is geometrically adjusted with the same cartographic projection of the GIS database. The modern aerial photography has developed digital image processing techniques, which have contributed for the integration between GIS and remote sensing. They provide a perfect basic mapping applicable to the remote sensing images.

The ilumination geometry may affect the quality of the image and the subsequent analysis. The ideal is the ilumination to be constant or close to constant through the whole image. In practice, although, the acquisition requires a definition on the total width of the viewing field, which results in a series of geometries and ilumination measures. The systems are dependent on the solar ilumination and azimute conditions may expressively limit the duration of the suitable acquisition windows.

Scale variations are inserted by the aerial camera leaning, ground plane moving and other factors. To make the image input easier in a GIS the image scale must be left constant.

The kind of sensors like the aerial cameras, the multispectral scanners and the airplanes side looking airbone radars influence on the geometric variation.

The geometric image depends upon the involved sensor. The hability of obtaining geometric accuracy is too much more developed in the conventional photogrametric, which is based on the aerial vertical photography. However, several other remote sensing types involve image continuous regeneration processes. These processes are more susceptible to geometric distortions and they can obstruct the integration with the Geographic Information Systems. The geometric error inserted by each one of these sensors can be qualified and removed, or properly minimized before the remote sensing data input into the GIS database.

The flight height or the orbit height of a remote sensing platform, together with the viewing fields and look angle and direction affect the image geometry, as seen before. In an additional interest, we have the platform speed and direction, orientation and height. The largest distinction of these parameters can be done between airplane platforms and satellite platforms. The airplane platforms are subject to uprightness and highness variations due to the atmospheric turbulences. The airplane highness and location are essential information for the remote sensing data to be accurately corrected and inserted into the Geographic Information System.

The precise location of the remote sensing corrected data or the final map depends on the control points on which positions are known in a space reference coordinate system. Besides the conventional survey techniques, procedures and recommendations of how taking photographies datuns, projects and precision patterns must be considered in the photogrametric applications.

Although it is fundamental for the error values determination, there isn't yet patterns and procedures developed to relate the geometric accuracy of remote sensing images. To directly input remote sensing data into a GIS it is necessary that some patterns are defined and adopted so that they allow the definition of control points and error ratios calculation. The development of the GPS technology, allowing high precision in the settlement of the ground coordinates can bring big contribution for the standardization.

Effects correction in some of the specific scenes are usually made during the photogrametric mapping. For instance, radial distortions due to the atmospheric refraction can be calculated and removed for a standard atmosphere, so as ground curvature effects. These kinds of effects are stronger at high altitudes, but they can change the location accuracy even in low altitudes.

Basic support points and atmosphere characteristics are pertinent to photogrametry, but they can be developed for other remote sensing applications. The understanding of these and the influence on films and digital products are important for the correct analysis and interpretation of these kind of data.

Data Processing

Digital image processing techniques are used for geometric corrections of remote sensing data and when the distortions are minimum, it is possible to use simple algorithms based on polynomials.

The geometric corrections of remote sensing digital data usually involve some kind of resampling, like contrast raise, cubic or linear convolution. These and other kinds of resampling affect the radiometric integrity of data and its space appearance has to be better knowed. Techniques for improving the geometric processing have been developed using different crossed correlation methods and multiple space images. Although, most of these techniques require developing of more and more sofisticated image processing environments.

The space digital data processing, sometimes, involves some kind of data conversion. It is possible to resample the data to such a degree that the geometric attributes resampled have little relation to the original data.

In terms of accumulated error, the largest impediment for the space data analysis comes from a lack of well documented methods and from a lack of statistic tools integrated with the existent software packages. Several software packages are organized in a hierarquical way, with limited statistic options, allowing the choice of one or two classifiers with limitations for the parameters defined by the user. As a result, inexperienced analysts may simply follow the software hierarch, using the standard options, not thinking on what is happening to the data. Flexible statistic tools must be developed for enumerating the particular difficulties inherent to the space data joints.

Data Analysis

The data analysis, in the remote sensing data flow and Geographic Information Systems, involves the relationship between the data variables and the subsequent interferences that can be developed. This error acumulation stage focus on the statistic techniques validation. Difficulties on the statistic analysis of space data source involves the typical adoption of the common linear model, composed by the space autocorrelation effects. The data analysis is also subject to errors proceeding from the variability of the analysis specialists. This variability involves the choice of pre-defined relevant variables or the synthesis of new multiple parameter variables, correlated or not.

The classification systems can be a source of error in the remote sensing data integration in Geographic Information System. Some of the potential error sources induced by the classification systems are related to the inability of the systems to categorize mixed classes, transition zones or ambigous classes definitions, the human subjectivity and the lack of compatibility between different classification systems used with remote sensing data and traditional data types. Thematic data levels created using general remote sensing data require the use of some kind of classification system to facilitate the data categorization for subsequent space data analysis by GIS. When we work with mixed pixels, or polygons and transition zones, or dynamic systems, inconsistent classification occur with the whole classification system. This inserts an error element which is particularly difficult to be quantified.

The data generalization is usually made during the remote sensing analysis by two reasons: space resolution and/or thematics or spectral data resolution. The space generalization involves the analysis to produce the minimum unit of the map. The sampling for a space resolution better than the original results in a substantial error. Spectral generalization can be executed by filters that enhance certain characteristic like the frame or homegenize similar pixels. Some filters preserve the frames, while reduce the noise. Although, because some filters change the original value of the pixels, some errors, like the precise frame location or the loss of spectral similar, still the only resource, may happen.

The data generalization after classification takes two patterns, spacial and thematics. The thematics generalization is the classes grouping to create significant homogeneus categories. Because of this, it is made with the analyst interference and the tendency errors can be introduced and information can be lost if the analyst does not recognize a unique resource.

Data Conversion

By the growth of the use of the Geographic Information Systems and the necessity of incorporating digital remote sensing data as a quick and reliable information source, it was inevitable the necessity of converting matrix data into vectorial format data.

Matrix format data are data arranged in a symmetrical spacing and same grid size. Satellite data are common examples of matrix data. These data are easily loaded into a computer, as a number matrix.

Vectorial data keep the actual shape of a polygon, using series of vertexes connected by lines. Vectorial data are preferred for the representation of the most thematic maps in the GIS, due to the uniform lines for the frame viewing. Most part of the mapped products, including the result of photo-interpretation is usually represented in vectorial format.

Unfortunatelly, there is significative error in the conversion from matrix to vectorial format and from vectorial to the matrix format. The magnitude of these errors depends on the algorithm used in the conversion process,on the cell grid used and on the orientation used for the matrix representation. By no considering this error it is possible to insert considerable problems in the analysis.

Error Avaliation

The quantitative error analysis can be done during any stage of the data processing, including data acquisition. Theoretically, an error value must be determined after each stage of the analysis, but in remote sensing projects, the error value ratios are determined only after the complete data analysis is made and usually directed to the thematic and location accuracy.

In determining the accuracy ratios an important point to be considered is the sample sizes. The high cost of each sample point carts the reduction of the sample size to its minimum, but keeping enough size to validate any statistic analysis. There are several recommendations and different equations to define the appropriate sample size.

The sample scheme used is an important factor in the accuracy evaluation. The error matrix must be representative for the whole classified image, and the choice of an unappropriated scheme (poor) may result in a tendentious error matrix, causing over or underestimation of the real accuracy.

It is said that space autocorrelation occurs when the presence or the absence, or the degree of a certain characteristic afects the presence, absence, or degree of the same characteristic in the neighbor units. This condition is particularly important in the evaluation of the accuracy for positive or negative error influences in neighbor locations.

The location accuracy in remote sensing can be expressed as RMSE (Root Mean Square Error), which is derived from the georeferencing algorithms that rectify images to the map coordinates. The RMSE is the square root of the average of the square of the errors and reflects the proportion or maximum or minimum pixels number that the control points of the image differ from the map or from the reference control points. Although,the RMSE not always reflects the accurate location of all the pixels in an image. The RMSE addresses only the control points and only related to the map. The most accurate way of examining the location accuracy is a topographic control with GPS data, which has a hig implementation cost.

The most common way to represent the thematic accuracy or classification generated by remote sensing is by an error matrix. An error matrix is a square arrengement of numbers belonging to a particular category related to the current category, as the one verifyed on the soil. The columns usually represent the reference data, while the lines indicate the classification generated by the remote sensing data. An error matrix is an effective way to represent the accuracy because the accuracy of each category is fully described with the inclusion and exclusion errors, present in the classification. The error matrix can be used as a point of departure for a series of statistic descriptive and analytics measures.

The two most common thematic accuracy measures use binomials probabilities or agreement Kappa ratio. Binomial probabilities are based on the correct percentage, so that they do not separate inclusion and exclusion errors. On the contrary, the Kappa ratio gives a different measure between the observed agreement of two maps and the agreement made by chance.

The Kappa advantages are that its calculation considers the elements out of the error matrix diagonal and that the Kappa conditional ratio can be calculated for individual categories. In order to standardize procedures for reports and for thematic statistical maps, the error matrix shall present and include the inclusion percentage error by category, the exclusion percentage error by category, the total correct percentage, number of sampled points, map accuracy and the statistic Kappa.

Final Product

The objective of most investigations of remote sensing and Geographic Information Systems is to produce a product which gives important accurate and quick information for scientists and administrators. The product may have several configurations, including thematic and statistic summary.

The thematic maps may contain statistic and dynamic information. A statistic thematic map is produced by analyzing the information collected in a unique date, while a dynamic map must produce the changes occured between succeeding observation dates.

In order to reduce the error of the final product there are important procedures for these maps generation. A substantial amount of error can be removed if the reader is provided with a complete cartobibliographic citation, i.e., the genealogy or lineage of the map products. In some remote sensing software packages there are methods for tracking the processing flow for a particular datafile. The general proposition has been to create a historical file by listing all the operations and parameters that have been applicable to a data join. Other kinds of error can be reduced by using good cartographic design principles in the generation of the map products, like the legends.

Geometrical error in final thematic map products can be inserted by the use of base maps with different grades, different national horizontal datum in the source materials and different minimum mapping units that are, then, resampled for a final minimum mapping unit.

It is commanding the improving of map legends that include cartobibliographic information of the geometrical nature of the original source material. This is the only way to allow the readers to judge the geometric accuracy of the thematic maps final products.

The final map must be uniform in its accuracy even being the addition of information from several sources. It is important for this map reader to know which of these sources are reliable thematic sources. There is a large necessity of standardization of the project and of the functions of the reliable thematic diagrams.

The fundamental principles of the cartographic projects must be followed specially in the building of the classes interval legends for thematic maps. More and more, the remote sensing and GIS information are presented in electronic viewing device and excessive classes intervals and poor colour selection variations produce poor cartographic communication on the CRT visualizers.

While a lot of progress has been done on the statistic thematic maps, dynamic changing detection maps almost always have poor legends. Too much research is necessary to make possible the report of the ocurred changes, accurately, to the reader.

Several scientists, nowadays, have superposed vectorial images with matrix images. This powerful technique gives a generic basic map that the reader can use to guide and evaluate the vectorial data. Unfortunately, there is no standardization in relation to the optimum viewing conditions for the bottom image or to the optimum project of the vectorial data. Researches are necessary to standardize and provide, as products, thematic maps which incorporate a matrix/vectorial integration.

Decision Making

The decision making is frequently presented with maps or statistic products, derived from remote sensing and GIS. In most situations, appropriate information concerned to the lineage of the thematic data layers and accuracy associated to thematic and geometry are not given. Besides this, the decision making needs the total precision estimate and the reliability of the products used in the process, although decision making is provided with little or no acquaintance about the potential error sources and no information concerning to final accuracy and reliability level.

There is a defined tendency between many decision makers of always accept products, maps and statistic reports as truthful. For the fact of several remote sensing and GIS products being thematic products, there is an enormous error potential in a decision making, for overestimating the accuracy and the thematic reliability level of the products. It is necessary that the remote sensing and GIS communities instruct the administrators to better understand the error sources associated to the products.

Implementation

Decisions based on innaccurate data and improper reliability level have a great probability of making incorrect decisions. The obvious implications of an incorrect decision are wrong actions of resource administration, which can mean serious consequences for the resource itself, causing its loss or degradation, adverse impacts on a particular ecosystem or its elements, impacts on the human health, besides the possibility of fines or other punitive actions.

As products derived from remote sensing and GIS are raisingly being used as decision basis for resources management and as problems regulator, there is a high potential for an explosion of the number of litigations. The biggest challenge for the remote sensing and GIS community will be the hability of properly portray and defend the accuracy and reliability of products used by administrators in the process implementation.

CONCLUSION

It is necessary a considerable amount of research and development to be made before the errors associated to remote sensing and GIS integration can be properly quantified and expressed in standard formats.

The objective of standard error reports is to provide an evaluation method of the adjustment of GIS products derived from remote sensing for specific applications and to facilitate the comparison between several research results.

The present procedures for remote sensing errors evaluation were adapted from statistic procedures which were not specifically developed for space data. These techniques have been adapted and have been reasonably good for small areas, but their application for regional or global grades is not economically viable. For the fact that the existent techniques refer to the global accuracy, the error space distribution is not evaluated. There must be developed techniques to evaluate the error space structure.

The philosophy and the recommendations for the acquisition of good field data to evaluate the maps accuracy has not been well addressed. Basic researchs must be developed on the accuracy levels associated with different ways of field verifying.

The classified satellite digital images (representation pixel-by-pixel) are easily filled in the matrix format but hardly converted into the vectorial format.

Numerous rules have been done in order to control the process of matrix data conversion into vectorial, but the effects on surface, size and accuracy of the polygons, when compared to the original matrix data, haven't been studied with severity yet. It is critical to research the effects of the conversion of remote sensing digital data.

Additional information is required on the characteristics of the location errors in remote sensing and the correlation between locations and classification errors. It is necessary more knowledge on the definition of alternatives for remote sensing platforms and on how much the GPS technology will contribute for the accuracy of remote sensing data location.

The final maps and statistical products must be standardized to provide information related to the accuracy and reliability associated to the specific data product.

REFERENCES

Aronoff, S., 1985. The minimum accuracy value as an index of classification accuracy. Photogrammetric Engineering & Remote Sensing 51(1):99-111.

Blakemore, M., 1984. Generalization and error in spatial databases. Cartographica 21:131-139.

Croswell, P.L., 1987. Map accuracy: what is, who needs it, and how much is enough? Papers from the Annual Conference of the Urban and Regional Information Systems Association Vol. II., Fort Lauderdale, FL, pp. 48-62.

Dicks, S.E., and Lo, T.H.C., 1990. Evaluation of thematic map accuracy in a land-use and land-cover mapping program. Photogrammetric Engineering & Remote Sensing 56(9):1247-1252.

Goodchild, M., and Gopal, S., 1989. Accuracy of spatial databases. Taylor & Francis, London.

Lunetta, R.S., Congalton, R.G., Fentersmaker, L.K., Jensen, J.R., McGwire, K.C., and Tinney, L.R., 1991. Remote sensing and geographic information system data integration: error sources and research issues. Photogrammetric Engineering & Remote Sensing 57(6):677-687.

Newcomer, J.A., and Szajgin,J., 1989. Accumulation of thematic map error in digital overlay analysis. American Cartographer 11(4):58-62.

Rickman, D., Ochoa, M.C., Holladay, K.W.,and Huh, O.K., 1989. Georeferencing airbone imagery over news deltas in Lousiana. Photogrammetric Engineering & Remote Sensing 55(8):1161-1165.