science for a changing world

Methods to Achieve Accurate Projection of Regional and Global Raster Databases

By Dr. E. Lynn Usery, Dr. Jeong Chang Seong,
Dan R. Steinwand, Michael P. Finn

Open-File Report 01-383

U.S. Department of the Interior
U.S. Geological Survey

METHODS TO ACHIEVE ACCURATE PROJECTION OF REGIONAL AND GLOBAL RASTER DATABASES

Principal Investigators

E. Lynn Usery, USGS, Rolla, MO
Jeong Chang Seong, Northern Michigan University, Marquette, MI
Dan R. Steinwand, USGS, EROS Data Center, Raytheon ITSS, Sioux Falls, SD
Michael P. Finn, USGS, Rolla, MO

This research aims at building a decision support system (DSS) for selecting an optimum projection considering various factors, such as pixel size, areal extent, number of categories, spatial pattern of categories, resampling methods, and error correction methods. Specifically, this research will investigate three goals theoretically and empirically and, using the already developed empirical base of knowledge with these results, develop an expert system for map projection of raster data for regional and global database modeling. The three theoretical goals are as follows:

(1) The development of a dynamic projection that adjusts projection formulas for latitude on the basis of raster cell size to maintain equal-sized cells.
(2) The investigation of the relationships between the raster representation and the distortion of features, number of categories, and spatial pattern.
(3) The development of an error correction and resampling procedure that is based on error analysis of raster projection.

HYPOTHESES

Regarding the first goal, we hypothesize that regional and global raster data can be accurately projected with appropriate equations that account for raster cell size and latitudinal position. For the second goal, we hypothesize that scale factors explain the impact of distortion on raster representation and that more categories and more complex spatial patterns cause more errors. Finally, we hypothesize for the third goal that error correction and resampling methods can be used for optimizing the projection accuracy of regional and global raster datasets. This proposed research potentially has impacts on all U.S. Geological Survey (USGS) programs involving the use of large regional and global raster data sets, such as Global Change Research and Place-Based Studies.

A complete description of this research project is included in USGS Open-File Report 01-181 (Usery and others, 2001). This current report will document project accomplishments during the first year, include the workplan for the second year, and document publications accomplished or planned from the research.

FY 2001 ACCOMPLISHMENTS

The research plan includes four major components: (1) develop a projection selection DSS; (2) develop a dynamic projection; (3) develop and implement an error theory model; and (4) develop a new categorical resampling method. Significant progress has been made on each of the four components of the research during the first year of the project. These accomplishments are detailed below and provide a basis for completing the research in fiscal year 2002 meeting all objectives stated in the original proposal. In addition, the project team saw the need to add a fifth component in the second year. That component will draw from the work of the other four components and examine the feasibility of designing a system for global raster data storage and analysis.

Decision Support System for Map Projection Selection

A Web-based system for supporting map projection selection has been designed and an initial prototype implemented. This prototype DSS is for projections of small-scale datasets, such as those with regional, continental, or global geospatial extent. The goal is to help users select a suitable projection when using a commercial geographic information system (GIS) software package by considering the characteristics of their data, particularly image or raster data, and their intended uses.

The prototype, in its current state, focuses on user input of geographic area (regional, continental, or global) and geometric characteristic preservation (shape or area) to drive the suggested projection or projections (fig.1). Somewhat similar to a design described by Jankowski and Nyerges (1989), further implementation of the DSS will expand the user input to include latitude (in regional cases), data type (raster or vector), volume or resolution, and other appropriate options. A tutorial on map projections will also be included in the final design and development.

The software for the current prototype is being developed as a combination of Java[1] programming language and Perl scripts, using Hyper-Text Mark-up Language and the Common Gateway Interface. This allows users to employ the browsers of their own choice and execute the DSS interactively across the Web independent of their hardware platform.

---

[1] Java is a trademark of Sun Microsystems, Inc

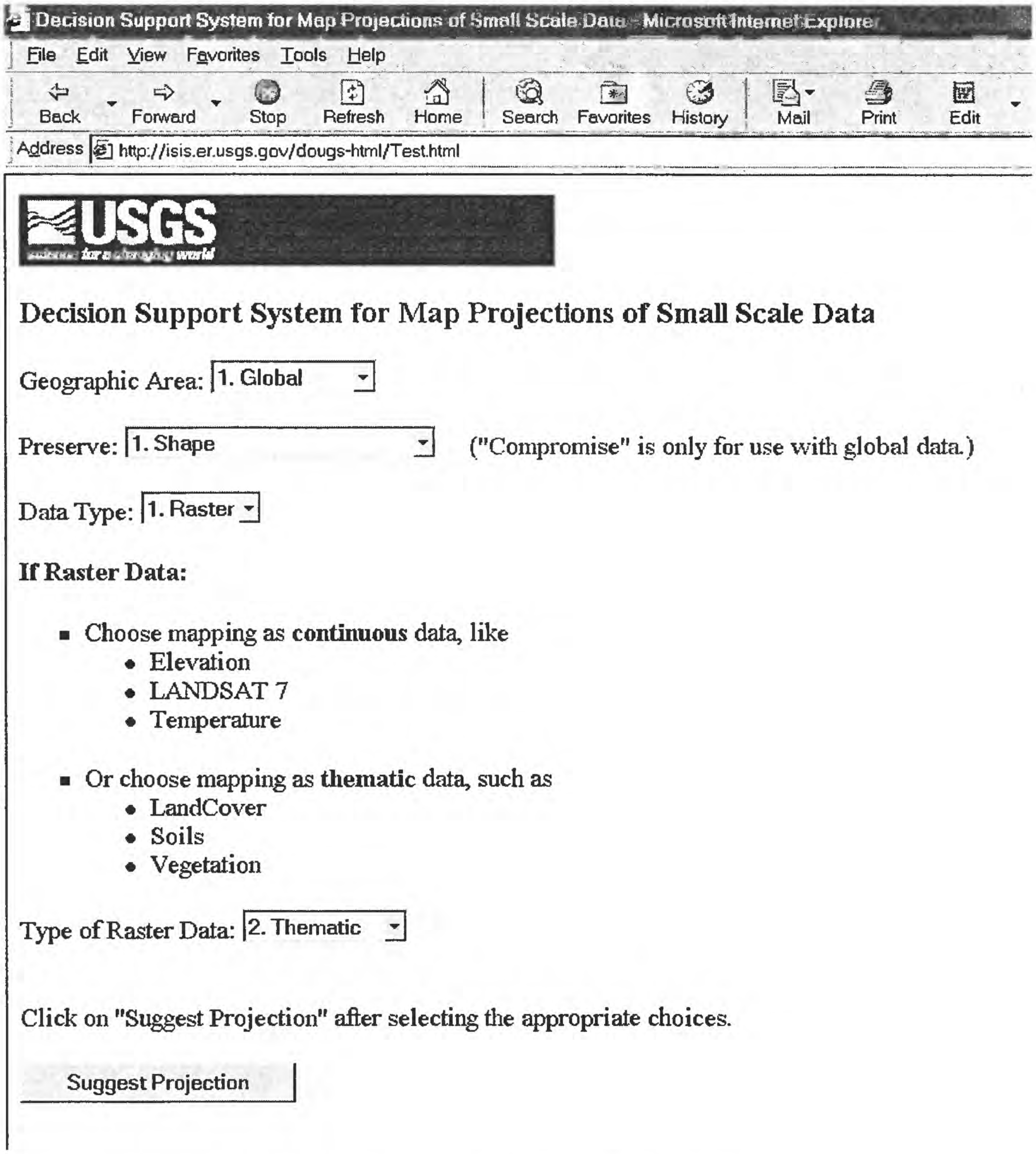


Figure 1. Screen shot from the prototype DSS for map projection.

Dynamic Projection

The approach to dynamic projection to preserve global areas in a raster representation began by establishing a standard of truth to judge the results of empirical projection transformations. That truth was established by developing global raster datasets in spherical coordinates with resolutions of 1-degree and 30 arc-seconds. The area of each cell in these datasets was then computed through a numerical integration procedure. The result is a dataset in spherical coordinates in which each cell contains a digital number value equivalent to the area of the cell. These cell values represent truth with respect to the actual ground area of the cells. In previous work (Usery and Seong, 2001), the accuracy of raster data was compared to a vector representation; thus, the current approach captures real world areas more precisely. To apply this concept to empirical data, such as vegetation categories, we summed the area values for occurrences of each vegetation type, thus yielding exact areas for each vegetation category in a worldwide distribution. These total values computed accurately from spherical coordinates are the basis for comparison with projected data of the same type. Figure 2 presents the results of tabulating global vegetation at 1-degree resolution for seven different projections compared with the global totals for each land cover obtained from the computation in spherical coordinates.

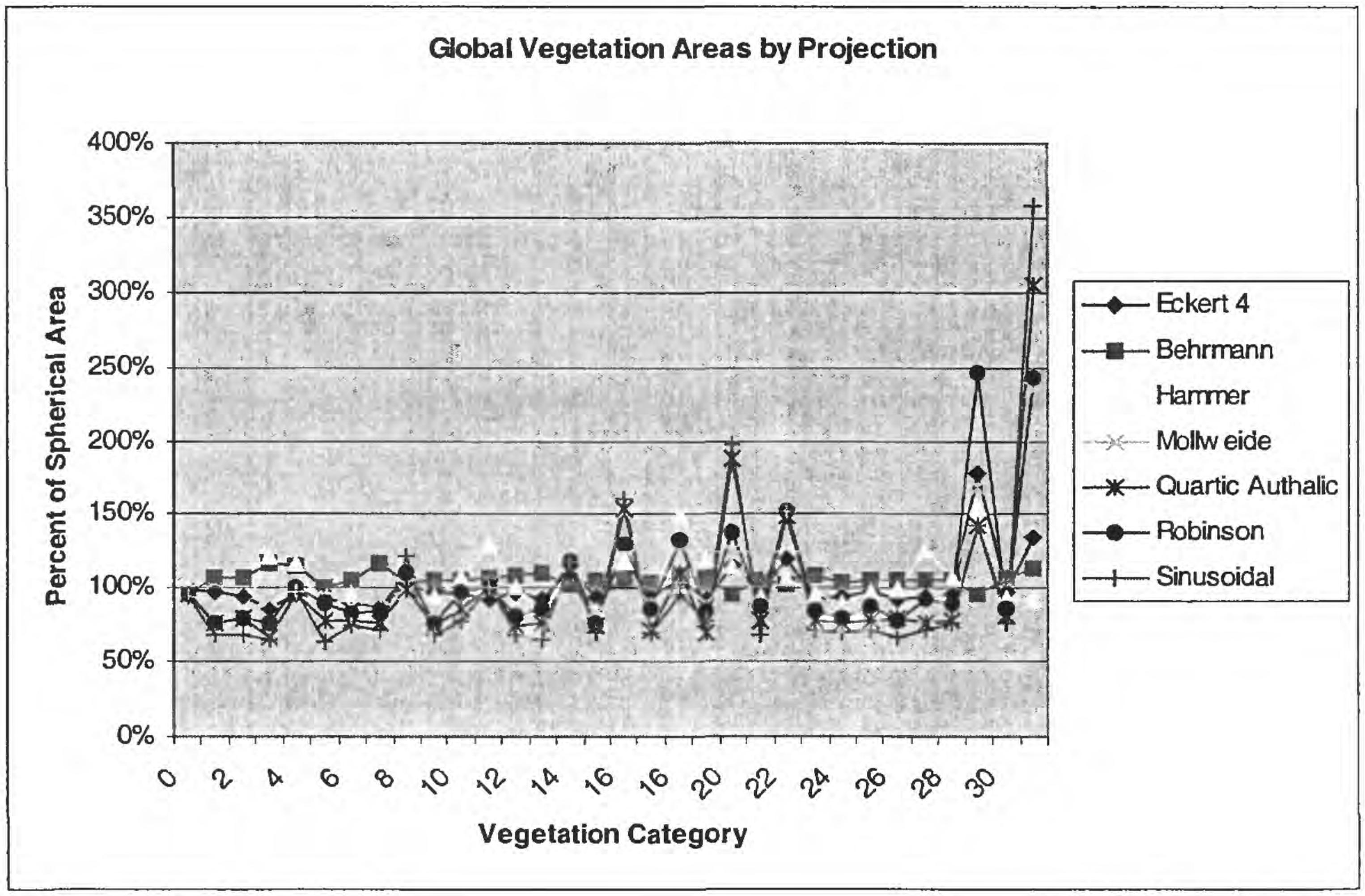


Figure 2. Variance of areas of global vegetation categories from actual values computed by numerical integration in spherical coordinates.

With the above basis, we developed an approach to project each raster line in spherical coordinates to a corresponding raster line in plane coordinates. While this approach allows maintaining the exact area from spherical coordinates in the plane representation, it prohibits joining adjacent raster lines since only pixels in a single raster line in plane space have the same areas. Thus, we can dynamically change the size of the projected pixel areas to match the corresponding spherical areas, but only single raster lines can be displayed. The logical completion of this approach is to resample all projected lines to achieve a common pixel size. The resampling can take advantage of accomplishments under the resampling task discussed below.

Error Theory

The scale factor model approach to error in raster projection has been evaluated for specific projections and shows promise in modeling error resulting from replicated and lost pixel values in the transformation process. The theoretical examination revealed that error results in two forms; areal size change of pixels and categorical error resulting from loss or duplication of pixels. A scale factor model, based on the horizontal and vertical scale factors of the projection, was developed to provide a computation of the resulting error from specific projections. According to the scale factor model, the raster representation accuracy is as follows:

$$\text{Raster Representation Accuracy of Distorted Features} = 1 / \text{Maximum Scale Factor}$$

The model was experimentally tested with the Equal-Area Cylindrical, Sinusoidal, and Mollweide projections. Results indicate that the model predicts error within 1 percent of actual values and that the Sinusoidal projection is subject to smaller errors in projecting raster data than the other projections tested.

The scale factor model was also applied in image reprojection in which significant pixel value change was observed. A scale factor model for reprojection was developed, and model results were compared with experimental results. Six possible reprojections among the Equal-Area Cylindrical, the Mollweide, and the Sinusoidal were tested. Results show that reprojection accuracy can be explained using the ratios of scale factor changes along vertical and horizontal axes between source and target projections. The reprojection accuracy was modeled as the reciprocal of the maximum scale factor change along either the vertical or horizontal axis:

$$\text{Reprojection Accuracy} = 1 / \text{Maximum} \left(X2 / X1,\ Y2 / Y1 \right)$$

where 1 and 2 represent the original and target projections, respectively. X is the horizontal scale factor, and Y is the vertical scale factor. The model explains reprojection accuracy very well. Figure 3 shows six reprojection situations and model and experimental accuracies. The model accuracy, however, is very sensitive to the skew effect that leads to significant error increase. Spatial autocorrelation and the number of unique pixels were found to affect the accuracy of reprojection.

Because the Sinusoidal projection appeared most accurate, it was investigated in more detail. The Sinusoidal projection showed 99.50 percent and 98.35 percent categorical accuracies when 54 sample datasets were reprojected from the Universal Transverse Mercator (UTM) to the Sinusoidal, and from the Sinusoidal to the UTM, respectively. The accuracies were much higher than other projections, such as Mollweide, Eckert IV, Hammer-Aitoff, and Geographic. Table 1 shows the result of reprojection accuracies between global projections and the UTM, a very accurate local projection.

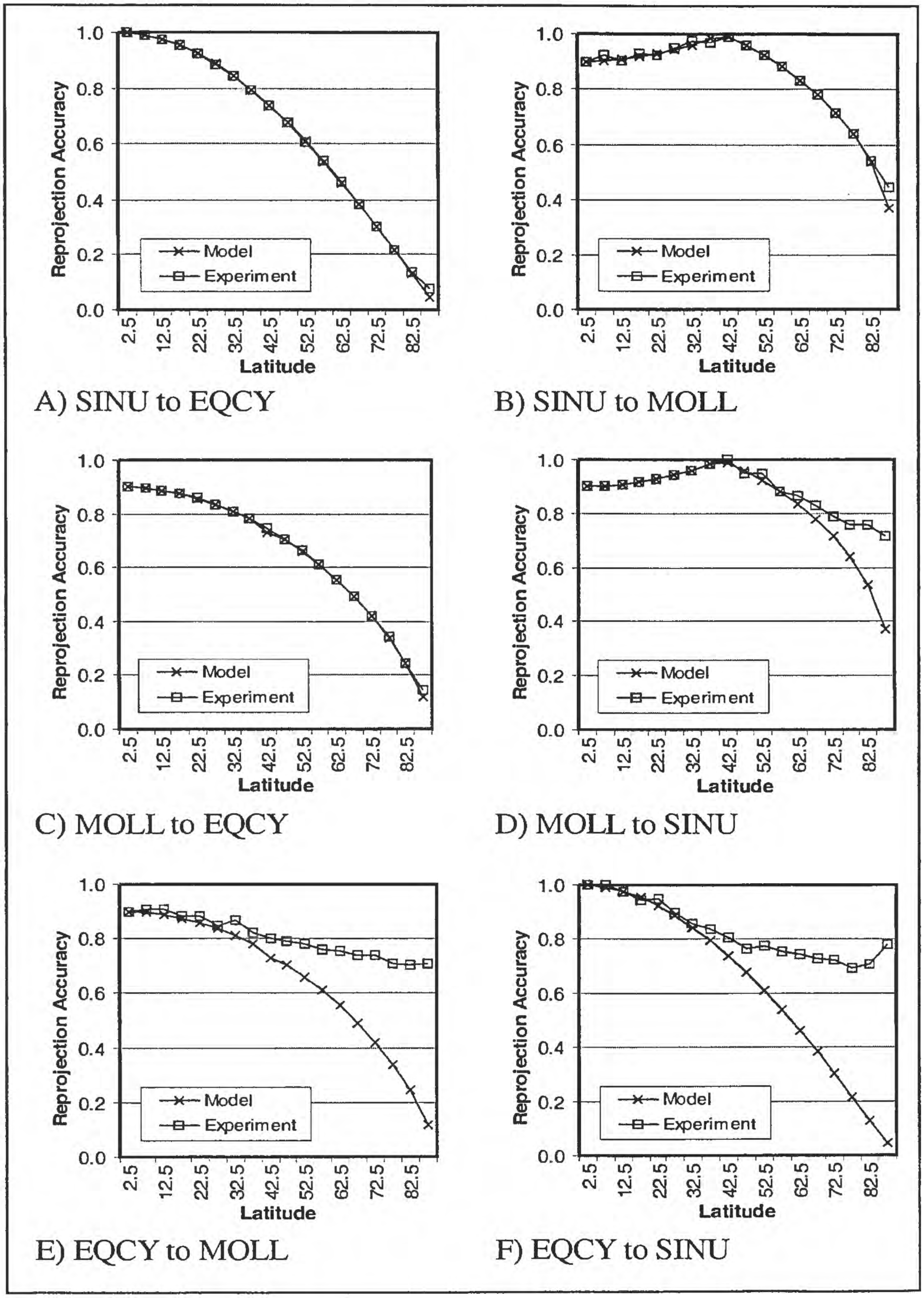


Figure 3. Pixel value changes during reprojection

Table 1. Reprojection accuracies between global and local projections

Original Projection	Target Projection	Minimum Accuracy (%)	Maximum Accuracy (%)	Average Accuracy (%)
UTM	Sinusoidal	98.5	99.9	99.5
UTM	Mollweide	48.2	99.6	85.8
UTM	Eckert IV	19.7	99.8	76.9
UTM	Hammer-Aitoff	53.2	100.0	87.0
UTM	Geographic	100.0	100.0	100.0
Sinusoidal	UTM	90.5	100.0	98.4
Mollweide	UTM	68.0	99.8	89.5
Eckert IV	UTM	29.9	99.3	85.2
Hammer-Aitoff	UTM	59.9	100.0	87.2
Geographic	UTM	9.2	99.9	65.3

In addition, the effect of skewness and number of categories was investigated with sample datasets. Figure 4 shows the extent of increase of raster representation error owning to skew effect. It shows that raster representation accuracy systematically and dynamically changes depending on skew angle and maximum scale factor. The findings were presented at the Association of American Geographer's (AAG) 2001 Annual Meeting and Conference (New York, N.Y.) and the American Society for Photogrammetry and Remote Sensing (ASPRS) 2001 Annual Conference (St. Louis, Mo.). Results of this research also are being made available through journal publication (Seong and Usery, 2001).

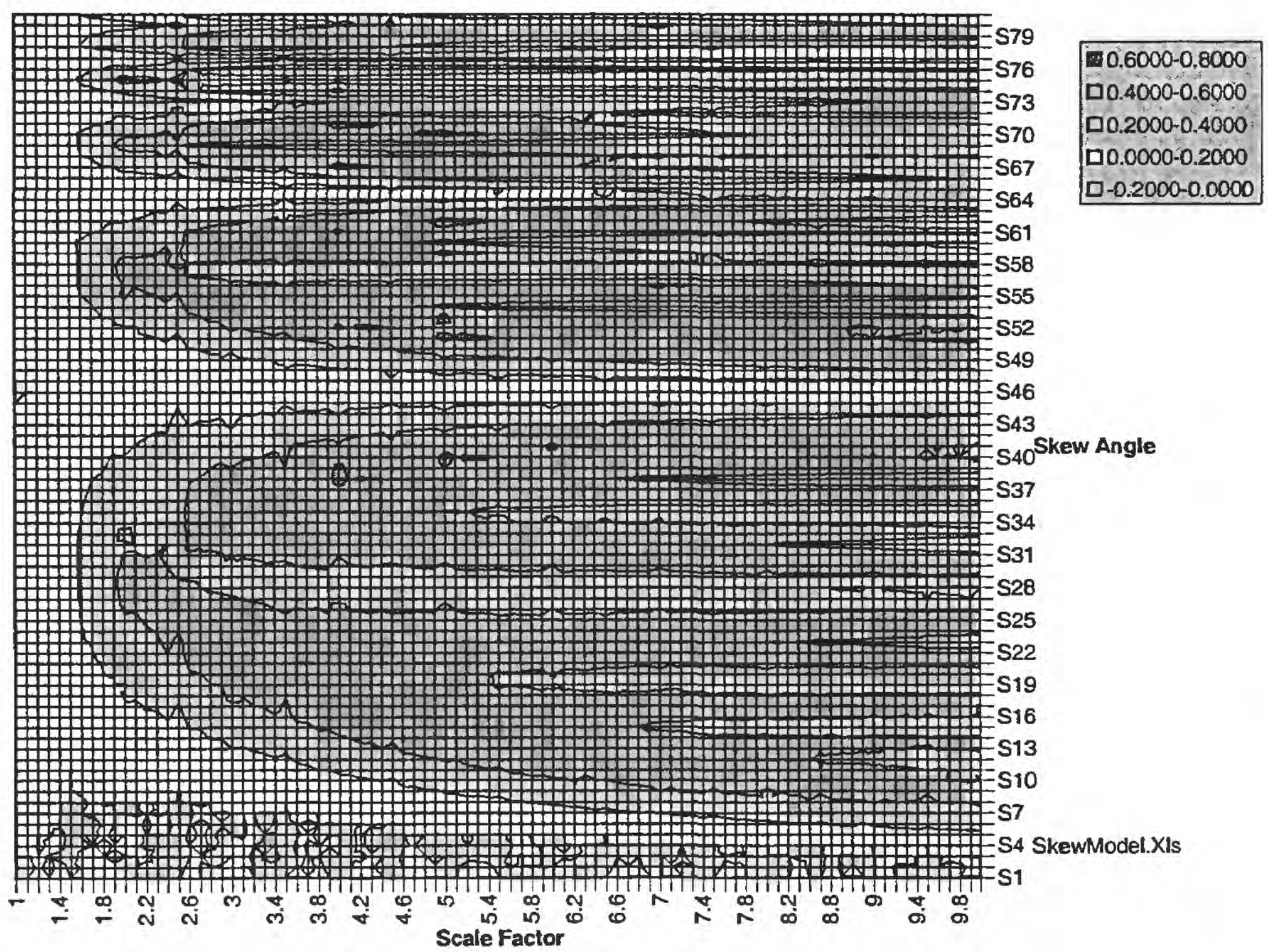


Figure 4. Accuracy increase due to the change of skew angle

Resampling

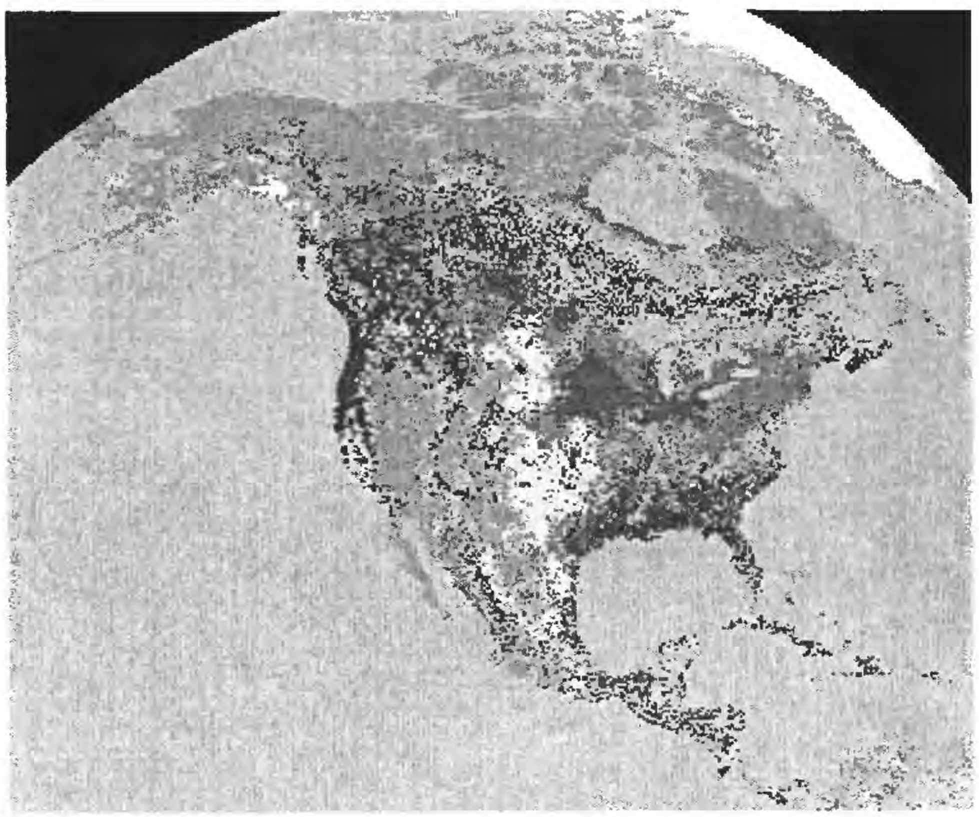
A new method for resampling categorical data has been developed. Commonly available methods for resampling categorical data (class or binned data, not signal-based, remote sensing data) typically consist of nearest neighbor-like resampling methods. These methods are chosen because their alternatives--cubic convolution, bilinear interpolation, and so on--are interpolating methods, which do not maintain categorical values. Furthermore, the geometric distortions present in the projection change for data of global extent are far greater than those that occur in moderate- to high-resolution remote sensing data. Indeed, most of the software tools available today were designed for single-scene, signal-based remote sensing image data, where the image extent usually extends only a few hundred kilometers, rather than for datasets of global or continental extent.

The typical nearest neighbor algorithm for categorical resampling takes the center (or upper left corner) of a pixel as point data and reprojects that coordinate into the new projection space. Because the resulting coordinate is often not at an exact pixel location, it is rounded to the nearest pixel position and that pixel's value is used to populate the resulting output image. Although this method is computationally efficient, it can result in imagery that is not representative of the original image owning to varying amounts of geometric distortions present in the transformation.

The new resampling method treats pixels not as points, but as areas. In summary, the new algorithm maps the corner coordinates of the image pixels between the two projections and determines the number of input image pixels that go into making the resulting output image pixel. If one pixel (or less than one complete pixel) goes into the making of the output image pixel, the nearest-neighbor approach is used. More often, multiple input image pixels go into one output image pixel. In this case, simple statistical methods (minimum, maximum, average, and so on) are used to determine the output pixel value. This many-to-one condition is present when the output image pixel size is greater than that of the input. This condition is also present in transformations of similar pixel sizes where geometric distortions are great (for example, an image in a projection where the pole is represented as a line, to a projection where the pole is represented as a point.) The algorithm also keeps track of pixel values that were present in the input area, but not used. These are written to a statistical image that can assist the end user in determining the extent of data loss or misrepresentation.

The two images in figure 5 illustrate the output of the algorithm. The first image was processed using nearest neighbor resampling and appears "noisy." The second image was processed with the maximum occurring pixel method and appears smoother.

Details of this algorithm and the tools developed to test and apply it are explained in the forthcoming article "A New Approach to Categorical Resampling."



(a) Nearest Neighbor Resampling

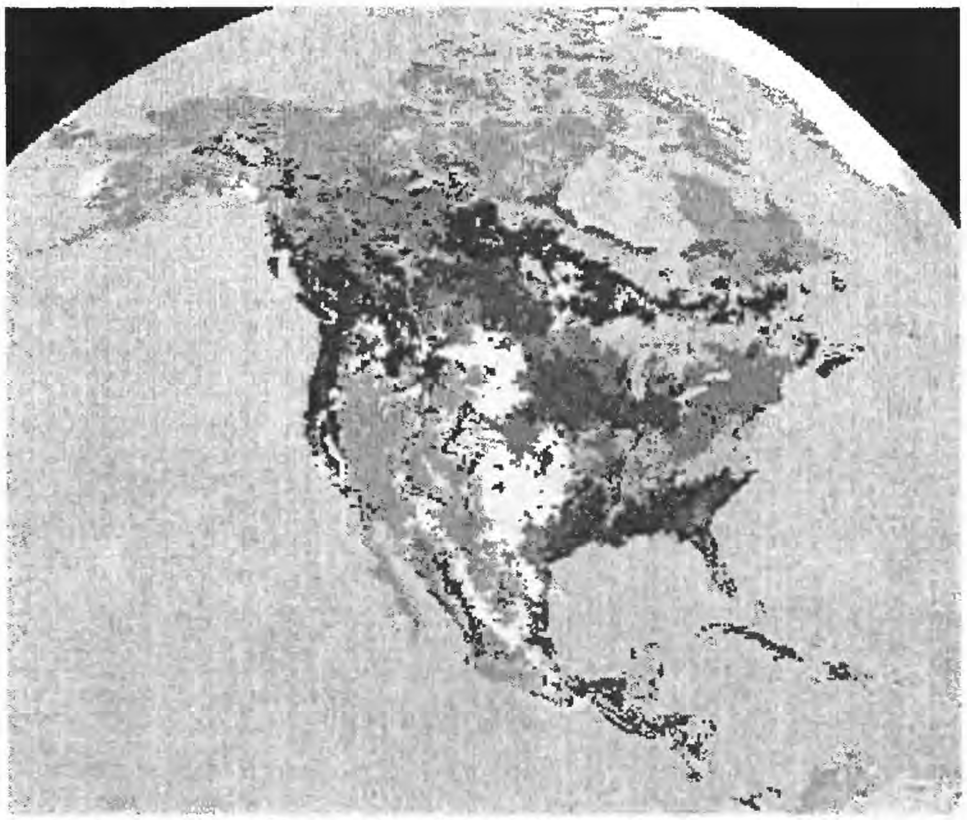

(b) Maximum Occurring Pixel Method

Figure 5. Example results from the new categorical resampling algorithm.

WORKPLAN FOR FISCAL YEAR 2002

Task 1: Projections Decision Support System (DSS)

Subtasks

(1) Refine the current prototype to integrate empirical research (from GRA Task 741) and arrive at a specific recommendation for a projection on the basis of a full suite of user input. (Version 0).
(2) Create an independent Web-based map projections tutorial.
(3) Integrate a map projections tutorial with DSS (Version Beta 1).
(4) Do public testing and obtain feedback of Version Beta 1.
(5) Incorporate feedback, modify Beta 1 DSS, and generate final Version 1.

Milestones/Products

(1) Version 0 DSS
(2) Map projections tutorial
(3) Version 1 Beta DSS
(4) Documentation of public comments
(5) Version 1 DSS

Task 2: Dynamic Raster Projection

Subtasks

(1) Apply pixel areas generated from spherical coordinates to empirical data (land cover, and so on) to compute the actual total areas to be preserved.
(2) Investigate the integerized sinusoidal projection.
(3) Develop a mathematical base for the data structure to preserve global raster areas computed in subtask 1.

Milestones/Products

(1) Actual areas of land cover by category.
(2) Recommendation of compatibility of integerized sinusoidal projection with dynamic projection goals.
(3) Mathematical formulas for the transformation of each raster line.

Task 3: Error Theory

Subtasks

(1) Complete the investigation of the effect of number of categories, pixel resolution, and skewing (publication of initial results in progress).
(2) Continue the application of scale factor model to other projections – global, continental, and the nonequal-area class of projections.
(3) Model the total minimum accuracy in each projection.

Milestones/Products

(1) Publication of Subtask 1 results.
(2) Documentation of the results of the application of the scale factor model.
(3) Table of accuracies.

Task 4: Categorical Resampling

(1) Complete the testing of the resampling algorithm developed in fiscal year 2001

Milestones/Products

(1) Publication of the resampling algorithm

Task 5: Global Raster Data Storage and Analysis System Design

Subtasks

(1) Integrate the results from Tasks 2, 3, 4 to establish a design for a Global Raster Data Storage and Analysis System.
   a. Investigate a user interface/visualization method
   b. Design an analysis system with a projection that is invisible to users

(2) Elicit public review and comment of the Global Raster Data Storage and Analysis System to determine its implementation feasibility.

Milestones/Products

(1) Visualization and analysis design.
(2) Documentation of public results.

PUBLICATIONS ASSOCIATED WITH THE RESEARCH PROJECT

Usery, E.L. and Seong, J.C., 2001, All equal-area projections are created equal, but some are more equal than others: Cartography and Geographic Information Science, 28/3, p.183-193.

Seong, J.C. and Usery, E.L., 2001, Modeling raster representation accuracy using a scale factor model, Photogrammetric Engineering and Remote Sensing, 67/10, p.1185-1191.

Seong, J.C., Mulcahy, K.A., and Usery, E.L., In press. The Sinusoidal Projection: A new meaning for global image data," The Professional Geographer.

Seong, J.C., Modeling the accuracy of image data reprojection," submitted to International Journal of Remote Sensing.

Usery, E.L., Beard, T., Bearden, M., Cox, J.D., Finn, M., and Ruhl, S., Projecting global databases to achieve equal areas for modeling applications," in work, to be submitted to Cartography and Geographic Information Science.

Planned:

Steinwand, D., A new approach to categorical resampling, to be submitted to Photogrammetruic Engineering and Remote Sensing.

Finn, M., Usery, E.L., Seong, J.C., and Steinwand, D., A decision support system (DSS) for global raster data projection, to be submitted to Geographical Systems

Usery, E.L., and Finn, M., Dynamic projection of raster data, to be submitted to The International Journal of Geographical Information Systems.

Seong, J.C., Projection effects of number of categories, pixel resolution, and skewing, to be submitted to The International Journal of Remote Sensing.

REFERENCES

Jankowski, P. and Nyerges, T., 1989, Design considerations for MaPKBS- Map Projection Knowledge-Based System: The American Cartographer, 16/2, p.85-95.

Seong, J.C. and Usery, E.L., 2001, Modeling raster representation accuracy using a scale factor model, Photogrammetric Engineering and Remote Sensing, 67/10, p.1185-1191.

Usery, E.L. and Seong, J.C., 2001, All equal-area projections are created equal, but some are more equal than others: Cartography and Geographic Information Science, 28/3, p.183-193.

Usery, E.L., Seong, J.C., Steinwand, D., 2001. "Methods to achieve accurate projection of regional and global raster databases," *U.S. Geological Survey Open-File Report 01-181.*