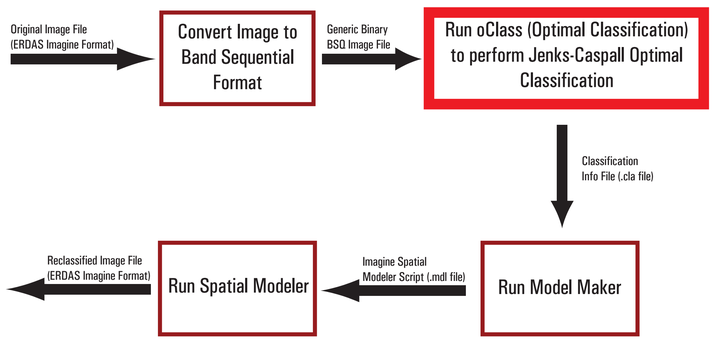
An Implementation of the Jenks-Caspall Algorithm for Optimal Classification of Data for Geographic Visualization

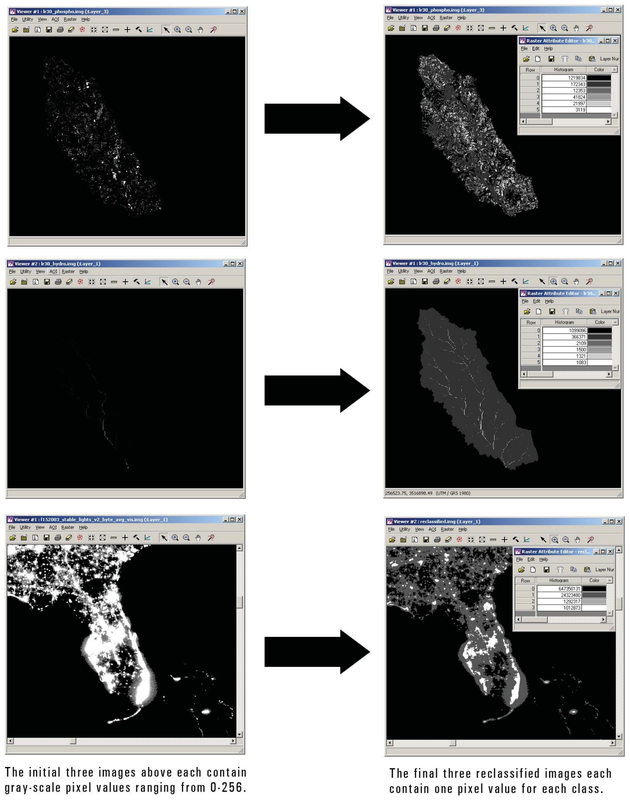
Michael P. Finn, Michael S. Williams, and E. Lynn Usery

We have developed a program to implement the Jenks-Caspall optimal data classification algorithm to assist researchers in creating geographic visualizations of spatial data, or in breaking down continuous datasets into classified sets to facilitate the complex analysis often associated with geospatial data analyses (Jenks and Caspall, 1971). Visualization is a form of graphical communication. Two of its sub-components, scientific visualization and information visualization, generally involve using computers to explore large multivariate data sets, or visual representations and analyses of abstract information (McCormick and others, 1987; Card and others, 1999; Slocum and others, 2005). In the geospatial sciences, geographers have noted the lineage of geographic visualization, or geovisualization, from the roots of cartography and have defined it either broadly as visual analysis of a paper map or a map created on a computer screen, or, more narrowly, as "private" activity in which unknowns are revealed in a highly interactive environment (MacEachren and others, 1992; MacEachren, 1994; MacEachren and others, 1999; Slocum and others, 2005).

Image Reclassification



MacEachren (1994) holds up Ferriea and Wiggins's (1990) "density dial" to manipulate class break points as an exemplar of this geovisualization where a user engages in "visual thinking" through private, high interaction, exploratory map use. Hallisey (2005) asserts that cartographic visualization must be included along with other analyses to achieve a comprehensive analysis of spatial data. Kraak (1999) categorizes functions within a geographic information systems environment for cartographic visualization as query, re-expression, multiple views, linked views, animation, and dimensionality. Cartographers have used data classification schemes for years to aid in the construction of thematic maps.

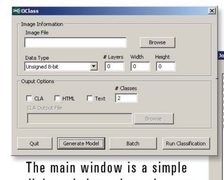


The initial three images above each contain gray-scale pixel values ranging from 0-256.

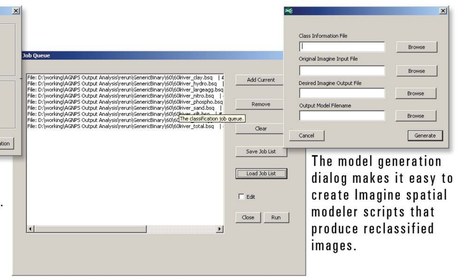
The final three reclassified images each contain one pixel value for each class.

The Jenks-Caspall algorithm is a method for improving thematic mapping as a communicative tool (Jenks and Caspall, 1971) and, therefore, as a visualization tool. The Jenks-Caspall algorithm is an optimal classification method that can assist geographers and cartographers in creating geographic visualizations of spatial data. Data classification involves combining raw data into groups that can be represented by a unique symbol. Optimal classification does the best job of evaluating how data are distributed along the number line. Cartographers traditionally have used the result of a data classification to produce choropleth maps. The objective of the Jenks-Caspall algorithm is to place similar data values in the same class by minimizing the sum of the absolute deviations about class means (Slocum and others, 2005). The algorithm iterates through the raw data and reclassifies them based on this objective.

Our program, optimal Classification (oClass), is a useful tool in geovisualization. It is consistent with Ferriea and Wiggins's "density dial," and can execute two of Kraak's important exploratory visualization functions, namely re-expression (by changing the display of data in some way) and multiple views (by using sets of maps for different classes to enable a holistic view of the data). We have used oClass successfully in research associated with optimally classifying output from environmental models.



The main window is a simple dialog window where the user enters all required parameters.



The batch window allows the user to perform multiple classification jobs in sequence.

The model generation dialog makes it easy to create Imagine spatial modeler scripts that produce reclassified images.

References

Card, S. K., J. D. Mackinlay, and B. Shneiderman (editors) (1999). Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, San Francisco, California.

Ferriea, J., and L. Wiggins (1990). The density dial: a visualization tool for thematic mapping. GeoInfo Systems, Volume 1, Pages: 69 – 71.

Hallisey, Elaine J. (2005). Cartographic Visualization: An Assessment and Epistemological Review. The Professional Geographer, Volume 57, Number 3, Pages: 350 - 364.

Jenks, George F., and Fred C. Caspall (1971). Error on choroplethic maps: Definition, measurement, reductions. Annals of the Association of American Geographers, Volume 61, Number 2, Pages: 217 – 244.

Kraak, Menno-Jan (1999). Visualising spatial distributions. In Geographical Information Systems (Volume 1: Principles and Technical Issues), Second Edition, edited by Paul A. Longley, Michael F. Goodchild, David J. Maguire, and David W. Rhind. Pages: 157 – 173. New York: John Wiley & Sons, Inc.

MacEachren, Alan M. (1994). Visualization in modern cartography: Setting the agenda. In Visualization in Modern Cartography, edited by Alan M. MacEachren and D. R. Fraser Taylor, pages 1 – 12. Pergamon, Oxford, England.

MacEachren, A. M., B. P. Buttenfield, J. B. Campbell, D. W. DiBiase, and M. Monmonier (1992). Visualization. In Geography's Inner Worlds: Pervasive Themes in Contemporary American Geography, edited by R. F. Abler, M. G. Marcus, and J. M. Olson, pages 99 – 137. Rutgers University Press, New Brunswick, New Jersey.

MacEachren, A. M., M. Wachowicz, R. Edsall, and D. Haug (1999). Constructing knowledge from multivariate spatiotemporal data: Integrating geographical visualization with knowledge discovery in database methods. International Journal of Geographical Information Science, Volume 13, Number 4, Pages: 311 – 334.

McCormick, B. H., T. A. DeFanti, and M. D. Brown (1987). Visualization in scientific computing. Computer Graphics, Volume 21, Number 6.

Slocum, Terry A., Robert B. McMaster, Fritz C. Kessler, and Hugh H. Howard. (2005). Thematic Cartography and Geographic Visualization, Second Edition. Prentice Hall, Upper Saddle River, New Jersey.