GLOBAL SCALE DATA MODEL COMPARISON

A. Jon Kimerling, Kevin Sahr, and Denis White
Department of Geosciences
Oregon State University
Corvallis, OR 97331, USA
(We acknowledge support from U.S. EPA -OSU Coop. Agreement CR 821672)

ABSTRACT

Transforming raw observations into globally regular sampling grids or surface tessellations is a fundamental data processing and storage problem underlying much of our global data analysis. The basic geometry of traditionally employed quadrilateral-based point or area grids, while well suited to array storage and matrix manipulation, may inherently hinder numerical and geostatistical modeling efforts. Several scientists have noted the superior performance of triangular point grids and associated hexagonal surface tessellations, although no thorough evaluation of global data model alternatives has been conducted. In this paper we present results from a global grid comparison study that focused on recursive tiling of polyhedral faces projected onto the globe. A set of evaluation criteria for global gridding methods were developed. Of these, metrics for spheroidal surface area, compactness, and centerpoint spacing were found to be of particular importance. We present examples of these metrics applied to compare different recursive map projection-based and quadrilateral spherical subdivision tilings. One map projection approach, the Icosahedral Snyder Equal Area (ISEA) recursive tiling, shows particular promise due to its production of equal area hexagonal tiles on the spheroid at all levels of recursive partitioning.

INTRODUCTION

A new era of high spatial and temporal resolution environmental data covering the entire globe is about to begin, ushered in by NASA's Earth Observation System (EOS) and other global data collection efforts such as the 1km AVHRR, land cover, and DEM data sets being compiled as part of the International Geosphere - Biosphere Program's Data and Information System (Eidenshink and Faundeen 1994, Hastings 1996). We should expect that earth scientists will accelerate their use of geographic information systems (GIS), numerical modeling approaches, and geostatistical methods, singly or in concert, to study global scale phenomena such as climate change and biodiversity loss. Such analyses will require both spatial and temporal integration of currently disparate data sets from a wide variety of data producers.

Transforming raw observations into global data models comprised of geometrically regular sampling grids or surface tessellations is a fundamental data processing and storage problem underlying global data analysis. One fundamental problem is that "regular" sampling grids or surface tessellations devised for the earth's surface, such as the ETOPO5 5 minute DEM or the NASA Earth Radiation Budget Experiment (ERBE) 2.5° global modeling grid, cannot be extended to the entire earth without losing regularity in both surface area and shape. Alternative approaches beg investigation.

An ancient realization is that subdividing a sphere with total regularity of surface area and polygonal shape within the tiles formed by the subdivision can be achieved only by projecting the faces of one of the five Platonic polyhedra (tetrahedron, hexahedron, octahedron, dodecahedron, icosahedron) onto the sphere. Further partitioning of any face will produce unavoidable variations in surface area, shape, or both.

Equally important is the realization that the basic geometry of commonly employed quadrilateral point grids or surface tessellations, while well suited to array storage and matrix manipulation, may inherently hinder numerical and geostatistical modeling efforts. Scientists have noted the superior performance of triangular point grids and associated hexagonal surface tessellations for numerical analyses central to studies of fluid dynamics, percolation theory, and self-avoiding walks. Additionally, hexagonal tessellations are favored by influential statisticians involved with developing survey sample designs and geostatistical methods such as Kriging .

It is clear that a thorough evaluation of alternative global data models is needed. We take a first step in this direction by presenting examples of results from a global grid comparison study funded by the U.S. Environmental Protection Agency (White et al. 1992). Comparisons are predicated on evaluation criteria, such as those presented below.

GLOBAL DATA MODEL COMPARISON CRITERIA

We believe that an ideal general purpose global data model would consist of n points and n areal cells on the globe and have the following properties:

1. Areal cells constitute a complete tiling of the globe, exhaustively covering the globe without overlapping.
2. Areal cells have equal areas.

3. Areal cells have the same topology.
4. Areal cells are the same shape.
5. Areal cells are compact.
6. Edges of cells are straight in some projection.
7. The edge between any two adjacent cells is a perpendicular bisector of the great circle arc connecting the centers of those two cells.
8. The points and areal cells of the various resolution grids which constitute the grid system form a hierarchy which displays a high degree of regularity.
9. A single areal cell contains only one point, i.e., each point lies in a different areal cell.
10. Points are maximally central within areal cells.
11. Points are equidistant from their neighbors.
12. Grid points and areal cells display regularities and other properties which allow them to addressed in an efficient manner.
13. The grid system has a simple relationship to the traditional latitude-longitude graticule.
14. The grid system contains grids of arbitrary resolution.

An early version of these criteria was formulated by Michael Goodchild, and we refer to this list as the "Goodchild Criteria". We have already noted that it is mathematically impossible for any discrete global point grid or surface tessellation to completely fulfill all of these criteria, since several are mutually exclusive. A good general purpose grid or tessellation might be expected to strike a balance among all criteria, whereas those tuned for specific applications or numerical methods might value certain of these criteria more highly. For example, geostatistical methods favor equal area tessellations that completely cover the globe and are compact.

Recursive Partitioning

Cells that can be partitioned recursively may form a tessellation system that is hierarchical and contains component sub-cells that may or may not exhibit a high degree of regularity. The terminology of recursive partitioning can best be understood from an illustration such as Figure 1. Two types of partitioning, sometimes called 4-fold and 9-fold, are shown in the top and bottom equilateral triangular cells. Each full triangle is at recursion level 0, and the initial partitioning into either 4 or 9 triangular sub-cells is termed recursion level 1. This 4-fold or 9-fold increase in the density of triangular sub-cells continues at recursion levels 2 and higher. Sets of six triangular sub-cells can be assembled into hexagons at each level of recursion, as shown in the right hand half of Figure 1. Notice that the hexagons as assembled are symmetrical about the three triangle vertices only with 9-fold partitioning, and such symmetry is a further advantage when assembling uniform global data models. This leads us to use 9-fold partitioning in our analyses. A similar illustration could be created for recursive partitioning of spherically rectangular quadrilaterals, with the desirable symmetry present for any n-fold partitioning.

Naturally, an infinite number of triangular, hexagonal, or quadrilateral recursion levels are possible, as recursive partitioning can continue indefinitely. However, we find 9-fold partitioning to recursion levels less than ten to be suitable for comparison purposes, since surface area, compactness, centerpoint spacing, and other metrics are still computable at the rapidly increasing cell densities. However, the computation effort quickly becomes immense at higher levels of recursion and the results may not add significantly to our understanding of the surface tesselation or point grid geometry.

Evaluation Criteria Metrics

Global data model evaluation criteria are of limited practical value until metrics are developed for each. Examining the criteria presented above, we see that both topological and geometrical metrics must be devised. We have focused on geometrical measures of surface area, compactness and centerpoint spacing for both triangular and hexagonal cells on a spheroid such as the GRS80 or WGS84, although we only present results for hexagonal sub-cells in this paper. All of these measures involve determination of geodesic distances using standard ellipsoidal distance equations. Spheroidal surface area for quadrilateral cells can be computed using standard equations found in Maling (1992) and other references. Computing the spheroidal surface area of more geometrically complex cells such as spheroidal hexagons requires the oriented triangle summation approach developed by Kimerling (1984).

Measurement of spheroidal compactness proved the greatest challenge. Many two-dimensional compactness measures are based on an area to perimeter ratio normalized to 1.0 for a circle. We have extended this idea to the spheroid by determining the spheroidal area to parallel of latitude perimeter ratio, normalized to a spheroidal cap of the same surface area as the cell.

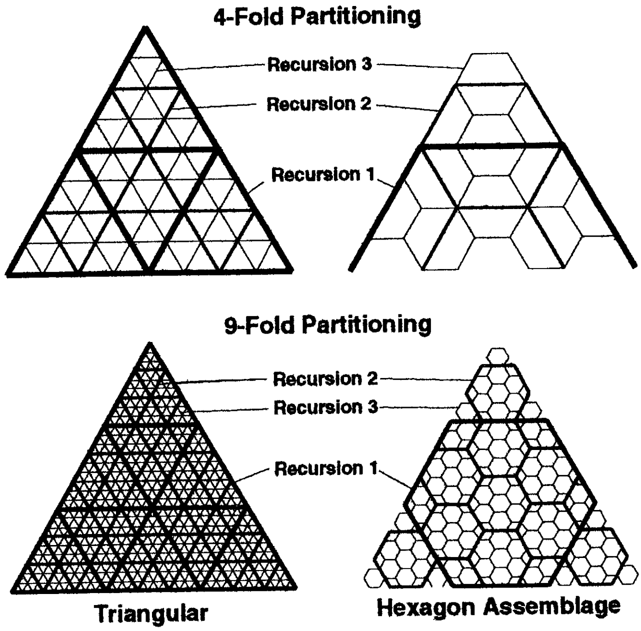

Figure 1. 4- and 9-fold triangular and hexagonal partitioning of an icosahedron face at the first three levels of recursion.

Our numerical analysis of sub-cell surface area, compactness and centerpoint distance focused on obtaining the average, range, and standard deviation for the population of sub-cells at each level of recursion. Range and standard deviation values were normalized as proportions of the average to allow direct comparison of values at different levels of recursion.

GLOBAL DATA MODEL COMPARISON EXAMPLES

Many spherical and fewer spheroidal point grids and surface tessellations have been devised as global scale data models, and in this paper we only compare a few commonly employed and/or potentially attractive tessellations. Our examples include two major classes of surface tessellations, namely quadrilateral approaches and polyhedral approaches based on map projection surfaces.

Equal Angle Quadrilateral Tessellations

Tessellations of the globe into quadrilateral cells of equal latitudinal and longitudinal extent are termed equal angle. Examples abound, including the 5' x 5' ETOPO5 global DEM, and the ERBE 2.5° x 2.5°, 5° x 5° and

10 °x 10 ° quadrilateral grids (Brooks 1981). Our example will be an initial partitioning of the globe into thirty-two 45° x 45° cells at recursion level 0, and subsequent 9-fold partitioning to recursion level 5. Hence, recursion level 2 corresponds to the 5° x 5° ERBE grid.

"Constant Area" Quadrilateral Tessellations

Constant area tessellations begin with an arbitrary sized quadrilateral cell at the equator, and then define the parallel and meridian cell boundaries across the globe so as to achieve approximately equal area cells. This is done either by keeping the latitude increment constant and adjusting the longitude increment as the pole is approached, or vice versa (Brooks, 1981). Our example is the Nimbus Earth Radiation Budget (ERB) Experiment grid, with initial 4.5° x 4.5° quadrilateral cells at the equator. The longitudinal increment increases in twelve discrete steps to 120° near each pole. Recursive subdivision into "constant area" sub-cells is more problematic, since 4-fold equal angle partitioning is commonly employed for simplicity. Breaking from tradition, we employ 9-fold equal angle partitioning to maintain consistency in our comparisons while using the same basic partitioning method. The initial 4.5° x 4.5° cells correspond approximately to recursion level 2 in the previous equal angle tessellation, and we carry the partitioning to recursion level 5.

Polyhedral Map Projection Surface Tessellations

The faces of a Platonic polyhedron are a natural starting point for a global data model, since each face is identical in surface area and is a regular spherical polygon when projected to the globe. Attention has been given to the octahedron (Dutton 1988, White et al. 1996) and the icosahedron (Baumgardner and Frederickson 1985), and we examine the latter in this paper. A convenient partitioning method for polyhedral faces is to create a map projection of each face that is of the same geometric form as the face, e.g., an equilateral triangle for each face of the icosahedron. We then partition the map projection surface recursively, producing 9 identical equilateral sub-triangles with 9-fold partitioning of the face. Sets of six sub-triangles can then be combined into hexagonal cells that are finally projected back onto the globe. Several map projections can be used, but we will examine two: the Snyder and Fuller-Gray.

The Snyder Polyhedral Equal Area projection (Snyder 1992) transforms each icosahedron face on the globe into an equilateral planar triangle while maintaining area equivalence throughout. The projection is made equal area by adjusting the scale outward from the center of each edge.

This results in increased shape distortion as each of three lines from the triangle center to corner vertices is approached.

The Fuller-Gray projection is based on the geometrical idea behind R. Buckminster Fuller's icosahedral world map projection. Fuller imagined the three edges of each icosahedron face as flexible bands curved to lie on the spherical surface. Each edge would be subdivided and holes drilled at n equally spaced increments, and flexible bands would be strung between corresponding holes on adjacent edges. This would create a triangular network of lines on the sphere, which could be flattened to create a regular grid of equilateral sub-triangles. Fuller imagined the vertices of each sub-triangle being the projection of the corresponding line intersection point on the sphere. These intersection points were later found physically impossible to achieve, since nearly all triplets of intersecting lines on the globe form small triangles in the plane, whose centerpoints are the best approximation of Fuller's idea. Gray (1994) has developed exact transformation equations for this approximation, producing a compromise projection having both small area and shape distortion. As with the Snyder projection, sub-triangles can be assembled into a hexagonal tesselation on the projection surface and globe.

GLOBAL DATA MODEL COMPARISON RESULTS

Variation in cell surface area is a major concern to geostatisticians and others. In Figure 2 we show area variation for the hexagonal and quadrilateral sub-cells produced by the four data models examined, using logarithmic scales of normalized cell areas at increasing levels of recursion with corresponding decreases in average cell area. Recognizing that the triangular partitioning of the icosahedron face performed here always creates 12 pentagons on the globe, we see that the surface area standard deviation for the Snyder projection model is always slightly greater than zero, even though there is no area variation among the hexagons forming the partition and all twelve pentagons are exactly 5/6th the area of each hexagon. However, at higher levels of recursion the 12 pentagons occupy progressively less of the total surface area and the standard deviation for the entire globe rapidly approaches zero. Hence , Figure 2 shows the Snyder model to be clearly superior for sub-cells less than 100,000 sq. km.

The variation among sub-cell centerpoint distances for the four models, seen in Figure 3, shows similar performance except for the poorly performing Equal Angle Quadrilateral model. The Fuller and Constant Area Quadrilateral models converge at higher recursion levels to essentially identical low variation, closely followed by the Snyder model.

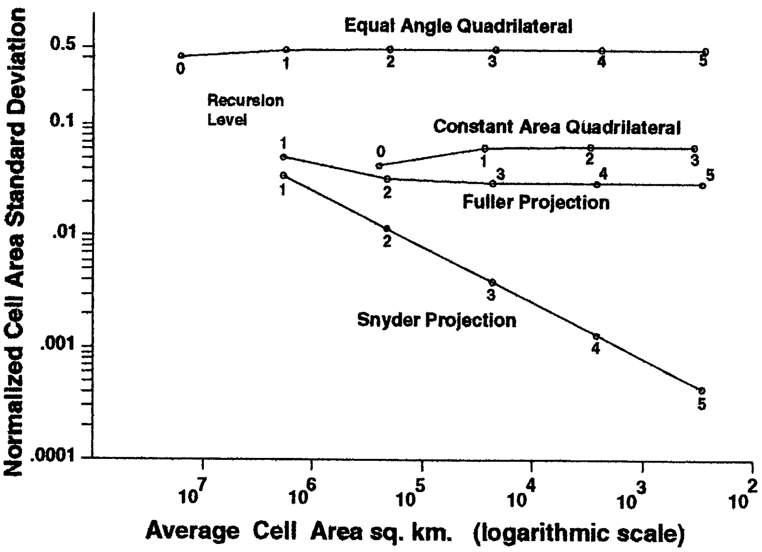

Figure 2. Normalized sub-cell area standard deviation vs. average cell area for four data models.

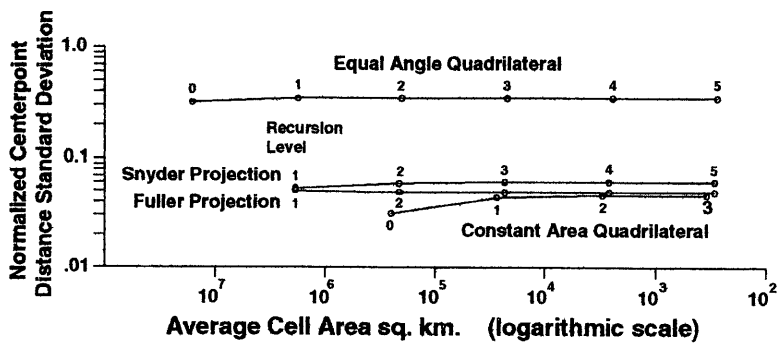

Figure 3. Normalized centerpoint distance standard deviation vs. average sub-cell area for four data models.

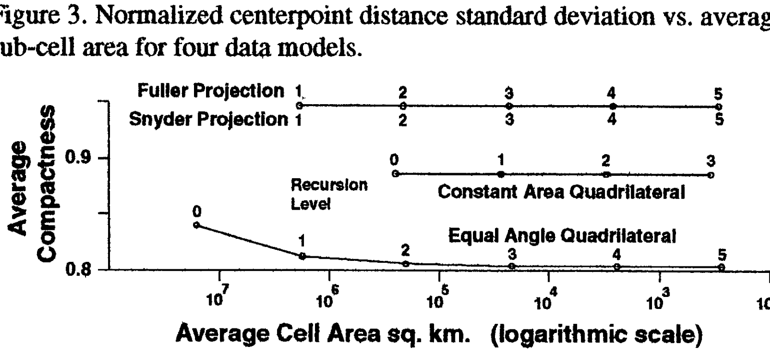

Figure 4. Average sub-cell compactness vs. average sub-cell area for four data models.

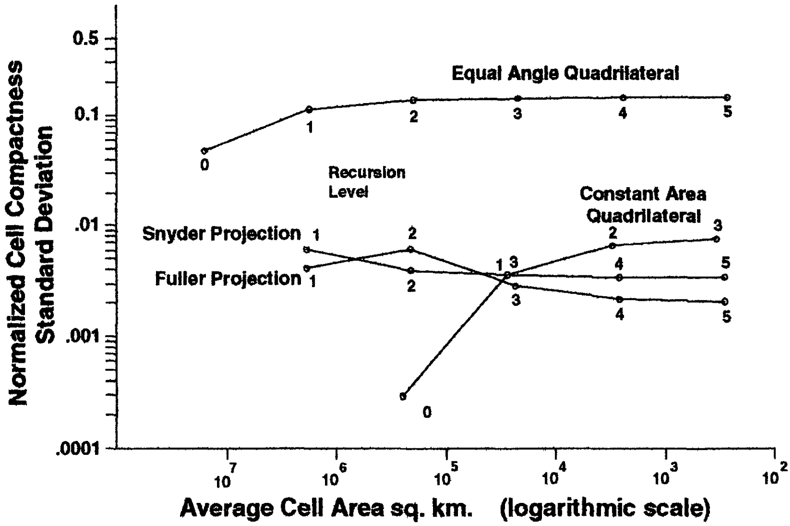

Figure 5. Normalized sub-cell compactness standard deviation vs. average sub-cell area for four data models.

Average sub-cell compactness values for the four data models (Figure 4) shows the superior performance of icosahedral models over quadrilateral, which is to be expected since hexagonal shapes are inherently more compact than rectangular. At all levels of recursion the Fuller data model produces hexagons slightly more compact than Snyder model hexagons, both being far more compact than the Constant Area and Equal Angle model quadrilaterals.

Variation in sub-cell compactness (Figure 5) shows the slightly better performance of the Fuller model over the Snyder at higher levels of recursion. The Constant Area Quadrilateral model produces the lowest variation at its initial tesselation, but compactness variation increases rapidly as the initial cells are partitioned in the equal angle manner.

CONCLUSION

Our global data model evaluation criteria and associated metrics have allowed us to compare data models varying widely in cell geometry and topology. Many more models than the four presented here as examples have been analyzed, and we have concluded that the Icosahedral Snyder Equal Area (ISEA) model recursive partitioning shows particular promise. This is due to its equal area hexagonal tiles on the spheroid, and to its high average cell compactness and low compactness variation relative to traditional quadrilateral tilings, especially at higher levels of recursion.

We are now working to develop an efficient ISEA tile addressing scheme and to demonstrate the advantages of this global data model when it is populated by data transformed from existing global data sets such as the ETOPO5 digital elevation model.

REFERENCES

Baumgardner, J.R. and P.O. Frederickson. (1985). Icosahedral Discretization of the Two-Sphere. *S.I.A.M. J. Num. Anal.* 22(6): 1107-15.

Brooks, D.R. (1981). Grid Systems for Earth Radiation Budget Experiment Applications. *NASA Technical Memorandum 83233.* 40 pp.

Dutton, G. (1988). Geodesic Modeling of Planetary Relief. *Cartographica.* 21:188-207.

Eidenshink, J.C. and J.L. Faundeen. (1994). The 1 km AVHRR Global Land Data Set: First Stages of Implementation. *Int. J. Rem. Sens.* 15:3443-62.

Gray, R.W. (1994). Exact Transformation Equations for Fuller's World Map. *Cartography and Geographic Information Systems.* 21(4): 243-46.

Hastings, D.A. (1996). The Global Land 1-km Base Elevation Digital Elevation Model: A Progress Report. *Global Change Newsletter.* No. 27: 11-12.

Kimerling, A.J. (1984). Area Computation from Geodetic Coordinates on the Spheroid. *Surveying and Mapping.* 44(4): 343-51.

Maling, D.H. (1992). *Coordinate Systems and Map Projections.* Oxford: Pergamon Press, 2nd ed.

Snyder, J.P. (1992). An Equal-Area Map Projection for Polyhedral Globes. *Cartographica.* 29(1):10-21.

White, D., Kimerling, A.J. and W.S. Overton. (1992). Cartographic and Geometric Components of a Global Sampling Design for Environmental Monitoring. *Cartography and Geographic Information Systems.* 19:5-22.

White, D., Kimerling, A.J., Sahr, K., and L. Song. (1996). Comparing Area and Shape Distortion on Polyhedral-based Recursive Partitions of the Sphere. Submitted to *Int. J. Geo. Info. Sys.*