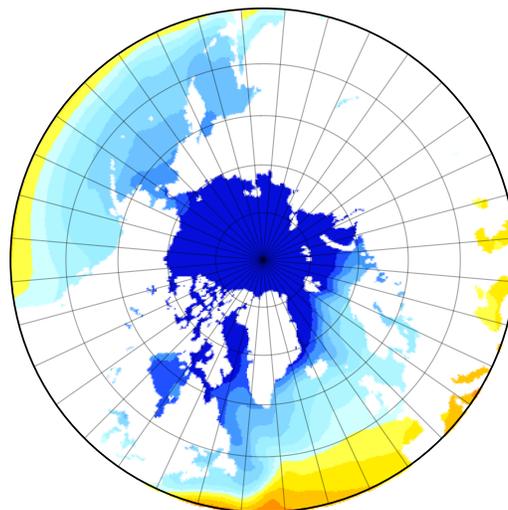
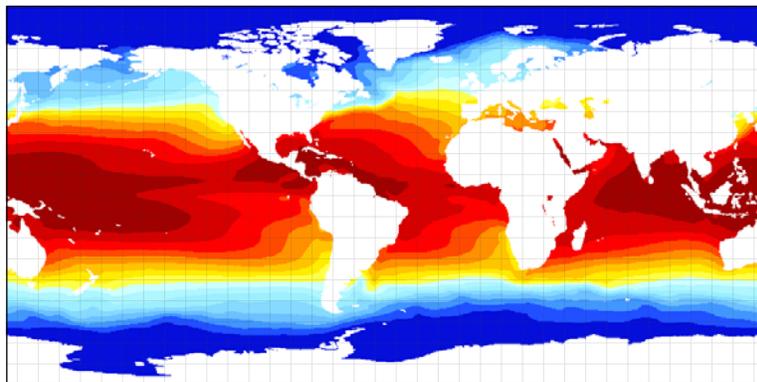


Computational Considerations for Collecting and Using Data in the Equidistant Cylindrical Map Projection and the Bounds of Sampling Geographic Data at Progressively Higher Resolution

Chapter 1 of Book 2,
Collection of Environmental Data, Section H, Environmental Modeling



Pliocene August Sea Surface Temperature (°C)
0 10 20 30

Techniques and Methods 2–H1

Computational Considerations for Collecting and Using Data in the Equidistant Cylindrical Map Projection and the Bounds of Sampling Geographic Data at Progressively Higher Resolution

By Kevin M. Foley

Chapter 1 of

**Book 2, Collection of Environmental Data, Section H, Environmental
Modeling**

Techniques and Methods 2–H1

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2011

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit <http://www.usgs.gov> or call 1-888-ASK-USGS

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod>

To order this and other USGS information products, visit <http://store.usgs.gov>

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:

Foley, K.M., 2011, Computational considerations for collecting and using data in the equidistant cylindrical map projection and the bounds of sampling geographic data at progressively higher resolution: U.S. Geological Survey Techniques and Methods 2–H1, 5 p.

Contents

Abstract.....	1
Introduction.....	1
Analysis	1
Conclusions.....	5
References.....	5

Figures

1–3. Graphs showing:	
1. Curve indicates the association between cell surface area in square kilometers and the position of the cell in degrees latitude for 0.5° cells.....	3
2. Curve indicates change in the number of cells per 1,000 square kilometers with change in latitude	3
3. Crosses indicate lakes and inland bodies of water, plotted by latitude and surface area. In each plot the lower curves represent half the area of a cell, and the upper curves represent double the area of a cell at the corresponding latitude	4

Table

1. Latitude and surface area of major world lakes and inland seas	3
---	---

Computational Considerations for Collecting and Using Data in the Equidistant Cylindrical Map Projection and the Bounds of Sampling Geographic Data at Progressively Higher Resolution

By Kevin M. Foley

Abstract

The Equidistant Cylindrical Map projection is popular with digital modelers and others for storing and processing worldwide data sets because of the simple association of latitude and longitude to cell values or pixels in the resulting grid. This projection does not accurately display area, and the diminished geographic area represented by cells at high latitudes is not often carefully considered. A simple mathematical analysis quantifies the discrepancy in area sampled by cells at different latitudes. The presence of this discrepancy indicates that the use of this projection can induce bias in data sets when both sampling and reporting data. It is demonstrated that as the resolution requirements of input data for models increase, the necessity of providing data to accurately describe smaller cells, particularly at high latitude, will be a challenge.

Introduction

The Equidistant Cylindrical Map projection, also known as the Plate Carrée or Simple Cylindrical Projection, is popular for storing and processing worldwide data sets because of the simple association of latitude and longitude to cell values or pixels in the resulting grid. This projection does not accurately display area (Snyder, 1987), portraying meridians that converge towards the poles as parallel, equidistant lines. The diminished geographic area represented by cells at high latitudes, while assumed by users of data, is not often carefully considered quantitatively. This relation is visualized by quantifying cell area compared to latitude and generating a plot. Basic geographic data (area) about world lakes is used as an example of information that might be sampled or presented in this map projection.

Analysis

The surface area of a cell can be determined by differencing the surface areas of two spherical caps, bound by the high (α_{hl}) and low (α_{ll}) latitude of the cell. This provides the surface area of a band that is then divided by the number of cells n that divide the circumference of the band. The Earth radius used for calculation is 6,367.4 kilometers (km), the average of minimum and maximum values for the Earth's spheroid (Snyder, 1987).

The formula for the spherical cap (Zwillinger, 1996):

$$S_{cap} = 2\pi Rh = \pi p^2 = \pi(a^2 + h^2)$$

$$a = R \cos \alpha, \quad h = R(1 - \sin \alpha)$$

Area of cell =

$$\frac{\pi((R \cos \alpha_{ll})^2 + R(1 - \sin \alpha_{ll})^2) - \pi((R \cos \alpha_{hl})^2 + R(1 - \sin \alpha_{hl})^2)}{n}$$

where:

- R = the radius of the sphere,
- a = the radius of the sphere at the base of the cap,
- h = the height of the cap,
- p = hypotenuse of the triangle formed a and h ,
- α_{hl} = the angle of the high-latitude boundary of the cell,

and

- α_{ll} = the angle of the low-latitude boundary of the cell.

The example calculation here will be for cells of 0.5° latitude \times 0.5° longitude. Calculations show that the surface area of a 0.5° cell is at a maximum of 3,088 square kilometers

2 Computational Considerations for Collecting and Using Data in the Equidistant Cylindrical Map Projection

(km²) near the equator bounded by latitudes 0° and 0.5°. The surface area of 0.5° cells decreases, as latitude increases, to a minimum surface area of 13 km² reached near the poles in cells bounded by latitudes 89.5° and 90°. The association between cell surface area and latitude is shown in figure 1. Values for the entire range of cells from the equator to the poles are presented in table 1.

When calculations are performed on data arranged in such a fashion, cell values can be weighted to normalize the effect of cell values that represent widely varying areas of the Earth's surface. The variation is more problematic in data generation or collection. In this case the Equidistant Cylindrical grid becomes a reticule through which data are sampled. Two types of data need to be considered because the method for determining the value that represents a cell is different in each case.

1. Continuous data—Data indicated by decimal numbers, where all values represent the magnitude single measured aspect of the Earth's surface. Examples would be elevation, surface temperature, insolation, and so on. When data of this type are processed to conform to the Equidistant Cylindrical grid, calculations ranging from simple averaging of the values that fall within the grid to more complex interpolations are performed to assign a best, representative value to each cell. If the data sampled are distributed evenly with respect to the Earth's surface, because of the variation in cell area, far fewer data points will be used in calculation of values for high latitude cells than low latitude cells so there is a variation in quality of data with latitude, quality decreasing towards the poles.
2. Categorical data—Data that can be indicated by a noncontinuous set of symbols. An aspect of the Earth's surface that is represented by categorical data can be segregated into polygonal areas containing like values. The data value for each polygon is a type or category. These values are not part of a continuum of values. Intermediate values or averages cannot be calculated. When categorical data are processed or binned to conform to the Equidistant Cylindrical grid, the value of the type or category that predominates the cell is selected to represent the cell. When the size of the geographic entities sampled approaches the size of cells in the grid, opportunities for sampling error arise and whether an entity is recorded in a data set is dependent not only on its surface area but also on the latitude of its location. Most geographic entities are highly irregular in shape, making precise calculations for potential sampling error impossible but thinking in broad terms allows for bracketing of conditions where problems can arise. When a grid is superimposed on a set of geographic features, an entity's value must predominate a cell to be represented in the resulting data set

(ESRI Inc., 2008). If the area of an entity is less than half the area of a cell at its particular latitude, the value of the entity will not be represented in the data set. Therefore the value of half the area of a cell is the lower limit of representation. Similarly, it would be unlikely that an entity with an area of double the area of a local cell could be positioned in a grid to not predominate at least one cell.

The variation by latitude of the number of 0.5° cells per an arbitrary 1,000 km² area is calculated and shown in figure 2. Note that at point *A*, as latitude increases, the curve first moves above the line denoting the surface area of one-half of a local cell. A 1,000 km² entity would not be recorded nearer the equator than this value of latitude. At point *B* the curve moves above the line denoting the surface area corresponding to twice that of a local cell. A 1,000 km² entity poleward of the point will likely be recorded in the sampled data set. Between points *A* and *B* a 1,000 km² entity will or will not be recorded dependent on its shape and where it falls in latitude, and in relation to the cells of the superimposed grid.

The ever-increasing processing speed of computers makes possible the modeling of climate at an ever-decreasing scale. The current spatial resolution of data sets processed by computer models range from 1–5° of latitude and 4–8° of longitude (Edwards, 2010). Data sets modeled at 4° × 5° resolution represent large mountain ranges by few cells, and peninsulas and islands disappear. Coarsely modeled topography and bathymetry do little to define corridors for, or barriers to, movements of the atmosphere and sea. The next iteration of U.S. Geological Survey PRISM Project climate model data sets will be based on 0.5° × 0.5° resolution (Harry J. Dowsett, personal communication, 2010). These will be better suited for modeling climate variables and movements of the atmosphere and oceans but also will require collection of data on a greatly expanded inventory of geographic features.

Using lakes and inland seas as an example, a list of the 50 largest is compiled in table 1 (data from National Oceanic and Atmospheric Administration, 1982) These have been plotted by surface area in square kilometers and degrees latitude, without distinction between north and south latitude, in figures 3*A*, 3*B*, and 3*C*. These are plotted with the curves defining the previously discussed horizons of feature representation, at two currently common resolutions: 4° × 5° and 2° × 2°, and at the proposed model resolution of 0.5° × 0.5°.

In all plots, the bottom curve represents half the area of a cell at the particular latitude. A lake or any geographic entity with an area that would plot below this curve would not be represented in a data set at the grid resolution represented by the plot. The upper curve represents twice the area of a cell at the particular latitude. A geographic entity with an area above this curve will almost certainly be represented in a data set at this grid resolution. The dashed curve represents the area of one cell in square kilometers at the particular latitude. Note that while compressed by the logarithmic scale of the y-axis, the effect of diminishing cell area with increasing latitude is

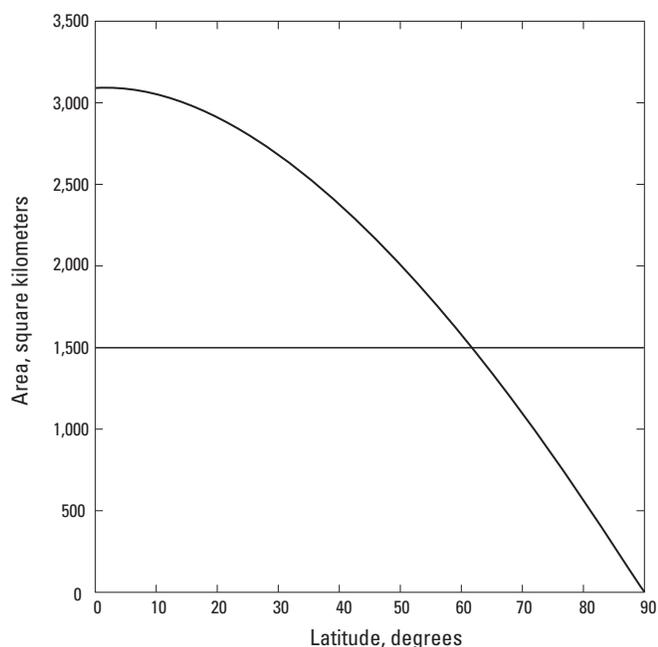


Figure 1. Curve indicates the association between cell surface area in square kilometers and the position of the cell in degrees latitude for 0.5° cells.

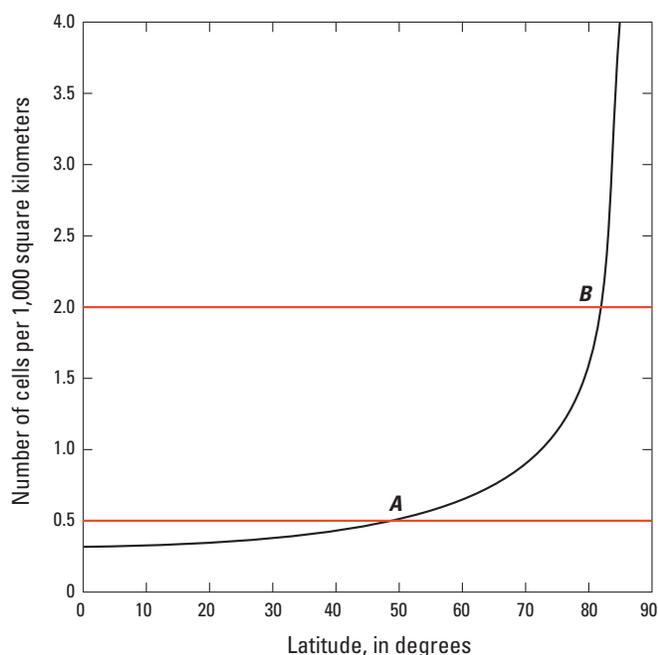


Figure 2. Curve indicates change in the number of cells per 1,000 square kilometers with change in latitude. *A* indicates the latitude where 1,000 square kilometers is contained in half cell. *B* indicates where 1,000 square kilometers is contained in two cells.

Table 1. Latitude and surface area of major world lakes and inland seas.

Body of water	Latitude (degrees)	Surface area (square kilometers)
Caspian Sea	40.0	371,000
Lake Michigan-Huron	44.0	117,585
Lake Superior	47.7	82,414
Lake Victoria	-1.0	68,800
Lake Tanganyika	-6.5	32,900
Lake Baikal	53.5	31,722
Great Bear Lake	66.0	31,080
Lake Malawi	-12.2	30,040
Great Slave Lake	61.7	27,200
Lake Erie	42.2	25,744
Lake Winnipeg	52.1	24,514
Lake Ontario	43.7	19,500
Lake Ladoga	61.0	17,700
Lake Balkhash	46.5	16,400
Lake Vostok	-77.0	15,690
Lake Maracaibo	10.7	13,210
Tonle Sap	13.0	13,000
Lake Onega	61.5	9,700
Lake Eyre	-28.7	9,500
Lake Titicaca	-15.8	8,372
Lake Nicaragua	11.6	8,264
Lake Athabasca	54.3	7,850
Reindeer Lake	57.6	6,500
Lake Turkana	3.1	6,405
Issyk Kul	42.5	6,236
Lake Vänern	58.9	5,650
Lake Winnipegosis	52.5	5,370
Lake Albert	1.7	5,300
Lake Urmia	37.7	5,200
Lake Mweru	-9.2	5,120
Nettilling Lake	66.3	5,066
Lake Nipigon	49.8	4,848
Lake Gairdner	-31.6	4,767
Lake Manitoba	51.0	4,624
Lake Taymyr	74.5	4,560
Qinghai Lake	37.0	4,489
Great Salt Lake	41.2	4,400
Lake Saimaa	61.3	4,400
Lake of the Woods	49.2	4,348
Khanka Lake	45.0	4,190
Dubawnt Lake	63.1	3,833
Lake Van	38.6	3,755
Lake Peipus	58.7	3,555
South Aral Sea	45.0	3,500
Lake Tana	12.0	3,500
North Aral Sea	46.9	3,300
Lake Bangweulu	-11.1	3,000
Dongting Lake	29.3	2,820
Lake Mai-Ndombe	-2.0	2,300
Lake Kyoga	1.5	1,720

4 Computational Considerations for Collecting and Using Data in the Equidistant Cylindrical Map Projection

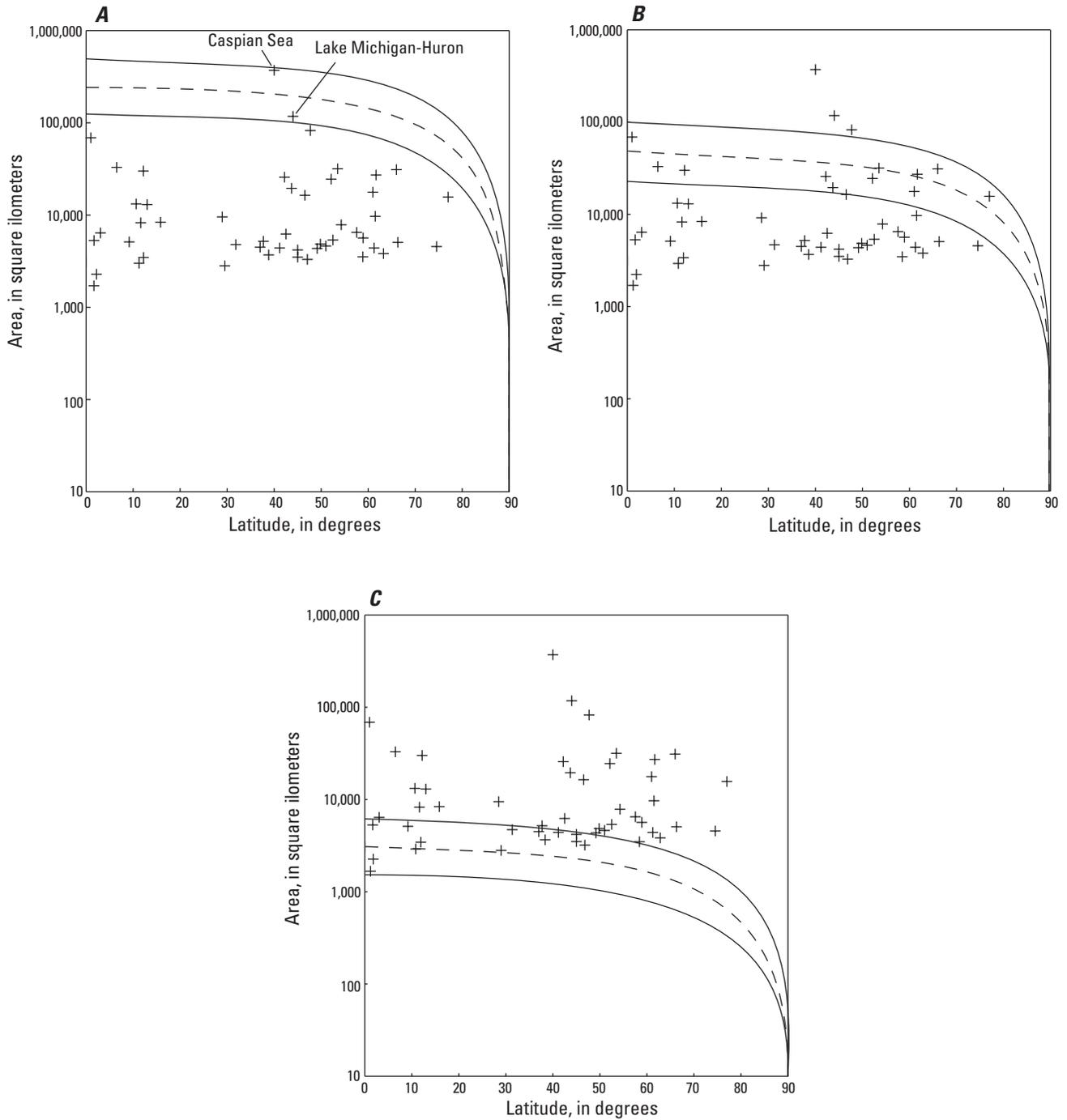


Figure 3. Crosses indicate lakes and inland bodies of water, plotted by latitude and surface area. In each plot the lower curves represent half the area of a cell, and the upper curves represent double the area of a cell at the corresponding latitude. The central dashed-line curves represent the area of one cell at the corresponding latitude. (A) Curves plotted for 4° x 5° cells. (B) Curves plotted for 2° cells. (C) Curves plotted for 0.5° cells.

apparent. It is possible to see ranges of area values that would be represented at high latitude but not low latitude.

However, the plots of the distribution of larger lakes show that they are absent at high latitudes where the effect is greatest. Figure 3A shows the resolution characteristics of a $4^\circ \times 5^\circ$ grid. The plot illustrates that only two bodies of water would possibly be represented at this resolution. The Caspian Sea plots in a range that shows high probability of representation, but what is now termed Lake Michigan-Huron shows a possibility but low probability of being represented. Former Lakes Michigan and Huron taken logically, individually, would not be represented. Figure 3B shows the resolution characteristics of a $2^\circ \times 2^\circ$ grid. The plot illustrates that 3 bodies of water will certainly be represented and 11 more now fall into a field of possibility-to-probability of being represented. Figure 3C shows the resolution characteristics of a $0.5^\circ \times 0.5^\circ$ grid. At this resolution, with the exception of a few, small, low latitude lakes, nearly all inland bodies of water will likely be represented as a grid cell in a derived data set. Note that the two or three lakes least likely to be represented would certainly be represented if situated at a higher latitude. At low latitude, the distribution of lakes by surface area reaches close to the lower limit of resolution. For low latitudes, the distribution generated by collecting data for the 50 largest lakes is sufficient.

However, at 45° latitude and higher, the curves representing the limits of resolution drop below the distribution of lakes. At these higher latitudes the gap between the point cloud representing lakes and the resolution curves strongly suggests that above 45° latitude the sample of 50 lakes is insufficient to document all the lakes that would be represented by cells. A broader threshold, including lakes of much smaller surface area, needs to be applied to higher latitudes to completely document lakes in the region.

Conclusions

Using a geographic grid of Equidistant Cylindrical projection for collection of data induces bias into the resulting data set. This bias is more pronounced where the surface area of grid cells is approximately in the range of the surface area of the geographic entities studied. Figure 3A demonstrates that at $4^\circ \times 5^\circ$ resolution few if any lakes would be represented in a derived data set. Little bias is possible. Nearly all are excluded. Conversely, figure 3C demonstrates that at $0.5^\circ \times 0.5^\circ$ resolution nearly all lakes will be represented. Here again, there is little possibility of bias. Figure 3B demonstrates the case where cell size approximates the size of the entities studied. At $2^\circ \times 2^\circ$ resolution, and other resolutions between the preceding cases, the size of many lakes, or other entities of a similar size, falls into range where representation is dependent upon latitude. It is important for both producers of data and consumers of data to take this inherent bias

into consideration. The case of providing lake data to fill $0.5^\circ \times 0.5^\circ$ cells shows that as the spatial resolution requirement for world models increases, the ability to provide accurate data, particularly paleoclimate data, will become an increasing challenge.

References

- Edwards, P.N., 2010, *A vast machine: Computer models, climate data, and the politics of global warming*: Cambridge Mass., MIT Press, 538 p.
- ESRI Inc., 2008, How polygon to raster (conversion) works: in ArcGIS 9.2 web help, accessed October 12, 2010, at <http://webhelp.esri.com/arcgisdesktop/9.2/>.
- National Oceanic and Atmospheric Administration, 1982, *Principal rivers and lakes of the world*: Washington, D.C., U.S. Government Printing Office, 52 p.
- Snyder, J.P., 1987, *Map projections—A working manual*: U.S. Geological Survey Professional Paper 1395, 289 p.
- Zwillinger, Daniel, ed., 1996, *CRC standard mathematical tables and formulae (30th ed.)*: Boca Raton, Fla., CRC Press, 832 p.

Publishing support provided by:
Denver Publishing Service Center

For more information concerning this publication, contact:
Center Director, USGS Geology and Environmental Change Science Center
Box 25046, Mail Stop 980
Denver, CO 80225
(303) 236-5344

Or visit the Geology and Environmental Change Science Center Web site at:
<http://esp.cr.usgs.gov/>

