Paper PO-27

Does SAS Distance Measurement Differ from ArcGIS Distance Measurement?

Imam Xierali, American Academy of Family Physicians

Abstract: Distance calculation on the surface of the earth is an important application for many fields of study such as geography and public health. This study examines the differences between the distance calculation with geodist function in SAS 9.2 and distance measures with ArcGIS 10. Distance measures between a sample of points were calculated with SAS 9.2 and ArcGIS 10, encompassing from very large to vary small scales. The correspondence between the distance measurements from the two software applications were then analyzed. Results suggest that there are significant differences between the SAS distance function and ArcGIS distance calculation between the points when map projection is a critical factor. When distance measures are important factors for one's analysis, consideration of different geodetic models and projection distortions could be important.

Keywords: ArcGIS 10; Euclidean distance; geodesic distance; map projection; SAS 9.2;

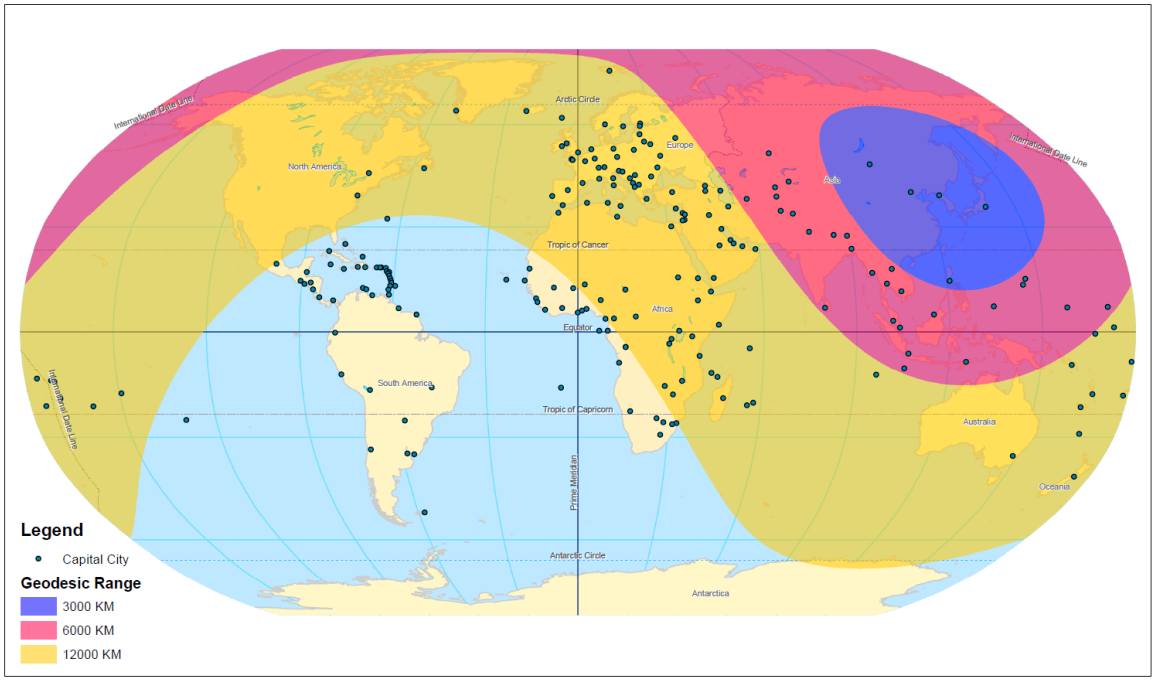
Introduction: Distance calculation on the surface of the earth is an important application for many fields of study such as geography and public health. Distance as the spatial separation can be viewed in both an absolute and relative sense (Fellmann, Getis, and Getis, 2007). In this study, distance is limited to the absolute distance between two points on the surface of the Earth. Many software applications can be used to calculate distance. In this study we examined the distance measure using the prominent statistical software SAS (version 9.2) and the popular Geographic Information System (GIS), ArcGIS desktop (version 10). The main purpose is to see if there is any difference between the distance calculated with geodist function in SAS 9.2 and distance measures in ArcGIS 10.

Background: Geodesy is the science of measuring the size and shape of Earth. A geodesic is the shortest route between two points on the Earth's surface, which is a segment of a great circle that passes through the two points. A sphere or ellipsoid is often used as a model representing the

Earth. However, the Earth is neither perfectly spherical nor perfectly smooth. Map projections can compensate for these irregularities, but they then contain certain spatial distortion.

Geodesic distance is used in airline flight path selection and weapon range studies. The ballistic missile weapon system range of North Korea may be the best publicized example (Asia Times, June 21, 2006: The long reach of North Korea's missiles). We recreated the scenario as described in this Asia Times article with ArcGIS 10. Three geodesic circles with 3000, 6000, 12000 kilometer radius from Pyongyang North Korea were created and layered over a world map. As the projected map shows, the geodesic circles do not look circular (see Figure - 1).

Figure - 1: Publicized North Korean ballistic missile range



Note: Map produced with ArcGIS 10 with Robinson Projection

Calculating the shortest distance between any two points on the surface of the earth in Base SAS 9.2 is easy with option to produce distance measures in different units. Here is the general form of the function (SAS Help Online):

distance = GEODIST(latitude-1, longitude-1, latitude-2, longitude-2 <,option(s)>)

Thus, knowing the latitudes and longitudes of the origin and destination, one can quickly calculate shortest distance between the two points. Previous release of SAS does not include this function and people have often used the Haversine formula to calculate the distance between two points as detailed by Ivis (2006) in his NESUG paper. One can also easily implement the Haversine equation in Excel (see *Movable* for other scripts) by using the latitudes and longitudes of origin and destination.

Distance =ACOS(COS(RADIANS(90-Latitude_1)) *COS(RADIANS(90-Latitude_2))
+SIN(RADIANS(90-Latitude_1)) *SIN(RADIANS(90-Latitude_2))
*COS(RADIANS(Longitude_1-Longitude2))) * 6371 kilometers.

Distance measure is an important concept in GIS applications. Many GIS functions are related to accurate distance measurement. However, the two dimension space represented in a typical GIS application has to deal with map projection issue. Map projection is important to properly align features and calculate the exact distances between features. Map projection is the process of projecting the spherical representation of the earth on a planar surface. However, this results in distortion of distances, directions, and or area sizes. With projection set, ArcGIS 10 by default calculates the Euclidean distance between the two points.

In the rest of paper, we address how we approached the comparison, results, and conclusion.

Data and Method: Distance measures between a sample of points were calculated using both SAS 9.2 and ArcGIS 10. The correspondence between the measurements from the two software applications were then analyzed. The sample of points are the coordinates of 215 capital cities of nation states (ESRI Data & Map 2010).

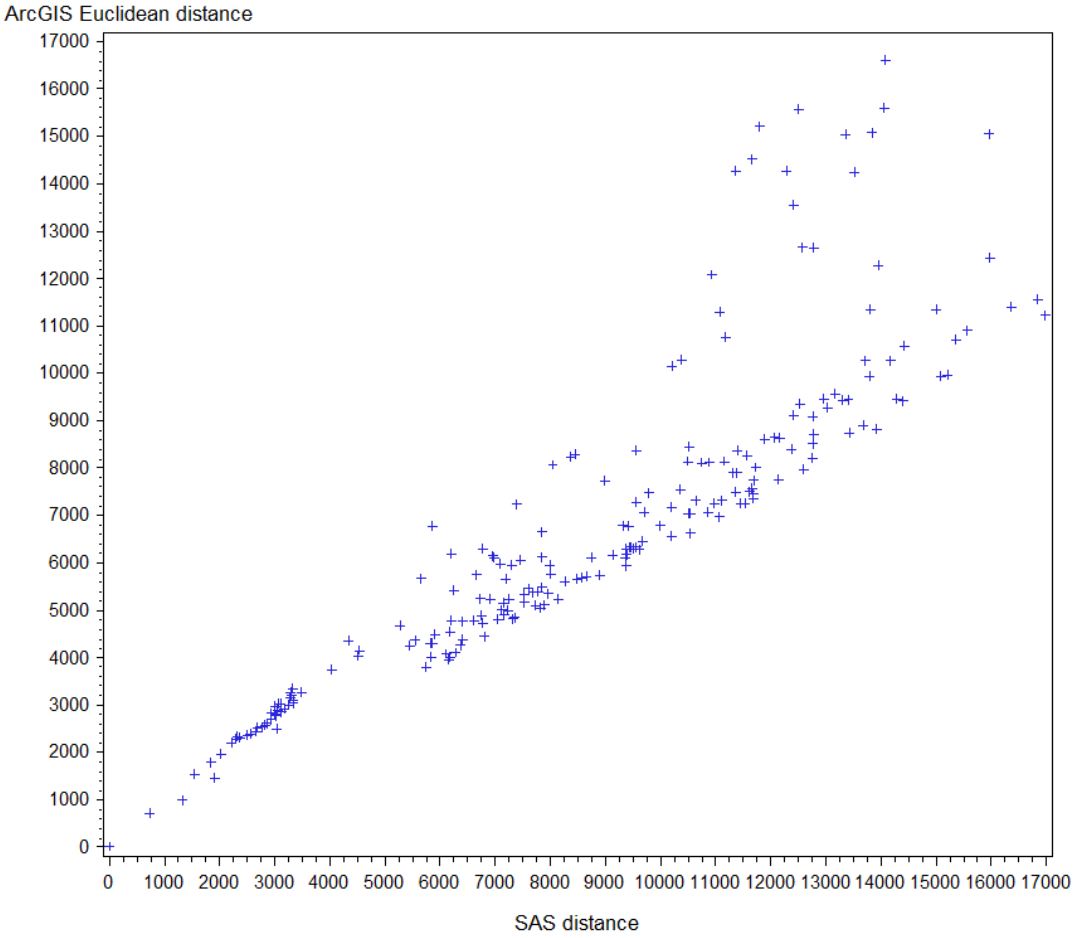
First we calculated Euclidean distance between Washington, DC and the capitals of other nation states in the world using the Near function in ArcGIS 10. In ArcGIS 10, the Capital Cities were projected to Equidistant Cylindrical projection. We used the Near function in ArcGIS toolbox to calculate the distance from Washington, DC to each of the 214 capital cities. We then calculated the geodesic distance between Washington, DC and each of the 214 capital cities. The geodesic

distance calculation is not yet straight forward in ArcGIS 10 without geoprocessing (ArcGIS Resource Center: Calculate Geodesic Distance Between Points).

After distance calculations in ArcGIS, we calculated the same distance in SAS 9.2 using geodist() function. Once all the distance measures were created by both software programs, we used paired t-test to examine if the distance means were significantly different.

Results: Euclidean distance calculated in ArcGIS from Washington, DC to capitals of world were significantly different from the distance calculated in SAS geodist() function (see Figure - 2).

Figure - 2: Scatter plot of SAS 9.2 distance measure and ArcGIS Euclidean distance measure



Paired t-test confirmed that the mean distance calculated by SAS geodist() function generated was 1926 kilometers longer than the mean Euclidean distance calculated using ArcGIS (see Table- 1).

Table - 1: Paired t-test of SAS 9.2 distance and ArcGIS 10 Euclidian distance

N	Mean	Std Dev	Std Err	Minimum	Maximum
214	1925.5	1817.6	124.3	-3405.2	5761.5

DF	t Value	Pr > |t|
213	15.50	<.0001

In ArcGIS 10, geodesic distance calculation is not very straight forward (ArcGIS Resource Center: Calculate Geodesic Distance Between Points). However, the distance measured between Washington, DC and the capitals of the nations was very similar to the distance calculated by SAS 9.2 geodist() function (see Figure - 3). The ArcGIS geodesic distance and SAS distance are statistically different but have practically insignificant differences as seen by Table - 2, mostly attributed to rounding errors.

Figure - 3: Scatter plot of SAS 9.2 distance measure and ArcGIS 10 geodesic distance measure

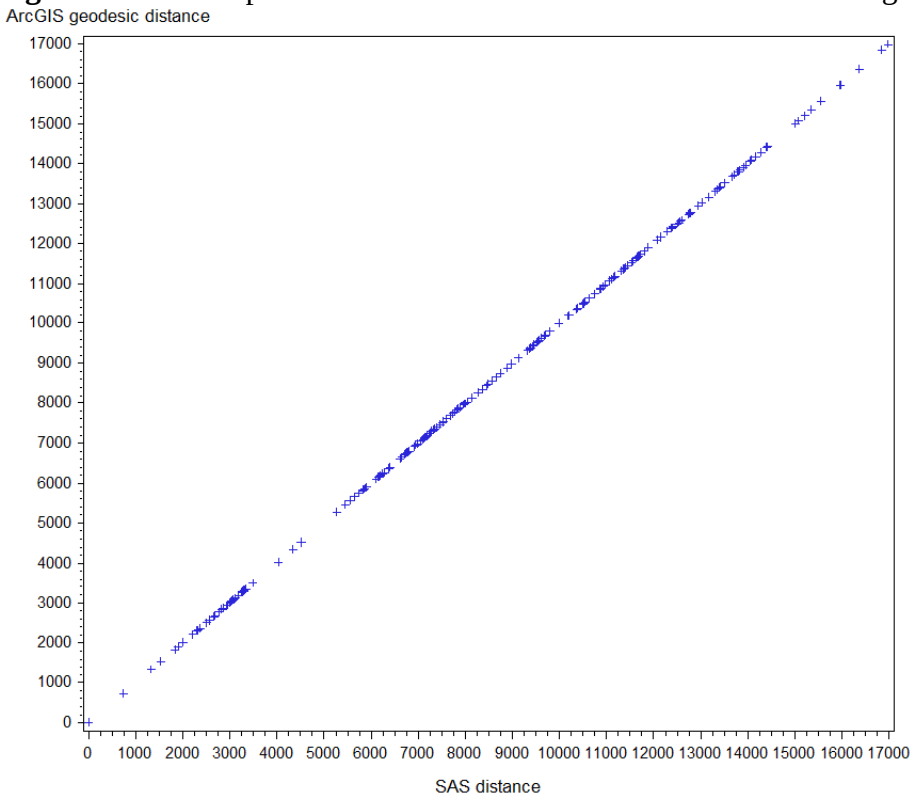


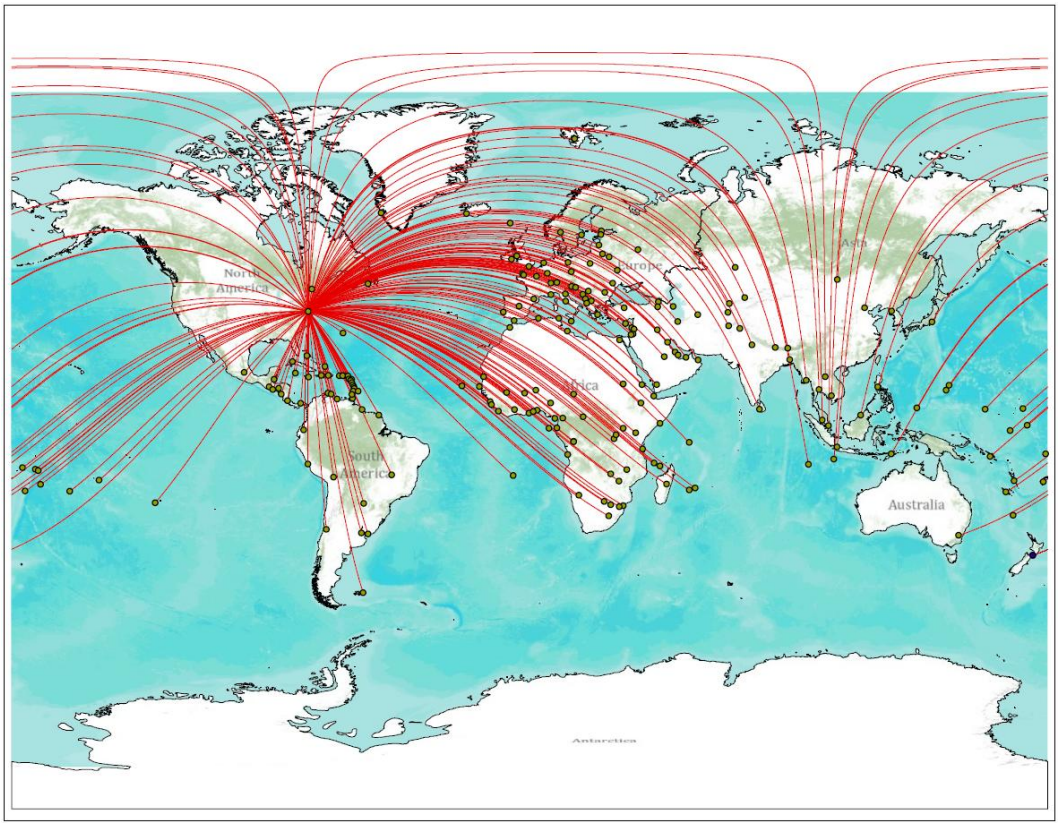
Table - 2: Paired t-test of SAS 9.2 distance and ArcGIS 10 geodesic distance

N	Mean	Std Dev	Std Err	Minimum	Maximum
214	0.000017	0.000026	1.77E-6	-0.00004	0.000039

DF	t Value	Pr > |t|
213	9.87	<.0001

The advantage of SAS 9.2 geodist() function is that it is very easy to calculate geodesic distance. It takes much less time than ArcGIS 10 in calculating the distance between two points. SAS 9.2 is a powerful statistical package. It is also developing GIS functions (see SAS 9.2 GIS). However, SAS 9.2 GIS is much less capable in terms of geoprocessing or spatial analysis.

Figure - 4: The shortest paths from Washington, DC to other capital cities in a planar map



Note: Map produced in ArcGIS 10 with World Miller Cylindrical projection

Figure - 5: The shortest path from Washington, DC to other capital cities in Google Earth

Note: Image created with Google Earth version 6.0.3.2197

ArcGIS 10 is a powerful GIS package. It can generate geodesic lines between two points for visualization (see Figure 4). It can also carry out sophisticated geoprocessing and spatial analysis. However, ArcGIS is still weak in terms of statistical functionality. It is still quite loosely coupled with statistical power. Nevertheless, ArcGIS 10 carries more spatial and geostatistical functions for raster-based data than any previous versions (See ArcGIS Spatial Analyst and Geostatistical Analyst extensions). It now also carries Spatial Statistics toolbox for vector-based data. By default, ArcGIS 10 produces Euclidean distance in the Near function when the layers are projected.

The Euclidean distance is different than the geodesic distance. ArcGIS 10 could also generate geodesic distance, with some model building. It produces geographic features that are visually compelling. For example, figure - 3 shows the shortest flight path from Washington, DC to the capital cities of other nations. In this planar map the lines look curved. When overlaying the same geodesic line feature in Google Earth they will look like the segments of the great circle that passes through the origin and destination points (see Figure - 5).

Conclusion: Results suggest that there are significant differences distances calculated by SAS distance function and by ArcGIS Euclidian distance. SAS 9.2 distance function is easy to carry out. ArcGIS geodesic distance will need users to run ArcGIS scripts/ models. The major difference is that, SAS geodist() function will not consider projection factors, whereas ArcGIS 10 has many ready options for map projections. But ArcGIS 10 does not yet give the user the function to calculate geodesic distance without some scripting or model building in ArcGIS environment. Thus when distance measures are important factors for analysis, careful consideration should be given to the differences in geodetic models and projection distortions.

Reference:

Fellmann JD, Getis A, and Getis J (2007). Human Geography: Landscapes of Human Activities. New York: McGraw-Hill.

Ivis F (2006). Calculating geographic distance: Concepts and methods. *NESUG*.

ArcGIS Resource Center: Calculate Geodesic Distance Between Points. http://resources.arcgis.com/gallery/file/geoprocessing/details?entryID=C0E67091-1422-2418-A09E-45FB0ACEDD6C. Last access August 16, 2011.

SAS 9.2 Help Online: http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a003113162.htm. Last accessed August 16, 2011.

Movable Type Scripts: Calculate distance, bearing and more between Latitude/Longitude points. http://www.movable-type.co.uk/scripts/latlong.html. Last accessed August 18, 2011.

Asia Times. August 21, 2006. The long reach of North Korea's missiles. By Bertil Lintner. http://www.atimes.com/atimes/Korea/HF21Dg02.html. Last accessed August 12, 2011.

Author Contact Information:

Imam Xierali
American Academy of Family Physicians
1350 Connecticut Avenue Suite 201
Washington DC 20036
Tel: 202-331-3360
Email: ixierali@aafp.org