

A SPREADSHEET MAPPING APPROACH FOR ERROR CHECKING AND SHARING COLLECTION POINT DATA

DESMOND H. FOLEY

Division of Entomology, Walter Reed Army Institute of Research, 503 Robert Grant Avenue, Silver Spring, MD, 20746, USA, foleydes@si.edu

Abstract. – The ready availability of online maps of plant and animal collection locations has drawn attention to the need for georeference accuracy. Many obvious georeference errors, for example, that map land animals over sea, the wrong hemisphere, or the wrong country, may be avoided if collectors and data providers could easily map their data points prior to publication. Various tools are available for quality control of georeference data, but many involve an investment of time to learn the software involved. This paper presents a method for the rapid map display of longitude and latitude data using the chart function in Microsoft Office Excel®, arguably the most ubiquitous spreadsheet software. Advantages of this method include: use of software that will be familiar to many data providers, immediate visual feedback to assess data point accuracy; and results that can be easily shared with others. Methods for making custom Excel chart maps are given, and we provide free charts for the world and a selection of countries at <http://www.vectormap.org/resources.htm>.

Key words. – Map, Excel; Chart; Scatter plot; georeference; collection data

The past decade has seen an increasing number of online databases designed to host collection records of a wide variety of organisms (e.g. Global Biodiversity Information Facility¹). These databases require georeferenced data, to map and compare collection points to other spatial information, sometimes in a geographical information system (GIS) setting (e.g.²). Guidelines for reporting georeference information for collection records have been developed (e.g. Chapman and Wieczorek 2006, MaNIS/HerpNet/ORNIS³). Georeference errors can occur, for example, when converting degrees minutes seconds to a decimal degree format, omitting the negative sign for the hemisphere, or typographical errors such as omitting, repeating, or transposing numbers. Various methods are available to automate the process of detecting errors in georeference information, such as using DIVA-GIS⁴, GeoLocate⁵, and tools available through GBIF⁶. For the former method, longitude and latitude point data are converted to shapefiles

and data cleaning is then undertaken by the ‘check coordinates’ option of DIVA-GIS, a “point-in-polygon” method (Chapman, 2005), which identifies points located outside all polygons (i.e. over the sea), and points that do not match relations for an administrative region (e.g. fall in another country, or state). However, this process can be labor-intensive and requires a certain familiarity with the software packages involved. Some databases (e.g. GBIF) screen data that is submitted to them, and will report back to the data provider a list of questionable georeferences before data is accepted. However, no such process exists, for example, for individuals who want to send location data as part of a journal article, or gene sequence metadata to GenBank, resulting in the frequent publication of imprecise or erroneous collection location details. If collectors and data providers could easily produce an orientation map for error checking, many georeference errors could be avoided at an early stage, when corrections tend to be easiest.

This paper presents a simple method for the rapid map display of longitude and latitude data using the chart function in Microsoft Office Excel®, arguably the most ubiquitous spreadsheet software.

1 <http://www.gbif.org/>

2 <http://www.mosquitomap.org/>

3 <http://manisnet.org/>

4 <http://www.diva-gis.org/>

5 <http://www.museum.tulane.edu/geolocate/>

6 <http://sourceforge.net/projects/gbif/files/Data%20Tester/>

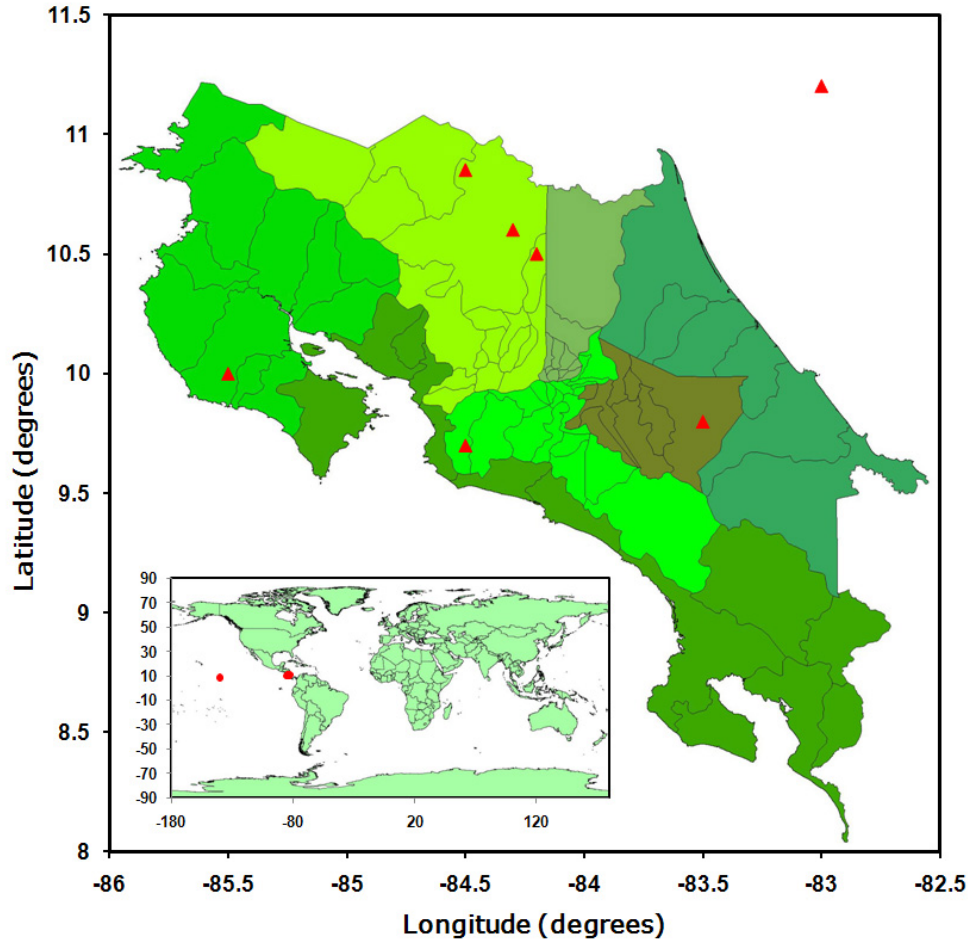


Figure 1. Microsoft Excel Chart (XY Scatter plot) with background picture of the world and test locations in Costa Rica shown as red points. Note the erroneous points over sea; these errors can easily be visually detected and corrected.

METHODS AND RESULTS

Converting the curved surface of the earth to a flat map requires a projection, and the geographic projection (also called the equidirectional, equidistant cylindrical, simple cylindrical, equirectangular, plate carrée or carte parallélogrammatique projection) maps longitude to equally spaced vertical straight lines, and latitude to evenly spread horizontal straight lines. Data defined on a geographic coordinate system is displayed as if a degree is a linear unit of measure. Latitude values are measured relative to the equator and range from -90° at the south pole to $+90^\circ$ at the north pole. Longitude values are measured relative to the prime meridian at Greenwich, United Kingdom, and ranges from $-$

180° when traveling west to 180° when traveling east. For example, Argentina, which is south of the equator and west of Greenwich, has negative longitude values and negative latitude values. The geographic projection is the standard for computer applications that process global maps, because of the simple relationship between the position of an image pixel on the map and its corresponding geographic location on Earth. At low resolution, such as at the level of countries, maps based on the earth as a sphere (e.g. Google Earth) display the lines of longitude and latitude as curvilinear rather than linear. However, at high resolution, these map lines approach linearity, and are therefore suitable for our purposes (see below).

The Excel chart option ‘Scatter’ allows the graphical display of Cartesian coordinates (XY) of

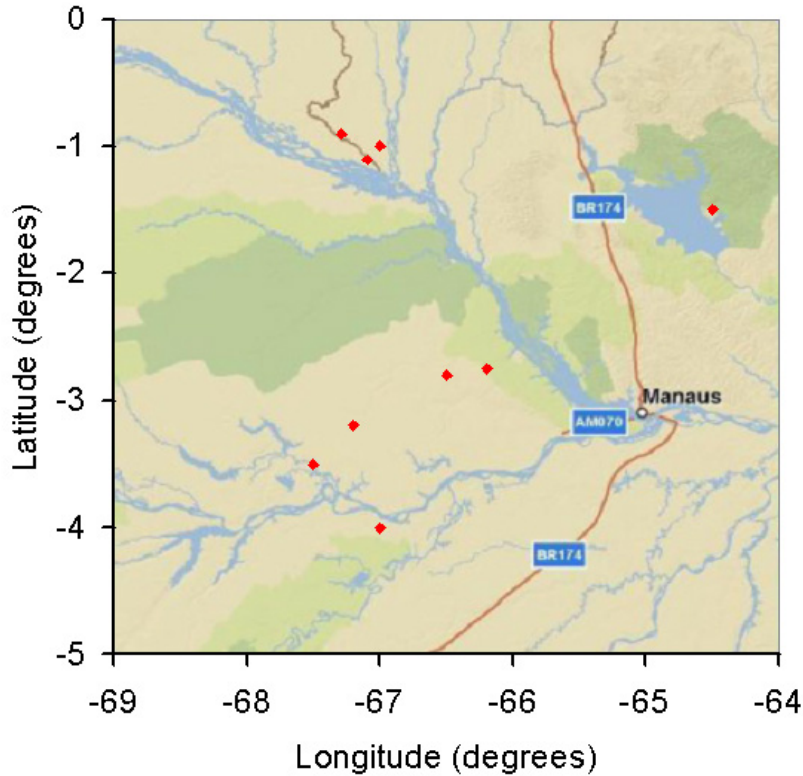


Figure 2. Microsoft Excel Chart (XY Scatter plot) with background picture from MosquitoMap and test locations shown as red points.

any parameter, including longitude (X-axis) and latitude (Y-axis). If the chart contains a background image of a map in a geographic projection, cropped to known longitude and latitude, and the axes of the chart are constrained to these longitudes and latitudes, then geographic correspondence between XY data and geographic location is established. The resulting thematic map can be used for basic orientation, and for identifying some georeference errors, by visual comparison of data point locations with coast and political boundaries. By mousing over points, information about these coordinates is displayed in Excel, allowing identification of particular coordinates that need checking. In the following I describe a method and provide examples for obtaining Excel chart maps from a shapefile, and from online maps.

CREATING MAPS FROM A SHAPEFILE

While various options for obtaining maps are possible, the GADM database of Global

Administrative Areas⁷ provides free country administrative region shapefiles. These shapefiles can be loaded in a GIS program such as the freely available DIVA-GIS or the commercially available ArcGIS®. Once loaded into a program capable of displaying a shapefile, an image file of the country or area of interest needs to be constructed, and this image is then cropped to the area of interest in a program such as MS Paint or Adobe Photoshop. A bounding box or corner points can be drawn in the GIS program to demarcate the area of interest, and the image exported as a file such as JPEG. For ease of presentation in Excel, we define the bounding box or corner points at positions that are easy to display. For example, a bounding box from 10-12.5 degrees latitude and 35-40 degrees longitude displays more easily in an Excel chart than one from 10.12445-12.45773 degrees latitude and 34.55689-39.66554 degrees longitude. We used the tool ‘Draw point’ in ArcView GIS 3.3, and

⁷ <http://www.gadm.org/country/>

then edited the shape properties (right click) to achieve the desired longitude and latitude. An upper left and lower right point was drawn to demarcate the area of interest. Graphics are selected, then graphics properties edited to make points small enough for accuracy and large enough to be visible (e.g. size 2). The zoom tool was then used to frame the view to just encompass the 2 points. This view was then exported as a JPEG format image file (200 dpi; 100% quality). Figure 1 shows an Excel scatter plot with a background picture of the world and test locations in Costa Rica shown as red dots. Note the erroneous points appearing over sea, which can be visually detected and corrected.

CREATING MAPS FROM AN ONLINE MAP

We used MosquitoMap and Google Earth to create the custom maps in Figure 2 and 3, respectively. In the MosquitoMap Viewer, we zoomed to the area of interest and demarcate this with Tools>Markup>Create New Item>Rectangle. The settings were for a fine solid line, with no fill (line width = 1, line transparency = 0, fill transparency = 100%). The corner overview map was hidden and a rectangle was drawn between

longitude and latitudes that are simple to chart. For example, Figure 2 was derived from a rectangle with an upper left corner of latitude = 0 degrees, longitude = -64 degrees; and a lower right corner of latitude = -5 degrees, longitude = -59 degrees. The map was then exported as an image file. The image was cropped according to the rectangle in Adobe Photoshop, and then loaded into a chart as a background image, and the axis' minimum and maximum adjusted to match the latitude and longitude. In the case of Figure 2, the Excel chart y-axis minimum was set to -5, maximum set to 0, and crossing the x-axis at -5; the x-axis was set to minimum -64, maximum -59, and crossing the y-axis at -64. The major unit was set to 1.

To create Excel charts using Google Earth images as backgrounds, the location in Google Earth must first be set to decimal degrees (set Tools>Options>3D View>Show Lat/Long to Decimal Degrees), and the area of interest must be at such a resolution that the lines of longitude and latitude are linear. These lines can be visualized by turning on Grid lines (View>Grid). A new placemark is activated for the upper left corner of the area of interest. In the New Placemark dialog

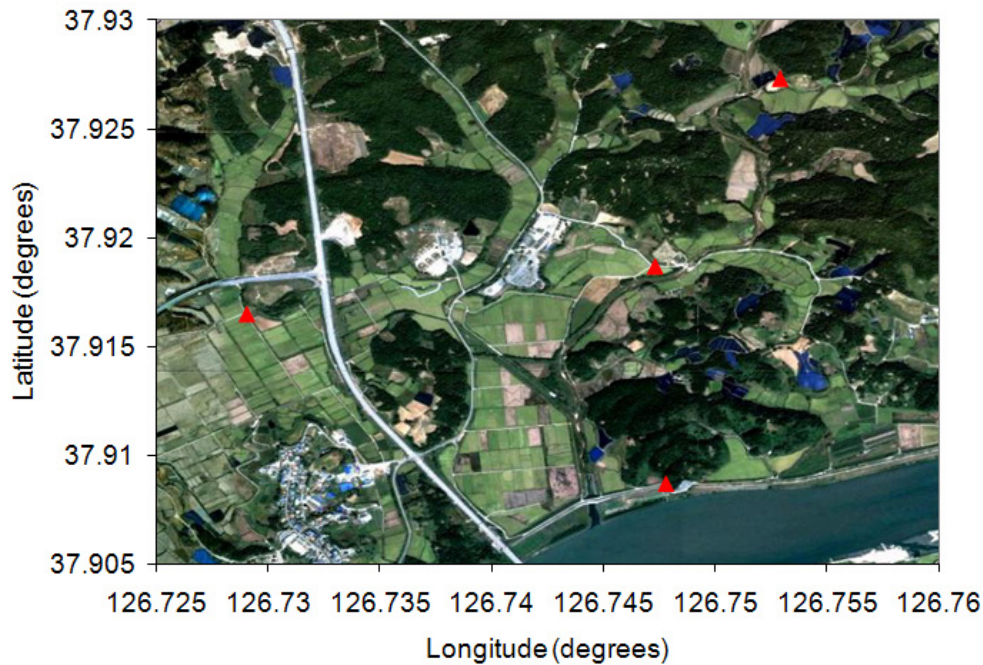


Figure 3. Microsoft Excel Chart (XY Scatter plot) with background picture from Google Earth and test locations shown as red points.

box, the latitude and longitude can be adjusted to remove unnecessary precision and simplify the coordinates for charting, for example, the corners of Figure 3 were limited to three decimal places. Within the New Placemark dialog box, under the View tab, the Heading and Tilt are changed to zero. The placemark Icon in the dialog box is clicked to change the symbology to a radio button (circle), of scale 0.5, and the name of the placeholder is removed. This ensures that the placeholder is unobtrusive in the final image. This process is repeated for the lower right corner of the area of interest. Under View, the grid lines, overview map, and sidebar are removed, and the area of interest positioned to allow an image to be taken that is clear of text. The image is then taken, either by Edit>Copy Image or File>Save>Save Image, and the image file cropped and loaded into an Excel chart, as detailed above. As a final step, locations in the Excel chart can be compared to the same coordinates in Google Earth to test accuracy.

Once a map is constructed in Excel, there is no setting that forces equally spaced horizontal and vertical gridlines and axis ticks. An Excel VBA project is available for making square gridlines⁸, but the simplest solution is to use a ruler and manually adjust the chart dimensions on the screen until the units are equal. Although points will appear at their correct longitude and latitude, printing from Excel may distort the map, unless the margins in Print Preview are adjusted so that the degrees longitude and latitude are the same size. Alternatively, the chart can be copied and pasted into another program, such as Photoshop or MS PowerPoint, with no distortion.

DISCUSSION

Ideally, once a suitable map is located, the area of interest in longitude and latitude would be defined by the data provider, and an image file ready for Excel would be produced in one action. Screenshots can be obtained in MS Windows XP and images pasted in MS Paint, then cropped and saved as image files. Windows Vista and Windows 7 has the screenshot function ‘Snipping tool’, which in the latter program can be found by selecting Start Menu>All Programs>Accessories. Freeware tools are available that have enhanced

screenshot functionality⁹, as well as various commercial programs. With Mac OS X, a snapshot of a selected area can be created as a file on the desktop by Command-Shift-4, then dragging and dropping the mouse over the selected area. Some of these screenshot options may replace reporting of the longitude and latitude of the underlying map application with screen coordinates of the mouse position. Although, we concentrate on the charting function of Excel because of the relative ubiquity of this program, enhanced mapping functionality for PDFs is now available in Adobe Reader 9 and Adobe Acrobat 9. As these and other product develop and become more widely available they may play a greater role in the cleaning and display of collection data.

One limitation of the Excel chart approach is that erroneous points that map beyond the boundaries of the map will not display on the chart, and could go undetected. The use of country and global Excel chart maps allow the detection of some but not all of these possible georeference errors. Although the size of charts can be expanded in Excel, zoom capability is limited, so points that fall just off the coast or in the wrong country may not be detected. The Excel Charting step should not be seen as a replacement for other more detailed georeference error detection methods. There are alternative and more developed tools for data cleaning, like the previously mentioned DIVA-GIS, and these will be indispensable when attempting to increase data quality for large collections.

During the writing of this paper I became aware that others have promoted mapping using the chart function of Excel, both for free (e.g.¹⁰,¹¹,¹²), and as part of software packages (e.g.¹³,¹⁴,¹⁵). However, I believe that this paper is the first to

⁹ http://www.inspire-soft.net/?nav=soft_screenshoter

¹⁰ <http://www.excelhero.com/blog/2010/04/excel-location-mapping.html>

¹¹ <http://forums.techguy.org/business-applications/917391-plot-latitude-longitude-points-world.html>

¹² <http://www.excelcharts.com/blog/how-to-create-thematic-map-excel/>

¹³ <http://www.excelhero.com/blog/2010/04/excel-location-mapping.html>

¹⁴ <http://www.mapcrossing.com/>

¹⁵ <http://office.microsoft.com/en-us/excel-help/create-a-map-with-mappoint-and-excel-HA010277101.aspx>

⁸ <http://peltiertech.com/Excel/Charts/SquareGrid.html>

describe in detail the Excel charting method specifically for those interested in error checking and sharing specimen location data. In addition, we provide Excel charts for a variety of countries for free download (see¹⁶). Users can cut and paste their collection coordinates into these files and immediately see global and country maps of their data. We have incorporated Excel charts for mapping collection points in our spreadsheet collection forms for arthropod disease vectors (e.g.¹⁷), and we encourage other groups to include a mapping function in their collection spreadsheets. Providing these simple tools is done in the hope that this will encourage mapping as a routine step before reporting geolocation data, and increase awareness of the importance of geolocation data quality.

ACKNOWLEDGEMENTS

Funding for this work was provided by the Global Emerging Infections Surveillance and Response System, a Division of the Armed Forces

Health Surveillance Center. Thanks go to L. M. Rueda, J. F. Ruiz, and R. C. Wilkerson for comments and review of this manuscript. This research was performed under a Memorandum of Understanding between the Walter Reed Army Institute of Research and the Smithsonian Institution, with institutional support provided by both organizations. The opinions and assertions contained herein are those of the authors and are not to be construed as official or reflecting the views of the Department of the Army or the Department of Defense.

LITERATURE CITED

- Chapman, A.D. (2005) *Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data*, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen.
- Chapman, A.D. and J. Wiczorek (eds), 2006. *Guide to Best Practices for Georeferencing*. Copenhagen: Global Biodiversity Information Facility. (accessed 2010-01-15)

¹⁶ <http://www.vectormap.org/resources.htm>

¹⁷ <http://www.mosquitomap.org/contribute.htm>