Seven possible states of geospatial data with respect to map projection and definition: A novel pedagogical device for GIS education

Steven E. Greco

Department of Human Ecology, University of California, Davis, CA, USA

ABSTRACT
Conceptually, the theory and implementation of “map projection” in geographic information system (GIS) technology is difficult to comprehend for most introductory students and novice users. Compounding this difficulty is the concept of a “map projection file” that defines map projection parameters of geo-spatial data. The problem of the “missing projection file” appears ubiquitous for all users, especially in practice where data is widely shared. Another common problem is inadvertent misapplication of the “Define Projection” tool that can result in a GIS dataset with an incorrectly defined map projection file. GIS education should provide more guidance in differentiating the concepts of map projection versus projection files by increasing understanding and minimizing common errors. A novel pedagogical device is introduced in this paper: the seven possible states of GIS data with respect to map projection and definition. The seven possible states are: (1) a projected coordinate system (PCS) that is correctly defined, (2) a PCS that is incorrectly defined, (3) a PCS that is undefined, (4) a geographic coordinate system (GCS) that is correctly defined, (5) a GCS that is incorrectly defined, (6) a GCS that is undefined, and (7) a non-GCS. Recently created automated troubleshooting tools to determine a missing map projection file are discussed.

1. Introduction

Underlying nearly all styles of geo-spatial data (e.g. vector, raster, GPS, LiDAR, etc.) are two systems of measurement: a quasi-spherical system for geographic coordinate system (GCS) and a planar system for projected coordinate system (PCS). This paper addresses a pedagogical need to help introductory students and novice users of geographic information systems (GIS) learn what these two measurement systems are, why they are used, how they are implemented in geo-spatial software, and how to avoid and troubleshoot potential problems arising from their use.

Over the past two decades there has been an incredible transformation of computer operating system user interfaces, GIS software itself, and the pedagogical methods for which to teach it. For instance, in my GIS laboratory manual (Greco 2016), geared for undergraduate students, my teaching assistants now teach in one 3 h lab session what was taught in 1990 over an entire quarter (ten 3 h lab sessions). In some cases what once were very steep learning curves have been greatly reduced with modern technological advances in software and hardware (Fagin and Wikle 2011; Foote et al. 2012). However, despite time, there are certain theoretical areas and technical concepts that do not change and are fundamental to understanding the theory and principles of mapping and geodesy, which is at the core of comprehending geo-spatial technology such as GIS. One of those core technical concepts is the “coordinate system” including map projection and all its intricacies.

Map projection is perhaps one the most challenging concepts for many beginning students to grasp and master when learning to use GIS technology due to its abstract conceptual density and technical complexity (Bampton 2012; Downs and Liben 1991). Good auxiliary teaching resources such as an explanatory textbook or a guide to map projection is important (Chang 2019; Kennedy and Kopp 2000; Lo and Yeung 2007; Maher 2010; McDonnell 1979; Shellito 2015). The sheer array of different PCSs (map projections) and differentiating it from GCS (i.e. non-projected data) can be daunting to the introductory student or novice user. Three such examples are shown in comparison in Figure 1 (http://www.colorado.edu/geography/gcraft/notes/mapproj/mapproj_f.html). It should be noted that Figure 1 illustrates a mix of superimposed geographic and projected coordinates of differing units of measurement that can only be produced outside of a GIS. As an instructor, my methods for teaching coordinate systems (both geographic and projected) have significantly evolved and I have devised some pedagogical tools to parse some of the most common mistakes students make when dealing with them. In this paper, I introduce a simple pedagogical device to improve GIS education and to fill a

CONTACT Steven E. Greco segreco@ucdavis.edu

© 2018 Wuhan University. Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
curricular gap missing in most, if not all, introductory GIS textbooks (to the author’s knowledge).

A fundamentally important topic that can be found in several introductory GIS textbooks is “map projection files” and the information they contain (Chang 2019; Shellito 2015). A map projection file defines the spatial reference parameters for geo-spatial data, such as for a vector-based feature class like a shapefile. In the form of a text file, the map projection file defines many variables such as the spheroid, GCS, horizontal lineal units, first and second standard parallels, primary meridian, latitude of origin, horizontal datum, any false eastings or northings, as well as the map projection itself and its name. The file extensions for projection files vary depending on the geo-spatial data type, but typically *.prj are designated for shapefiles, *.aux for some image files or *.tfw for some Tiff files, and there are several others.

Early on in GIS curricular development, it was noted that sharing data is an important pedagogical tool (Kemp and Wright 1997). Exchange of data between students and colleagues (of various professions) has been and continues to be a very common and widespread practice. In this endeavor GIS metadata assumes great importance and projection files are an important part of that story. However, a seemingly ubiquitous error made by novice users and students in introductory GIS courses is differentiating a projection file from the map projection of the geo-spatial data itself. As shown in Figure 2, the example projection file (Figure 2(a)) defines the variables in the PCS of the map (Figure 2(b)). Initially this can be a vexing concept to many students but some computer savvy students pick it up readily, especially if they have been exposed to programming. For the vast majority, however, the problem of the “missing projection file” can be the source of countless hours of frustration or sheer non-understanding why things are going so wrong. How many times have GIS instructors heard the question, “Why are my data invisible or in the Antarctic?!”. Frequently projection files are inadvertently not copied or possibly do not exist and are not passed on in the data exchange process.

Even GIS professionals have to deal with the problem of missing projection files when sharing data among colleagues of various skill levels. GIS software developers are constantly trying to improve and minimize this problem, for instance, by forcing the user to define a projection file when initially creating a shapefile. But whether or not it is included with the other essential components of a shapefile (i.e. *.shp, *.shx, and *.dbf) is dependent on the technical knowledge of the person copying the data files. More contemporary data structures that are self-defining file formats such as feature classes created within a geodatabase and GeoTiffs (rather than Tiffs with *.aux files) will help to solve this problem in the long term. However, it should be noted that it is possible to import shapefiles that lack a projection file into a geodatabase, thus putting the user in the situation of the missing projection file problem. This issue is discussed further in Section 3.2.

Another very common problem is that map projection files can be corrupted through misapplication of the “Define Projection” tool (found in the Data Management Toolbox of Esri’s ArcGIS software) or its equivalent in other GIS software packages. The Define Projection tool allows the user to change the definition of a projection without actually changing the map projection of the GIS dataset itself. In contrast to this tool, the “Project” tool uses the current map projection definition (as defined in the projection file) to project the geo-spatial dataset to a different map projection or PCS. This is a routine operation for all GIS professionals. However, when a novice user or beginning student in
degrees, minutes, and seconds or from grads) from the center of the Earth (Kennedy and Kopp 2000). Thus, GCS data are georeferenced but are not projected in Cartesian coordinate map space and therefore, by definition, these data cannot be measured for area or length of features – they first must be converted to a PCS to do so. GIS data that are projected or in a GCS both share the common errors made for projection file information described above. All map projections are based on a GCS (Kennedy and Kopp 2000).

2. Seven possible states of geo-spatial data with respect to map projection and definition

The pedagogical device introduced in this paper is: the seven possible states of geo-spatial data with respect to map projection and definition (Table 1). This section will review each respective state. In Section 3, some troubleshooting techniques will be discussed to deal with some of these situations.

2.1. A PCS that is correctly defined

In a perfect world this is what we hope for that the GIS dataset is projected with no errors and correctly defined. Unfortunately we do not live in a perfect world.

2.2. A PCS that is incorrectly defined

This situation is described above in the introduction and can lead to great confusion and frustration for novice or advanced GIS professionals. Through various methods the map projection file can become unintentionally corrupted or inadvertently incorrectly defined through the Define Projection tool or other means.

### Table 1. Seven possible states of geo-spatial data with respect to map projection and definition.

| Definition status | Projection status |
|-------------------|-------------------|
| Correctly defined | 1. Projected (PCS) (e.g. Albers) | 4. Not projected (GCS) |
| Incorrectly defined | 2. Projected (PCS) (e.g. Albers, but definition is wrong) | 5. Not projected (GCS, but definition is wrong) |
| No definition | 3. Projected (PCS) (e.g. Albers and missing projection file) | 6. Not projected (GCS and missing projection file) |
| | | 7. Non-geographic coordinates; not defined (e.g. image or scanned map in page units) |

Note: “na” means not applicable.

GIS misapplies the Define Projection tool chaos and utter confusion can ensue; perhaps flabbergasted is a good description. Redefining a map projection file without changing the actual map projection of the data itself is a state of GIS data that I term in this paper “incorrectly defined” and this commonly committed unintentional error can have cascading effects in subsequent use of the dataset.

Finally, at the beginning of learning GIS, students and novice GIS users struggle with the concept of the difference between map projection (or PCS) and GCS. The coordinate system for GCS data is represented as points of latitude and longitude measured as angles in decimal degrees (converted from

---

**Figure 2.** Distinguishing between (a) the projection file and (b) the projected map of California.
2.3. A PCS that is undefined

This is the case of the missing projection file as described above in the introduction. This is a common problem in data exchange among GIS users, both professionally and in educational settings. This is often caused by an oversight, possibly due to ignorance or carelessness, or inadvertence in nature. The file may not have been copied either by the user of a website that provides (serves) data or by a colleague who provides data but is unaware of the function or importance.

2.4. A GCS that is correctly defined

Again in an ideal world this is what we hope for – no errors – this is similar to the case in Section 2.1, but it deals with GCS such as decimal degrees, not map projections (i.e. PCS).

2.5. A GCS that is incorrectly defined

This is similar to the case in Section 2.2, but again, it deals with GCS such as decimal degrees, not map projections (i.e. PCS). To reiterate, this type of error can lead to data display errors and can generate great frustration and confusion. If the GIS dataset in question has a PCS (i.e. a map projection) and then subsequently it is misdefined as a GCS, these data can disappear (become extremely small in scale relative to the projected data) when added to a data frame defined with a map projection (i.e. a PCS). Sometimes this is colloquially referred to as disappearing into a "black hole" because the data are invisible due to the illogical nature of the definition.

2.6. A GCS that is undefined

This is similar to the case in Section 2.3, the missing projection file, but it deals with GCS such as decimal degrees, not map projections (i.e. PCS). Again, this is a common error in geo-spatial data exchange among GIS students and colleagues.

2.7. A non-GCS

This is the state of geo-spatial data that is not yet georeferenced. It could be a raw scan of a paper map or an image file of a map captured by a camera (e.g. *.jpg or *.tif). These data have coordinates in page units (in mm or inches) and must first be put through a geo-referencing process to place it in projected (Cartesian) coordinate map space, or a GCS, before they are usable in a GIS.

3. Improving GIS education in the future and troubleshooting

3.1. Model curricula and GIScience

PCS or map projection is among the most important topics in GIS curricula as identified in a survey of GIS educators in higher education (Fagin and Wikle 2011). Numerous model curricula identify map projections as an essential component of and a fundamental topic in GIS courses (Unwin 1990). This includes the National Center for Geographic Information and Analysis core curriculum in the “technical area” (Goodchild and Kemp 1992; Kemp 2012; Kemp and Wright 1997) and the First Edition of Geographic Information Science (GIScience) and Technology Body of Knowledge in the area of geospatial data (DiBiase et al. 2007). Fagin and Wikle (2011) argue more emphasis on pedagogy is needed to train graduate students to be effective future GIS instructors. Thus, more pedagogical devices such as the one presented in this paper are valuable teaching tools. Goodchild (1992) notes the importance of map projections and that GIScience needs to revive the orthographic projection to conduct analyses at the global level “to understand geographical processes at the global scale.” Understanding the nuances of PCS (i.e. map projections) and GCS will become more important as theory and practice in global sciences increase.

Students and professionals learning about GIS now have some excellent resources for map projections and their parameters in the European Petroleum Survey Group (EPSG) catalog (http://spatialreference.org/) and utilities that use the coordinate system database (http://epsg.io). Esri also has several useful documents (see http://pro.arcgis.com/en/proapp/arcpy/classes/pdf/geographic_coordinate_systems.pdf and http://pro.arcgis.com/en/pro-app/arcpy/classes/pdf/projected_coordinate_systems.pdf).

3.2. Best practices and troubleshooting map projection errors

When exchanging geo-spatial data with a colleague the projection file (correctly defined) should be copied and zipped along with all the other essential files. Another best practice is to use self-defining file formats such as GeoTiffs rather than Tiffs with *.aux files or using feature classes in geodatabases rather than shapefiles, as discussed above in the introduction. In the future these types of formats will hopefully prevail, however, the mass proliferation of shapefiles (especially in Internet-based data repositories and archives) will require that students learn how to effectively deal with them. The open specification shapefile format is also commonly employed for open source GIS, such as the Quantum GIS software.
package. It is plausible that the shapefile format will be with us for several decades necessitating the continued education in the use of projection files.

What can be done to troubleshoot a missing projection file? First, the data user should contact the person or website the data was originally obtained from to inquire whether one exists or what the map projection is supposed to be, so the user can define it themselves. If this fails the user should then look at the form of the map coordinates (i.e. how the coordinates appear) to see if they match the coordinates of a dataset in the same geographic area with a known map projection or to rule out/in certain coordinate systems (e.g. Universal Transverse Mercator (UTM) eastings are all positive and less than one million). If the coordinates appear the same or very similar then try defining the undefined dataset and test if the new definition corrects the problem (i.e. it allows the datasets to properly align and behave as expected). The user should start by looking at commonly used coordinate systems and map projections for their area of interest, such as UTM and State Plane systems. Other specific tests can also be performed such as examining min/max coordinates in the Properties dialogue box.

If the manual techniques described above fail, there are free new stand-alone software and web-based utilities being developed to determine or best “guess” an unknown coordinate system. First released in 2016, one such software utility is called “Shapefile ProjectionFinder” developed by Manfred Egger (https://www.egger-gis.at/automatic-projection-detection/shapefile-projectionfinder/) (Egger 2016). Another tool is “Projectionuesser” (http://jhnkly.github.io/projectionuesser/) developed by cartographer John Kelly at the GreenInfo Network, a non-profit GIS consulting group. Using this method the user uploads a simplified shapefile (with only *.shp and *.shx files) that contains just one polygon. The utility then uses PostGIS to project the data using all the projections in its database. Each of these coordinate system iterations is mapped to a global map and the user is asked to select the correctly positioned polygon which then results in a hyperlink to the ESPG database (discussed above) of map projections and their parameters. The user then selects a file format from a list of specific GIS software programs to copy the projection text file. These two tools are the first known free and publicly accessible utilities to attempt the function of determining a map projection due to a missing projection file that the author is aware of, and others may evolve in the near future.

What can be done to troubleshoot a misdefined or incorrectly defined projection file? As discussed above, oftentimes a novice user will change the definition of a map projection (using “Define Project”) without changing the actual map projection of the geo-spatial data. If this problem is suspected then the user should first re-examine the original file(s) received from the website or colleague that provided the data to see if a projection file is present or not. If one is present then compare the coordinate systems to see if there is an error and if one is not present then use the techniques described above. If the projection file is still generating misalignment then delete the projection file and use the map projection utilities described above.

All of the above techniques underscore the importance of geo-spatial data stewardship. It is critically important that a copy of the original data should always be kept and that modifications to a dataset should be made to a subsequent copy of the dataset. This would allow for the comparisons described above. Communication with the people or website that provided the original geo-spatial data can be critically important, however, sometimes this is not possible, especially if metadata are lacking sources, and therefore troubleshooting can be a user’s last resort.

The “project-on-the-fly” utility of Esri software can confound understanding these map projection problems. ArcMap sets the default datum and map projection based on the first layer added to a new map window. Subsequently added data layers are automatically converted to the first layer’s map projection. Most of the time this utility is very helpful and improves workflow efficiencies by not having to pre-convert all layers to the same projection for viewing purposes. However, to troubleshoot map projection errors this utility can cause confusion, especially if an incorrectly defined data layer is added to a data frame as the first layer. It would be helpful if there was an option to turn off the “project-on-the-fly” utility for map projection error troubleshooting purposes.

3.3. GIS textbooks

In future introductory GIS textbooks, and future editions of existing textbooks, it would be helpful if common map projection errors are discussed. Textbooks (Chang 2019; Shellito 2015) have exercises to define a dataset without a projection file and then to project it to a new projection. These are fundamental and useful exercises, but to alert novice and new learners to the potential errors in map projections could prevent many frustrating experiences trying to cope with these errors. The textbook by Shellito (2015) has the best description of potential errors in map projections and has a brief description about misdefined or incorrectly defined projection files and non-GCSs. All seven states of geo-spatial data with respect to map projection and definition along with troubleshooting should be made known to students and novice users.
4. Conclusions

The complexity of map projections and GCSs, and with respect to their definition in geo-spatial data, is a major academic challenge to teach and to learn as a beginning student. This paper presents a novel pedagogical device to help current and future GIS instructors deal with a part of this complexity, as it relates to common technical errors propagated in projection files, their definition, or missing projection files. The commonly made errors by beginner and novice GIS users discussed in this paper are not comprehensively covered in any GIS textbook (to the author’s knowledge). In the future, hopefully, the seven possible states of GIS data with respect to map projection and definition described in this paper will be included to help students and GIS professionals parse these problems out and effectively troubleshoot them. There are new tools for troubleshooting map projection file problems and can help geo-spatial analysts (novice and advanced) work more effectively with their data.

Acknowledgments

The author would like to thank A. Manfree for information about troubleshooting. This article was improved by three anonymous reviewers.

Funding

This research was funded by the USDA National Institute of Food and Agriculture, Hatch Project 1011533.

Notes on contributor

Steven E. Greco (PhD) is a professor at the University of California, Davis, in the Department of Human Ecology where he conducts research and teaching in the topics of GIS, landscape ecology, and conservation planning.

References

Bampton, M. 2012. “Addressing Misconceptions, Threshold Concepts, and Troublesome Knowledge in GIScience Education.” In Teaching Geographic Information Science and Technology in Higher Education, edited by D. Unwin, 117–132. Hoboken, NJ: Wiley.

Chang, K. 2019. Introduction to Geographic Information Systems. 9th ed. New York, NY: McGraw Hill.

DiBiase, D., M. DeMers, A. Johnson, K. Kemp, A. T. Luck, B. Plewe, and E. Wentz. 2007. “Introducing the First Edition of Geographic Information Science and Technology Body of Knowledge.” Cartography and Geographic Information Science 34 (2): 113–120. doi:10.1559/152304007781002253.

Downs, R. M., and L. S. Liben. 1991. “The Development of Expertise in Geography: A Cognitive-Developmental Approach to Geographic Education.” Annals of the Association of American Geographers 81 (2): 304–327. doi:10.1111/j.1467-8306.1991.tb01692.x.

Egger, M. 2016. “Shapefile Projectionfnder.” In Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings: Volume 16, Article 11. Bonn, Germany. https://scholarworks.umass.edu/foss4g/vol16/iss1/11. doi: 10.7275/R54T6GJ9.

Fagin, T. D., and T. A. Wikle. 2011. “The Instructor Element of GIS Instruction at US Colleges and Universities.” Transactions in GIS 15 (1): 1–15. doi:10.1111/tgis.2011.15.issue-1.

Foote, K. E., D. Unwin, N. J. Tate, and D. DiBiase. 2012. “GIS&T in Higher Education: Challenges for Educators, Opportunities for Education.” In Teaching Geographic Information Science and Technology in Higher Education, edited by D. Unwin, 1–15. Hoboken, NJ: Wiley.

Goodchild, M. F. 1992. “Geographical Information Science.” International Journal of Geographical Information Systems 6 (1): 31–45. doi:10.1080/02693799208901893.

Goodchild, M. F., and K. K. Kemp. 1992. “NGCIA, Education Activities: The Core Curriculum and Beyond.” International Journal of Geographical Information Systems 6 (4): 309–320. doi:10.1080/02693799208901915.

Greco, S. E. 2016. Learning GIS: A Lab Manual for Learning ArcGIS Desktop. 4th ed. Dubuque, IA: Kendall Hunt.

Kemp, K. K. 2012. “Reflections on Curriculum Development in the US and Abroad: From Core Curriculum to Body of Knowledge.” In Teaching Geographic Information Science and Technology in Higher Education, edited by D. Unwin, 47–59. Hoboken, NJ: Wiley.

Kemp, K. K., and R. Wright. 1997. “UCGIS Identifies GIScience Education Priorities.” Geo Info Systems 7 (9): 16–18. 20.

Kennedy, M., and S. Kopp. 2000. Understanding Map Projections. Redlands, CA: ESRI Press.

Lo, C. P., and A. K. W. Yeung. 2007. Concepts and Techniques of Geographic Information Systems. 2nd ed. Upper Saddle River: Prentice Hall.

Maher, M. M. 2010. Lining up Data in ArcGIS: A Guide to Map Projections. Redlands, CA: ESRI Press.

McDonnell, P. W., Jr. 1979. Introduction to Map Projections. New York, NY: M. Dekker.

Hullito, B. A. 2015. Discovering GIS and ArcGIS. New York, NY: W. H. Freeman.

Unwin, D. 1990. “A Syllabus for Teaching Geographical Information Systems.” International Journal of Geographical Information Systems 4 (4): 457–465. doi:10.1080/02693799008941559.