Georeferencing and the Distribution of *Argia vivida* in California from 1910 to 1996

Jesson Go
December 16, 2010

Abstract

Georeferencing assigns longitude and latitude, among other spatial information, to data in a particular study; however, the georeferencing process entails a large amount of subjectivity and potential error. Using collections from the UC Berkeley Essig Museum of Entomology and the California Academy of Sciences, the georeferencing process is analyzed based on *Argia vivida* specimens across the state of California collected between 1910 and 1996.

Introduction

Many museum collections house thousands of specimens gathered across previous decades but have yet to be catalogued in digital databases or made available to researchers and the public in an accessible and meaningful way. In order to accomplish this, every specimen in a museum collection is numbered and labeled with information, particularly the taxonomic name, date collected, and location collected. This information can then be digitized and georeferenced; georeferencing is the process that assigns latitude and longitude coordinates to each specimen based on the locations listed on these labels, allowing these specimens to be spatially mapped for distribution studies, comparisons to climate data, and so on.

For this small study, specimens of the damselfly species *Argia vivida* collected between 1910 and 1996 in California were taken from the Essig Museum at UC Berkeley and the California Academy of Sciences in San Francisco. Like many odonates, *Argia vivida* generally spawns in aquatic environments; the larvae are mostly carnivorous[1]. The adults are distinguished by their vivid blue color and are largely collected around pools and streams that are sustained by springs, as well as around hot springs in some locations[2]. Georeferencing the *Argia vivida* collections will give an idea of the species' distribution in California over a period of several decades.

Methods

All of the specimens from the collections had labels with location information that was manually entered verbatim into the Essig Museum database; information such as the date collected and the California county of collection was entered into corresponding fields in the database. Each specimen was given a unique identification number that was referred to its database record. After data entry, the information in the database was downloaded and served as the basis for georeferencing; none of these records had a longitude and latitude coordinate pair or

---

[1] Koslucher, D. G., Minshal, G. W.
[2] Westfall, M. J., Jr., May, M. L., p 290

an area of uncertainty to describe the extent of the location. Determining these values was the georeferencing process itself.

The point-radius method[3] was used as the basis for georeferencing *Argia vivida*, where general guidelines were used to determine the latitude, longitude, and extent of the location. A location was first determined by searching for the location in a gazetteer such as AcmeMapper, which usually returned a latitude and longitude pair. If this pair was not the geographic center of the location, a new center and corresponding coordinate pair was manually selected. A radius was drawn from the location center to the furthest point in that given location; this may have been the city limits or the edge of a lake. The resulting circular area was intended to reflect the uncertainties with the extent the location, since most locations are not perfect circles and may have clearly defined boundaries. By convention, the extent of a named place was measured from the center of the location to its furthest point on the map. However, these boundaries were not always clear, and so for a more conservative estimate the extent was measured halfway between its own center to the center of the nearest named place. If there were two named places in the overall locality of the specimen, or if there are multiple possibilities for a location within a close proximity, the extent of each named place was measured as if it were an independent area. These extents were added to each other and then added to the distance between the centers of both named places; the center of only one named place was chosen as the georeferenced location. The whole procedure was done to account for the amount of error and uncertainty involved in the location specified on the specimen label.

The georeferencing calculator was then used to correct for errors in the latitude and longitude coordinates as well as the extent[4]. Georeferencing required a reference such as a gazetteer for determining longitude and latitude coordinates; the calculator factors in any potential error from the gazetteer in terms of direction and distance. The resulting calculation produced a corrected longitude and latitude and a corrected uncertainty. The main inputs were the initial latitude and longitude, any offset distance given (i.e. 3 miles from), any direction given (i.e. 3 miles north of), and the extent of the named place; the units for the coordinates and the distances were chosen (decimal degrees for latitude and longitude, miles for the distances). To correspond with the datum of BioGeomancer and AcmeMapper, the WGS84 datum was used for all calculations; the coordinate precision was set to "exact" and the distance precision used was ½ mile.

Once all of the specimens had a latitude and longitude coordinate and a net uncertainty, the Excel table containing all of the information was uploaded into ArcMap, where the points were mapped onto a county map of California projected to a Geographic Coordinate System (GCS_North_American_1983)[5]. Two more maps were produced by grouping the data by year: one map shows the distributions of points from 1910 to 1950 and the other shows the distributions from 1950 to 1996.

Results

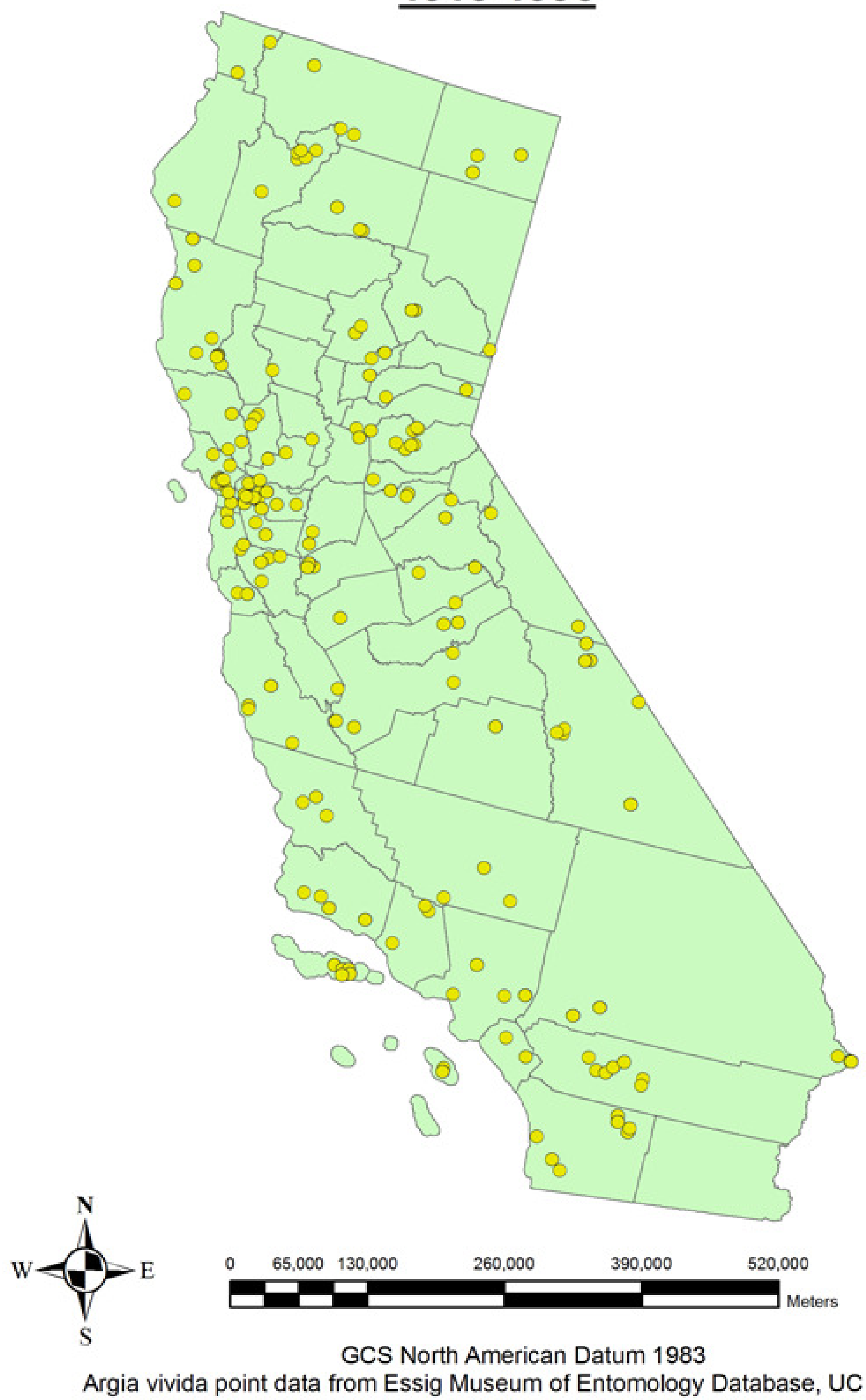
The uncertainties for the specimen locations ranged from 95 meters for "EMEC304140: UC Berkeley, near Wellman Hall", to 44,000 meters for "EMEC304564: Argus Mountains". The

[3] Wieczorek, J., Guo, Q., Hijmans, R. J.
[4] http://manisnet.org/gci2.html
[5] U.S. Census Bureau

smaller uncertainties tended to be associated with smaller locations with more easily defined boundaries, such as a lake or creek, while the largest uncertainties were associated with mountain ranges, label locations that listed multiple named places, or locations that had a large offset distance between the named places.

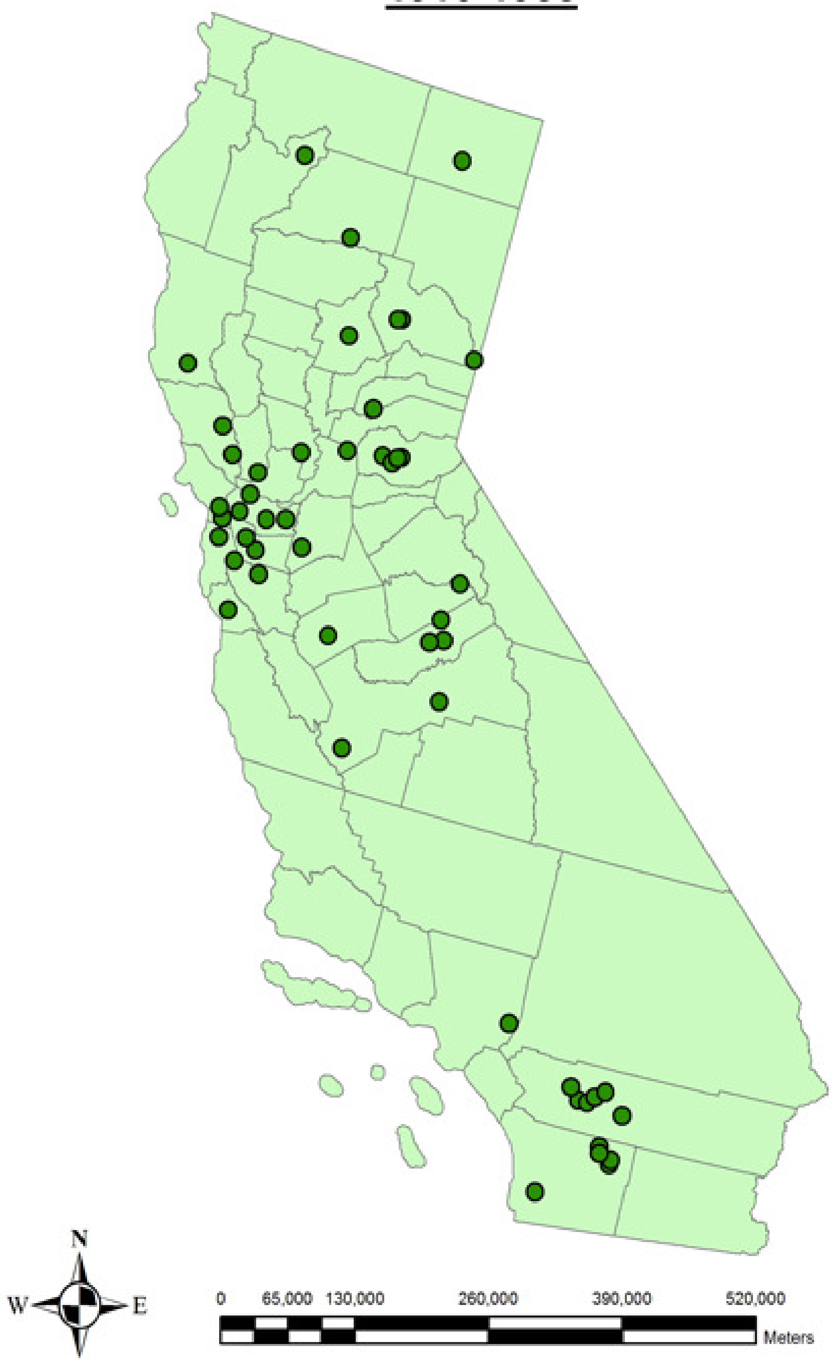
From the 625 specimens used in the following figures, the distribution of *Argia vivida* seems to be widespread across the state, although it should be noted that multiple specimens were found in the same location. Two other maps show the distribution of *Argia vivida* between 1910 and 1996: the first map shows distributions before (and during) 1950 and the second map shows distributions after 1950. Argia vivida distributions seem to have spread out across California in the period after 1950. The maps only display the location of each data point and do not depict the extent associated with each point.

<u>Distribution of *Argia vivida* across California 1910-1996</u>



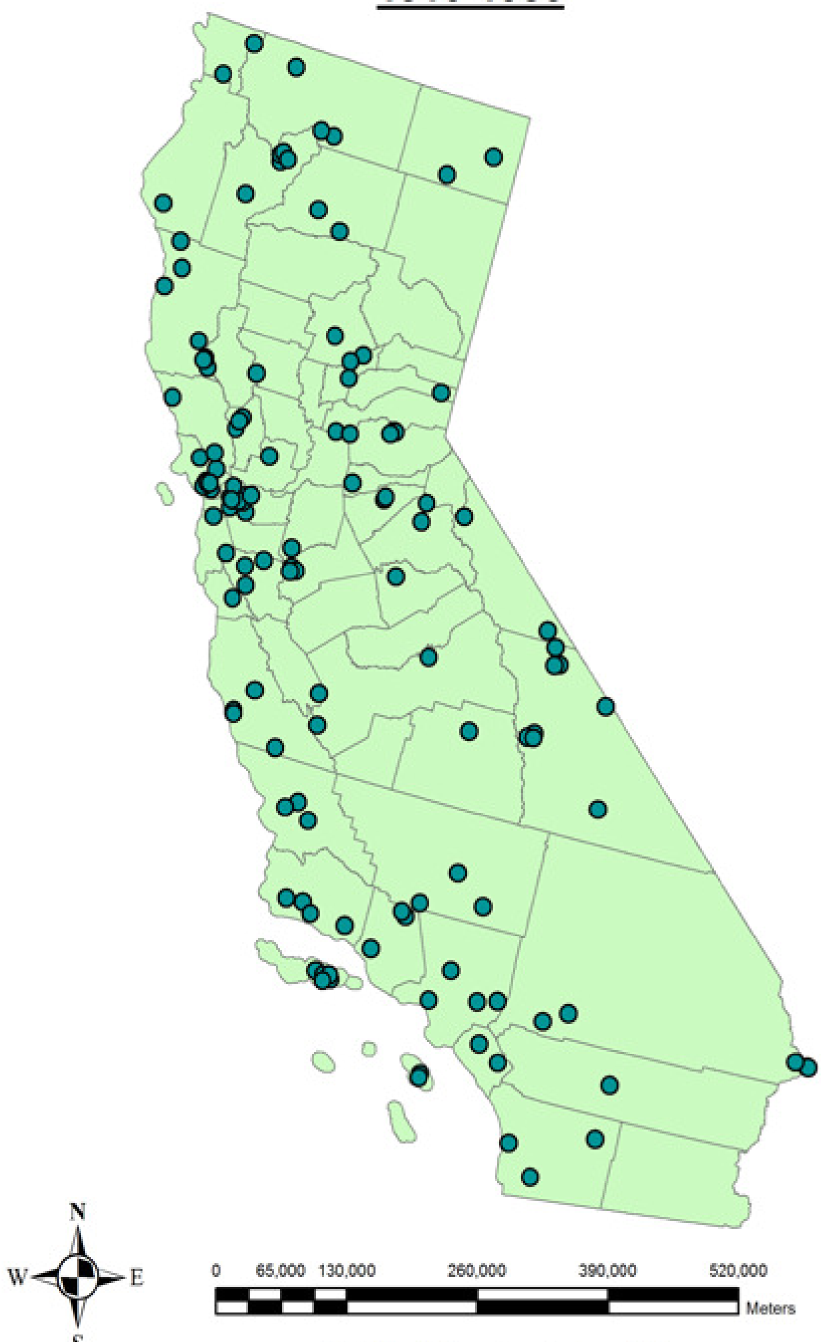
GCS North American Datum 1983
Argia vivida point data from Essig Museum of Entomology Database, UC Berkeley
County Map of California from the US Census Bureau

Distribution of *Argia vivida* across California 1910-1950



GCS North American Datum 1983
Argia vivida point data from Essig Museum of Entomology Database, UC Berkeley
County Map of California from the US Census Bureau

Distribution of *Argia vivida* across California 1910-1950



GCS North American Datum 1983
Argia vivida point data from Essig Museum of Entomology Database, UC Berkeley
County Map of California from the US Census Bureau

Discussion

The main issue throughout the georeferencing process was the subjectivity involved in interpreting the named places associated with each specimen and determining the extent of each of the named places. While most of the specimens had county names in addition to a named place, which narrowed the possibilities of the intended locality, areas within the specified county still had different areas with the same name. An area referred to as “Sagehen” in Nevada County could have possibly referred to “Sagehen Field Station,” “Sagehen Meadows,” or “Sagehen Springs,” among a few other possibilities since “Sagehen” was not explicitly stated as one type of geographic feature or another. The locations noted on the labels were sometimes outdated or not widely used, and so these locations could not be found in a search engine or by inquiring AcmeMapper or BioGeomancer. For locations such as “Mokel Hill”, a close match was found (in this case “Mokelumne Hill”) and an assumption was made to use treat the closest match as the georeferenced locality, though this clearly has a degree of uncertainty involved. Even if a named place was identified and located with latitude and longitude coordinates, the extent of the locality could have been subject for debate. Other errors could have originated from the manual data entry, especially for specimens that had handwritten and illegible labels. For these reasons alone, georeferencing has proven to be a very slow and time-consuming process.

Initially, the BioGeomancer application was used to produce latitude and longitude coordinates, as well as extents of named places, an option allowed one to switch between a boundary geospatial map, a satellite map, a hybrid map (combining the satellite and geospatial maps), and a terrain map. However, BioGeomancer consistently returned the same values for different location queries, including values such as 5062 meters. The database used by BioGeomancer is also incomplete; either an error message will appear for an undocumented named place, or a completely irrelevant location will be returned from the query. Several California locations were searched with the “CA” abbreviation but returned a location in Canada with an extremely large extent.

AcmeMapper was the main gazetteer used to manually determine the extent of named places and offered multiple maps for analysis in addition to the maps provided in BioGeomancer, AcmeMapper utilized more map options, particularly the topographic map. The topographic map was especially helpful in locating and determining the extent of more obscure or historical named places. Like BioGeomancer, AcmeMapper sometimes returned erroneous locations for the queries provided, especially if the searched name was not well known. A number of such queries consistently returned a location at 36.77826, -119.4179 near Kings River and Avocado Lake in Fresno County.

Georeferencing is still a useful tool for projecting species distribution over time, provided that there are a large number of specimens to georeference. However, this study examined a statewide distribution of *Argia vivida*; a larger scale entails more potential error due to less precision in the data points. The above figures seem to show a difference in distributions, but such observation may be due to a difference in the number of specimens, or in this case, the number of georeferenced locations corresponding to specimens. The larger scale could be used to select smaller areas for a more regional study but itself would not provide the most accurate representation of *Argia vivida* distributions in California.

Works Cited

Koslucher, D.G., Minshall, G.W., 1973, "Food Habits of some Benthic Invertebrates in a Northern Cool-Desert Stream." Transactions of the American Microscopical Society, 92 (3), 441-452.

U.S. Census Bureau. "2008 TIGER/Line Shapefiles for: California." http://www2.census.gov/cgi-bin/shapefiles/state-files?state=06

Westfall, Minter J., Jr., May, Michael L. Damselflies of North America. Gainesville, FL: Scientific Publishers, 1996.

Wieczorek, J., Guo, Q., Hijmans, R. J., 2004. "The Point Radius method for georeferencing locality descriptions and calculating associated uncertainty." International Journal of Geographical Information Science, 18 (8), 745-767.

Wieczorek, J., 2007. "Georeferencing Calculator." http://manisnet.org/gci2.html