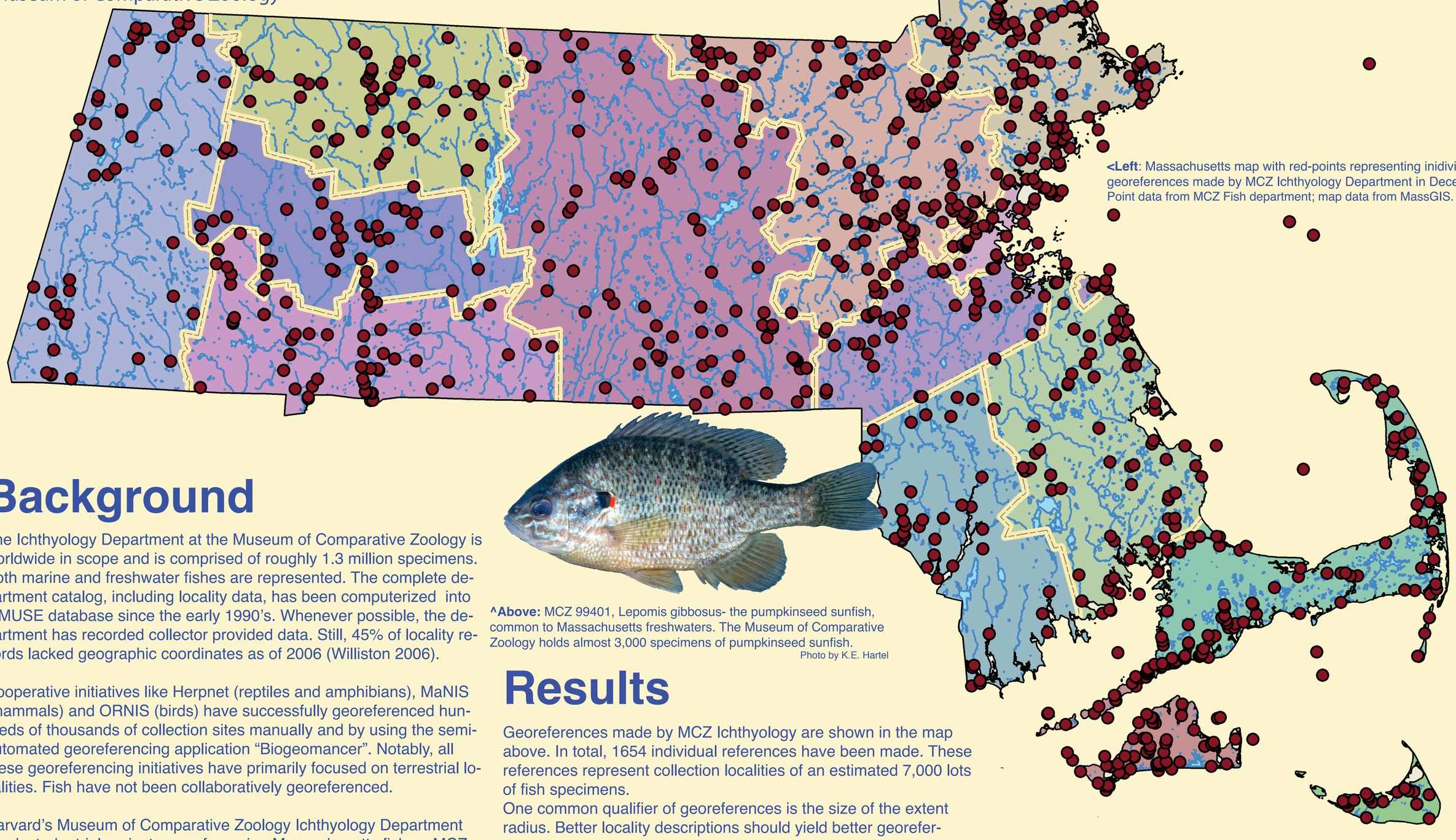
Georeferencing Massachusetts Fishes

Andrew D. Williston and Karsten E. Hartel
Ichthyology Department
Museum of Comparative Zoology



<Left: Massachusetts map with red-points representing inidividual georeferences made by MCZ Ichthyology Department in December, 2007. Point data from MCZ Fish department; map data from MassGIS.

^Above: MCZ 99401, Lepomis gibbosus- the pumpkinseed sunfish, common to Massachusetts freshwaters. The Museum of Comparative Zoology holds almost 3,000 specimens of pumpkinseed sunfish.
Photo by K.E. Hartel

Background

The Ichthyology Department at the Museum of Comparative Zoology is worldwide in scope and is comprised of roughly 1.3 million specimens. Both marine and freshwater fishes are represented. The complete department catalog, including locality data, has been computerized into a MUSE database since the early 1990's. Whenever possible, the department has recorded collector provided data. Still, 45% of locality records lacked geographic coordinates as of 2006 (Williston 2006).

Cooperative initiatives like Herpnet (reptiles and amphibians), MaNIS (mammals) and ORNIS (birds) have successfully georeferenced hundreds of thousands of collection sites manually and by using the semi-automated georeferencing application "Biogeomancer". Notably, all these georeferencing initiatives have primarily focused on terrestrial localities. Fish have not been collaboratively georeferenced.

Harvard's Museum of Comparative Zoology Ichthyology Department conducted a trial project: georeferencing Massachusetts fishes. MCZ Ichthyology has an extensive collection of Massachusetts fishes. Only about 219 of 2110 of Massachusetts inshore localities have collector provided coordinate data. Now, over 1600 georeferences have been made using the point radius methods. The department notes a number of considerations that are important in georeferencing a primarily aquatic collection.

Methods

The most common and widespread collection georeferencing method is the point-radius method. According to an established set of standards (Chapman and Wieczorek, 2006), the georeferencer assigns a coordinate point to a known location and an accompanying radial distance that expresses the probable uncertainty and extent of the reference. Recorded data included: longitude, latitude, coordinate determiner, determined date, maximum error, datum, method, locality extent, verification status and remarks.

Several applications were used in the point radius method, including Google Earth, on-line gazetteers (Topozone, Fuzzy-G), but primarily Biogeomancer.

References

Biogeomancer Workbench. Version 1.10. 2007-2008. <http://bg.berkeley.edu/latest/>

Chapman, A.D. and J. Wieczorek (eds). 2006. Guide to Best Practices for Georeferencing. Copenhagen: Global Biodiversity Information Facility.

Fuzzy-G, the fuzzy gazetteer. 2002-2005. FuzzyG v2.1-dev, Christian Kohlschütter. <http://tomcat-dmaweb1.jrc.it/fuzzyg/query/>

Google Earth. 2007. Google. <www.earth.google.com>

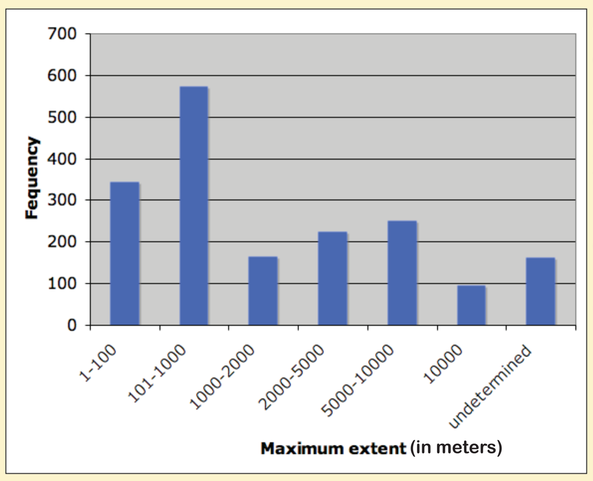
TopoZone.com. 1999-2007. Demand Media, Inc. <www.topozone.com>

Williston, A.D. 2006. GIS and Curation in the MCZ Ichthyology Department. Harvard Returns to Geography Poster Session.

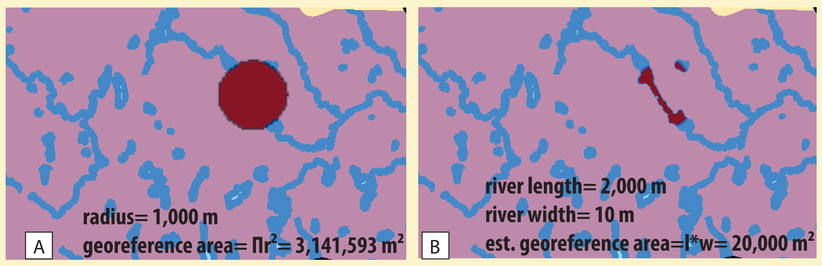
Results

Georeferences made by MCZ Ichthyology are shown in the map above. In total, 1654 individual references have been made. These references represent collection localities of an estimated 7,000 lots of fish specimens.
One common qualifier of georeferences is the size of the extent radius. Better locality descriptions should yield better georeferences with smaller maximum extents. The graph below is a frequency distribution evaluation of MCZ Ichthyology's Massachusetts georeferences based on the Uncertainty Index of Chapman and Wieczorek (2006).

The most commonly cited reason for not georeferencing a description was that the description was too vague and the resulting georeference would have limited or confusing biological significance. The georeferences with the least uncertainty often represented intersections of terrestrial and aquatic paths (i.e. bridges). Potential improvements to the Massachusetts data set includes determining datum associated with collector provided coordinates



<Left: Frequency distribution of maximum extents in georeferenced Massachusetts records. The majority of Massachusetts records could be georeferenced within an extent of 1km or less.



^Above: Two figures demonstrating the problem of spatial fit in georeferencing aquatic data. **A**. Point radius method. **B.** Extent trimmed to the outline of the georeferenced feature.

Future Directions

Georeferencing MCZ Massachusetts fishes has increased the MCZ's number of Massachusetts fish georeferences 8-fold in the course of three weeks. Judging from this initial success, georeferencing will be very productive in the MCZ Fish collection. Still, there are still challenges in georeferencing aquatic data that may not have been fully addressed in previously established georeferencing standards.

Not all workbench applications may be presently best suited for aquatic locality types. It was noted that the most frequent and precise locality descriptions were often road/river intersections. This might reflect common collecting practices in ichthyology. Road/river junctions were not recognized by Biogeomancer workbench during the time of this project. It is suggested here that future georeferencing applications for aquatic data account for river and road junctions.

Aquatic localities, like ponds and stream sections, are often not well represented by the point radius method. Assigning a circular area to a winding river is inaccurate- most of the extent area is dry land where there is often no possibility to collect a live fish (Figure 2A). This general problem has been referred to as "spatial fit" (Chapman and Wieczorek (eds). 2006). GIS "intersect" geoprocessing of extent circles against known aquatic features shape files might produce more accurate, trimmed extents (Figure 2B). The difference in accuracy between radii and these trimmed extents might be dramatic- as much as a 99% reduction in area, based on a 1km extent that represents a straight, 10 meter wide river. While trimming methods could continue to use the point radius method, special considerations, especially recording the feature's name, might facilitate geoprocessing.

Georeferencing has been considered "discipline independent." In theory this is true- a place can be a coordinate, whether that coordinate represents a fish or a tree. In practice, place descriptions can be best understood within context of the object represented. In a sense, the object can represent some part of the locality desciption and greatly effect the georeferencer's judgment. To its credit, the zoological community has considered this by forming defined initiatives which can evaluate localities in the context of the collection: Herpnet, Manis and Ornis. A similar community approach amongst aquatic disciplines- ichthyology, malacology, aquatic invertebrates, etc.- might be the most efficient way to address issues specific to aquatic data.