INT. J. GEOGRAPHICAL INFORMATION SCIENCE
VOL. 18, NO. 8, DECEMBER 2004, 745–767

Taylor & Francis
Taylor & Francis Group

Research Article

The point-radius method for georeferencing locality descriptions and calculating associated uncertainty

JOHN WIECZOREK*

Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building,
University of California, Berkeley, CA 94720, USA;
e-mail: tuco@socrates.berkeley.edu

QINGHUA GUO

Department of Environmental Sciences, Policy & Management, 151 Hilgard
Hall #3110, University of California, Berkeley, CA 94720, USA

and ROBERT J. HIJMANS

Museum of Vertebrate Zoology, 3101 Valley Life Sciences Building,
University of California, Berkeley, CA 94720, USA

Natural history museums store millions of specimens of geological, biological, and cultural entities. Data related to these objects are in increasing demand for investigations of biodiversity and its relationship to the environment and anthropogenic disturbance. A major barrier to the use of these data in GIS is that collecting localities have typically been recorded as textual descriptions, without geographic coordinates. We describe a method for georeferencing locality descriptions that accounts for the idiosyncrasies, sources of uncertainty, and practical maintenance requirements encountered when working with natural history collections. Each locality is described as a circle, with a point to mark the position most closely described by the locality description, and a radius to describe the maximum distance from that point within which the locality is expected to occur. The calculation of the radius takes into account aspects of the precision and specificity of the locality description, as well as the map scale, datum, precision and accuracy of the sources used to determine coordinates. This method minimizes the subjectivity involved in the georeferencing process. The resulting georeferences are consistent, reproducible, and allow for the use of uncertainty in analyses that use these data.

1. Introduction

Natural history collections contain more than 2500 million specimens of geological, biological, and cultural entities (Duckworth *et al.* 1993). These resources constitute a foundation for numerous scientific disciplines, such as anthropology, biogeography, biosystematics, conservation biology, ecology, and paleontology. The data associated with natural history specimens vary widely in nature and content between disciplines as well as between institutions, including everything

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810412331280211

from hand-written notes taken in the field at the time of collection (field notes) to databases and published articles in professional journals. Underlying this variation, however, is a core set of concepts common to all natural history collections, one of the most important of which is the 'collecting event' - a description of the time and place (locality) where a specimen was collected. The collecting event is an essential association between the specimen and its natural context and is required for quantitative analyses of specimen data together with other spatial data using geographical information systems (GIS).

Despite increasing interest in natural history collection data, there remain considerable obstacles to their use in GIS. The most prevalent of these obstacles is that locality descriptions are often not georeferenced. Traditionally, localities have been recorded as textual descriptions, often based on names and situations that can change over time. This tradition is slowly changing to document localities with supplementary geographic coordinates, the value of which are now widely recognized (Krishtalka and Humphrey 2000, GBIF 2002) and the collection of which has been greatly facilitated by the availability of the Global Positioning System (GPS). Nevertheless, researchers interested in spatial analysis using museum specimen data face a daunting legacy of data without coordinates. For example, at the beginning of the 'Mammal Networked Information System' Project (MaNIS 2001), which consists of a distributed database network for mammal collections, 17 North American mammal collections pooled their specimen locality data for a collaborative georeferencing effort. 87.8% of the 296 737 distinct collecting localities from these collections had no coordinates. As of March 2003, 61.2% of the 3 260 453 specimens accessible through Lifemapper (KU-BRC 2002) did not have georeferenced localities. These statistics are typical of natural history collections data that are in digital media today, and indicate the magnitude of the georeferencing challenge.

In the relatively few cases in which localities have been assigned coordinates, there is seldom any documentation of the method used to determine those coordinates. For example, of the localities for which coordinates had already been determined at the outset of the MaNIS project, 78.4% of 36 197 records had no associated metadata regarding the areas encompassed by the localities, nor did they include information about the methods and assumptions used in assigning the coordinates and uncertainties associated with them. Thus, even where present, georeferenced localities may be of limited utility since we have no knowledge of how they were generated.

To the best of our knowledge, there are currently no published, comprehensive guidelines for georeferencing descriptive locality data. In the absence of such guidelines, it has been common practise to assign a single point to a locality, without estimates of how well that point represents the actual locality. Some authors call for the capture of categorical measures of uncertainty (McLaren *et al.* 1996, Knyazhnitskiy *et al.* 2000), but do not investigate the nature of uncertainties, their magnitudes, or how different sources of uncertainty combine. Given the nature of locality descriptions and the variation in quality of coordinate sources (maps and gazetteers, for example), uncertainty must be estimated under rigorous guidelines. Whereas the coordinates of some localities can be determined with great precision, others can only be roughly approximated. If these differences are not taken into account, uncertainties cannot be incorporated into analyses and it becomes impossible to determine whether a given record is appropriate for a particular application. Spatial analysis without consideration of data uncertainty may be of limited utility (Fisher 1999).

Numerous studies have investigated the positional accuracy of spatial data (Goodchild and Hunter 1997; Leung and Yan 1998; Veregin 2000; Van Niel and McVicar 2002), which is defined as the difference between test data and corresponding "true" data of demonstrably higher accuracy (Goodchild and Hunter 1997; FGDC 1998) and which is expressed as a standard error for a set of points in a GIS layer (Stanislawski *et al.* 1996; Bonner *et al.* 2003).

The approaches used in these studies cannot be directly applied to estimating uncertainty in georeferenced localities. In contrast to many spatial data sets, which consist of unambiguously identifiable objects that can be directly and repeatedly measured, it is difficult to provide true data against which to test for many of the types of potential errors ("uncertainties") that plague descriptive localities.

Here we present a simple, practical method for computing and recording coordinates for a locality. We identify the potential sources of uncertainty, present methods for determining their magnitudes, and provide a procedure for combining uncertainties into a single estimate of "maximum" uncertainty associated with the coordinates.

We propose that the method presented here provides a framework for producing consistent and accurate interpretations of the locality descriptions and represents a substantial improvement over current practices. Efficiency, accuracy, and repeatability are our primary goals.

2. Georeferencing methods

2.1. *Point method*

There are various methods by which locality descriptions can be georeferenced. The most commonly used is the 'Point' method, by which a single coordinate pair is assigned to each location. This method ignores the fact that a locality record always describes an area rather than a dimensionless point and that collecting may have occurred anywhere within the area denoted. The specificity (that is, how well the description constrains the interpretation of the area) with which a locality is recorded directly influences the range of research questions to which the data can be applied. For example, recording only the state from which a specimen was collected will not be of much utility in the compilation of a species list for a National Park in that state. By providing only a point for a georeferenced record, the distinction is lost between locality descriptions that are specific and those that are not.

2.2. *Shape method*

The shape method is a conceptually simple method that delineates a locality using one or more polygons, buffered points, and buffered polylines. A combination of these shapes can represent a town, park, river, junction, or any other feature or combination of features found on a map. While simple to describe, the task of generating these shapes can be difficult. Creating shapes is impractical without the aid of digital maps, GIS software, and expertise, all of which can be relatively expensive. Also, storing a shape in a database is considerably more complicated than storing a single pair of coordinates. Particular challenges to making this method practical for georeferencing natural history collections data include assembling freely accessible digital cartographic resources and developing tools for automation of the georeferencing process. Nevertheless, of all of the approaches discussed here, this method has the potential to generate the most complete digital spatial descriptions of localities.

2.3. *Bounding box method*

A common way to describe a geographic feature is to use a bounding box – a set of two pairs of coordinates that together form a rectangle (in the appropriate projection) that encompasses the locality being described. Geographic features in the Alexandria Digital Library Gazetteer Server (ADL 2001) are sometimes described using bounding boxes. The bounding box method is a limited shape method by which only points or projected rectangles can be described. This method offers some advantages over the shape method. For example, bounding boxes are much easier to produce and store than arbitrary shapes, particularly in the absence of digital cartographic tools. In addition, database queries can be performed on bounding boxes without the need for a spatial database engine. However, describing a locality with a bounding box tends to be less specific than describing it with a more complicated shape.

2.4. *Point-radius method*

The point-radius method describes a locality as a coordinate pair and a distance from that point (that is, a circle), the combination of which encompasses the full locality description and its associated uncertainties. The key advantage of this method is that the uncertainties can be readily combined into one attribute, whereas the bounding box method requires contributions to uncertainty to be represented independently in each of the two dimensions. This simple difference can have a profound effect on the economy of georeferencing. Recognizing the practical advantages for natural history collections, for which the economy of producing and maintaining data are critical concerns, the guidelines for georeferencing descriptive localities presented here will be described in terms of the point-radius method. Nevertheless, the discussions of the sources of uncertainty are relevant to the 'Shape' and 'Bounding box' methods as well.

3. Applying the point-radius method

3.1. *Step one: classify the locality description*

Locality descriptions among natural history collection data encompass a wide range of content in a baffling array of formats. From the perspective of georeferencing, however, there are effectively only nine different categories of descriptions (table 1). The locality type will determine the process of calculating coordinates and uncertainties.

A locality description can contain multiple clauses and can match more than one of the categories given in table 1. If any one of the parts falls into one of the first three categories, the locality should not be georeferenced. Instead, an annotation should be made to the locality record giving the reason why it is not being georeferenced. In this way, anyone who encounters the locality in the future will benefit from previous effort to diagnose problems with georeferencing the locality description.

If the locality description does not fall into any of the first three categories in

Table 1. Types of locality descriptions commonly found in natural history collections.

Type	Description	Examples
1) dubious	The locality explicitly states that the information contained therein is in question.	‘Isla Boca Brava?’, ‘presumably central Chile’
2) can not be located	Either the locality data are missing, or they contain other than locality information, or the locality cannot be distinguished from among multiple possible candidates, or the locality cannot be found with available references.	‘locality not recorded’, ‘Bob Jones’, ‘lab born’, ‘summit’, ‘San Jose, Mexico’
3) demonstrably inaccurate	The locality contains irreconcilable inconsistencies.	‘Sonoma County side of the Gualala River, Mendocino County’
4) coordinates	The locality consists of a point represented with coordinate information.	‘42.4532 84.8429’, ‘UTM 553160 4077280’
5) named place	The locality consists of a reference to a geographic feature (e.g., town, cave, spring, island, reef, etc.) having a spatial extent.	‘Alice Springs’, ‘junction of Dwight Avenue and Derby Street’
6) offset	The locality consists of an offset (usually a distance) from a named place.	‘5 km outside Calgary’
7) offset along a path	The locality describes a route from a named place.	‘1 km S of Missoula via Route 93” “600 m up the W Fork of Willow Creek’
8) offsets in orthogonal directions	The locality consists of a linear distance in each of two orthogonal directions from a named place.	‘6 km N and 4 km W of Welna’
9) offset at a heading	The locality contains a distance in a given direction.	‘50 km NE Mombasa’

table 1, the most specific part of the locality description should be used for georeferencing. For example, a locality written as 'bridge over the St. Croix River, 4 km N of Somerset' should be georeferenced based on the bridge rather than on Somerset as the named place with an offset at a heading. The locality should be annotated to reflect that the bridge was the locality that was georeferenced. If the more specific part of the locality cannot be unambiguously identified, then the less specific part of the locality should be georeferenced and annotated accordingly.

3.2. *Step two: determine coordinates*

The first key to consistent georeferencing using the point-radius method is to have well-defined rules for determining the coordinates of the point. Coordinates may be retrieved from gazetteers, geographic name databases, maps, or even from other locality descriptions that have coordinates (for example, from localities recorded in the field using a GPS receiver). The source and precision of the coordinates should be recorded so that the validity of the georeferenced locality can be checked at any time. The original coordinate system (for example, decimal degrees, degrees minutes seconds, UTM) and geodetic datum (for example, WGS84, NAD27) used in the coordinate source should also be recorded. This information helps to determine sources and degree of uncertainty, especially with respect to the original coordinate precision (section 3.3.3.3). We recognize that specific projects may require particular coordinate systems, but we find geographic coordinates in decimal degrees to be the most convenient system for georeferencing. Since this format relies on just two attributes, one for latitude and the other for longitude, it provides a succinct coordinate description with global applicability that is readily transformed to other coordinate systems as well as from one datum to another. By keeping the number of recorded attributes to a minimum, the chances for transcription errors are minimized.

When transforming coordinates from one system or datum to another, it is important to preserve as much precision as possible. Coordinate precision is not a measure of accuracy – it does not imply specific knowledge of the locality represented by the coordinates; that role is assumed by the uncertainty measurements, as described in section 3.3. Every coordinate transformation has the potential to introduce error. The greater the precision with which the coordinates are captured, the less the error that is propagated when further coordinate transformations are made.

3.2.1. *Identify named places and determine their extents*

The first step in determining the coordinates for a locality description is to identify the most specific named place within the description. Gazetteers and geographic name databases provide coordinates for named places (commonly referred to as 'features'). However, we use the term 'named place' to refer not only to traditional features, but also to places that may not have proper names, such as road junctions, stream confluences, and cells in grid systems (for example, Townships).

Every named place occupies a finite space, or 'extent'. In some sources, places may be given in the form of bounding box coordinates for larger features (ADL 2001), but in general only a coordinate pair, not an extent, is given. Some coordinate sources are accompanied by rules governing the placement of the coordinates within a named place. For example, the US Geographic Names Information Service (USGS 1981) places the coordinates of towns at the main post office unless the town is a county seat, in which case the coordinates refer to the county courthouse. Similarly, the same source places the coordinates of a river at its mouth. In the absence of one of these specific points of reference, the geographic centre of the named place is usually recorded. Because of these inconsistencies in assigning coordinates for named places, including inconsistencies within a single data source, the extent of the named place becomes an important consideration in determining uncertainty.

The geographic centre (that is, the midpoint of the extremes of latitude and longitude) of the named place is recommended as the location of the coordinates because it describes a point where the uncertainty due to the extent of the named place is minimized. If the locality describes an irregular shape (for example, a winding road or river) and the geographic centre of that shape does not lie within the locality, then the point nearest the geographic centre that lies within the shape is the preferred reference for the named place and represents the point from which the extent of and offsets from that named place should be calculated.

3.2.2. *Determine offsets*

Offsets consist of combinations of distances and directions from a named place. Some locality descriptions explicitly state the path to follow when measuring the offset (for example, 'by road', 'by river', 'by air', 'up the valley'). In such cases the georeferencer should follow the path designated in the description using a map with the largest available scale to find the coordinates of the offset from the named place. The smaller the scale of the map used, the more the measured distance on the map is likely to overshoot the intended target.

It is sometimes possible to infer the offset path from additional supporting evidence in the locality description. For example, in the locality '58 km NW of Haines Junction, Kluane Lake' supports a measurement by road since the final coordinates by that path are nearer to the lake than going 58 km NW in a straight line. Altitudes given with the locality description may also support one offset path over another. By convention, localities containing two offsets in orthogonal directions (for example, '10 km S and 5 km W of Bikini Atoll') are always linear measurements.

Sometimes the environmental constraints of the collected specimen can imply the method of measurement of the offset. For example, '30 km W of Boonville, California' if taken as a linear measurement, would lie in the Pacific Ocean. If this locality is supposed to refer to the collection site of a terrestrial mammal, it is likely that the collector followed the road heading west out of Boonville, winding toward the coast, in which case the animal was collected on land.

If either of the above methods fail to distinguish the offset method, it may be necessary to refer to more detailed supplementary sources, such as field notes or itineraries, to determine this information. Supplementary sources do not always exist or they may not contain additional information, making it difficult to distinguish between offsets meant to be along a path and those meant to be along a straight line. A particularly conservative approach is to not georeference localities that fall into this category and instead record a comment explaining the reasoning. However, value can still be derived by georeferencing localities that suffer from this ambiguity. One solution for dealing with these localities is to determine the coordinates based on one or the other of the offset paths. Another solution is use the midpoint between all possible paths. There may be discipline-specific reasons to choose one solution over another, but the georeferencer should always document the choice and accommodate the ambiguity in the uncertainty calculations.

3.3. *Step three: Calculate uncertainties*

The second key to consistent georeferencing using the point-radius method (after determining the coordinates of the point) is to have well-defined rules for determining the radius of the circle that encompasses the locality and all of its associated uncertainties. Whenever subjectivity is involved, it is preferable to overestimate uncertainty. We have identified the following six sources of uncertainty inherent in descriptive localities or the resources used to georeference them:

1) extent of the locality
2) unknown datum
3) imprecision in distance measurements
4) imprecision in direction measurements
5) imprecision in coordinate measurements
6) map scale

3.3.1. *Uncertainty due to the extent of the locality*

The extents of named places mentioned in locality descriptions are an important source of uncertainty. Not only are the rules for assigning coordinates to named places largely undocumented in most coordinate data sources, but also the points of reference may change over time – post offices and courthouses are relocated, towns change in size, and so on. Moreover, there is no guarantee that the collector paid attention to any particular convention when reporting a locality as an offset from a named place. For example, '4 km E of Bariloche' may have been measured from the post office at the civic plaza, or from the bus station on the eastern edge of town, or from anywhere else in Bariloche. In most cases we no longer have a way of knowing the actual location used to anchor the offset.

The maximum uncertainty due to the extent of the named place (figure 1) is the maximum distance between any two points within the named place (the 'span'). If we have coordinates for a named place from a gazetteer, for example, without knowing where in the named place those coordinates lie, then the span is the uncertainty due to the extent of the named place. If we have a map of the named place, then a more refined uncertainty estimate can be made by measuring the distance from the point marked by the coordinates to the point in the named place furthest from those coordinates. The magnitude of the uncertainty value is minimized if the coordinates mark the geographic centre of the named place and is generally about half the span of the locality.

Many localities are based on named places that have changed in size over time; current maps might not reflect the extents of those places at the time specimens were collected there. If possible, extents should be determined using maps dating from the same period as the specimen collecting events. In most cases, the current extent of a named place will be greater than its historical extent and the uncertainty

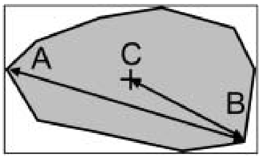


Figure 1. The maximum (AB) and minimum (BC) uncertainties due to the extent of a named place (shaded area).

will be somewhat overestimated if current maps are used. It is recommended to record the named place, its extent, and the source of these data while georeferencing so that users of the data can verify this important component of the uncertainty calculation.

3.3.2. *Uncertainty due to an unknown datum*

A geodetic datum is a mathematical description of the size and shape of the earth and of the origin and orientation of coordinate systems. Seldom in natural history collections have geographic coordinates been recorded together with geodetic datum information. Even now, with GPS coordinates being recorded as definitive locations, the geodetic datum is typically ignored. A missing datum reference introduces a complicated ambiguity, which varies geographically (Welch and Homsey 1997).

Many currently available maps of North America are based on the North American Datum of 1927 (NAD27), but the North American Datum of 1983 (NAD83) is being used increasingly more often among newer maps. NAD83 is essentially the same as the World Geodetic System of 1984 datum (WGS84), a standard reference datum for the Global Positioning Systems (Defense Mapping Agency 1991). We calculated the magnitude of uncertainty for North America (Canada, USA, and Mexico) based on the differences between NAD27 and NAD83/WGS84 (figure 2) using transformation functions in ArcGIS (ESRI, Redlands, CA, USA). The uncertainty from not knowing which of these datums was used to determine the coordinates varies in the contiguous USA from 0–104 m. In the extreme western Aleutian Islands of Alaska, the discrepancy can be as much as 237 m, while in Hawaii the differences are consistently ca. 500 m. On the global scale, we calculated a maximum uncertainty of 3552 m due to an unknown datum. This value was obtained by comparing pairwise distances between all combinations of datums listed in the WGS84 definition (NIMA 2000) at one degree intervals in both latitude and longitude. Given the potential magnitude of this uncertainty, every effort should be made to use coordinate sources that provide datum information and to record the datum of those sources as a routine part of data collection.

3.3.3. *Imprecision as a source of uncertainty*

Precision is a measure of the specificity with which a measurement is recorded. Precision can be difficult to gauge from a locality description; it is seldom, if ever, explicitly recorded. Further, a database record may not reflect, or may reflect incorrectly, the precision inherent in the original measurements, especially if the

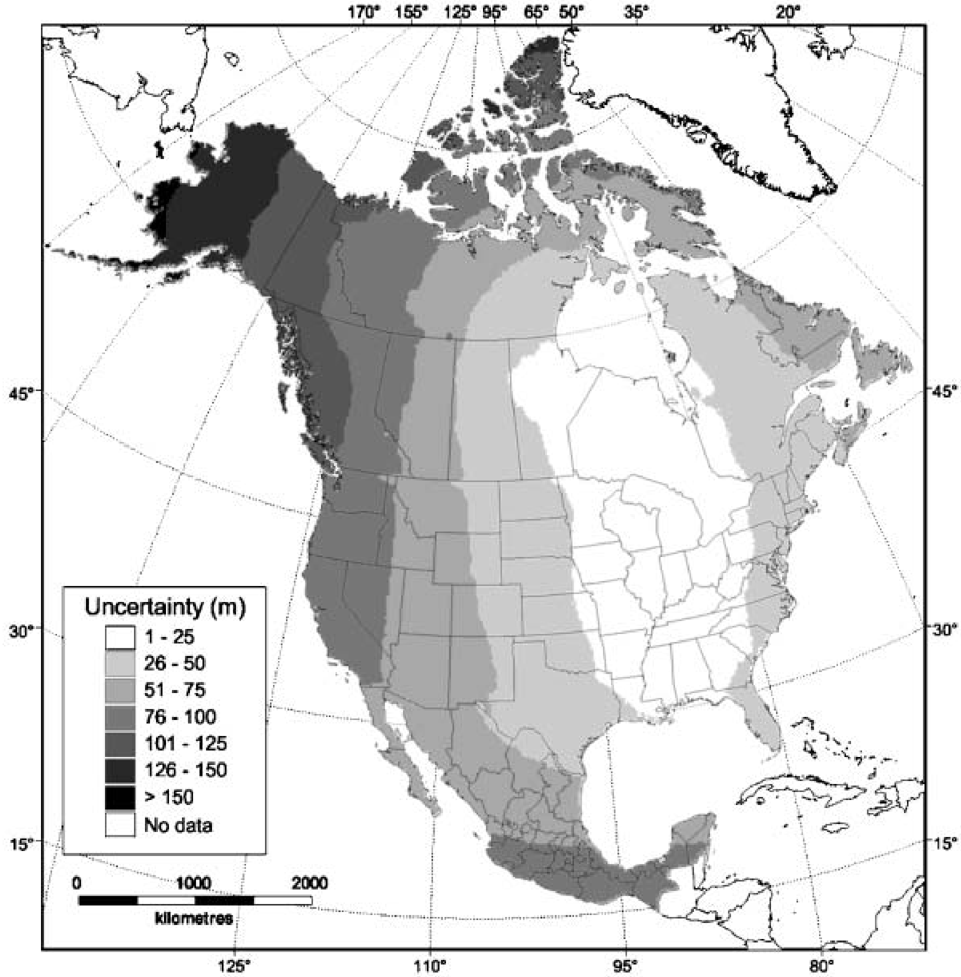


Figure 2. Uncertainty from not knowing whether coordinates were taken from a source using NAD27 or NAD83 – the geodetic datums most commonly used on maps in Canada, the USA, and Mexico.

locality description in the database has undergone standardization, reformatting, or secondary interpretation of the original description. There are distinct implications that arise from the level of precision in distance measurements, directions (headings), and coordinates. These are addressed in the subsections below.

3.3.3.1. *Uncertainty associated with distance precision.* Distance may be recorded in a locality description with or without significant digits, and those digits may or may not be warranted. Distances are commonly recorded with few or no significant digits, or even with fractions. Locality descriptions may also have undergone reformatting to remove fractions or significant digits. For example, suppose a specimen label was written in the field as locality '¾ km W of Inverness', which was entered into a database as '0.75 km W of Inverness'. In the original, it is clear that the collector was confident of recording the distance with one quarter km precision. Without consulting the specimen tag it may be difficult to determine how much distance precision is warranted. If the original tag is not consulted, then a conservative way to ensure that distance precision is not inflated is to treat distance measurements as integers with fractional remainders, thus 10.25 becomes 10 ¼, thus accounting for the possible (and not uncommon) transformation of a fraction in the original data to a real number in the database record. The uncertainty for these distances should be calculated based on the fractional part of the distance, using 1 divided by the denominator of the fraction.

Examples: '**9 km N of Bakersfield**' (fraction is 1/1, uncertainty should be 1 km)
'**9.5 km N of Bakersfield**' (fraction is ½, uncertainty should be 0.5 km)
'**9.75 km N of Bakersfield**' (fraction is ¾, uncertainty should be 0.25 km)
'**9.6 km N of Bakersfield**' (fraction is 6/10, uncertainty should be 0.1 km)

For measurements that appear as integer multiples of powers of 10 (for example, 10, 20, 300, 4000), use 0.5 times ten to that power for the uncertainty.

Examples: '**140 km N of Bakersfield**' (uncertainty should be 5 km)
'**100 km N of Bakersfield**' (uncertainty should be 50 km)
'**2000 m N of Bakersfield**' (uncertainty should be 500 m)

3.3.3.2. *Uncertainty associated with directional precision.* Direction is almost always expressed in locality descriptions using cardinal or inter-cardinal directions rather than degree headings. This practise can introduce uncertainty due to directional imprecision. The problem arises from the fact that we don't know, out of context, what the recorder meant by 'north' except that it is distinct from the other cardinal directions. Hence, 'north' is not 'east' or 'west', but it could be any direction between northeast and northwest. The directional uncertainty in these cases is 45 degrees in either direction from the given heading.

Example: '**10 mi N of Bakersfield**'

If a related set of locality descriptions (for example, those by a collector on a given expedition) contain any directions more specific than the cardinal directions (for example, 'NE'), then the person recording the data was demonstrably sensitive to inter-cardinal directions. Thus, 'NE' could mean any direction between ENE and NNE. The directional uncertainty in these cases is 22.5 degrees in either direction from the given heading.

Example: '**10 mi NE of Bakersfield**'

A locality description that contains further refined directions is correspondingly more precise. Thus, in the following example the directional uncertainty is 11.25 degrees.

Example: '**10 mi ENE of Bakersfield**'

If the locality description contains two orthogonal directions, convention holds that the measurements are linear in exactly those directions. In this case there is no directional imprecision.

Example: '**10 mi N and 5 mi E of Bakersfield**'

3.3.3.3. *Uncertainty associated with coordinate precision.* Recording coordinates with insufficient precision can result in unnecessary uncertainties. Therefore, as many digits of precision as are reported by the source should be retained when recording geographic coordinates. The magnitude of the uncertainty due to coordinate imprecision is a function not only of the precision with which the data are recorded, but also a function of the datum and the coordinates themselves. Uncertainty due to the imprecision with which the original coordinates were recorded can be estimated as follows:

$$uncertainty = \sqrt{lat\_error^2 + long\_error^2} \tag{1}$$

where

$$lat\_error = \pi\ R \times (coordinate\ precision)/180.0$$

and

$$long\_error = \pi\ X \times (coordinate\ precision)/180.0$$

where R is the radius of curvature of the meridian at the given latitude, X is the distance from the point to the polar axis, orthogonal to the polar axis, and *coordinate precision* is the precision with which the coordinates were recorded, as a fraction of one degree. R is given by Equation 2.

$$R = a\left(1 - e^2\right) \Big/ \left(1 - e^2 \sin^2(latitude)\right)^{3/2} \tag{2}$$

where a is the semi-major axis of the reference ellipsoid (the radius at the equator) and e is the first eccentricity of the reference ellipsoid, defined by Equation 3.

$$e^2 = 2f - f^2 \tag{3}$$

where f is the flattening of the reference ellipsoid. X is also a function of geodetic latitude and is given by Equation 4.

$$X = N \cos(latitude) \tag{4}$$

where N is the radius of curvature in the prime vertical at the given latitude and is defined by Equation 5.

$$N = a \Big/ \sqrt{1 - e^2 \sin^2(latitude)} \tag{5}$$

Example: Latitude = 10.27; Longitude = −123.6; Datum = WGS84

In this example the *coordinate precision* is 0.01 degrees. Thus, *lat_error* = 1.1061 km, *long_error* = 1.0955 km, and the uncertainty resulting from the combination of the two is 1.5568 km. These calculations use a semi-major axis (a) of 6378137.0 m and a flattening (f) of 1/298.25722356 based on the WGS84 datum.

Examples of error contributions for different levels of precision in the original coordinates (using the WGS84 reference ellipsoid) are given in table 2. Calculations are based on the same degree of imprecision in both coordinates and are given for several different latitudes.

3.3.4. *Uncertainty due to map scale*

Maps have an inherent level of accuracy. Unfortunately, the accuracy of many maps, particularly old ones, is undocumented. Accuracy standards generally explain the physical error tolerance on a printed map, so that the net uncertainty is dependent on the map scale. Following are the map accuracy standards published by the US Geological Survey: '*For maps on publication scales larger than 1:20,000, not more than 10 percent of the points tested shall be in error by more than 1/30 inch,*

Table 2. Uncertainty in meters as a function of latitude. Estimates of uncertainty are based on coordinate precision measured in degrees using the WGS84 reference ellipsoid and are rounded up to the next greater integer value.

	Latitude			
Precision	0 degrees	30 degrees	60 degrees	85 degrees
1.0	156904	146962	124605	112109
0.1	15691	14697	12461	11211
0.01	1570	1470	1247	1122
0.001	157	147	125	113
0.0001	16	15	13	12
0.00001	2	2	2	2

measured on the publication scale; for maps on publication scales of 1:20,000 or smaller, 1/50 inch' (USGS 1999).

It is important to note that a digital map is never more accurate than the original from which it was derived, nor is it more accurate when you zoom in on it. The accuracy is strictly a function of the scale and digitizing errors of the original map. A value of 1 mm of error can be used on maps for which the standards are not published. This corresponds to about three times the detectable graphical error and should serve well as an uncertainty estimate for most maps. By this rule, the uncertainty for a map of scale 1:500 000, for example, is 500 m.

3.4. *Step four: calculate combined uncertainties*

The uncertainties associated with a given locality description depend on the coordinate source, of which we identify four categories: GPS, locality record, gazetteer, and map. Table 3 shows the potential sources of uncertainty that may be relevant for each of the four categories. We describe how to calculate the various combinations of uncertainties in the subsections below.

3.4.1. *Calculating uncertainties having no directional imprecision*

Distance uncertainties in any given direction are linear and additive. Following is an example of a simple locality description and an explanation of the manner in which multiple sources of uncertainty interact.

Example: '**6 km E (via Highway 58) of Bakersfield**'

The potential sources of uncertainty for this example are 1) the extent of

Table 3. Potential sources of uncertainty inherent in georeferencing descriptive localities using four common sources of coordinates.

	Source of uncertainty						
Coordinate source	GPS inaccuracy	locality extent	unknown datum	coordinate imprecision	distance imprecision	map scale	direction imprecision
GPS	X	X	X	X			
locality record		X	X	X			
map		X	X	X	X	X	X
gazetteer		X	X	X	X		X

Bakersfield, 2) an unknown datum, 3) distance imprecision, and 4) map scale. Suppose the centre of Bakersfield is 3 km from the eastern city limit and the distance is being measured on a USGS map at 1:100,000 scale with the NAD27 datum. The uncertainty due to the extent of Bakersfield is 3 km, there is no uncertainty due to an unknown datum, the distance imprecision is 1 km, and the uncertainty due to map scale is 51 m (167 ft). The overall uncertainty for this locality is the sum of these, or 4.051 km.

If there are two orthogonal offsets from a named place in the locality description, uncertainties apply to each of the directions and the combination of them is non-linear.

Example: **'6 km E and 8 km N of Bakersfield'**

For the example above, ignore, for the moment, all sources of uncertainty except those arising from distance imprecision. Under this simplification, a proper description of the uncertainty is a bounding box centred on the point 6 km E and 8 km N of Bakersfield. Each side of the box is 2 km in length (1 km uncertainty in each cardinal direction from the centre). In order to characterize the net uncertainty with a single distance measurement, we need to calculate the radius of the circle that circumscribes the above-mentioned bounding box. The radius could either be measured on a map or calculated using a right triangle, the hypotenuse of which is the line between the centre of the bounding box and a corner. Given the rule that the distance precision is the same in both cardinal directions, the triangle will always be a right isosceles triangle and the hypotenuse will always be $\sqrt{2}$ times the distance precision. So, for the above example the uncertainty associated with the distance precision alone is 1.414 km (figure 3).

Thus far we have accounted only for distance precision in this example. To incorporate the uncertainty due to extent, determine the distance from the geographic centre of the named place to the furthest point within the named place in either of the two cardinal directions mentioned in the locality description. Add this distance to the uncertainty due to the distance precision and multiply the sum by $\sqrt{2}$. Suppose the furthest extent of the city limits of Bakersfield either east or north from the geographic centre is 3 km. There is a total of 4 km of uncertainty in each of the two directions and the radius of the circumscribing circle is 4 km times $\sqrt{2}$, or 5.657 km (figure 4).

Suppose the coordinates for Bakersfield (35°22′24″N, 119°01′04″W) are taken from the GNIS database (USGS 1981), in which the datum is either NAD27 or NAD83, and the coordinates are given with precision to the nearest second. At this location the uncertainty due to an unknown datum is 79 m. The datum uncertainty contributes in each of the orthogonal directions. Thus, the summed uncertainty in each direction is 4.079 km and the net uncertainty is this number times $\sqrt{2}$, or 5.769 km.

The coordinates in the GNIS database are given to the nearest second. The uncertainty due to coordinate precision alone is about 39 m at the latitude of Bakersfield based on Equation 1. This number already accounts for the contributions in both cardinal directions, so it must not be multiplied by $\sqrt{2}$. Instead, simply add the coordinate precision uncertainty to the calculated sum of uncertainties from the other sources. For the example above, the net uncertainty is $5.769 + 0.039 = 5.808$ km.

If the coordinates for Bakersfield had been taken from a USGS map with a

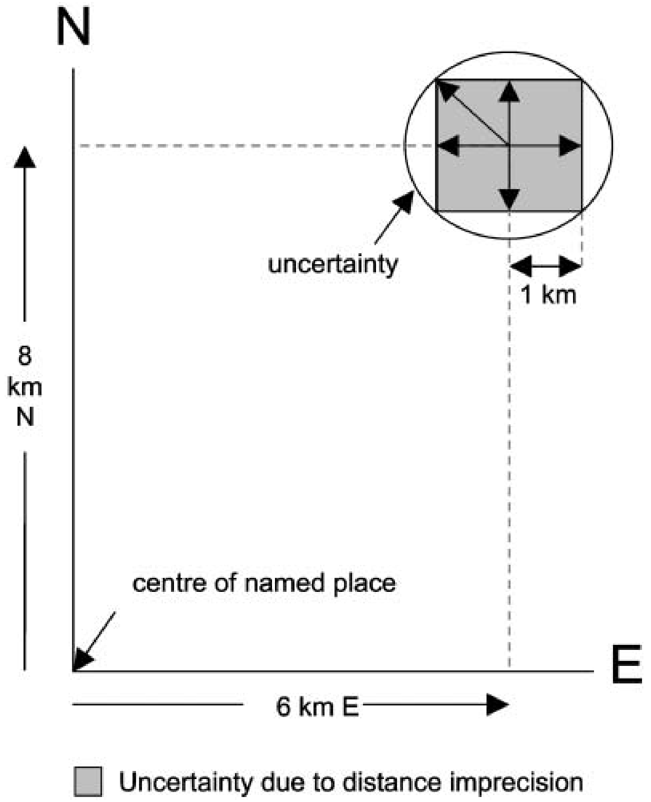


Figure 3. Uncertainty due to distance imprecision for two orthogonal offsets from the centre of a named place.

scale of 1:100 000, the datum would be on the map, so there would be no contribution to the error from an unknown datum (assuming the georeferencer records the datum with the coordinates). However, the uncertainty due to the map scale would have to be considered. For a USGS map at 1:100 000 scale, the uncertainty is 167 ft, or 51 m (based on the USGS map accuracy standards). In the above example, the uncertainty in each direction is 4.051 km. When multiplied by $\sqrt{2}$, their combination is 5.729 km. Add the uncertainty due to coordinate imprecision to this value to get the net uncertainty. Suppose the minutes are marked on the margin of the map and we interpolated to get coordinates to the nearest tenth of a minute. The coordinate precision is 0.1 minutes and the uncertainty is 0.239 km from this source, therefore the maximum error distance is $5.769 + 0.239 = 5.968$ km.

3.4.2. *Calculating combined distance and direction uncertainties*

The distance uncertainties in a given direction are linear and additive, but their sum contributes non-linearly to the uncertainty arising from directional imprecision. An additional technique is required to account for the correlation between these two types of imprecision.

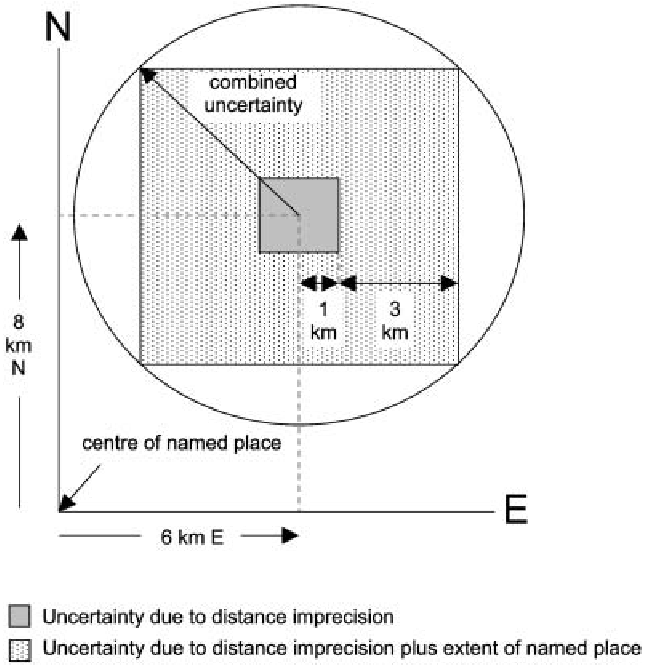

N
combined
uncertainty
8
km
N
1
km
3
km
centre of named place
E
6 km E
Uncertainty due to distance imprecision
Uncertainty due to distance imprecision plus extent of named place

Figure 4. Uncertainty due to the combination of distance imprecision and the extent of a named place.

Example: **‘9 km NE of Bakersfield’**

Without considering distance precision, the directional uncertainty (figure 5) is encompassed by an arc centred (at the coordinates *x,y*) 10 km (*d*) from the centre of Bakersfield at a heading of 45 degrees (θ), extending 22.5 degrees in either direction

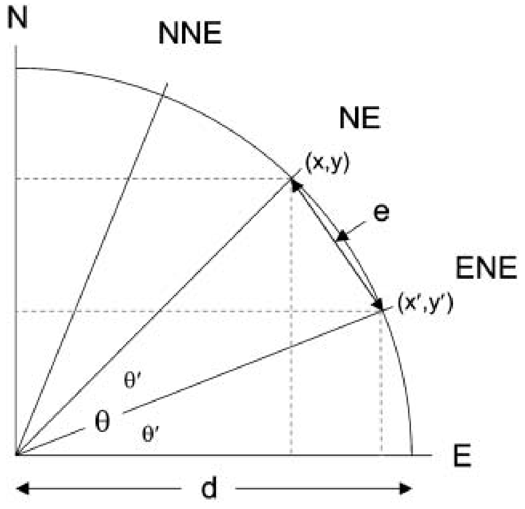


Figure 5. Uncertainty (*e*) due to direction imprecision for a direction specified as northeast (NE). The actual direction could be anywhere between ENE and NNE; *e* represents the maximum distance by which the actual locality could vary from reported locality.

from that point. At this scale the distance (e) from the centre of the arc to the furthest extent of the arc (at x',y') at a heading of 22.5 degrees (θ') from the centre of Bakersfield is given by Equation 6.

$$e=\sqrt{(x'-x)^2+(y'-y)^2} \qquad (6)$$

where $x=d\cos(\theta)$, $y=d\sin(\theta)$, $x'=d\cos(\theta')$, and $y'=d\sin(\theta')$. For the example above, the uncertainty (e) due to the direction imprecision is 3.512 km.

Now consider the distance uncertainties in this example. Suppose the contributions to distance uncertainty are 3 km (extent of Bakersfield), 1 km (distance precision for '9 km'), 0.079 km (unknown datum), and 0.040 km (gazetteer data are recorded to the nearest second) for a sum of 4.119 km. The shape of the region describing the combination of distance and direction uncertainties will be a band twice this width (2 × 4.119 = 8.238 km) centred (at the coordinates x,y) on an arc offset from the origin by 9 km, spanning 22.5 degrees on either side of the NE heading (figure 6). Uncertainty is still calculated with Equation 6, but now $x'=(d+d')\cos(\theta')$, and $y'=(d+d')\sin(\theta')$, where d' is the sum of the distance uncertainties.

The geometry can be generalized and simplified, by rotating the image in figure 6 so that the point (x',y') is on the x axis (figure 7). After rotation, Equation 6 still holds, but now $x=d\cos(\alpha)$, $y=d\sin(\alpha)$, $x'=d+d'$, and $y'=0$, where d' is still the sum of the distance uncertainties and a is an angle equal to the magnitude of the direction uncertainty. For the example above, the distance uncertainty is 4.119 km and the direction uncertainty is 22.5 degrees. Given these values, the maximum error distance is 5.918 km.

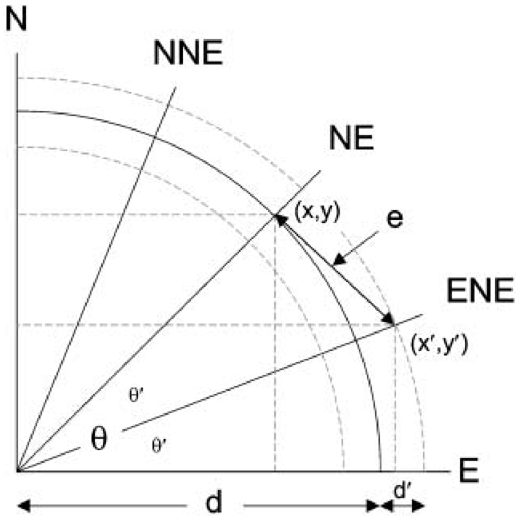


Figure 6. Uncertainty (e) due to the combination of distance imprecision (d') and direction imprecision (θ') for a locality specifying an offset (d) northeast (NE) of the centre of a named place. The actual locality could be anywhere between ENE and NNE and up to a distance d' either side of the offset d.

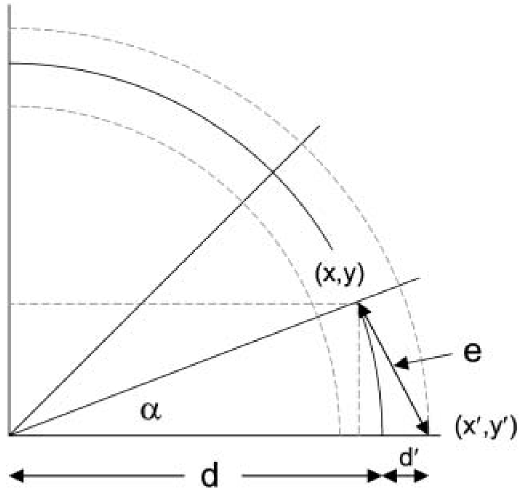


Figure 7. Uncertainty diagram rotated to simplify the equation for the net uncertainty (*e*) – the combination of distance and direction uncertainties.

3.5. *Step five: calculating overall error*

Thapa and Bossler (1992) distinguish between primary and secondary data collection. Primary data are taken directly from the field (ground surveying, remotely sensed imagery, GPS readings). Secondary data are derived from existing documents (maps, charts, graphs, gazetteers). Errors in secondary data consist not only of those introduced in primary data collection (such as human and instrumental errors), but also of those introduced from secondary data collection (such as errors due to map inaccuracy). The *post facto* process of georeferencing specimen locality descriptions relies heavily on secondary data. Thapa and Bossler (1992) conclude that it is difficult, if not impossible, to calculate the total error introduced by secondary data collection, because the functional relationships among the various sources of error are unknown. They assume a linear relationship between the total error and individual errors (e_i, typically Root Mean Square [RMS] is used), and apply the law of error propagation (Equation 7).

$$Total\ error = \sqrt{e_1^2 + e_2^2 + \ldots + e_n^2} \tag{7}$$

where e_n^2 is the standard error for source of error *n*.

There are a number of ambiguities that arise in locality descriptions to which root mean square errors and the law of error propagation cannot be readily applied. For example, how does one find the RMS error in the interpretation of "west"? In addition, many of our individual error components, such as the error from having an unknown datum, do not have a Normal distribution. For these reasons, we have calculated maximum potential errors. The error propagation law does not apply to this type of error. Instead, we calculate total error as the sum of individual error components (Equation 8), and not as the square root of the sum of the squared errors (Equation 7; which would always leads to a lower estimate than Equation 8).

$$Maximum\ uncertainty = \sum u_i + \sum u_d \tag{8}$$

where u is the maximum uncertainty for independent (i) or dependent (d) sources of error.

Like Thapa and Bossler (1992), we assume a linear relationship between total error and individual errors for which there is no known functional relationship (all 'independent' uncertainties u_i). Uncertainties that do have known relationships (all 'dependent' uncertainties u_d; for example, uncertainty due to distance and directional imprecision) are combined first on the basis of their relationships and are then combined linearly to achieve the overall maximum uncertainty.

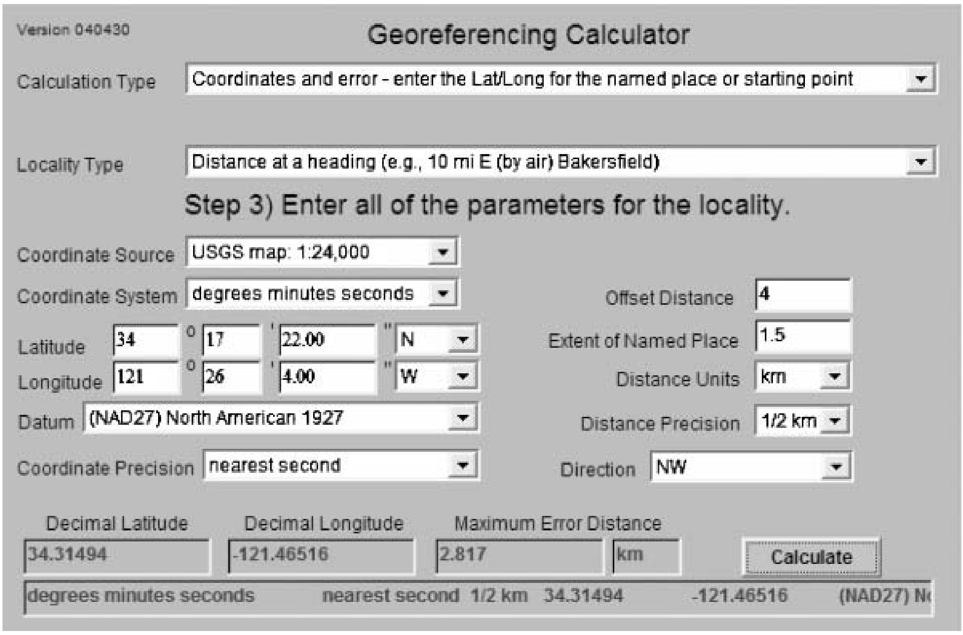
3.6. *Step six: document the georeferencing process*

When georeferencing a locality description, it is important to document the process by which the data were determined and record this information with each locality record so that anyone who encounters the data will benefit from the effort expended in providing a high-quality georeference. We recommend that the list of attributes recorded for each georeferenced locality include decimal latitude, decimal longitude, horizontal datum, net uncertainty (distance and units), original coordinate system, name of the person, organization, or software version that georeferenced the locality, georeferencing date, references used, reason if not georeferenced, named place, extent of the named place, determination method (for example, the point-radius method), verification status, and the assumptions made. With completely documented georeferenced localities, researchers who use the data can quickly verify that the georeferencing was done correctly.

4. Discussion

The point-radius method described here was developed to meet the georeferencing challenges of the MaNIS project, in which more than 40 individuals have used these methods in a collaborative georeferencing effort covering locality descriptions from all over the world. Localities were grouped by geographic region for the MaNIS project, with each participating institution georeferencing all of the localities within a given region for all participating institutions. A Java applet to calculate coordinates and uncertainties (figure 8) for the point-radius method was created by the first author and is freely available for use in Internet web browsers (Wieczorek 2001). Uncertainty calculations using this tool are simple, fast, and yield consistent results. Georeferencing rates for geographic regions varied, depending heavily on the resources that were available to the georeferencers. Where digital maps were available for a geographic region, the mean (±1 SD) georeferencing rate was 16.6 (±8.3) localities per hour ($n=14$ data sets from 14 institutions). The mean georeferencing rate for regions where printed maps were used instead of digital media was 9.6 (±6.8) localities per hour ($n=39$ data sets from four institutions). These rates include the determinations of both coordinates and uncertainties, with full documentation as recommended in section 3.5.

The georeferencing rates reported for MaNIS include only those data sets that were georeferenced manually, without the benefit of automated techniques. Preliminary tests suggest that the efficiency of georeferencing can be increased through automation, but that the resulting georeferences need to go through an extra verification step to ensure that the interpretation of the descriptive locality was made correctly. Even without automation, systematic error checking is necessary to find inaccurate locality descriptions or incorrectly georeferenced



Version 040430

Georeferencing Calculator

Calculation Type: Coordinates and error - enter the Lat/Long for the named place or starting point

Locality Type: Distance at a heading (e.g., 10 mi E (by air) Bakersfield)

Step 3) Enter all of the parameters for the locality.

Coordinate Source: USGS map: 1:24,000

Coordinate System: degrees minutes seconds

Latitude: 34 ° 17 ' 22.00 " N

Longitude: 121 ° 26 ' 4.00 " W

Datum: (NAD27) North American 1927

Coordinate Precision: nearest second

Offset Distance: 4

Extent of Named Place: 1.5

Distance Units: km

Distance Precision: 1/2 km

Direction: NW

Decimal Latitude: 34.31494

Decimal Longitude: -121.46516

Maximum Error Distance: 2.817 km

degrees minutes seconds nearest second 1/2 km 34.31494 -121.46516 (NAD27) N

Figure 8. Screen shot of the Georeferencing Calculator after coordinates and uncertainty for a locality comprised of an offset at a heading have been calculated.

localities. Some errors can be exposed by analyses that include complementary data sets. One test for georeferenced localities is to determine if the coordinates for the locality lie within the correct administrative boundaries, such as a country or lower level geographic unit (Hijmans *et al.* 1999). A more interesting test can be made by combining locality data for a given species with environmental data for those localities to reveal ecological 'outliers' that may have resulted from inaccuracies in the locality description or from the misidentification of the specimen. Another example is to plot the collecting events of an expedition in temporal order; localities that lie outside of the normal patterns in the expedition may be in error. These examples illustrate that GIS can be used *post-hoc* to improve the quality of the original data as well as to validate georeferences.

We have identified individual sources of error associated with the coordinates of a point that represents a collection locality, and we have provided methods for quantifying these individual error components in terms of maximum potential error. We suggest summing the individual maximum error components because commonly used alternative approaches, such as the law of error propagation, do not readily apply.

Without baseline test data, it is also difficult to produce error descriptions using alternative, fuzzy models (Altman 1994; Cross and Firat 2000), because this approach also relies on functions to describe error distributions. However, the law of error propagation, using standard errors, as well as fuzzy methods would be useful for determining error contributions for different coordinate sources (maps, gazetteer, and GIS layers, for example) where test data are available. These methods could even prove viable under limited circumstances for the much more difficult case of georeferencing locality descriptions. Appropriate error functions would have to be built from sets of locality descriptions for which the true localities were known. These functions might then be applied to localities of similar syntax.

Nevertheless, the one goal of this study is to provide an effective means to filter individual records based on the upper bound of the combination of all uncertainties inherent in the assignment of coordinates to a place with a spatial extent. By careful specification of the assumptions and of the techniques for combining uncertainties, we present a simple, practical method for computing and recording geographic coordinates and assigning this "maximum" uncertainty to each individual locality description.

The methods in this study provide an effective means to filter individual records based on the upper bound of the combination of all uncertainties inherent in the assignment of coordinates to a place with a spatial extent. In addition, more elaborate methods could be developed to use the uncertainty associated with a georeference in analysis, using fuzzy logic or other approaches (Burrough and McDonnel, 1998). Every georeference is a hypothesis. Before georeferenced data are used in analyses, every effort should be made to ensure that the locality description accurately describes the place where the specimen was collected. This is particularly true of localities reported with coordinates; even though the coordinates may accurately refer to a specific location such as beginning of a trap line, the specimens may have been collected over a considerably greater area. Collectors should also be aware of this problem and annotate their localities to avoid underestimations of the extent of the locality.

5. Summary

The point-radius method provides a practical solution for georeferencing descriptive localities that can be widely implemented, especially in communities where sophisticated GIS expertise is lacking. By accounting for the size of the locality, the point-radius method provides a more accurate description of a locality than is possible with the point method. By providing a single measure of the combination of uncertainties inherent in the locality description, the applicability of a locality for a given analysis can be more readily discerned than with the bounding box method. By capturing the spatial attributes of the locality in a simple, consistent set of parameters, the point-radius method offers a solution that is practical for natural history collections without the need for spatial databases that would be necessary to store georeferences created using the shape method.

Checking for and correcting errors can be time consuming. With a well-defined georeferencing method, appropriate tools, and proper documentation of the resulting data, the number of errors will be minimized and the results of effort expended to georeference the locality will be available in perpetuity.

Acknowledgements

The authors would like to thank Stan Blum, Elizabeth Proctor, and George Chaplin for their early inspiration to develop rigorous georeferencing methods. Larry Speers encouraged us to develop and document methods that are practical for natural history collections. Gary Shugart and Reed Beaman have provided critical analysis of the point-radius method and have investigated means to automate the process. Eileen Lacey provided useful discussion and criticism. Craig Wieczorek provided programming assistance. We extend special thanks to Barbara

Stein and the numerous participants in the MaNIS Project, without whose practical feedback and encouragement these methods would not likely have been elaborated. Funding leading to this publication was generously provided by the National Science Foundation (DBI #0108161) and the UC Berkeley Museum of Vertebrate Zoology.

References

ADL (Alexandria Digital Library), 2001, Alexandria Digital Library Gazetteer Server (http://fat-albert.alexandria.ucsb.edu:8827/gazetteer/).

Altman, D., 1994, Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems*, **8**, 271–289.

Bonner, M. R., Han, D., Nie, J., Rogerson, P., Vena, J. E., and Freudenheim, A. L., 2003, Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology*, **14**, 408–412.

Burrough, P. A., and McDonnel, R. A., 1998, Fuzzy sets and fuzzy geographical objects. In *Principles of Geographical Information Systems* (Oxford, U.K.: Oxford University Press), pp. 265–291.

Cross, V., and Firat, A., 2000, Fuzzy objects for geographical information systems. *Fuzzy Sets and Systems*, **113**, 19–36.

Drummond, J., 1990, A framework for handling error in Geographic Data manipulation. In *Fundamentals of Geographic Information Systems: A Compendium*, ASPRS, pp. 109–118.

Defense Mapping Agency, 1991, *Department of Defense World Geodetic System 1984, Its Definition and Relationships with Local Geodetic Systems (2nd edition), DMA Technical Report 8350.2*, Defense Mapping Agency, Fairfax, Virginia.

Duckworth, W. D., Genoways, H. H., and Rose, C. L., 1993, Preserving natural science collections: chronicle of our environmental heritage. National Institute for the Conservation of Cultural Property, Washington, D.C.

FGDC (Federal Geographic Data Committee), 1998, *Geospatial positioning accuracy standards. Part 3. National standard for spatial data accuracy*. Federal Geographic Data Committee, FGDC-STD-007.3-1998, Virginia, USA.

Fisher, P. F., 1999, Models of Uncertainty in Spatial Data. In *Geographical Information Systems*, edited by P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind (New York: John Wiley & Sons), pp.191–205.

GBIF (Global Biodiversity Information Facility), 2002, *Draft Report of the Meeting of the Digitization of Natural History Collections Scientific and Technical Advisory Group of the Global Biodiversity Information Facility*. (Kopenhagen: GBIF).

Goodchild, M. F., and Hunter, G. J., 1997, A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science*, **11**, 299–306.

Hijmans, R. J., Schreuder, M., De la Cruz, J., and Guarino, L., 1999, Using GIS to check co-ordinates of germplasm accessions. *Genetic Resources and Crop Evolution*, **46**, 291–296.

Knyazhnitskiy, O. V., Monk, R. R., Parker, N. C., and Baker, R. J., 2000, Assignment of global information system coordinates to classical museum localities for relational database analyses. *Occasional Papers, Museum of Texas Tech University*, **199**, 1–15.

Krishtalka, L., and Humphrey, P. S., 2000, Can natural history museums capture the future? *BioScience*, **50**, 611–617.

KU-BRC (University of Kansas Biodiversity Research Centre), 2002, Lifemapper (http://www.lifemapper.org).

Leung, Y., and Yan, J. P., 1998, A locational error model for spatial features. *International Journal of Geographical Information Science*, **12**, 607–620.

MaNIS (Mammal Networked Information System), 2001, (http://elib.cs.berkeley.edu/manis/).

McLaren, S. B., August, P. V., Carraway, L. N., Cato, P. S., Gannon, W. L., Lawrence, M. A., Slade, N. A., Sudman, P. D., Thorington, R. D., Williams, S. L., and Woodward, S. M., 1999, Documentation standards for automatic data processing in mammalogy, Version 2. Committee on Information Retrieval, American Society of Mammalogists.

NIMA (United States National Imagery and Mapping Agency), 2000, Department of Defense World Geodetic System 1984. Its Definition and Relationships with Local Geodetic Systems. TR8350.2, Third Edition, (Bethesda, Maryland: NIMA).

STANISLAWSKI, L. V., DEWITT, B. A., and SHRESTHA, R. L., 1996, Estimating positional accuracy of data layers within a GIS through error propagation. *Photogrammetric Engineering and Remote Sensing*, **62**, 429–433.

THAPA, K., and BOSSLER, J., 1992, Accuracy of Spatial Data Used in Geographic Information-Systems. *Photogrammetric Engineering and Remote Sensing*, **58**, 835–841.

USGS (United States Geological Survey), 1981, Geographic Names Information System. (http://nsdi.usgs.gov/products/gnis.html).

USGS (United States Geological Survey), 1999, National Mapping Program Technical Instructions. Part 2. Specifications. Standards for Digital Line Graphs. (Reston, Virginia: USGS).

VAN NIEL, T. G., and MCVICAR, T. R., 2002, Experimental evaluation of positional accuracy estimates from a linear network using point- and line-based testing methods. *International Journal of Geographical Information Science*, **16**, 455–473.

VEREGIN, H., 2000, Quantifying positional error induced by line simplification. *International Journal of Geographical Information Science*, **14**, 113–130.

WELCH, R., and HOMSEY, A., 1997, Datum shifts for UTM coordinates. *Photogrammetic Engineering and Remote Sensing*, **63**, 371–375.

WIECZOREK, J. R., 2001, *Georeferencing Calculator* (http://bnhm.berkeley.museum/manis/GC.html).