

Department of Fish and Wildlife
Biogeographic Data Branch
1416 9th Street, Suite 1266
Sacramento, CA 95814

<http://atlas.dfg.ca.gov>

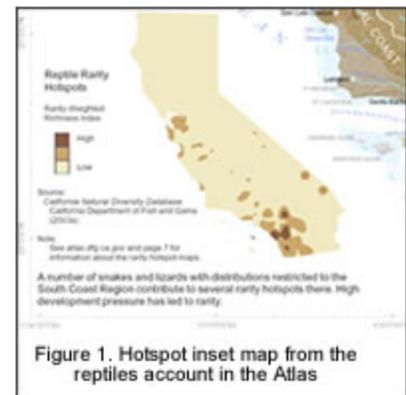
About the Rarity-Weighted Richness Index Maps

By Kevin Hunting
California Department of Fish and Game
October 2003

Describing and portraying the distribution of rare species on maps is a challenging task. On the one hand, it is important to present the best available information graphically as this venue has the greatest impact and can be readily absorbed by a broad audience. On the other hand, map creation requires interpretation of the data and acceptance of data biases and assumptions which are not always apparent when viewing the map. This paper provides the analytical background used in the development of the rarity maps produced for the Atlas of the Biodiversity of California (Atlas) and describe known biases associated with the data upon which the maps rely. The following description of our methods and approach also attempts to guide the Atlas user's decisions on proper use of Atlas information.

Creation of rarity maps for California was largely inspired by national rarity maps created as part of the nationwide analysis of rarity "hotspots" completed by NatureServe and described in the Atlas. NatureServe's publication [Precious Heritage \(Stein et al. 2000\)](#) seeks to highlight areas in the United States supporting relatively high numbers of rare taxa. The results of their effort point to California as supporting several rare taxa "hotspots" in the United States. (See the [NatureServe web site](#) for a series of rarity and endemism "hotspot" maps from Precious Heritage). In particular, the nationwide Rarity-Weighted Richness Index (RWRI) map emphasizes the importance of California for conserving rare species on a national scale.

The Atlas modifies the analysis conducted in Precious Heritage by focusing it on California and attempting to address the limitations of using positive sighting data for rarity mapping. The Atlas's RWRI analysis and mapping are subject to many of the same limitations and assumptions as the national effort described in Precious Heritage. The Mapping Species Heritage-Style and Novel Views of the World sections of Precious Heritage illuminate the limitations and assumptions of using positive sighting ("Heritage") data for depicting rarity. In summary, the discussions in these sections of Precious Heritage point to the advantages of combining several data sources and, in some cases, modeling a taxon's distribution to better describe a taxon's range and improve on the data biases described below. Furthermore, they describe how observation quality and attribute depth are essential for inferring distribution in areas where either data were not collected or the collected data were incomplete.

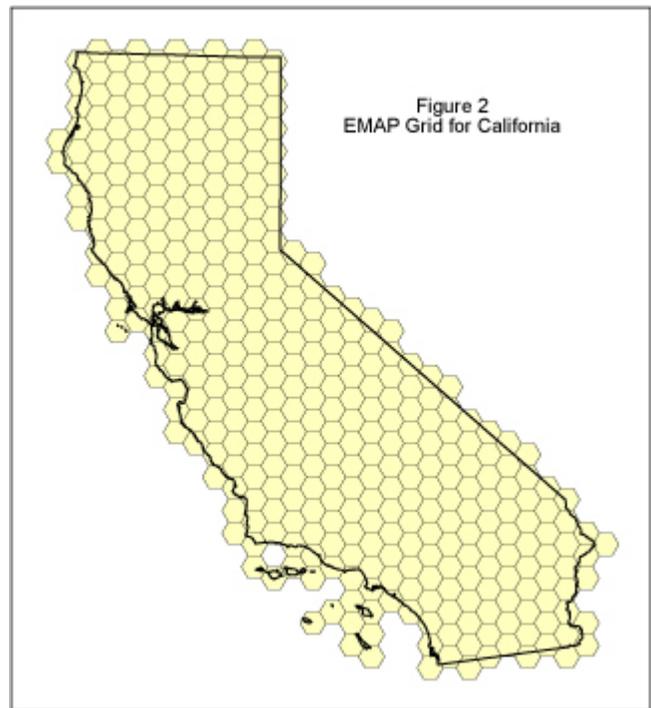


Limitations relating to mapping at a statewide (instead of nationwide) scale and the use of a relatively large polygon size for analysis diminish the applicability of the Atlas maps for conservation planning purposes. However, the maps and analysis do represent the patterns of diversity of rare species as we know them today and serve to identify areas where additional information or research is needed to uncover areas supporting rare taxa. In addition, display of heritage data using the RWRI method highlights the extent and scope of California's natural heritage information.

Selecting an Approach

As described above and in the Atlas, we selected the "rarity-weighted richness index" (RWRI) used by the NatureServe's Natural Heritage Network as a measure of rarity of taxa on a national scale. This method was employed by NatureServe as part of a nation-wide analysis of rarity hotspots depicting the highest concentrations of "rare" species in the nation. Page 7 of the Atlas describes the approach in greater detail.

The foundation of this approach is the use of a grid of adjacent equal-area hexagons covering the entire state as the analytical unit for describing rarity (Figure 2). This grid of hexagons was developed by the U.S. Environmental Protection Agency (USEPA) Environmental Monitoring and Assessment Program (EMAP) as a standard unit for conducting area-based analyses. We considered alternative units of analysis including U.S. Geological Survey quadrangles (approximately 57 square miles, 147 square kilometers). However, the EMAP grid system seemed best suited for our needs because 1) the results are ostensibly comparable to NatureServe's nation-wide rarity hotspot analysis and 2) the hexagonal design of the grid results in each grid cell sharing the same length of common boundary with every other cell. This design minimizes the possibility of occurrences falling into multiple grid cells, offers minimal distortion at multiple spatial scales, and supports hierarchical data systems (White et al 1992).



We also considered how EMAP grid cell size might affect RWRI results. Small cells would result in a large number of cells containing no data due to the patchy extent of the heritage dataset. Use of large grid cells would reduce the sensitivity of the weighting system and possibly mask patterns important to our analysis. After considering a few sizes, we selected a cell size of 250.4 square miles (648.5 square kilometers) which was small enough to meet our analysis sensitivity needs and large enough to support at least some data in every cell. This grid size also highlights regional variation and supports pattern contrast on regional and inset maps.

Using Positive Sighting Occurrence Data

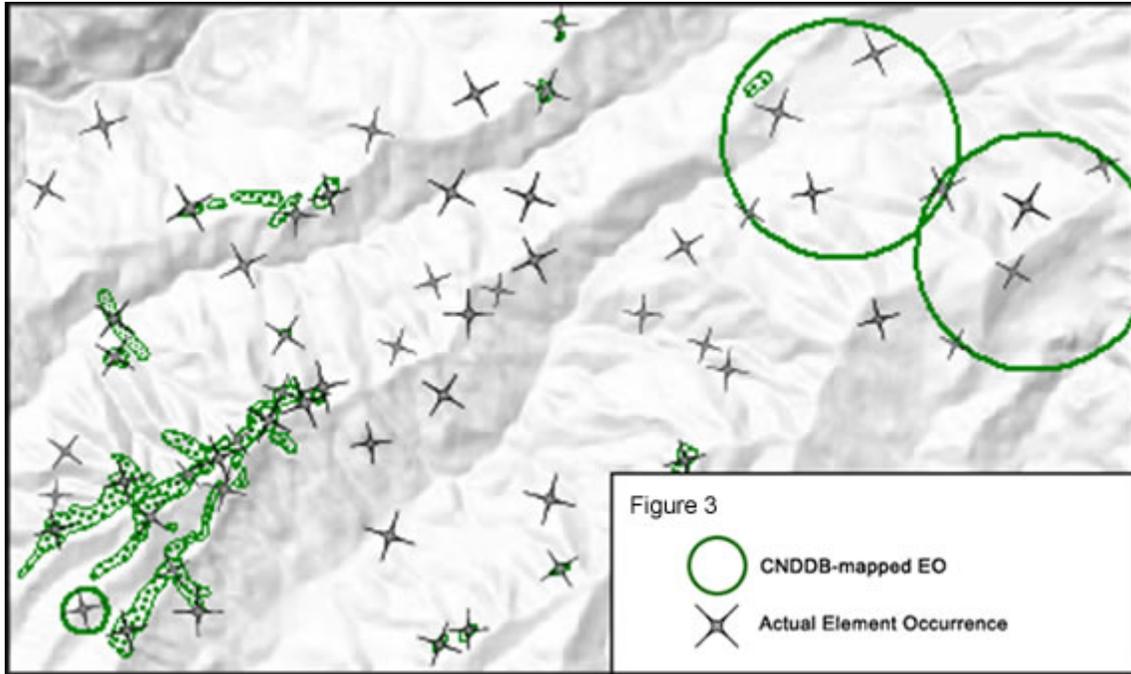
As mentioned above, creating a map that shows areas of the highest concentrations of the rarest species in California is a challenging task. What data should be used for this kind of analysis? What analytical methods should be used to most accurately describe rarity? Do we have enough information to create a rarity map that can be used for multiple purposes? The following description of our methods and approach attempts to answer these questions and guide the Atlas user's decisions on proper use of Atlas information.

The first step was deciding what data would best reflect rarity on a statewide scale. Following the approach adopted by NatureServe for a nationwide analysis of "rarity hotspots," we selected the Department's [California Natural Diversity Database \(CNDDDB\)](#). The CNDDDB is a collection of positive sighting data from a variety of sources maintained over a relatively long period of time (1979 – 2003). These data comprise the most complete set of information on the state's declining and/or vulnerable taxa. We found that despite the large data set (>40,000 records) and long time frame, there were both strengths and limitations of using the CNDDDB data for this purpose. Table 1 summarizes these strengths and limitations, and the implications of using these data for conservation planning purposes.

Table 1. Evaluation of CNDDDB Data for Rarity Mapping and Conservation Planning Uses

Strengths	Limitations	Conservation Planning Use
Statewide scope	Incomplete extent	May assist with fine-scale land use decisions. Insufficient for determining relative rarity (i.e., how many rare taxa occur in one area compared with another).
Rigorous quality control	Data entry priorities independent of decline or vulnerability status	Excellent data quality for extant occurrences. Support for regulatory and local government planning (i.e., can trigger requirements for avoidance or further information gathering. Number of occurrences not an appropriate index for relative rarity.
Long time frame	Includes only special status taxa	Can infer occurrence broadly from historic data. Should be used in conjunction with other data and models for reserve planning or to establish conservation priorities. Data set limited to taxa designated as special status and may not support data for some declining or imperiled taxa.

Use of the CNDDDB data for rarity mapping resulted in a visual manifestation of CNDDDB data distribution by taxa; the maps are as much a view of frequency of reports and survey intensity as they are the representations of actual distribution of rare taxa. This limitation stems from the ambiguity surrounding areas where there are no CNDDDB detections. This means that lack of an observation might be due to 1) no surveys or inventories in that area, 2) surveys or inventories were completed in the area but reports of occurrences were not submitted to the CNDDDB or, 3) surveys or inventories were completed in the area but yielded negative results. In all three cases, areas without a CNDDDB detection could support greater or fewer rare taxa than neighboring detection areas (Figure 3).

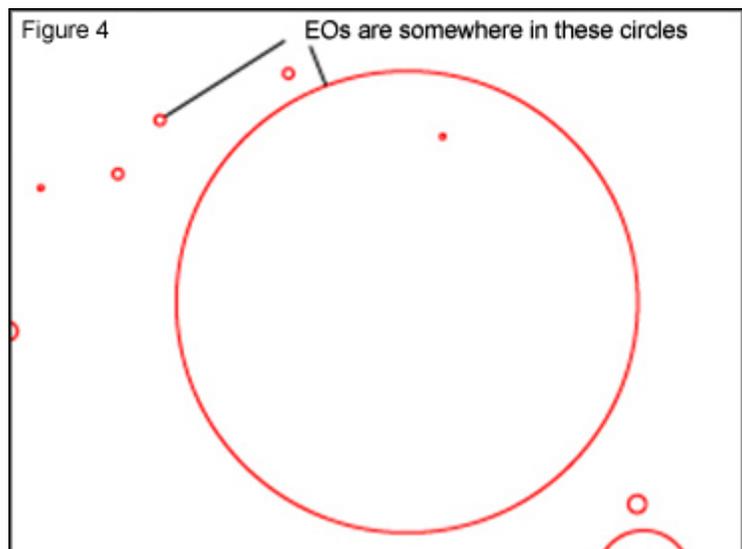


Many rarity mapping processes use range instead of positive sighting occurrence data because range data reflect an occurrence decision for every point on the map regardless of scale. While range maps may either overstate (commit) or understate (omit) actual range, they indicate where a taxon might occur and where it might not occur. This differs from positive sighting data, which describe where a taxon was observed but makes no assessment of areas without observations. Evaluating the limitations of both methods, we used the CNDDDB database and decided the RWRI would be an appropriate tool for showcasing the heritage database.

Selecting Data Attributes to Minimize Effects of Limitations

The CNDDDB supports considerable information about each taxon's occurrence. This attribute information includes dates of observation, the precision of the location, an assessment of current occurrence status, the source of the data, and more. Occurrence attributes can be valuable tools for "filtering" a data set to include records matching a given set of criteria.

To improve the quality of the RWRI maps, we selected a subset of CNDDDB records that matched criteria designed to minimize known data limitations and narrow selection of CNDDDB records to those that suited our analytical needs. We determined the criteria best suited for this objective and decided on the following criteria for selecting CNDDDB records for use in the RWRI map process:



1. Filtering the CNDDDB data set to include only those occurrences attributed as "extant" or presumed currently present, and remove occurrences categorized as "extirpated" or "possibly extirpated".
2. Discarding EOs with exceedingly low precision. The CNDDDB uses distance from a center point as a measure of specificity of an occurrence. In other words, the larger the EO circle, the less precise the location (Figure 4). EOs with an "accuracy" class of 10 were removed from consideration to improve overall EO location specificity. Accuracy class 10 occurrences are represented by a 5-mile radius circle and often include older museum records with little or no specific location information.

CNDDDB Data Input Bias

One of the limitations of the CNDDDB identified in Table 1 is the prioritization of data entry. The amount of data received by the CNDDDB staff usually exceeds their capability to enter the data into the database. The resulting backlog is addressed by setting priorities for entering the data based on internal Department needs, anticipated or realized listing or special status actions, conservation planning needs, and other factors. This approach results in an inequity between the status and distribution of the taxa and the number of records in the CNDDDB. For example, taxon A may be exceedingly rare and known from only a dozen or so locations in California. The CNDDDB may have 12 EOs for this taxon. Taxon B may be more widespread with relatively stable populations and be represented by only 1 or 2 CNDDDB records due to data entry priorities. Because the RWRI weights taxa based on the number of EMAP grid cells in which a taxon occurs (see *Atlas* page 7 for a discussion of how index values are assigned to a grid cell), taxon B would receive a higher score than taxon A, thus skewing the RWRI map results.

We considered options for addressing this inequity and decided to more closely examine the relationship between range size, EMAP grid occurrence, and statewide status ("S") rank for each major taxa group. S rank, or CNDDDB State Ranks, are a ".measure of a [taxa's] rarity throughout its range in California." (Table 2). We considered the possibility that either range or S rank might be a suitable surrogate for number of occurrences and provide a means to address the CNDDDB data entry bias.

Table 2. CNDDDB S Rank Rating System

S Rank	Description
1	Extremely endangered: <6 viable occurrences (EOs) or < 1,000 individuals, or 2,000 acres of occupied habitat.
2	Endangered: about 6-20 EOs or 1-3,000 individuals, or 2-10,000 acres of occupied habitat.
3	Restricted Range, rare: about 21-100 EOs or 3-10,000 individuals, or 10-50,000 acres of occupied habitat.
4	Apparently Secure: some factors exist to cause some concern such as narrow habitat or continuing threats.
5	Demonstrably Secure: commonly found throughout its historic range.

Since the number of EMAP grids in which a taxa occurs should relate to a taxon's range size, we addressed potential data entry biases by comparing range area from maps developed by the California Wildlife and Habitat Relationship (CWHR) system to the number of EMAP grid occurrences generated by the RWRI process. In addition, to answer our question regarding use of the S rank system as a surrogate for EMAP grid occurrence, we compared S rank to both range size and number of CNDDDB occurrences.

For each major animal (vertebrate) group represented by EOs in the CNDDDB (i.e., amphibians, reptiles, birds, and mammals), we calculated a range area (square meters) and the number of EMAP grid occurrences. We then compared the S rank value for each taxon to the calculated range and occurrence values using a Pearson's correlation to measure the degree to which the values were related. Sorting these tables by range size, grid occurrence, and S rank revealed obvious discrepancies between these factors and highlighted those taxa under-represented by EOs in the CNDDDB as a result of data entry priorities. We then selectively removed taxa from analytical consideration, recalculated correlation values, and reconsidered the resulting taxa list. We attempted to balance the need to represent all CNDDDB taxa with our desire to maximize applicability by removing taxa which skewed RWRI results. The analysis revealed that range size and EMAP grid occurrences were most closely related so we attempted to maximize the correlation between these factors during our selective removal of taxa from the RWRI process. Table 3 summarizes the results of this analysis and lists the taxa removed from consideration in the final RWRI map.

Table 3. Results of Data Entry Bias Analysis

Taxa Group	Initial r^2 value	Final r^2 value	Taxa Removed from Consideration
Amphibians ¹ (n=15)	n/a	0.95987	None
Reptiles (n=33)	0.65641	0.76247	Chuckwalla, rosy boa
Mammals (n=68)	0.08861	0.35237	American badger, Yuma myotis, long-legged myotis, fringed-myotis, spotted bat, long-eared myotis, small-footed myotis
Birds (n=85)	0.29659	0.44564	Sharp-shinned hawk, loggerhead shrike, black-crowned night heron, ferruginous hawk, American white pelican, snowy egret, great egret, short-eared owl.
¹ All 15 taxa represented with EOs in the CNDDDB included in RWRI map.			

Appropriate Use of the RWRI Maps

The RWRI maps should be considered a way in which to view data in the CNDDDB. The maps provide an interesting and provocative depiction of the data and describe where rare species have been observed. However, they should not be misconstrued as rarity maps generated from either detailed range data or complete positive occurrence data. Table 4 describes the implications of mapping designations between RWRI maps created using range *versus* positive occurrence data. Limitations appear to favor use of range data for this type of analysis. However, as mentioned previously, range data are not available for all taxa, are limited by errors of omission and commission and therefore may poorly reflect potential taxa occurrence at scales used in the Atlas. Range data are often generated using decision support methods and often rely on a combination of professional judgment, anecdotal information, and occurrence data.

Table 4. RWRI Source Data Comparison

Data Source	Map Designation		
	Low Rarity	Moderate Rarity	High Rarity
Positive Occurrence Data	Includes "no data" areas, which may or may not support rare taxa.	In relation to "High" designation only. Absolute rarity compared to other locations in state cannot be determined.	May reflect number of surveys, survey effort, or data entry priorities and not abundance of rare species. For comparison with other CNDDDB data only.
Range or Complete Occurrence Data	Occurrence determination for every point on the map	In relation to both "High" and "Low" designations.	Measure of absolute high rarity in state.

While the maps appear at first glance to describe areas rich in rare taxa, and therefore a high priority for conservation or preservation, this is not necessarily the case. Vast areas of the state have not been inventoried and suitable habitat for many declining taxa remains unsurveyed. Similarly, areas supporting numerous EOs may reflect survey effort or intensity. To test the independence of CNDDDB EOs, we examined the relationship between the percentage of all occurrences for each major animal taxonomic group, the area of ecological regions in the state, and the number of [habitat types \(CWHR\)](#) in each region. If the percentage of occurrences in each taxonomic group within an ecological region more or less reflected the diversity of habitats, we could expect to see a high correlation between the number of habitat types in an ecological region and the percentage of reported occurrences. While there was a moderate correlation between these factors, there was a stronger correlation between the percentages of occurrences between each group. This means that the percentage of occurrences for any vertebrate taxonomic group was more or less independent of the size of the ecoregion subsection or its diversity as measured by the number of CWHR habitat types. If the data were collected systematically, we would expect to see a correlation between the diversity of habitats within an ecoregion subsection and the number of different taxa occurring in that region. This illustrates the survey, reporting, or data entry bias inherent in the CNDDDB data.

The quality and accuracy of the RWRI maps will improve as the CNDDDB database continues to grow and additional data sources are included in the analysis. The Department is currently updating or creating range maps for almost all terrestrial vertebrates in California using CNDDDB and other observation-based data. When completed, the range

maps, which are developed using a tested and repeatable approach with clearly identified assumptions, will open the door to improved analysis and better tools for making land use and conservation decisions.

References

Stein, B.A., L.S. Kutner, and J.S. Adams, eds. (2000) *Precious Heritage: The Status of Biodiversity in the United States*. Oxford University Press, Oxford, Mass. 416 pp.

White, D., Kimerling, A.J., Overton, W.S. (1992) Cartographic and geometric components of a global sampling design for environmental monitoring, *Cartography and Geographic Information Systems*, 19(1),5-22.