Environmental Modelling & Software 53 (2014) 13–26

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A programmable information system for management and analysis of aquatic species range data in California

Nicholas R. Santos [a,*], Jacob V.E. Katz [a,b], Peter B. Moyle [a,c], Joshua H. Viers [a,d]

[a] Center for Watershed Sciences, University of California, Davis, United States
[b] California Trout, United States
[c] Department of Wildlife, Fisheries, & Conservation Biology, University of California, Davis, United States
[d] School of Engineering, University of California, Merced, United States

ARTICLE INFO

Article history:
Received 18 September 2012
Received in revised form
26 October 2013
Accepted 30 October 2013
Available online 3 December 2013

Keywords:
Expert opinion
Biological observation
Species distribution
Geographic information systems (GIS)
Database
Arcpy

ABSTRACT

The decline of species worldwide is both alarming and difficult to document due to a lack of reliable information on the geospatial extent and corresponding status of a given taxon. Freshwater habitats are disproportionately degraded globally with resultant declines in populations in freshwater fishes and subsequent retractions in biogeographic ranges. Conservation challenges in freshwater are compounded because aquatic taxa are inherently difficult to map. We addressed this problem for California freshwater fishes by developing the software and underlying database. The software consists of a Python program, database, and suite of tools using ESRI ArcGIS scripting interfaces to translate species range data into an electronic record set of occurrences housed in Microsoft Access. The system was designed to capture, store, map, and report on the spatial and temporal dynamics of targeted species by using standard spatial units as primary indexing objects to meet current natural resource management objectives. However, the software not only tracks the provenance of underlying empirical records through space and time, but also is robust to inferential modeling results and expert knowledge, which allows for future empirical discovery and validation. After importing and standardizing 274,555 records from 154 data layers, we found that most existing records are highly concentrated spatially, representing only 39% of the mapping domain. We also determined that most empirical records are skewed toward recreational fisheries, with few records documenting the range of native species found in California. Future biogeographic mapping efforts will be aided by the baseline data and updated range maps contained in the database. Although the system is currently used for the inventory and mapping of native freshwater fish species in California, the underlying informatics framework is agnostic to biological taxonomy or spatial realm allowing other to adapt the computer code and database for their own needs.

Software availability

Name of software: PISCES
Developer: Nicholas Santos
Contact information: Nicholas Santos – nrsantos@ucdavis.edu; +1 530 754 9362
Year first available: 2012
Availability and cost: Open source licensed under Creative Commons Attribution NonCommercial ShareAlike 3.0 Unported license. Software and data available at https://bitbucket.org/nickrsan/pisces
Program language: Python, SQL
Program size: 6.5 GB (including supporting data)

1. Introduction

1.1. Background

Global environmental change from human activities is both widespread and rapid (Millenium Ecosystem Assessment, 2005). This is especially true in freshwater ecosystems, which are disproportionately affected by human activities (Dudgeon et al., 2006). Freshwater ecosystems are affected by pollution, water diversion and flood protection infrastructure, habitat alteration, and invasive species, among many stressors. Further, freshwater ecosystems and freshwater supplies are increasingly vulnerable to

* Corresponding author. Tel.: +1 530 754 9362.
E-mail address: nrsantos@ucdavis.edu (N.R. Santos).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.10.024

the combined effects of human abstraction and hydroclimatic alteration (Vörösmarty et al., 2000).

Efforts to minimize the negative effects of anthropogenic change through conservation often rely on maps to depict subjects of interest (Wilson et al., 2007), such as the location of a species or a collection of species, or alternately places of refuge, such as a park or preserve. To initiate a process or program to conserve a species or a suite of species, it is necessary to know where to act. While this need is common in resource management, few systematic tools have been developed to collect, store, and map aquatic biodiversity through space and time. This may be due to the inherent challenges of mapping ranges of aquatic species including: animals in aqueous environments are often difficult to detect; methods of detection are comparatively poor; field observations are highly localized; present day distributions are confounded by human introductions; and human activities have extirpated species from suitable habitats (Naiman and Dudgeon, 2011; Nel et al., 2009; Pittock et al., 2008; Strayer and Dudgeon, 2010; Thieme et al., 2007; Viers and Rheinheimer, 2011; Vorosmarty et al., 2010). This paper introduces a software system designed to function within these constraints to produce high resolution, easily updated range maps for resource managers.

1.2. California

Like other regions of the Mediterranean Biome, California is disproportionately impacted by human activities (Grantham et al., 2010; Underwood et al., 2009). The state has the combined challenges of being ecologically unique – 82% of fish taxa are regionally endemic (Moyle et al., 2011) – and containing 38 million people. Freshwater ecosystems of California are threatened by several compounding factors, including but not limited to river regulation, pollution, and invasive species (Marchetti et al., 2004). The number of the state's native fish taxa in decline is worrisome. Of 129 native fish taxa 83% of are in sharp decline or have already gone extinct (Moyle et al., 2011). Indicative of the critical nature of this problem is the impending extinction of several California runs of anadromous salmonid (Katz et al., 2012).

The United States Department of Agriculture Forest Service (FS) is California's largest land holder, administering 9,526,000 ha across 18 National Forests, accounting for approximately 20% of the state's land surface. Management objectives, coupled with regulatory mandates, such as the federal Endangered Species Act, obligate the FS to identify, catalog, monitor and account for aquatic species. To date, FS and other resource managers have relied on digital maps generated as part of a statewide mapping effort of California fishes conducted in 1998 (Moyle and Randall, 1998). While these maps have proved helpful in better understanding the distribution of fishes in California, they were developed with early geographic information system (GIS) technologies at coarse scales and did not portray empirical data but instead were created using expert knowledge. Furthermore, these maps could not be readily updated as new information became available or combined with other mapping approaches, such as inferential spatial modeling – issues that afflict many similar efforts (Graham and Hijmans, 2006).

1.3. PISCES

To more effectively understand and manage the sensitive fishes of California's National Forests, we developed a GIS dependent database and accompanying software. The software, termed PISCES, was intended to robustly manage mapping efforts of sensitive fish species distributions across forests in ways that previous efforts could not. Designed as a decision support system for resource managers, PISCES incorporates and catalogs disparate data types of empirical and inferred species observations. Subsequent encoding of these observations relies on a standardized, yet generic, data framework that overcomes issues of spatial scale, temporal discontinuity, data format discrepancies, and regional context.

Standard outputs from PISCES map or tabulate phenomena that allow managers to identify species ranges, patterns of biodiversity, and areas of where biological data is lacking. Although intended to solve acute management problems associated with California's sensitive fish species on FS lands, the PISCES platform, with its enhanced GIS database and data provenance and transaction capabilities, can be applied to other natural resource management systems.

PISCES fits in with a general class of research software that attempts to organize, store, and present disparate sets of data (See (Guzman et al., 2013; Horsburgh et al., 2009; Rangel et al., 2010; Souza Muñoz et al., 2011; Villa et al., 2009)) – a common need as organizations gather overlapping data in different forms. This very basic question of "*where*" is foundational to much other work in ecology and environmental science and is the reason we have GIS. Systems such as PISCES aid in that question of where by translating, organizing, and collecting data into functional units with problem-domain specific tools. In fact, while our instance of PISCES is California focused, the software and database are a generalized system for tracking and analyzing species range *anywhere* using a predefined set of zones.

1.4. Design requirements

We imposed several formal design requirements on the development of PISCES that were intended to address issues of species data collection, storage, and analysis within existing desktop GIS software.

1. **Resource managers are often in a standalone computing environment with "stock" desktop software; thus, outcome must be serverless and portable and not require additional investment in hardware or software.** PISCES uses Python 2.6, ArcGIS and a Microsoft Access geodatabase for data processing and storage.
2. **Resource managers collect new information regularly; thus, data and results must be dynamic.** PISCES can process and import many new datasets and will update maps and tables as data are updated.
3. **Resource managers are not database engineers; thus, data must be easily updateable and flexible to unknown and non-standard data formats.** Data import occurs via extensible classes that can adapt to datasets regardless of format, design, scale, quality, or quantity. New classes can be added as needed to handle foreign data.
4. **Resource managers are faced with ever-changing mandates and management challenges; thus, resultant frameworks must be compatible with future questions and needs.** Flexible configuration and callback functions allow for runtime customization of maps to meet future needs. Stock functions and map configurations handle major outputs of interest and new functions can be added.
5. **Resource managers have inherent, place-based knowledge; thus, software must be able to capture verbal and written accounts.** PISCES includes tools that support interactive range viewing, input, and editing with experts in ArcMap 10 to capture data not currently in any file.
6. **Resource managers are human and mistakes can happen; thus, all data and changes must be traceable.** PISCES logs all transactions and stores all inputs unmodified so that changes can be traced. In many cases, a change can be reverted.