Biological Conservation 211 (2017) 161–171

Contents lists available at ScienceDirect

Biological Conservation

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

Integrating multiple datasets with species distribution models to inform conservation of the poorly-recorded Chinese seahorses

Xiong Zhang*, Amanda C.J. Vincent

Project Seahorse, Institute for the Oceans and Fisheries, University of British Columbia, 2202 Main Mall, Vancouver, BC V6T 1Z4, Canada

ARTICLE INFO

Keywords:
Data-deficient species
Species distribution models
Local ecological knowledge
Coarse-grain maps
Seahorses

ABSTRACT

Modeling and mapping species distributions are vital to biodiversity conservation, but challenging for data-limited species whose localities are poorly recorded. Here we examine the utility of three datasets and species distribution models in conservation of seahorses (*Hippocampus* spp.), a genus of poorly-recorded marine fishes. We collated occurrences from field data of species sightings (SS), peer-reviewed literature (PRL), and fishers local ecological knowledge (LEK) for five seahorse species in China. We modelled seahorse distributions using different combinations of these datasets. We first compared model performance and predictions between PRL and LEK, and then evaluated the impact of adding LEK and/or PRL to SS. Our results indicated that LEK provided higher-resolution maps than PRL and tended to generate slightly better models. There is more predictive consistency between LEK and PRL on presence-probability maps than on presence/absence maps. Adding LEK and/or PRL to SS improved model performance across species. Our study suggests that integrating LEK (and PRL) and limited SS with species distribution models can usefully inform conservation for poorly-recorded species.

1. Introduction

Species distribution maps are vital to biodiversity conservation (Pimm et al., 2014). Anthropogenic activities have driven incredible biodiversity loss, which in turn has significant impact on human society. To protect the threatened wildlife, we need biogeographic information to assess their conservation status (Mace et al., 2008), and design nature reserves (Lourie and Vincent, 2004; Micheli et al., 2013). Wildlife habitat maps are also indispensable for resource management, as new development projects expand across land and the sea (McShane et al., 2011; Reis et al., 2012).

Mapping species distributions is challenging for poorly-recorded species, whose population localities are poorly documented in peer-reviewed literature or other sources. This difficulty often necessitates the use of multiple datasets, including new field data. Fine-resolution (e.g. $10 \times 10\ \mathrm{m}^2$) species sightings (SS, in the form of GPS coordinates) from natural history collections or other sources (e.g. citizen science) are the most frequently-used datasets. But SS collection is often biased towards easily-accessed regions and common taxa (Phillips et al., 2009; Robinson et al., 2011). Peer-reviewed literature (PRL) can be a second dataset, but it may only contribute coarse range maps for poorly-recorded species. A third source of species data is local ecological knowledge (LEK), which refers to the knowledge system learnt by people through interactions with their local environment (Berkes, 1993). Compared with traditional surveys (e.g. transect sampling), interview-based LEK research can generate cost-effective but often coarse-resolution (e.g. $10 \times 10\ \mathrm{km}^2$) datasets (Carter and Nielsen, 2011; Laze and Gordon, 2016).

Species distribution models (SDMs), which predict presence probability of focal species based on limited species presences/absences and environmental data, might provide a powerful way to derive spatially-explicit maps and to inform conservation for poorly-recorded species (Guisan and Thuiller, 2005; Franklin, 2010). The predictive maps based on SDMs have facilitated population surveys for rare species (Guisan et al., 2006; Stirling et al., 2016), and are useful for conservation planning (Guisan et al., 2013). Some SDMs contain techniques to examine species-habitat relationships, which are central to ecology (Guisan and Thuiller, 2005). In literature, there are basically two types of SDMs regarding the availability of species-absence data: presence-absence models, and presence-only models (see Franklin, 2010 for a review). Presence-only models are more suitable to poorly-recorded species since their absences are hard to determine.

Mapping and modeling species distributions is particularly challenging for poorly-recorded marine species. Marine biota and environmental surveys have historically fallen behind the terrestrial counterparts (Costello et al., 2010). Scuba-diving has only been used for collecting site-level species data since ~1960s (Caddy, 1968), and remote sensing techniques have only contributed spatial data for

---

* Corresponding author.
E-mail address: x.zhang@projectseahorse.org (X. Zhang).

http://dx.doi.org/10.1016/j.biocon.2017.05.020
Received 5 November 2016; Received in revised form 10 May 2017; Accepted 21 May 2017

marine environments since 1980s (Bernstein, 1982; Wentz and Schabel, 2000). The utility of survey techniques (e.g. underwater visual census) can be restricted by the unique features of marine environment (water clarity, depth, etc.). These characteristics of ocean systems make it more difficult to study geographic distributions of marine organisms.

Seahorses (*Hippocampus* spp.) provide a typical example of rarely-recorded marine organisms whose distributions are difficult to determine. These relatively rare, cryptic, and small fishes, are difficult to detect or survey (Vincent et al., 2011; Aylesworth et al., 2017). Additionally, seahorses can raft with holdfasts (e.g. seaweed) and disperse over long distances, although they are generally stationary (Lourie et al., 2005; Caldwell and Vincent, 2013). Our knowledge about their distribution ranges is still developing. About 15% of the current total sightings from our citizen science database (iSeahorse, iseahorse.org) are located beyond the ranges that we previously knew. To date, seahorse localities are poorly-recorded in many regions.

China is among the countries where seahorses are poorly-documented and threatened. Seahorses are distinguished by their heavy use in Traditional Chinese Medicine (TCM). Every year, millions of dried seahorses are used in TCM by Chinese people (Vincent et al., 2011). To date, formal seahorse biogeographic research is rare in China. Six seahorse species are purportedly present, and probably all are threatened (Wang and Xie, 2009). One of these species, *H. kelloggi* (great seahorse), is on China's List of Wildlife under National Protection, mandating a nationwide ban on its catch and trade by law (MEP, 2002). The other five species have been proposed to be added to the List, which is under review (Zhang Chun-Guang, per. comm.). However, the lack of distribution knowledge of seahorse populations in China's vast marine territory impedes the protection of these poorly-known animals.

Here we present the first biogeographic study of seahorses in China, with an aim to inform their conservation. We collate multiple species datasets (i.e. SS, PRL, and LEK) and environmental data to build and compare species distribution models. We test whether species data from PRL and LEK can generate similar predictions of seahorse distributions. We examine if adding information from LEK and PRL to SS can improve model performance and predictions. By doing so, our study provides insights on species data collection and analyzing techniques for distribution modeling studies on poorly-recorded species.

2. Materials and methods

2.1. Study area

Our study area spans China's coastal waters (17° to 41°N; 106° to 125°E, Fig. 1), which are fringed by the Bohai Sea, Yellow Sea, East China Sea, and the northern South China Sea. The coastline stretches across 18,000 km from temperate to tropical zones (see details in Liu, 2013).

2.2. Species distribution model

We used a typical presence-only model, maximum entropy (Maxent, Phillips et al., 2006), to analyze our data and to predict seahorse distributions. Maxent produces a habitat suitability map for the focal species based on a set of related variables (model predictors) and a set of georeferenced occurrences. Maxent is considered as one of the most powerful modeling techniques (Hernandez et al., 2006; Phillips et al., 2006), as it is 1) robust to positional uncertainty/errors in species occurrences (Graham et al., 2008; Fernandez et al., 2009), 2) suitable for limited occurrences (e.g. SS dataset in our case), and 3) reliable for deriving predictive maps with coarse-grain data (Osborne and Leitao, 2009).

2.3. Model predictors

We compiled data for twenty-one variables belonging to three categories: 1) climate and geophysical suitability (Tyberghein et al., 2012), 2) food availability, and 3) macro-habitat availability from online databases (Table S1 in Appendix A). Original data were interpolated with resolution of 1/12° in latitude and longitude (~10 km) using Inverse Distance Weighting in an ArcMap (Cheung et al., 2009). We chose 1/12° as our standard resolution because the majority of the original data were at this resolution, and it also represents cells explicit enough for mapping seahorses at the broad spatial scale of our study area. Since seahorses are typically found in shallow waters, we used a 200-m depth envelope (commonly considered to be continental shelf) as the geographic boundary for all environmental data. By doing so, we can prevent model over-prediction. We then used Pearson correlation coefficients to identify and exclude highly correlated variables ($|r| > 0.7$), which were not used in the model.

2.4. Species data

2.4.1. Species sightings (SS)

We first obtained a total of 33 species sightings (SS) from five online databases: Global Biodiversity Information Faculty (GBIF, www.gbif.org), Oceanic Biodiversity Information System (OBIS, www.iobis.org), FishNet2 (www.fishnet2.net), FishBase (www.fishbase.org), and iSeahorse (www.iseahorse.org). We then obtained new sightings records of seahorses from Chinese colleagues, divers, and fishers during our interview-based research in China (see next paragraph of local ecological knowledge). We validated the species identification for all records by checking specimens where possible, using a standard identification textbook (Lourie et al., 2004). To ensure data quality, sightings located on land or out of our defined range (i.e. 200-m depth of China's seas) were not used.

2.4.2. Peer-reviewed literature (PRL)

We extracted data from peer-reviewed literature (PRL) drawn from the China Knowledge Resource Integrated Database (www.eng.oversea.cnki.net, see Appendix A), having found little information in western literature. We emailed authors to request photos of the specimens to validate their identifications. If specific localities were not documented, we included the entire study/sampling area described in the paper as part of the species' range. All species maps from the validated records in literature were digitalized in an ArcMap.

2.4.3. Local ecological knowledge (LEK)

To derive local ecological knowledge (LEK), we conducted semi-structured interviews (Huntington, 2000) at 79 fishing ports (Fig. 1) along the entire coast of China from April to September 2015 (see protocol in Appendix A). The choice of these sites was based on comprehensive consultation with four Chinese colleagues and 28 fishers in the field. At each fishing port, we first chose participants recommended by local fisheries scientists, community leaders, and interviewed fishers. We also haphazardly reached out to other fishers who were available and knowledgeable (e.g. skippers). We conducted each semi-structured interview on board a vessel allowing all fishers working on the boat to participate. This group setting allowed us to cross-validate data among the fishers. Our interviews covered fishers using different types of fishing gears ($n = 10$) in situ.

In each interview, we first identified the seahorses (Fig. 2). We evaluated available specimens in situ then presented a collection of seahorse photographs to help participants recall seahorses that they had sighted. After the interview, we validated the interviews by checking specimens from other sources at the same site. These sources included other participants, local seafood landings and markets, and stores at the same fishing port.

After the taxonomic portion of the interview, we worked with participants to generate distribution maps of each species (Fig. 3). Local commercial fishers often use China's fishing-zone maps (Fig. S1 in