



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. $10 \times 10 \text{ m}^2$) 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. $10 \times 10 \text{ km}^2$) 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).

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 Facility (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

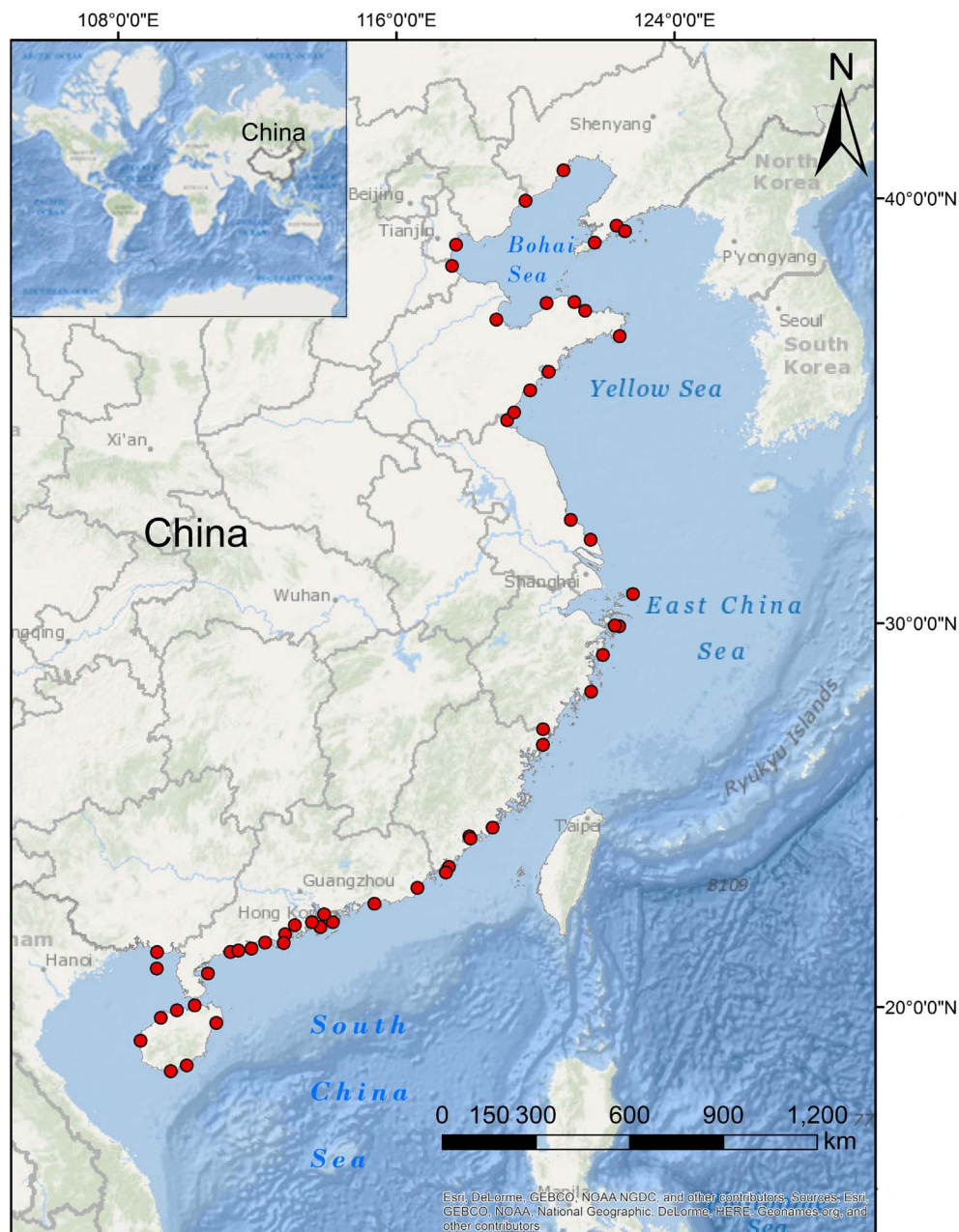


Fig. 1. Study area displaying the sampled fishing ports (red points) along the coast (including islands) of the People's Republic of China. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix A), sometimes in a digital version, to guide their fishing activities, such as locating fishing grounds. Therefore, these maps were ideal tools to help fishers describe (without drawing) species locations. We also asked fishers to recall habitat (depth, substrate) or geographic (distance to landmarks) features of each location if they could. These additional data were checked against nautical charts prior to analyses as a manner to test the reliability of fishers' knowledge. If fishers were not familiar with fishing-zone maps, we presented nautical charts instead for them to recall species distributions. We used an iPad with iGIS software to facilitate fishers' mapping in the field, and digitalized their narrative data in an ArcMap later. As a final step to ensure data quality, we overlaid all fishers' maps of the same species and only retained areas including at least two observations.

2.4.4. Occurrence-points sampling from PRL and LEK coarse maps

Given that species distribution models can only use species point data rather than polygon maps (PRL and LEK original datasets), we

systematically sampled presence points from the PRL and LEK maps. To do so, we first refined the original irregular polygons (i.e. PRL or LEK maps) to range maps consisting of cells with our standard resolution ($1/12^\circ$) in an ArcMap (see Appendix A).

To generate occurrence points from the coarse range maps, we adapted a probability-based sampling approach based on habitat suitability (Niamir et al., 2011). This approach performed better than another technique, random sampling, in an initial trial (see Appendix A). The probability-based approach assigns denser occurrence points to more suitable habitats within the range map, based on the ecological principle that higher abundance was expected in more suitable areas. The required habitat suitability was derived from the Maxent model based on the SS dataset (see 1st scenario in 2.5). We then employed the spatially-balanced sampling in an ArcMap to generate occurrences from the range maps of LEK and PRL for each dataset. This probability-based sampling tool considers sampling probability (i.e. habitat suitability, in this case) and minimizes spatial autocorrelations among the generated

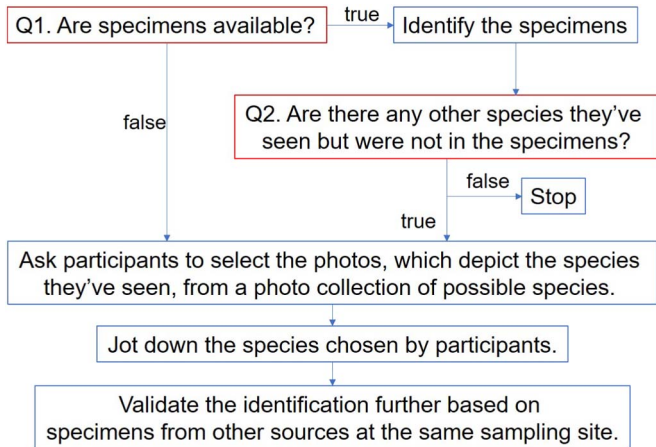


Fig. 2. Process of seahorse taxonomic data collection and validation based on fishers' local ecological knowledge.

points (Theobald et al., 2007).

2.5. Model scenarios and settings

We executed three model scenarios based on different datasets (and combinations) for each species (Table 1): 1) a model with only SS data (1st scenario), 2) three models respectively based on LEK, PRL, and their combination (LEK & PRL, 2nd scenario), and 3) three models separately with SS & LEK, SS & PRL, and ALL (i.e. all sources of datasets, 3rd scenario).

To generate pseudo-absence points for each model, we created explicit bias files to determine sampling background for Maxent (Phillips et al., 2006). For SS datasets, we defined a buffer zone around sightings of all species as the common sampling background for each species. By doing so, the model will generate pseudo-absences only within a certain distance from seahorse presence points. This can advance model's ability to discriminate "highly suitable" from "suitable" habitats (i.e. minimizing over-prediction, Mateo et al., 2010), which is important for mapping distributions for rare species like seahorses (Zarnetske et al., 2007). We chose 1/4° (~15 nautical miles) as the buffer size, as it produced models with the smallest variation among model parameter estimates using different buffer sizes in an initial trial. Similarly, for LEK (or PRL) datasets, we used the union of range maps from LEK (or PRL) of all species as the same sampling background for each species. For each combination of different datasets, we overlaid the sampling backgrounds from the member datasets as the new sampling background. Then we created the bias file based on the sampling background for each model. All the above

processes were done by using the SMDtoolbox (Brown, 2014).

We conducted all models by using the Maxent software (version 3.3.3 k, Phillips et al., 2006). For each model, the number of pseudo-absence points were kept at default (n < 10,000). We randomly divided the species data into training and test subsets (75% and 25% respectively), and replicated the randomizations 15 times. We applied the regularization multiplier to control for over-parameterization (Crall et al., 2015).

2.6. Model evaluation and statistics analyses

We used the mean and standard deviation of three statistics as measures of model performance and variability. The first statistic was the area under the curve (AUC) of the receiver-operating characteristic (ROC) plot (Hanley and McNeil, 1982), which is one of outputs of the Maxent (i.e. test AUC). The ROC plot demonstrates presence/absence prediction accuracy with all possible thresholds of the probability value predicted by the model. The AUC of the ROC plot measures model's general accuracy of both presence and absence predictions. Although the use of AUC has been criticized (Lobo et al., 2008), it is reliable for evaluating presence-only models without applying any probability threshold (Lawson et al., 2014). The second and third statistics were Sensitivity (true presence rate) and Specificity (true absence rate, Altman and Bland, 1994). These two are threshold-dependent and could be better indicators of a model's discriminatory power than the AUC (Lobo et al., 2008). We used the probability threshold at which training Sensitivity plus Specificity was maximized (hereafter SSM threshold) to calculate these two statistics (Liu et al., 2005). In the 2nd scenario of comparison between LEK and PRL, we used SS datasets as independent test data to estimate Sensitivity and Specificity. The calculation was realized in R (R Core Team, 2016) based on the output data from Maxent.

We used Quade tests (Quade, 1979) to examine the effect of using different datasets (LEK, PRL, LEK & PRL) on model performance and variability by controlling for species (n = 5) in the 2nd scenario. For each species, we employed the I similarity statistic (ISS, Warren et al., 2008) to measure the strength of agreement on presence-probability maps between LEK and PRL. We derived the presence/absence maps (SSM thresholds, see above) for LEK and PRL to calculate an agreement ratio (spatial overlap to spatial union) respectively for the predicted presence and absence. Both ISS and the two agreement ratios range from 0 (unmatched) to 1 (fully matched). We then examine the effect of adding LEK and/or PRL to SS on model performance and variability by Quade tests based on model results of the 1st and the 3rd scenarios (Table 1).

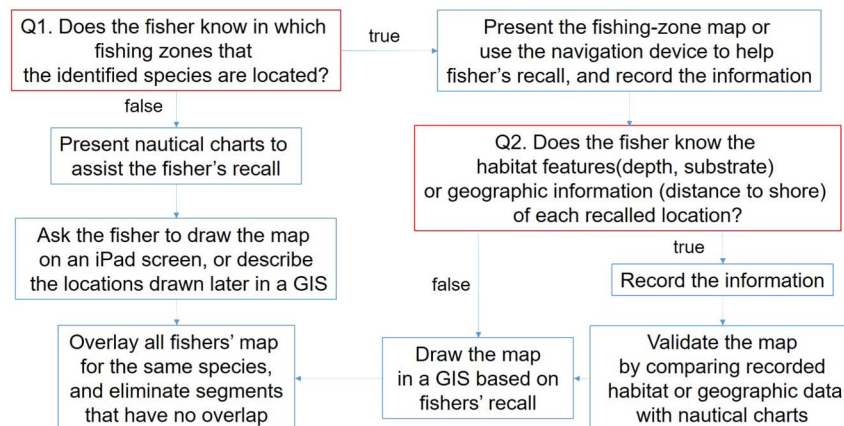


Fig. 3. Processes of mapping and validation of seahorse species distributions based on fishers' local ecological knowledge.

Table 1

The three model scenarios tested in the study. Abbreviation: SS, species sightings; LEK, local ecological knowledge; PRL, peer-reviewed literature; ALL, all data combined (SS, LEK, and PRL); AUC, area under the curve.

| Model scenario | Dataset | Model measure | Statistical analysis | Aim |
|----------------|-------------------------|--|----------------------|---|
| 1st scenario | SS | AUC, sensitivity, specificity | – | To derive probability map; to examine effects of adding LEK and/or PRL data to SS when combined with the 3rd Scenario |
| 2nd scenario | LEK, PRL, LEK & PRL | AUC, sensitivity, specificity; I similarity statistic, presence agreement, absence agreement | Quade test | To compare model performance and results between LEK and PRL |
| 3rd scenario | SS & LEK, SS & PRL, ALL | AUC, sensitivity, specificity | Quade test | To examine effects of adding LEK and/or PRL data to SS when combined with the 1st Scenario |

2.7. Model prediction and ecological interpretation

We followed three steps to determine the presence/absence map for each species. First, we only considered the four models in the first and third scenarios based on datasets including SS, given that SS were more precise than those points downscaled from LEK and PRL maps. Second, we rejected poorly fitting models (AUC lower than 0.7, Manel et al., 2001) and derived the presence/absence maps (by SSM threshold) for the remaining models. Third, we derived an original-data map by overlaying original SS, LEK, and PRL datasets, and then compared the original-data map with each of the predictive maps. We finally selected the predictive map that had a higher number of overlaps with the original-data map and covered less area. If some of the original sightings/ranges were not represented by the selected map, we added the cells occupied by these sightings/ranges to the “selected” map and labeled them as “omission ranges”.

We derived a species-richness map by stacking the presence-probability maps instead of the binary presence/absence maps, as recommended by Calabrese et al. (2014). The presence-probability map for each species was from the logistic-probability predictions of the model that generated the “selected” map (hereafter, selected model). The species-richness map was derived by overlaying these probability maps in an ArcMap.

We applied the selected model's permutation importance to identify key environmental factors (Searcy et al., 2016), and the partial response curves to interpret species ecological niches (Stirling et al., 2016).

3. Results

3.1. Species data

We obtained a total of 55 species sightings (SS), 463 fishers' maps (LEK), and 42 literature maps (PRL) across the same five species (Table 2, see Figs. S2–S4 in Appendix B). Four other species were reported but not validated (see details in Appendix B and Table S2 therein). We excluded these species and their maps (21% of total maps) from model datasets. Among the valid species, *H. trimaculatus* was the most frequently sighted in SS dataset and the most frequently reported in the LEK dataset. *Hippocampus mohnikei* was the most frequently recorded in PRL dataset. All LEK and PRL data were range maps coarser than our mapping resolution, but LEK were finer (LEK:

$1177 \pm 1115 \text{ km}^2$, PRL: $12,453 \pm 11,630 \text{ km}^2$). We generated occurrence points ($n = 50\text{--}200$, Table 2) from the coarse-grain maps (LEK and PRL) by PBS approach. The total number of points differed among species given the different sizes of total area covered by original maps. A minimum of 50 points and a maximum of 200 points were chosen to build robust models and minimize spatial autocorrelation after initial trials.

3.2. Local ecological knowledge (LEK) vs. peer-reviewed literature (PRL)

We found that although LEK tended to produce better models than PRL and LEK & PRL (Fig. 4), the differences were not statistically significant. The highest AUC was obtained by LEK for all species except *H. trimaculatus* (Fig. S5 in Appendix B). The highest values of true presence rate (Sensitivity) were derived from LEK for all species but *H. kelloggi* (Fig. S5). The highest values of true absence rate (Specificity) were also generated by LEK datasets for all species except *H. mohnikei* and *H. kuda* (Fig. S5). All model performance measures (AUC mean and SD, Sensitivity, and Specificity) across species were not statistically different among the three datasets (LEK vs. PRL vs. LEK & PRL, Quade tests, all $p > 0.05$).

Our result indicated that LEK and PRL were generally consistent on model predictions when tested on probability maps, but they did not match well on predicted distributions. The I similarity statistics were medium to high across the five species (0.571–0.853, Table 3), suggesting a good agreement on the predictive probabilities between LEK and PRL. In contrast, the Presence agreements between model results of the two datasets were generally low (Mean = 21.3%), although the Absence agreements were relatively high (Mean = 71.9%, Table 3).

3.3. Effects of adding species locations from coarse-grain data (LEK and PRL) to species sightings (SS)

We found that adding LEK and/or PRL to SS generally derived better predictive models than using SS alone (Fig. 5 left), although the effects differed among the measures. We detected significant differences on Sensitivity (true presence rate) among datasets (Quade tests, $p < 0.05$, $n = 5$), although not on AUC ($p = 0.115$) or Specificity (true absence rate; $p = 0.108$). Sensitivity mean value for SS dataset was significantly lower than those for SS & PRL and SS & LEK (both $p < 0.05$), but not

Table 2

Summary of five Chinese seahorse species with the frequency of records (Frequency), the total number of maps (Maps), and the total amount of points (Points) from three sources: species sightings, local ecological knowledge, and peer-reviewed literature. For the latter two, the points were generated by the downscaling technique based on original maps.

| Seahorse species | Species sightings | | Local ecological knowledge | | | Peer-reviewed literature | | |
|-------------------------|-------------------|--------|----------------------------|------|--------|--------------------------|------|--------|
| | Frequency | Points | Frequency | Maps | Points | Frequency | Maps | Points |
| <i>H. kelloggi</i> | 10.9% | 6 | 13.4% | 37 | 100 | 11.8% | 4 | 100 |
| <i>H. kuda</i> | 27.3% | 15 | 11.2% | 31 | 50 | 5.9% | 2 | 50 |
| <i>H. mohnikei</i> | 14.5% | 8 | 37.1% | 103 | 100 | 82.4% | 28 | 100 |
| <i>H. spinosissimus</i> | 9.1% | 5 | 18.3% | 51 | 100 | 2.9% | 1 | 100 |
| <i>H. trimaculatus</i> | 38.2% | 21 | 87.0% | 241 | 200 | 8.8% | 3 | 200 |

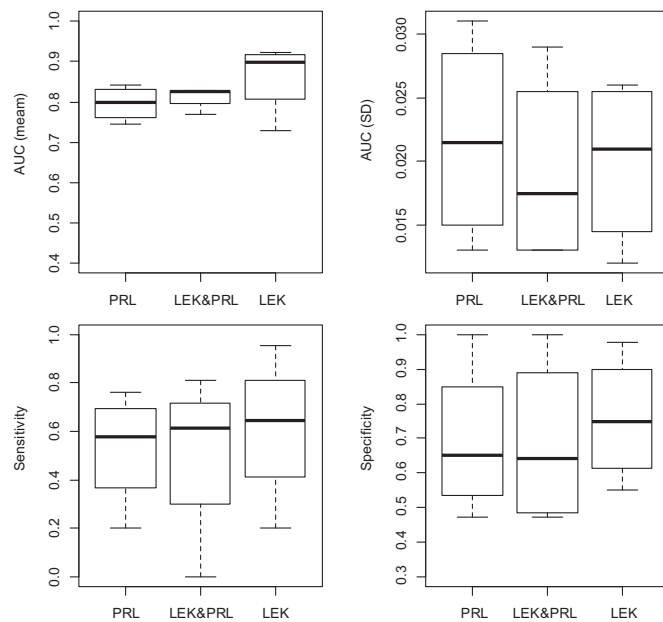


Fig. 4. Boxplots of model performance and variability of Maxent models generated from three different datasets (PRL, LEK, LEK & PRL) based on four measures: AUC (area under the curve) mean and SD (standard deviation), Sensitivity (true presence rate), and Specificity (true absence rate). PRL, peer-reviewed literature data; LEK, local ecological knowledge data; LEK & PRL, the combination of the two.

Table 3

Estimates of agreements on predictive maps between models of local ecological knowledge and models of peer-reviewed literature across the five seahorse species, based on three different measures: I similarity statistic, presence agreement, and absence agreement.

| Species | I similarity statistic | Presence agreement | Absence agreement |
|-------------------------|------------------------|--------------------|-------------------|
| <i>H. kelloggi</i> | 0.571 | 12.0% | 64.7% |
| <i>H. kuda</i> | 0.804 | 7.1% | 86.1% |
| <i>H. mohnikei</i> | 0.681 | 26.7% | 54.0% |
| <i>H. spinosissimus</i> | 0.778 | 16.6% | 75.6% |
| <i>H. trimaculatus</i> | 0.853 | 44.3% | 78.9% |
| Mean | 0.737 | 21.3% | 71.9% |

for ALL ($p = 0.07$). Therefore, adding LEK or PRL, but not both, to SS could significantly improve presence prediction. Model general performance (AUC mean) tended to improve by adding LEK and/or PRL (Fig. 5 left), while absence prediction (Specificity mean) only tended to improve by adding LEK alone (Fig. 5 left).

For model variability, we obtained similar but more consistent results on different measures (Fig. 5 right). We examined statistical differences on standard deviations of all measures (AUC, Sensitivity, and Specificity) among the compared datasets (Quade tests, all $p < 0.05$, $n = 5$). The standard deviations of AUC and Sensitivity for SS dataset were significantly higher than those for the others (posthoc-Quade test, all $p < 0.05$). The standard deviation of Specificity for SS dataset was significantly higher than SS & PRL and ALL (both $p < 0.05$), but not SS & LEK ($p = 0.06$). These results revealed that adding LEK and/or PRL reduced model variability on predicting presence; while only adding PRL or LEK & PRL data decreased model variability on predicting absence.

3.4. Model prediction

The predictive maps indicated that these five seahorse species were generally divergent in spatial distributions, with more species located in the south (Fig. 6). We derived presence/absence maps (Fig. 6a–e) and presence-probability maps (Fig. S6 in Appendix B) based on SS & LEK

datasets for all species but *H. mohnikei*, for which ALL dataset was used. Stacking the presence-probability maps derived a species richness ranging from 0.05 to 3.13, which were then rounded to the nearest integer (i.e. 0 to 3, Fig. 6f). That resulted in an area of 70.3% of our defined shallow seas was occupied by only one seahorse species, 2.1% by two species together, and 0.1% by three species together. Different seahorse species mainly co-existed in the South, especially the eastern Hainan Island and the Penghu Archipelago (Taiwan Province, Fig. 6f).

3.5. Parameter estimation and species ecological niches

We selected eight predictors from the original twenty-one factors (Table 4; Table S1 in Appendix A). The rest eleven predictors were removed since they were highly correlated with one of remaining eight predictors (Pearson correlation test, $|r| > 0.7$; see details in Table S1). The ecological niches of different species were reflected in the response curves on the selected environmental predictors (Figs. S7–S14 in Appendix B). Spatial autocorrelation was low (Moran's $I = 0.01$ to 0.09) in the model datasets for all species, which justified the use the model results to interpret predictor importance and species ecological niches.

The importance of different predictors varied among species, with sea surface temperature (SST mean) generally the most influential variable (Table 4). *Hippocampus mohnikei* was the only species that was more likely to occur in colder waters (Fig. S7), which might explain the significant spatial separation between *H. mohnikei* and the other four species. Species niche separation was also reflected on other predictors. For example, compared with other species, the presence probabilities of *H. kuda*, *H. mohnikei*, *H. trimaculatus* were higher in shallower water closer to the shore (Figs. S8 & S9).

4. Discussion

Our study demonstrated that spatially-explicit species maps of poorly-documented species can be derived by integrating readily available data with species distribution models. Lacking explicit occurrence maps is a common challenge for conservation planning for poorly-recorded organisms (Rondinini et al., 2006; Levin et al., 2014). We indicated that valuable species data could be derived from local ecological knowledge (LEK) and peer-reviewed literature (PRL), when species sightings (SS) were rare. By integrating different species datasets with a presence-only model (i.e. Maxent), we illustrated that it is beneficial to add LEK and/or PRL to the limited SS. These findings are encouraging given the need to inform conservation actions for rarely-studied species (Rondinini et al., 2006), but the financial and temporal constraints on data collection (Anadón et al., 2009).

We demonstrated that fishers can provide species-level maps for seahorse species in China, a capacity which is not universally held by fishers (Aylesworth et al., 2017). The divergence might relate to fisher interest, seahorse morphology, and overlaps in seahorse distributions. For instance, in Thailand, fishers were poor at distinguishing species and may have lacked incentives to care (Aylesworth et al., 2017). In contrast, Chinese fishers may benefit financially from species identification, for large or smooth seahorses fetched higher prices according to respondents. In Thailand, morphological divergences among the seven species may be difficult to differentiate, whereas the five Chinese species have apparent differences on species-specific features, body size, and smoothness (see details in Appendix B and Figs. S15 & S16 therein, and Lourie et al., 2004). In the Thai study, there was likely considerable overlap in seven seahorse species' distributions across the five degrees of latitude (5–10°N) in a tropic zone, whereas the five species in China were apparently more dispersed across the 24° of latitude (17° to 41°N). Fishers in our study only mentioned a maximum of two species each, suggesting little overlap.

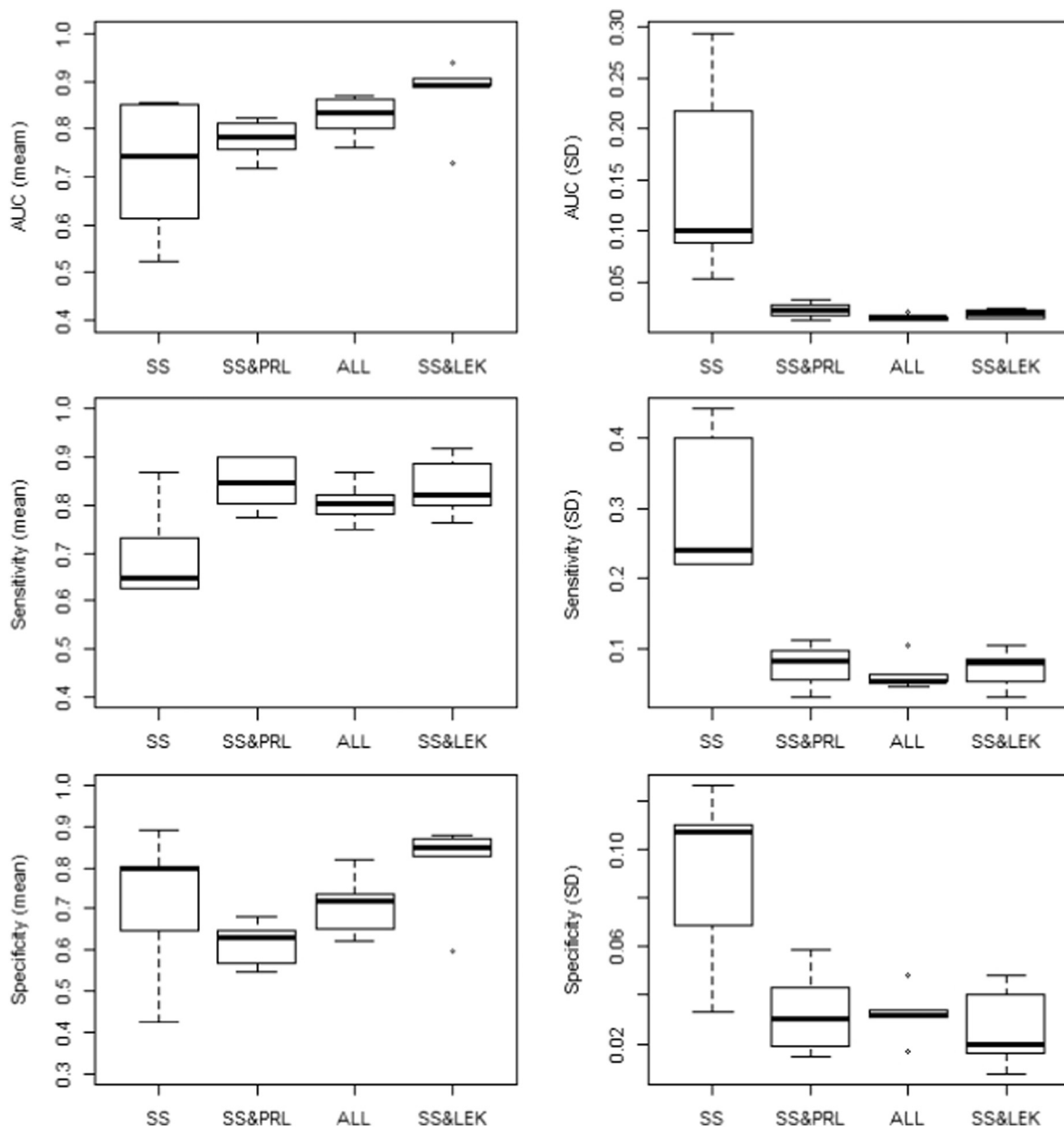


Fig. 5. Boxplots of model performance (mean, left panel) and variability (SD, standard deviation; right panel) measured with the AUC (area under the curve), Sensitivity (true presence rate), and Specificity (true absence rate) among all five seahorse species, based on four different datasets: SS, species sightings; SS & PRL, sightings plus peer-reviewed literature; ALL, sightings plus peer-reviewed literature plus local ecological knowledge; and SS & LEK, sightings plus local ecological knowledge.

4.1. Seahorse species distributions

The distribution patterns for the five species in China are generally consistent with counterparts in other regions. As in Peninsular Malaysia (Choo and Liew, 2003), we found that *H. kelloggi* were more likely to occur in deep (> 30 m) offshore waters, and *H. kuda* was patchily restricted to shallow inshore waters in China. *Hippocampus mohnikei*, was largely clumped in China's temperate zone, as in Japan (Lourie et al., 2004), and Korea (Choi et al., 2012). But this species could occasionally occur in China's warmer regions extending to Southeast Asian countries (see Aylesworth et al., 2016 and references therein). *Hippocampus spinosissimus* and *H. trimaculatus* were more likely to be sympatric, and less patchy than the other three species (Choo and Liew, 2003; Lawson et al., 2015). *Hippocampus trimaculatus* also had the widest habitat, and was relatively more abundant than the other species in bycatch (Choo and Liew, 2003). We noted that Chinese seahorse populations were not likely to be found in estuaries, in contrast to its

Malaysian counterparts. A possible explanation might be the relatively higher environmental stress in China's estuaries (e.g. pollution, SOA, 2012).

Our study indicated that seahorse distributions are highly correlated with ocean temperature at a large spatial scale, in line with many other marine taxa (Tittensor et al., 2010). As ectothermic fish, seahorses are expected to require suitable water temperatures to sustain their metabolism and reproduce (Beitinger and Fitzpatrick, 1979). This temperature suitability could vary among species as indicated in our study and reported from studies in seahorse aquaculture (Koldewey and Martin-Smith, 2010). Given China's seas are among the world's most rapid warming zones (Belkin, 2009), a further study to explore the impact of global warming on seahorse distributions in China would be interesting and meaningful.

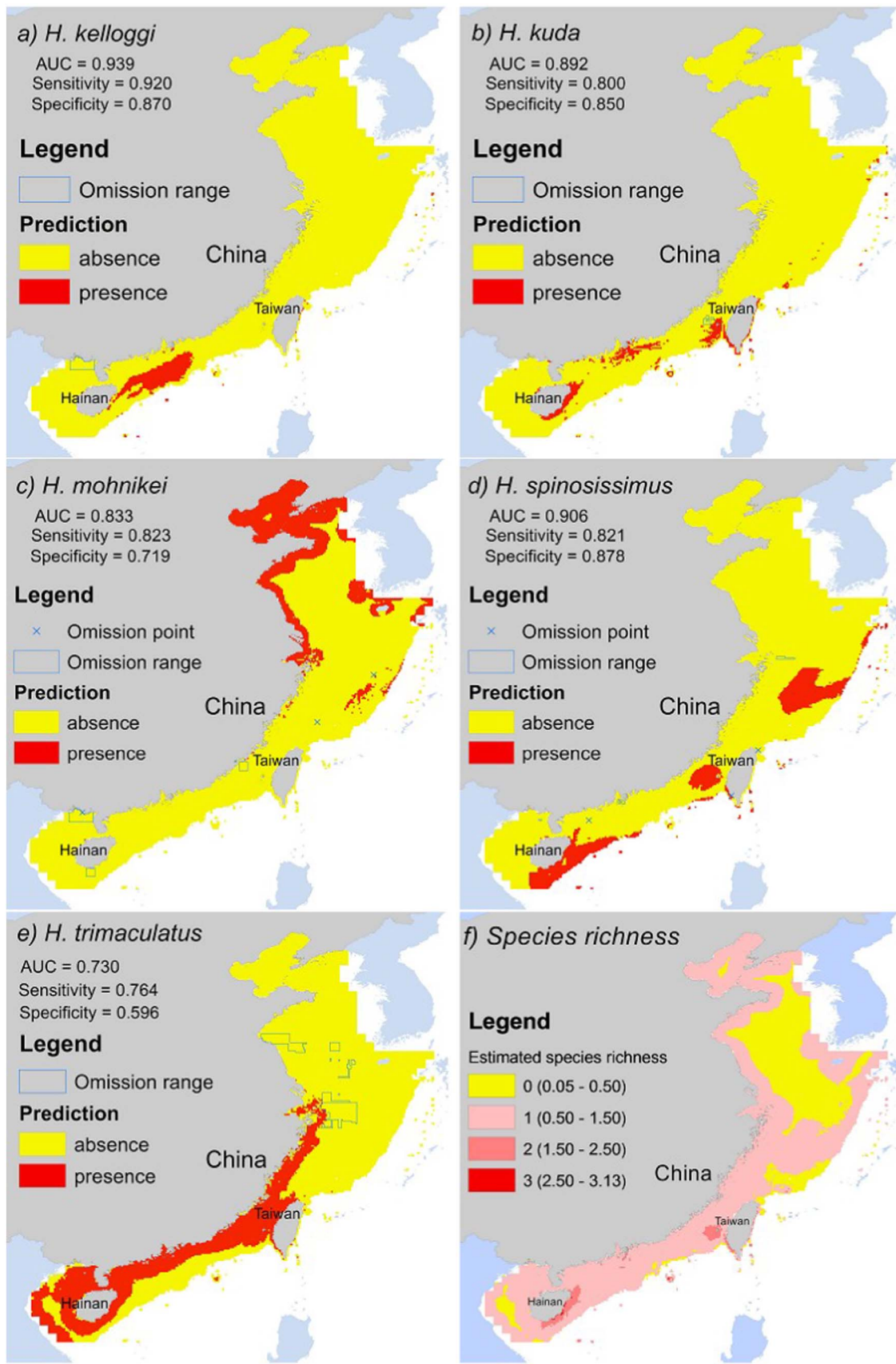


Fig. 6. Predicted distribution maps of a) *H. kelloggi*, b) *H. kuda*, c) *H. mohnikei*, d) *H. spinosissimus*, e) *H. trimaculatus*, and f) species richness of seahorses in China. Omission range or point represent model predictive errors. AUC, area under the curve; Sensitivity, true presence rate; Specificity, true absence rate.

Table 4

Selected model predictors and their relative importance (%) for the model of each species, based on analysis of variable contributions in the Maxent model. Predictors were ranked by the average importance across the five species.

| Predictors | Predictor importance % in modeling for each species | | | | | Mean |
|-------------------------|---|----------------|--------------------|-------------------------|------------------------|------|
| | <i>H. kelloggi</i> | <i>H. kuda</i> | <i>H. mohnikei</i> | <i>H. spinosissimus</i> | <i>H. trimaculatus</i> | |
| Sea surface temperature | 65.6 | 46.2 | 72.2 | 47.6 | 58.2 | 58.0 |
| Distance to shore | 2.4 | 31 | 16.8 | 2.2 | 17.6 | 14.0 |
| Calcite concentration | 11.3 | 14.3 | 2 | 20.3 | 7.4 | 11.1 |
| Silicate concentration | 16.1 | 4.4 | 2.8 | 10.7 | 5.1 | 7.8 |
| Depth | 0.7 | 1.9 | 4.4 | 1.1 | 7 | 3.0 |
| pH | 1.9 | 1.3 | 0.5 | 10.6 | 0.3 | 2.9 |
| Primary productivity | 2.2 | 0.3 | 0.6 | 6.7 | 3.1 | 2.6 |
| Macro-habitat | 0.6 | 0.6 | 0.8 | 0.9 | 1.3 | 0.8 |

4.2. Limitations

The low modeling importance of macro-habitat and the derived species-habitat relations should be viewed with caution, given the constraints on our data availability and spatial scale. First, the macro-habitat datasets we used here were largely drawn from global sources, and thus they might not have good coverage or resolutions in China's seas. Second, we learnt from fishers that Chinese seahorses use various other macro-habitats, including macro-algae, sea fans, and artificial structures (e.g. mussel farms, also see Aylesworth et al., 2016). However, we were unable to obtain these habitat data for China. On the other hand, the importance of macro-habitats might be more prominent at a smaller spatial scale (e.g. a lagoon), as found in European seahorses (Curtis and Vincent, 2005) and reef fishes (Komyakova et al., 2013). To clarify the truth, a study in multiple spatial scales with more comprehensive and explicit macro-habitat variables could be helpful.

4.3. Implications for conservation and management in China

Our new spatially-explicit maps for five seahorse species across China can be used for conservation and management in at least four ways. First, the predictive maps can guide local authorities and researchers to identify more specific locations of seahorse populations for further conservation actions (e.g. protected areas). Second, Chinese authorities can use our maps to protect seahorses through fisheries management. For instance, fisheries officers can explore overlap between seahorse distribution maps and fishing-zone maps to identify locations, where reducing or relocating fishing activities could maximize the benefit for seahorse conservation (Foster and Arreguin-Sánchez, 2014). Third, local law enforcement officers can identify localities (e.g. fishing ports) that are close to seahorse habitats to effectively strike against illegal fishing and trade in seahorses. Fourth, our maps come timely to serve China's ongoing coastal ecological-redline planning, which aims to protect coastal waters with significant ecological importance (Peng et al., 2016). Given that seahorses are found in various coastal ecosystems, protecting seahorses means protecting these critical waters. We think a planning includes seahorse habitats could be ecologically meaningful and beneficial to China's marine systems.

4.4. Integrating multiple datasets in SDM research

Among the approaches used in this study, we adapted a technique especially applicable for poorly-recorded species to address the coarse-grain maps in SDM research. The use of LEK in modeling distributions is a common challenge in both terrestrial and marine systems. Local citizens could only provide coarse-range maps for the focal organisms (Bergmann et al., 2004; Carter and Nielsen, 2011; Aylesworth et al., 2017; Laze and Gordon, 2016). Downscaling techniques are thus vital to generate finer-resolution data from coarse-grain maps like LEK. We

adapted the downscaling technique of Niamir et al. (2011), which used expert knowledge to determine habitat suitability for each species. Expert knowledge may be only available for well-known organisms (Murray et al., 2009; Niamir et al., 2011), and not for rarely-studied ones like Chinese seahorses. To fill the gap, we used the Maxent model to generate habitat suitability, as it only requires a small number of species sightings ($n \geq 5$). This advantage may make our technique more applicable to poorly-documented species, which was rarely addressed in literature.

Our study indicated that integrating multiple datasets in predicting distributions for poorly-recorded species is beneficial, but it should be done with appropriate datasets combinations. We demonstrated that integrating LEK with PRL should be taken with caution, as it may not necessarily improve model performance. In our study, both LEK and PRL occurrences were sourced from coarse-grain maps, which were not very consistent in space. Previous studies of comparing LEK with data from traditional biological surveys have also shown similar mismatch on other species (Thornton and Scheer, 2012). Combining species data from "inconsistent" sources might cumulate spatial variance. Besides, given both LEK and PRL are likely coarse maps, the derived occurrences could include some spatial errors. This is especially true for PRL maps in our study (10 times coarser than LEK maps, on average). Both reasons above could increase model residual and degrade model performance (Graham et al., 2008). We suggest future studies, which aim to integrate datasets from different sources (especially coarse-grain data), better compare different scenarios as we executed to identify the best datasets combinations and model predictions.

We highlight the possible utility of LEK in species distribution modeling for poorly-recorded marine species, in line with similar studies on terrestrial counterparts (Irvine et al., 2009; Anadón et al., 2010; Laze and Gordon, 2016). As acknowledged in terrestrial studies (Anadón et al., 2010), we think that LEK can be a cost-effective data source for modeling poorly-recorded species in marine systems. In addition, LEK can provide information to potentially improve model predictions. For instance, with the key information about physical features (e.g. fence, roads) and weightings of environment factors, Irvine et al. (2009) significantly improved model predictions of deer (*Cervus elaphus*). In our study, local fishers have sighted seahorses often co-occurring with sea fans in bycatch, and clumps of juvenile seahorses drifting in water column with branches of macro-algae (Fig. B16 in Appendix B). These LEK data suggest that other related macro-habitats variables if available might derive better predictions.

4.5. The importance of spatially-explicit maps of poorly-recorded species

Spatially-explicit biogeographic maps for poorly-recorded species can help inform conservation actions. In particular, greater detail in spatially-explicit biogeographic maps can improve conservation planning (Rondinini et al., 2006). For instance, species of concern can benefit from fine-grain resolution when exploring overlaps between their distributions and protected areas (Rondinini et al., 2006; Pimm

et al., 2014). The corollary is that fine-resolution distribution maps of focal species can inform us of conservation gaps, where actions should be taken. The ability to develop such spatially-explicit maps for poorly-known species should allow protective measures even as knowledge is being improved. Such potential may be particularly important in the many biodiverse countries (e.g. China) with limited biogeographic data and resources (Liu, 2013). In such instances tapping into diverse sources of information (including local ecological knowledge) can create valuable species distribution models and predictive maps.

Acknowledgements

This is a contribution from Project Seahorse. We thank Dr. Shan Xiujuan, Dr. Du Jianguo, Dr. Wang Jiangyong, and Dr. Lin Qiang for their suggestions on sampling sites and information about local context. We thank Dr. Wang Jiangyong, Dr. Wang Shifeng, Mr. Lin Haisheng for providing contact information of local fishers. We thank Dr. Guo Huayang, Dr. Dai Fangchun in assisting interviews with fishers. We are grateful to Dr. William Cheung, Dr. Sarah Gergel, Dr. Lindsay Aylesworth, Jennifer Selgrath, Riley Pollom, Tanvi Vaidyanathan, Kyle Gillespie, and Ting-Chun Kuo, and two anonymous referees for their constructive comments. XZ was supported by a Chinese Doctorial Scholarship from the China Scholarship Council. This project was funded by Disney Conservation Fund (2015). Project Seahorse is supported by Guylian Chocolates (Belgium) in a partnership for marine conservation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.biocon.2017.05.020>.

References

- Altman, D.G., Bland, J.M., 1994. Diagnostic tests. 1: sensitivity and specificity. *BMJ [Br. Med. J.]* 308, 1552.
- Anadón, J.D., Giménez, A., Ballestar, R., 2010. Linking local ecological knowledge and habitat modelling to predict absolute species abundance on large scales. *Biodivers. Conserv.* 19, 1443–1454.
- Anadón, J.D., Giménez, A., Ballestar, R., Pérez, I., 2009. Evaluation of local ecological knowledge as a method for collecting extensive data on animal abundance. *Conserv. Biol.* 23, 617–625.
- Aylesworth, L., Lawson, J., Laksanawimol, P., Ferber, P., Loh, T.L., 2016. New records of the Japanese seahorse *Hippocampus mohnikei* in Southeast Asia lead to updates in range, habitat and threats. *J. Fish Biol.* 88, 1620–1630.
- Aylesworth, L., Phoosawat, R., Suvanachai, P., Vincent, A., 2017. Generating spatial data for marine conservation and management. *Biodivers. Conserv.* 26, 383–399.
- Beitinger, T.L., Fitzpatrick, L.C., 1979. Physiological and ecological correlates of preferred temperature in fish. *Am. Zool.* 19, 319–329.
- Belkin, I.M., 2009. Rapid warming of large marine ecosystems. *Prog. Oceanogr.* 81, 207–213.
- Bergmann, M., Hinz, H., Blyth, R., Kaiser, M., Rogers, S., Armstrong, M., 2004. Using knowledge from fishers and fisheries scientists to identify possible groundfish 'essential fish habitats'. *Fish. Res.* 66, 373–379.
- Berkes, F., 1993. Traditional ecological knowledge in perspective. In: *Traditional Ecological Knowledge: Concepts and Cases*. 1.
- Bernstein, R., 1982. Sea surface temperature estimation using the NOAA 6 satellite advanced very high resolution radiometer. *J. Geophys. Res. Oceans* 87, 9455–9465.
- Brown, J.L., 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods Ecol. Evol.* 5, 694–700.
- Caddy, J., 1968. Underwater observations on scallop (*Placopecten magellanicus*) behaviour and drag efficiency. *Journal of the Fisheries Board of Canada* 25, 2123–2141.
- Calabrese, J.M., Certain, G., Kraan, C., Dormann, C.F., 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models. *Glob. Ecol. Biogeogr.* 23, 99–112.
- Caldwell, I.R., Vincent, A.C., 2013. A sedentary fish on the move: effects of displacement on long-snouted seahorse (*Hippocampus guttulatus* Cuvier) movement and habitat use. *Environ. Biol. Fish.* 96, 67–75.
- Carter, B.T., Nielsen, E.A., 2011. Exploring ecological changes in Cook Inlet beluga whale habitat through traditional and local ecological knowledge of contributing factors for population decline. *Mar. Policy* 35, 299–308.
- Cheung, W.W., Lam, V.W., Sarmiento, J.L., Kearney, K., Watson, R., Pauly, D., 2009. Projecting global marine biodiversity impacts under climate change scenarios. *Fish. Fish.* 10, 235–251.
- Choi, Y.-U., Rho, S., Park, H.-S., Kang, D.-H., 2012. Population characteristics of two seahorses, *Hippocampus coronatus* and *Hippocampus mohnikei*, around seagrass beds in the southern coastal waters of Korea. *Ichthyol. Res.* 59, 235–241.
- Choo, C., Liew, H., 2003. Spatial distribution, substrate assemblages and size composition of sea horses (family *Syngnathidae*) in the coastal waters of peninsular Malaysia. *J. Mar. Biol. Assoc. UK* 83, 271–276.
- Costello, M.J., Coll, M., Danovaro, R., Halpin, P., Ojaveer, H., Miloslavich, P., 2010. A census of marine biodiversity knowledge, resources, and future challenges. *PLoS One* 5, e12110.
- Crall, A.W., Jarnevič, C.S., Young, N.E., Panke, B.J., Renz, M., Stohlgren, T.J., 2015. Citizen science contributes to our knowledge of invasive plant species distributions. *Biol. Invasions* 17, 2415–2427.
- Curtis, J.M., Vincent, A.C., 2005. Distribution of sympatric seahorse species along a gradient of habitat complexity in a seagrass-dominated community. *Mar. Ecol. Prog. Ser.* 291, 81–91.
- Fernandez, M., Blum, S., Reichle, S., Guo, Q., Holzman, B., Hamilton, H., 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. *Biodivers. Inform.* 6.
- Foster, S., Arreguin-Sánchez, F., 2014. Using distribution patterns of small fishes to assess small fish by-catch in tropical shrimp trawl fisheries. *Anim. Conserv.* 17, 217–224.
- Franklin, J., 2010. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Presspp. 11–12.
- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Townsend Peterson, A., Loisele, B.A., 2008. The influence of spatial errors in species occurrence data used in distribution models. *J. Appl. Ecol.* 45, 239–247.
- Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N.G., Lehmann, A., Zimmermann, N.E., 2006. Using niche-based models to improve the sampling of rare species. *Conserv. Biol.* 20, 501–511.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009.
- Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* 16, 1424–1435.
- Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29–36.
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29, 773–785.
- Huntington, H.P., 2000. Using traditional ecological knowledge in science: methods and applications. *Ecol. Appl.* 10, 1270–1274.
- Irvine, R., Fiorini, S., Yearley, S., McLeod, J., Turner, A., Armstrong, H., White, P., Van Der Wal, R., 2009. Can managers inform models? Integrating local knowledge into models of red deer habitat use. *J. Appl. Ecol.* 46, 344–352.
- Koldewey, H.J., Martin-Smith, K.M., 2010. A global review of seahorse aquaculture. *Aquaculture* 302, 131–152.
- Komyakova, V., Munday, P.L., Jones, G.P., 2013. Relative importance of coral cover, habitat complexity and diversity in determining the structure of reef fish communities. *PLoS One* 8, e83178.
- Lawson, C.R., Hodgson, J.A., Wilson, R.J., Richards, S.A., 2014. Prevalence, thresholds and the performance of presence-absence models. *Methods Ecol. Evol.* 5, 54–64.
- Lawson, J., Foster, S., Lim, A., Chong, V., Vincent, A., 2015. Novel life-history data for threatened seahorses provide insight into fishery effects. *J. Fish Biol.* 86, 1–15.
- Laze, K., Gordon, A., 2016. Incorporating natural and human factors in habitat modelling and spatial prioritisation for the *Lynx lynx martinoi*. *Web Ecol.* 16, 17.
- Levin, N., Coll, M., Frascchetti, S., Gal, G., Giakoumi, S., Göke, C., Heymans, J.J., Katsanevakis, S., Mazor, T., Öztürk, B., 2014. Biodiversity data requirements for systematic conservation planning in the Mediterranean Sea. *Mar. Ecol. Prog. Ser.* 508, 261–281.
- Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28, 385–393.
- Liu, J., 2013. Status of marine biodiversity of the China Seas. *PLoS One* 8, 1–24.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–151.
- Lourie, S., Green, D., Vincent, A., 2005. Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (*Syngnathidae*: *Hippocampus*). *Mol. Ecol.* 14, 1073–1094.
- Lourie, S.A., Foster, S.J., Cooper, E.W., Vincent, A.C., 2004. *A Guide to the Identification of Seahorses*. Project Seahorse and TRAFFIC North America. University of British Columbia and World Wildlife Fund, Washington, DC.
- Lourie, S.A., Vincent, A.C., 2004. Using biogeography to help set priorities in marine conservation. *Conserv. Biol.* 18, 1004–1020.
- Mace, G.M., Collar, N.J., Gaston, K.J., Hilton-Taylor, C., Akçakaya, H.R., Leader-Williams, N., Milner-Gulland, E.J., Stuart, S.N., 2008. Quantification of extinction risk: IUCN's system for classifying threatened species. *Conserv. Biol.* 22, 1424–1442.
- Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* 38, 921–931.
- Mateo, R.G., Croat, T.B., Felicísimo, Á.M., Muñoz, J., 2010. Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Divers. Distrib.* 16, 84–94.
- McShane, T.O., Hirsch, P.D., Trung, T.C., Songorwa, A.N., Kinzig, A., Monteferri, B., Mutekanga, D., Van Thang, H., Dammert, J.L., Pulgar-Vidal, M., 2011. Hard choices: making trade-offs between biodiversity conservation and human well-being. *Biol. Conserv.* 144, 966–972.
- MEP, 2002. *National List of Wildlife with Protection Priority*. http://sts.mep.gov.cn/swdxy1/swwzyzbh/200211/t20021118_83384.htm Higher Education Presspp. 746.

- Micheli, F., Levin, N., Giakoumi, S., Katsanevakis, S., Abdulla, A., Coll, M., Fraschetti, S., Kark, S., Koutsoubas, D., Mackelworth, P., 2013. Setting priorities for regional conservation planning in the Mediterranean Sea. *PLoS One* 8, e59038.
- Murray, J.V., Goldizen, A.W., O'Leary, R.A., McAlpine, C.A., Possingham, H.P., Choy, S.L., 2009. How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. *J. Appl. Ecol.* 46, 842–851.
- Niamir, A., Skidmore, A.K., Toxopeus, A.G., Muñoz, A.R., Real, R., 2011. Finessing atlas data for species distribution models. *Divers. Distrib.* 17, 1173–1185.
- Osborne, P.E., Leitao, P.J., 2009. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. *Divers. Distrib.* 15, 671–681.
- Peng, Z., Wenhui, L., Jun, S., Pearson, S., Hongsheng, Y., 2016. Natural coast protection and use in China: Implications of resource protection “Redline” policies. *Coast. Manag.* 44, 21–35.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197.
- Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., Sexton, J.O., 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344, 1246–1252.
- Quade, D., 1979. Using weighted rankings in the analysis of complete blocks with additive block effects. *J. Am. Stat. Assoc.* 74, 680–683.
- Reis, E., López-Iborra, G.M., Pinheiro, R.T., 2012. Changes in bird species richness through different levels of urbanization: implications for biodiversity conservation and garden design in Central Brazil. *Landsc. Urban Plan.* 107, 31–42.
- Robinson, L., Elith, J., Hobday, A., Pearson, R., Kendall, B., Possingham, H., Richardson, A., 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Glob. Ecol. Biogeogr.* 20, 789–802.
- Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H., Possingham, H.P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecol. Lett.* 9, 1136–1145.
- Searcy, C.A., Shaffer, H.B., Ackerly, D.D., Bronstein, J.L., 2016. Do ecological niche models accurately identify climatic determinants of species ranges? *Am. Nat.* 187, 423–435.
- SOA, 2012. Bulletin of Marine Environmental Quality of China in 2011. China Ocean Press, Beijing, China (in Chinese).
- Stirling, D.A., Boulcott, P., Scott, B.E., Wright, P.J., 2016. Using verified species distribution models to inform the conservation of a rare marine species. *Divers. Distrib.* 22, 808–822.
- R Core Team, 2016. R: a language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria, URL: <https://www.R-project.org/>.
- Theobald, D.M., Stevens Jr., D.L., White, D., Urquhart, N.S., Olsen, A.R., Norman, J.B., 2007. Using GIS to generate spatially balanced random survey designs for natural resource applications. *Environ. Manag.* 40, 134–146.
- Tittensor, D.P., Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V., Worm, B., 2010. Global patterns and predictors of marine biodiversity across taxa. *Nature* 466, 1098–1101.
- Thornton, T.F., Scheer, A.M., 2012. Collaborative engagement of local and traditional knowledge and science in marine environments: a review. *Ecol. Soc.* 17, 8.
- Tyberghein, L., Verbruggen, H., Pauly, K., Troupin, C., Mineur, F., De Clerck, O., 2012. Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Glob. Ecol. Biogeogr.* 21, 272–281.
- Vincent, A., Foster, S., Koldewey, H., 2011. Conservation and management of seahorses and other *Syngnathidae*. *J. Fish Biol.* 78, 1681–1724.
- Wang, S., Xie, Y., 2009. China species red list. In: Vertebrates. Vol. 2 Higher Education Press.
- Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868–2883.
- Wentz, F.J., Schabel, M., 2000. Precise climate monitoring using complementary satellite data sets. *Nature* 403, 414–416.
- Zarnetske, P.L., Edwards, T.C., Moisen, G.G., 2007. Habitat classification modeling with incomplete data: pushing the habitat envelope. *Ecol. Appl.* 17, 1714–1726.