Spatial pattern and determinants of global invasion risk of an invasive species, sharpbelly *Hemiculter leucisculus* (Basilesky, 1855)

Xianghong Dong a,b,c,d, Tao Ju f, Gaël Grenouillet c,e, Pascal Laffaille d, Sovan Lek c, Jiashou Liu a,*

a State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, PR China  
b University of Chinese Academy of Sciences, Beijing 100190, PR China  
c Laboratoire Evolution & Diversité Biologique (EDB), UMR5174, CNRS, IRS, UPS, 118 route de Narbonne, 31062 Toulouse Cedex 9, France  
d Ecolab, Université de Toulouse, CNRS, INPT, UPS, Toulouse 31062, France  
e Institut Universitaire de France, Paris, France  
f Key Laboratory of Freshwater Biodiversity Conservation, Ministry of Agriculture and Rural Affairs of China, Yangtze River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Wuhan 430223, PR China

**Highlights**

- Invasion risk and its determinants are important for invasive species' management.  
- Sharpbelly's global habitat suitability and its determinants were first forecasted.  
- A global risk map was drawn with habitat suitability and introduction likelihood.  
- Areas with high invasion risk by sharpbelly spread over the world except Antarctica.  
- Human Influence Index was the most critical factor shaping sharpbelly's invasion.

**Abstract**

Invasive species have imposed huge negative impacts on worldwide aquatic ecosystems and are generally difficult or impossible to be eradicated once established. Consequently, it becomes particularly important to ascertain their invasion risk and its determinants since such information can help us formulate more effective preventive or management actions and direct these measures to those areas where they are truly needed so as to ease regulatory burdens. Here, we examined the global invasion risk and its determinants of sharpbelly (*Hemiculter leucisculus*), one freshwater fish which has a high invasive potential, by using species distribution models (SDMs) and a layer overlay method. Specifically, first an ensemble species distribution model and its basal models (developed from seven machine learning algorithms) were explored to forecast the global habitat-suitability and variables importance for this species, and then a global invasion risk map was created by combining habitat-suitability with a proxy for introduction likelihood (entailing propagule pressure and dispersal constraints) of exotic sharpbelly. The results revealed that (1) the ensemble model had the highest predictive power in forecasting sharpbelly's global habitat-suitability; (2) areas with high invasion risk by sharpbelly patchily spread over the world except Antarctica; and (3) the Human Influence Index (HII), rather than any of the bioclimatic variables, was the most important factor influencing sharpbelly's future invasion. Based on these results, the present
1. Introduction

During the past two centuries, the number of non-indigenous species (NIS) worldwide has increased considerably (Sala et al., 2000; Seebens et al., 2017), causing severe ecological and economic hazards to their recipient regions (Millennium Ecosystem Assessment, 2005). Among these NIS, the sharpbelly Hemiculter leucisculus (Basilewsky, 1855), one small freshwater fish with minor commercial importance, is of particular concern and an iconic example that can be used to illustrate these deleterious impacts. This fish is natively distributed in Far East Russia, Mongolia, Korean Peninsula, China, and Vietnam (CABI, 2016), but has successfully established invasive populations in many countries outside its native range, including Afghanistan, Iran, Iraq, Kazakhstan, Uzbekistan, Turkmenistan, and Azerbaijan (Mustafayev et al., 2015; CABI, 2016). In these novel habitats, this species has been found to harm local aquaculture activities either by competing with the juveniles of farmed fish or directly feeding on their eggs and fry (CABI, 2016), trigger taxonomic homogenization of indigenous ichthyofauna through displacing native small fish species (Rosenthal, 1976), and even put public health into a dangerous situation via transmitting the parasite Diplostomum spathaceum (Rudolphi, 1819) of human eye fluke disease (Palmieri et al., 1976; Sattari et al., 2007). Due to these detrimental effects, this species has been recently listed as one of the most dangerous and potential invaders to Turkey waters (Tarkan et al., 2014; Tarkan et al., 2015; Tarkan et al., 2017). Worse still, after scrutinizing the propagation pathways of H. leucisculus, Smith et al. (2014) presumed that this species is probable to follow the footprint of the notorious small Asian cyprinid fish Pseudorasbora parva (Temminck & Schlegel, 1846), which has made trans-continental invasion within 50 years (Gozlan et al., 2010) and is deemed to be difficult or impossible to be eradicated from those invaded regions (Britton et al., 2010; Aparicio et al., 2012). This status exacerbates the urgency of taking preventive measures to halt or restrain the expansion of H. leucisculus into new water bodies, especially where the environmental conditions are suitable for its establishment and subsequent invasion. In this context, it becomes particularly imperative to select a proactive tool to ascertain this undesirable species’ global invasion risk and its determinants since such information can help us formulate more effective preventive or managerial actions and direct these measures to those areas where they are truly needed (Reshetnikov and Ficetola, 2011; Poulos et al., 2012; Guisan et al., 2013).

Species distribution models (SDMs), which depict the associations between species observation data and underlying environmental parameters, have been testified to be quite a useful proactive tool to discern invasion potential and its determinants of invasive organisms (Peterson, 2003; Guisan and Thuiller, 2005; Reshetnikov and Ficetola, 2011; Poulos et al., 2012; Banha et al., 2017). There are many algorithms available for a specific application scenario, but each of them has some uncertainties (Grenouillet et al., 2011). Hence, in practice, many scientists recommended ensemble model (i.e. to merge or join the outputs of several basal models) for reducing the uncertainties of individual algorithms (Araújo and New, 2007; Buisson et al., 2010; Grenouillet et al., 2011; Li and Wang, 2013). Hitherto, this consensus method has been applied to a variety of aquatic invasive species to delineate their invasion potential in those areas where invasion has not yet been made or observed, such as the eastern mosquito fish Gambusia holbrooki Girard, 1859 (Murphy et al., 2015), the Chinese mitten crab Eriocheir sinensis (H. Milne-Edwards, 1853) (Zhang et al., 2019), the golden apple snail Pomacea canaliculata (Lamarck, 1819) (Lei et al., 2017), and the Asian clam Corbicula fluminea (O. F. Müller, 1774) (Gama et al., 2015).

However, reviewing the assumptions of SDMs, we can easily find that these aforementioned studies only presented current habitat-suitability for these aquatic invasive species, but did not treat bio-invasion as a process (Banha et al., 2017) and paid little attention on those key factors related to their invasion success, such as propagule pressure (i.e. individuals number into a non-native environment), geographical dispersal constraints (e.g. anthropogenic or natural barriers), biotic interactions, and species traits (Araújo and Guisan, 2006; Ficetola et al., 2007; Sax et al., 2007; Filipe et al., 2013; Banha et al., 2017). Actually, whether or not one invasive species can appear in a particular location clearly more depends on these key factors rather than habitat-suitability (Araújo and Pearson, 2005; Araújo and Guisan, 2006; Williams et al., 2008; Václavík and Meentemeyer, 2009; Banha et al., 2017). Although some invasion biologists have demonstrated that taking these elements into consideration can make the invasion risk assessment closer to the true one (Gallardo and Aldridge, 2013; Gallardo, 2014), only a limited number of studies to date have followed this advice (To our knowledge: Gallardo and Aldridge, 2013; Gallardo, 2014; Bradie and Leung, 2015; Fletcher et al., 2016; Banha et al., 2017).

With this in view, being different from two previous works carried out in the conterminous USA, which both only adopted one single algorithm: the Genetic Algorithm for Rule-Set Production (GARP) (Chen, 2008) and the Euclidian algorithm (Sanders et al., 2014), the present study used an ensemble model and its basal models to forecast the global habitat-suitability and its determinants for the invasive species H. leucisculus, and then overlaid the habitat-suitability with a proxy for introduction likelihood (entailing two key factors: propagule pressure and dispersal constraints) to produce a global risk map for this invader (Hulme, 2009). We sought to answer the following questions: (1) what is the spatial pattern of invasion risk of H. leucisculus at a global scale; (2) which are the most influential environmental variables that drive the invasion of H. leucisculus; and (3) how can we impede or mitigate further expansion of this problematic fish in the future.

2. Materials and methods

2.1. Species presence/pseudo-absence dataset

Presence records of H. leucisculus complied with the literature review (n = 362), FishBase (http://www.fishbase.org; retrieved in February 2019), and the Global Biodiversity Information Facility (GBIF; http://www.data.gbif.org; retrieved in February 2019). The latest presence was recorded in 2019. After removing the duplicate, incomplete, and apparently-erroneous entries (e.g. terrestrial records), we preliminarily derived 614 inhomogeneously-distributed records, with higher density in the native range and lower density in the invasive range (Fig. 1). In order to diminish the effect of spatial auto-correlation originated from spatial clustering of presence records (i.e. more than one presence records dropped in one environmental grid cell with 2.5 arc-minutes spa-
tial resolution, ca. 4.5 km² at the equator), a spatial thinning or rarefying approach similar to that used by Riul et al. (2013) and Boria et al. (2014) was operated on our presence dataset, resulting in that at most one record was contained in each environmental grid cell. Correspondingly, 541 presence records of *H. leucisculus* were finally retained. Seeing that presence-absence algorithms should perform better than presence-only algorithms (Elith et al., 2006) and the true absence records of *H. leucisculus* at a global scale are impossible to be accessed confidently, we instead generated 541 pseudo-absence records (PAs) spanning the globe using a random method conditioned by excluding pixels where presences are known (Capinha et al., 2011; Barbet-Massin et al., 2012; Jensen et al., 2017). Afterwards, we combined these pseudo-absence records with those retained presence records, and then shuffled the whole dataset to obtain the ultima species presence and pseudo-absence dataset for posterior analysis.

2.2. Environmental variables

Based on the previous studies (Chen, 2008; Sanders et al., 2014) and our expert opinions, three types of environmental variables considered ecologically meaningful for spatial distribution of *H. leucisculus* were initially selected to compose the candidate predictors of the present study: (1) the Global Human Influence Index (HII; an unitless synthetic variable consisting of four data types’ proxies for human influence: human settlement, land transformation, infrastructure, and human access), which represents direct human disturbance over the land surface (Sanderson et al., 2002) and ranges from 0 (no impact) to 64 (maximum impact) (SEDAC, 2005; Sandel and Svenning, 2013; Yohe et al., 2014); (2) Global Reservoirs and Dams (ResDam; water surfaces) (Lehner et al., 2011); and (3) 19 bioclimatic variables (BIO1–BIO19), which are commonplace in the field of ecological niche modeling (Hijmans et al., 2005). All the 21 raster-based variable layers were resampled using the method of bilinear interpolation, and then standardized, aiming to make themselves share the same resolution (2.5 arc-minutes, ca. 4.5 km² at the equator), spatial extent, and geographic coordinate system (WGS 1984 datum), as well as reduce dimension effects on the models (Vander Zanden et al., 2004; Fletcher et al., 2016). Prior to further processing, the variance inflation factor (VIF) for each candidate predictor variable layer was calculated and the superfluous predictor variable layer (whose VIF value is more than 10) was removed (Conoscenti et al., 2016; Duque-Lazo et al., 2018) so as to avoid over-fitting, obtain more succinct or interpretable models and eliminate the effects of multicollinearity (Dormann et al., 2013; Li et al., 2015; Fletcher et al., 2016; Júnio and Nóbrega, 2018). As a result, six non-redundant predictor variable layers which consisted of the final set of predictive variables for subsequent modeling were obtained: mean temperature of driest quarter (BIO8); precipitation seasonality (BIO15); precipitation of warmest quarter (BIO18); precipitation of coldest quarter (BIO19); HII; and ResDam (Table 1).

2.3. Ensemble modeling

Hybrid or ensemble models have been proved superior to single algorithms as this framework can reduce uncertainties of the latter, and provide more robust and reliable projections (Araújo and New, 2007; Thuiller et al., 2009; Grenouillet et al., 2011; Poulos et al., 2012; Fletcher et al., 2016; Bae et al., 2018). Here, we opted for seven different algorithms that have been frequently applied in the field of ecological niche modeling to evolve the basal models required by the ensemble model of the present study. These algorithms were Generalized linear model (GLM), Generalized boosting model (GBM), Generalized additive model (GAM), Flexible discriminant analysis (FDA), Artificial neural network (ANN), Multiple adaptive regression splines (MARS), and Random forest (RF). All these algorithms were fitted using the default settings on biomod2-platform (Thuiller et al., 2016) in an open-source statistical software R 3.5.0 (R Core Team, 2018), as these configurations are suitable for most scenarios (Li et al., 2013; Zhang et al., 2019). In fitting these algorithms, a random 70% of the records (n = 757) were chosen as the training set to calibrate these algorithms and the remaining 30% (n = 325) were withheld for evaluating these algorithms’ performance, i.e. predictive power (Thuiller et al., 2016; Molloy et al., 2017). This process was replicated 10 times to account for individual algorithms variabilities, avoid bias from the dataset split, and add rigour to the results (Molloy et al., 2017; Singer et al., 2017; Bae et al., 2018). Thus, 10 different basal models were yielded for each algorithm. Basal models and subsequent ensemble model performances and idiosyncrasies were assessed by the scores of the area under the curve (AUC) of receiver operating characteristic (ROC), which has the advantage of being prevalent and threshold-independent (Fielding and Bell, 1997; Franklin, 2010; Duque-Lazo et al., 2018). According to Swets (1988) and Duan et al. (2014), this metric ranges from 0 to

---

Please cite this article as: X. Dong, T. Ju, G. Grenouillet et al., Spatial pattern and determinants of global invasion risk of an invasive species, sharpbelly *Hemiculter leucisculus* (Basilewsky, 1855), Science of the Total Environment, https://doi.org/10.1016/j.scitotenv.2019.134661
Table 1
Description of alternative abiotic factors used in the present study. Variables with bold font were retained for modeling procedures.

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Code</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>WorldClim</td>
<td>Mean Annual Temperature</td>
<td>BIO1</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Mean Diurnal Range (Mean of monthly (max temp - min temp))</td>
<td>BIO2</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Isothermality (BIO2/BIO7) (* 100)</td>
<td>BIO3</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Temperature Seasonality (standard deviation *100)</td>
<td>BIO4</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Max Temperature of Warmest Month</td>
<td>BIO5</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Min Temperature of Coldest Month</td>
<td>BIO6</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Temperature Annual Range (BIO5-BIO6)</td>
<td>BIO7</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Mean Temperature of Wettest Quarter</td>
<td>BIO8</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Mean Temperature of Driest Quarter</td>
<td>BIO9</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Mean Temperature of Warmest Quarter</td>
<td>BIO10</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Mean Temperature of Coldest Quarter</td>
<td>BIO11</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Annual Precipitation</td>
<td>BIO12</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Precipitation of Wettest Month</td>
<td>BIO13</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Precipitation of Driest Month</td>
<td>BIO14</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Precipitation Seasonality (Coefficient of Variation)</td>
<td>BIO15</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Precipitation of Wettest Quarter</td>
<td>BIO16</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Precipitation of Driest Quarter</td>
<td>BIO17</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Precipitation of Warmest Quarter</td>
<td>BIO18</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>WorldClim</td>
<td>Precipitation of Coldest Quarter</td>
<td>BIO19</td>
<td><a href="http://www.worldclim.org/bioclim">http://www.worldclim.org/bioclim</a></td>
</tr>
<tr>
<td>SEDAC</td>
<td>Global Human Influence Index</td>
<td>HII</td>
<td><a href="http://sedac.ciesin.columbia.edu/data/sets/browse">http://sedac.ciesin.columbia.edu/data/sets/browse</a></td>
</tr>
<tr>
<td>SEDAC</td>
<td>Global Reservoir and Dam</td>
<td>ResDam</td>
<td><a href="http://sedac.ciesin.columbia.edu/data/sets/browse">http://sedac.ciesin.columbia.edu/data/sets/browse</a></td>
</tr>
</tbody>
</table>

*Note: SEDAC is the abbreviation of socioeconomic data and applications center.

1 and the evaluation criteria are as follows: poor model (0.5–0.7), fair model (0.7–0.9), and excellent model (0.9–1). Following this standard, only basal models with enough discriminatory capacity (i.e. AUC score greater than 0.9) were finally retained to develop the ensemble model as this can ensure that our ensemble model has comparatively the strongest power to predict the occurrence probability of *H. leucisculus* across the world (Zhang et al., 2019).

Specifically, we built our ensemble model using a simple weighted approach. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Predictive model performance

The predictive power of the seven modeling algorithms was compared using the area under the curve (AUC) of the receiver-operating characteristic (ROC). GLM, GBM, GAM, FDA, ANN, MARS, RF, and Enwm are Generalized linear model, Generalized boosting model, Generalized additive model, Flexible discriminant analysis, Artificial neural network, Multiple adaptive regression splines, Random forest, and mean-weighted Ensemble model, respectively. Blue solid circles represent the AUC scores of the 10 repetitions of the seven different modeling algorithms used in the present study. Numerical values in red are the median of the AUC scores of each approach. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.) The predictive performance of the seven modeling algorithms was consistently excellent (no AUC scores less than 0.9) though there existed some variation among the 10 basal models within the same algorithm (Fig. 2). Based on the median of the AUC scores, RF was the most predictive algorithm (AUC = 0.955), followed by GBM (0.951), GAM (0.947), MARS (0.944), FDA (0.942), ANN (0.941), and GLM (0.934). As expected, the ensemble model surpassed all of the individual algorithms, with the highest AUC score of 0.972.

3.2. Variables importance and response curves

Of the six environmental variables, HII (mean ± SE = 0.329 ± 0.030), BIO18 (0.267 ± 0.026), BIO15 (0.080 ± 0.009), and BIO8 (0.040 ± 0.015) were the four most influential factors to the occurrence probability of *H. leucisculus*, whereas ResDam (0.010 ± 0.004)
played a less pivotal role (Fig. 3). The shapes of the response curves of different algorithms for the two most important variables were similar to each other (Fig. 4 A, B). More specifically, the occurrence probability of *H. leucisculus* showed a positive logistic response and a single-valley curve to HII and BIO18, respectively (Fig. 4 A, B). In contrast, there did not exist a consensus cross-algorithm pattern between the occurrence probability of *H. leucisculus* and BIO15, as well as BIO8 (Fig. 4 C, D).

#### 3.3. Spatial patterns of global habitat-suitability and invasion risk

Almost all invasive and native presence records of *H. leucisculus* were successfully projected by our ensemble model (Fig. 5 A). Outputs from this model also suggested that *H. leucisculus* could occupy a broader extent than it has been documented across the world. More specifically, several areas of Midwest, West Coast, and Florida of USA, Mexico, Central America, east-central part and northwest coast of South America, southern part and northwest coast of Africa, Central, Western, and Southern Europe, Southeast Asia, southern foot of the Himalayas, east coast of Australia and southeast of Papua New Guinea showed higher habitat-suitability for *H. leucisculus* though no relevant records have been made yet in these regions (Fig. 5A). While taking introduction likelihood into account (i.e. considering the invasion of *H. leucisculus* as a process) rather than just habitat-suitability, we found that some portions of Midwest and Florida of USA, Mexico, Honduras, Guatemala, Salvador, Cuba, Jamaica, Dominica, southeastern Brazil, Ecuador, Madagascar, Central, Western, and Southern Europe, Turkey, Southeast Asia, southern foot of the Himalayas as well as Japan possessed higher risk of being invaded by *H. leucisculus* in the future (Fig. 5B). In contrast, although Haiti, Nicaragua, Costa Rica, Panama, Central South America, southern part and northwest coast of Africa, Southwestern Asia (most territories of eastern Iran, Afghanistan and northwestern Pakistan), southeast of Papua New Guinea as well as east coast of Australia presented higher habitat-suitability, these regions harbored lower risk of being invaded by *H. leucisculus* in the future (Fig. 5B).

### 4. Discussion

#### 4.1. Spatial patterns of global habitat-suitability and invasion risk

In our study, the global habitat-suitability of invasive *H. leucisculus* was projected by using an ensemble forecasting model since such a model possessed the strongest generalization ability and, consequently is expected to provide the most precise localization of the regions which are suitable for this species. The results indicated that our ensemble model not only successfully identified almost all sharpbelly’s presence records used in this study and those suitable areas in the conterminous USA (i.e. northeast of the Mississippi river basin, West Coast, and Florida Peninsula) anticipated by two previous studies (Chen, 2008; Sanders et al., 2014), but also mined a considerable amount of new habitats suit-
able for *H. leuciscus* that have not yet been detected formerly, such as Central America, Cuba Island, Hispaniola Island, Brazilian Plateau, Circum-Mediterranean, as well as central and southeastern Africa (including Madagascar Island). Despite this, it is particularly notable that our predictions for the conterminous USA are apparently inconsistent with the anterior works. More specifically, our assessment did not support the whole east or large portions of the west of the conterminous USA were inhabitable for *H. leuciscus*, whereas Chen (2008) and Sanders et al. (2014) prognosticated they were, respectively. Of course, the discrepancies in predictions among different studies can be attributed to several reasons, for instance, the use of different modeling algorithms (Pearson et al., 2006; Araújo and New, 2007) and/or data sources (Steiner et al., 2008; Poulos et al., 2012; Verbruggen et al., 2013). However, given the current study (1) adopted an ensemble model yet both Chen (2008) and Sanders et al. (2014) only used an individual modeling algorithm; (2) calibrated the models with presence records from both native and invasive ranges while Chen (2008) simply considered that from native range; (3) included more constraints which may influence the species’ distribution than Sanders et al. (2014) (21 versus 16); (4) assembled a larger dataset than both Chen (2008) and Sanders et al. (2014), it is logical to conclude that our results for the conterminous USA or even the entire world are more accurate and reliable.

More than predicting the worldwide habitat-suitability, the present study also created a global invasion risk map for this invasive species, which differed from many previous SDMs studies. Compared to the global habitat-suitability, the invasion risk map generally focalizes smaller geographical ranges where both species’ habitat-suitability and introduction likelihood are high, and hence informs the hotspots possessing higher invasion risk (Hulme, 2009; Fletcher et al., 2016; Banha et al., 2017). The risk map revealed that albeit several areas were exceptions (e.g. Haiti, Panama, and southeast of Papua New Guinea), most of the regions inhabitable for this pest were concurrently vulnerable to being invaded in the future, for instance, Brazilian Plateau, Southeast Asia, and Japan where this species was first recorded recently in Okayama Prefecture (Nitta et al., 2017). Among all these areas with higher invasion risk, it needs to be emphasized that our prediction for Turkey was highly consistent with two previous assessments for the same country (Tarkan et al., 2014; Tarkan et al., 2017), which in turn also corroborates that our risk map has a good performance.

### 4.2. Influential factors and their response curves

HII was ranked as the most relevant variable to sharpbells’ global invasion and all algorithms indicated that the invasion risk of *H. leuciscus* exhibited a positive logistic response to HII: the higher the level of human disturbance is, the greater the invasion risk of *H. leuciscus* is, which was also observed in many other taxa, including aquatic plants, birds, and insects (Gallardo and Aldridge, 2013; Obenauer et al., 2017; Cardador and Blackburn, 2019). This is not surprising because on the one hand, disturbed habitats that are commonly characterized by simplified communities, little competition or predation, and abundant organic matter, are usually more vulnerable to invasion (Lozon and MacIsaac, 1997). On the other hand, *H. leuciscus* is an ecological generalist (Chen, 2008) and has remarkable tolerance to a wide range of environmental conditions (e.g. thermal and water pollution) (Townsend et al., 2003; Thinh et al., 2012; Mousavi-Sabet et al., 2019). However, by contrast, ResDam contributed the least to sharpbells’ invasion success although we are convinced that it exerts a huge influence on aquatic ecosystems, possibly because this element has been partially accounted for within HII.

Apart from human-related features, the bioclimatic variables (BIO18, BIO15, BIO8, and BIO19) were also indispensable in assessing sharpbells’ global invasion risk. For instance, BIO18 (i.e. precipitation of warmest quarter) played a key role in determining whether *H. leuciscus* can make a successful future invasion in India. It is well known that India is under intense human disturbance, but our risk map showed that the invasion risk of *H. leuciscus* there is not very high. One reasonable explanation is that BIO18, which affects species seasonal distributions (O’Donnell...
and Ignizio, 2012), is at a relatively low level in this country (BIO18 = 293.74 mm). As for why a trough occurred in the response curves of BIO18, two previous studies have provided a plausible explanation: at least one spatial cluster of *H. leuciscus*, which came from the areas where precipitation is extremely insufficient but greatly depended on the rainfall of the warmest quarter of the year, was entailed in our presence dataset (Stelmashchuck and Stadnichenko, 2011; Tytar and Makarova, 2015).

4.3. Strategies to control further expansion

Numerous studies have suggested that one of the major problems caused by aquatic invasive species is that they are difficult or impossible to be eradicated once established a viable population in a new habitat (Williams and Meffe, 2000; Zambrano et al., 2006; Chen, 2008). In addition, eradication tends to be cost-intensive and can only eliminate or attenuate further harm rather than restore the invaded systems to its pristine conditions (Myers et al., 2000). Hence, the best way to minimize the impairments by aquatic invasive species is prevention or preclusion since such a strategy is eco-friendly, cheaper, and more effective (Leung et al., 2002; Hulme, 2006; Patimar, 2007). Unquestionably, this approach is also the first and best line to control *H. leuciscus* expansion. To be specific to application-level, some targeted actions can be taken to preclude the torrent of new introductions of this fish: (1) prohibiting negligent life-release of *H. leuciscus* into non-native habitats; (2) checking all transported fish seed to exclude possible mingled *H. leuciscus*; (3) educating related practitioners, including fish-farm staff, fishermen, aquarists, and even anglers, to enhance their awareness of the jeopardizing of *H. leuciscus* invasion; (4) abstaining *H. leuciscus* escapes from aquaculture facilities; (5) disposing ships’ ballast water in accordance with some widely recognized international standards, such as D2, G2, and G8 (Basurko and Mesbah, 2011); and (6) reducing human disturbance on those susceptible aquatic ecosystems (i.e. areas with high invasion risk forecasted by the present study). Beyond that, long-term prudent monitoring on those susceptible water bodies is also very necessary as this residuary to early detecting new introductions and taking timely remedial measures when those aforementioned actions fail. According to Reshetnikov and Ficetola (2011), once the preventive measures mentioned-above don’t work, the rest of what we can do mainly includes: (1) isolating public from invaded water bodies; (2) severing the links between invaded and uninvaded water bodies (e.g. establishing electrical barriers); and (3) eradicating new invasive populations by releasing local predatory fish or by using piscicides. For (3), a successful example is known in some carp aquaculture ponds in Iran, where 13 unwanted fishes including *H. leuciscus* were almost completely (88.9%) eliminated by releasing apex predator northern pike *Esox lucius* (Linnaeus, 1758) (Khalaf et al., 2010). Despite this positive experience, reliable and practical methods for preventing invasive fishes from large-scale geographical expansion are still quite lacking at the moment. This means that more water bodies will be probably colonized by *H. leuciscus* before new more effective strategies are found and implemented.

5. Conclusion

In this paper, we used a set of SDMs (including an ensemble model) to forecast the worldwide habitat-suitability and its determinants of an avowed invasive species sharpbelly, and then overlaid the habitat-suitability with a proxy for this fish’s introduction likelihood (entailing propague pressure and dispersal constraints) to create a global invasion risk map. The results reveal that the ensemble model has the best performance in predicting sharpbelly’s habitat-suitability and HI is the most-influential factor that favors the future invasion of exotic sharpbelly. In addition, the risk screening of this invader also suggests that there exist some regions in all continents except Antarctica which are at medium-high future invasion risk. Nonetheless, reliable prevention and management strategies for this pest are currently absent. In light of this, a series of targeted measures were advised to stem or slow this species’ further expansion. Incidentally, as an added value, the approach used in this work might be general enough to be transferred to other invasive species’ risk assessment, especially for those whose dispersal capacities are very limited.

Conflict of interest statement

The authors stated that there is no conflict of interest.

Acknowledgements

We would like to express our heartfelt thanks to those researchers who contributed to the occurrence records of *H. leuciscus*, Lei Shi and Tao Xiang for their assistance in literature retrieval, Jingyao Su for her assistance in graphics, as well as Ruojing Li, Mantang Xiong, and Shuyi Yang for their valuable comments and recommendations, which led to the notable improvement of the earlier manuscript of this paper. This study was financially supported by the National Science and Technology Supporting Program of China (2012BAD25B08); the Earmarked Fund for China Agriculture Research System (CARS-45); the State Key Laboratory of Freshwater Ecology and Biotechnology (2014FB204); and the China Scholarship Council (CSC).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2019.134661.

References
