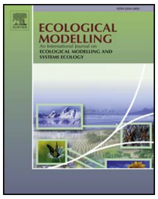
Ecological Modelling 371 (2018) 37–49

Contents lists available at ScienceDirect

Ecological Modelling

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



Forecasting the ongoing invasion of *Lagocephalus sceleratus* in the Mediterranean Sea

Gianpaolo Coro[a,*], Luis Gonzalez Vilas[c], Chiara Magliozzi[d], Anton Ellenbroek[b], Paolo Scarponi[a], Pasquale Pagano[a]

[a] *Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" – CNR, Pisa, Italy*
[b] *Food and Agriculture Organization of the United Nations (FAO), Rome, Italy*
[c] *University of Vigo, Pontevedra, Spain*
[d] *Cranfield Water Science Institute, Cranfield University, Cranfield, United Kingdom*

ARTICLE INFO

Keywords:
Ecological niche modelling
AquaMaps
Artificial Neural Networks
Invasive species
Maximum Entropy
Support Vector Machines
Lagocephalus sceleratus
Species distribution models

ABSTRACT

Invasive species from the Suez Canal, also named "Lessepsian species", often have an ecological and financial impact on marine life, fisheries, human well-being and health in the Mediterranean Sea. Among these, the silver-cheeked toad-fish *Lagocephalus sceleratus* (Gmelin, 1789) has rapidly colonised the eastern Mediterranean basin and is currently moving westwards. This pufferfish has a highly opportunistic behaviour, it attacks fish captured in nets and lines and seriously damages fishing gears and catch. It is a highly-toxic species with no immediate economic value for the Mediterranean market, although it currently represents 4% of the weight of the total artisanal catches. Consequently, the possible effects on Mediterranean fisheries and health require to enhance our understanding about the future geographical distribution of this pufferfish in the whole basin.

In this paper, an overall habitat suitability map and an effective geographical spread map for *L. sceleratus* at Mediterranean scale are produced by using cloud computing-based algorithms to merge seven machine learning approaches. Further, the potential impact of the species is estimated for several Mediterranean Sea subdivisions: The major fishing areas of the Food and Agriculture Organization of the United Nations, the Economic Exclusive Zones, and the subdivisions of the General Fisheries Commission for the Mediterranean Sea. Our results suggest that without an intervention, *L. sceleratus* will continue its rapid spread and will likely have a high impact on fisheries. The presented method is generic and can be applied to other invasive species. It is based on an Open Science approach and all processes are freely available as Web services.

1. Introduction

The number of species in the Mediterranean Sea arriving through the Suez Canal (also named "Lessepsian" species) continues to increase (Nader et al., 2012; Golani, 2010). Recent studies estimate that more than 5% of the marine species are non-native and 13.5% are invasive, including fish, invertebrates, and macrophytes (Galil, 2009; Zenetos, 2010; Fricke et al., 2015; Zenetos et al., 2015; Golani, 2010). These "invasive species" (Shine et al., 2000) settle in the new habitat, increase in number, and spread in the area, potentially threatening native biological diversity (Galil et al., 2015; Coll et al., 2010) and economy (Galil, 2008). Thus, they require particular effort by supervising organisations in order to monitor and predict their spread.

Among these species, the silver-cheeked toad-fish *Lagocephalus sceleratus* (Gmelin, 1789) is of particular concern. The first reliable records in the Mediterranean Sea date back to 2003, but the number of observations has rapidly grown so that it is considered one of the fastest expanding invasive species in the basin (Akyol et al., 2005; Peristeraki et al., 2006). It owes its success to the high growth and reproduction rate, the lack of natural predators, the ability to exploit food resources, and the capacity to tolerate a wide range of environmental conditions (Yaglioglu et al., 2011).

It has a skin without scales, with dark spots on top, and lateral silver bands. This species is common in the Red Sea, belongs to the *Tetraodontidae* family, is extremely poisonous, and can be lethal to humans if eaten, due to high level of Tetrodotoxin neurotoxin (TTX) present in several organs (e.g. the liver) and excreted from the skin as a repellent after swelling (Yaglioglu et al., 2011; Nader et al., 2012). It usually prefers shallow waters and medium-high water temperature, which is correlated to faster TTX uptake. Thus, climate change could be

---

[*] Corresponding author.

E-mail addresses: coro@isti.cnr.it (G. Coro), luisgv@uvigo.es (L.G. Vilas), c.magliozzi@cranfield.ac.uk (C. Magliozzi), anton.ellenbroek@fao.org (A. Ellenbroek), paolo.scarponi@isti.cnr.it (P. Scarponi), pagano@isti.cnr.it (P. Pagano).

https://doi.org/10.1016/j.ecolmodel.2018.01.007
Received 15 September 2017; Received in revised form 25 December 2017; Accepted 12 January 2018

beneficial for this species, particularly in the Mediterranean Sea (Nader et al., 2012).

Scientific studies have estimated the potential impact of *L. sceleratus* on economic and human health in the eastern Mediterranean Sea (Ünal et al., 2015, 2017). In this region this is now one of the most important species (in biomass) on *Posidonia oceanica* meadows, being a major problem to artisanal fisheries considering that it damages fishing gears (e.g. nets and lines) and predates heavily on local stocks of squids and octopuses (Kalogirou et al., 2010). However, these studies do not report definitive ecological and economic future impact assessments and usually involve more qualitative than quantitative predictions. Overall, they indicate that the fish currently represents the 4% of the weight of the total artisanal catches (Nader et al., 2012) and has already negatively impacted the economy of some Mediterranean countries (Ünal et al., 2017). Also, since 2003 several episodes of death and serious illness have been recorded after fish consumption, since fishermen and other people usually cannot identify this relatively new species (Bentur et al., 2008; Kheifets et al., 2012).

This scenario calls for priority actions to prevent, detect and possibly eradicate *L. sceleratus* (Zenetos et al., 2016), especially considering that the Suez Canal capacity is being enlarged (Searight, 2016) and climate change is facilitating the invasion (Galil et al., 2014; ICES, 2007; FAO, 2007). One approach could be to use selective fishing especially on big individuals and localised precautionary actions in those areas where the pufferfish will possibly move and settle in the next years (Ünal et al., 2017). Therefore, a map of the ongoing invasion pattern could guide the development of preventive and corrective actions (Zenetos et al., 2015, 2016) and could also help filling a gap between research and management about this fish (Ünal et al., 2015).

In the past decade, there has been a growing interest in the application of ecological niche models (ENMs) to predict the distribution of invasive species (Guisan et al., 2014). Different approaches have been used based on the evaluation of the niche differences between a species' native region and the invaded region (Peterson, 2003; Barbosa et al., 2012; Leidenberger et al., 2015). In some cases, these approaches also take into account how climate change facilitates the species' spread into the invaded region (Sax et al., 2007; Thuiller et al., 2005). ENMs-based approaches to invasive species modelling use a varied range of models, including envelope-based (Sutherst, 2000; Jeschke and Strayer, 2008), statistical (Ficetola et al., 2007; Bidegain et al., 2015), and machine learning models (Peterson and Robins, 2003). Most of these models estimate an association between a species' presence and a number of environmental parameters, and produce a probability distribution. This is then projected onto a certain area (over time) to get a dynamic visualisation of the invasion (Mellin et al., 2016; Carlos-Júnior et al., 2015). The most used ENM in this context is the "Genetic Algorithm for Rule-set Production", GARP (Stockwell, 1999), which uses a machine learning approach (Peterson and Vieglais, 2001; Ganeshaiah et al., 2003; Sanchez-Flores et al., 2008; Underwood et al., 2004). Another widely used model is the Maximum Entropy presence-only model (Ficetola et al., 2007; West et al., 2016), whereas presence-absence models, e.g. Artificial Neural Networks (Kulhanek et al., 2011) and Support Vector Machines (Pouteau et al., 2011; Sadeghi et al., 2012), are less frequent because of the scarcity of reliable absence data. Usually, alternative ENMs-based approaches have complementary features which capture different characteristics of a species' invasion (Elith and Graham, 2009). Thus, it is common to compare or merge the output of different models in order to produce a final spread estimate (Castelar et al., 2015; Farashi and Najafabadi, 2015; Padalia et al., 2014; Sobek-Swant et al., 2012).

Most of the cited studies assume climate niche conservatism (Pearman et al., 2008; Peterson and Vieglais, 2001), i.e. the ENM calculated using data from the species' native environment is supposed to successfully predict invasion in exotic areas (Petitpierre et al., 2012; Strubbe et al., 2013; Castelar et al., 2015). However, other works have highlighted that the climatic niche may change during the invasion (Broennimann et al., 2007; Lauzeral et al., 2011), which can overturn the conservatism assumption (Shabani and Kumar, 2015). Further, ENMs usually do not account for the effects of species interactions and possible geographical dispersal limitations, thus the results of approaches purely based on ENMs should be interpreted and used with caution (Sax et al., 2007).

In this paper, an approach is proposed to estimate the potential ecological niche and the potential geographical distribution of *L.sceleratus* in the Mediterranean Sea: its spread is predicted up to a stable distribution, based on the algorithmic merge of the output of seven machine-learning models that each estimate the potential niche or habitat suitability. Further, an effective geographical distribution is estimated and a potential impact indicator is produced for different subdivisions of the Mediterranean Sea, which include the major fishing areas of the Food and Agriculture Organization of the United Nations (FAO), the Economic Exclusive Zones (Attard, 1987), and the general subdivisions of the General Fisheries Commission for the Mediterranean Sea. The predictive value of the generated geographical distribution is assessed in a comparison with real observation records in the Mediterranean Sea and with a *dynamic* model that simulates the spread of the pufferfish over time. Our analysis follows an Open Science approach (Hey et al., 2009), and all models are available as-a-Service under a representational standard. Every step can be reproduced, repeated and reused for other invasive species, or with different ancillary data.

2. Material and methods

In this section, the used technology and data (Section 2.1) and the baseline models that constitute our method are described (Section 2.2). Further, our method to estimate the *geographical reachability* distribution of *L. sceleratus* is presented (Sections 2.3 and 2.4). A *dynamic* model is also described (Section 2.5), which is used in Section 3 as a reference to assess the performance of our method. Moreover, the metrics used to calculate the models' performance, their mutual similarities, and a risk indicator for the Mediterranean Sea are presented (Section 2.6). Finally, the generality of our method and its applicability to other invasive species is discussed (Section 2.7).

2.1. Technology and data

2.1.1. Computational and data access platform

Our method requires the training of machine learning models with a large set of alternative parametrisations. The goal is to find the "optimal" model, i.e. the model with the best performance on a test set. The experiment reported in this paper required to train ~150,000 parametrisations, which was very time-consuming and computationally demanding. To overcome this, a cloud computing platform was used to train many alternative parametrisations of a given machine learning model at the same time. In particular, the *gCube DataMiner* open-source system[1] (Coro et al., 2017) was used for this Big Data processing and to interoperate with the services of the D4Science distributed e-Infrastructure (National Research Council of Italy, 2016). D4Science facilitates data preparation and processing, and fosters collaboration among scientists according to Open Science paradigms (Hey et al., 2009). This set-up includes (Assante et al., 2016): (i) collaborative experimentation spaces, where processes can be re-executed and parametrised several times by others, (ii) services for data sharing between users, and (iii) application of standards for data and processes representation. The DataMiner represents and stores all the trained models and their respective parametrisations in a standard and exportable ontological format (Prov-O, Lebo et al., 2013), which summarises the set of input/output data and metadata that enable any other authorised user to

---

[1] Freely accessible and usable after registration at https://services.d4science.org/group/biodiversitylab/data-miner.