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.

reproduce and repeat the experiment (*provenance* of the computation). All models are published in the D4Science e-Infrastructure as open-source and free-to-use Web services under the Web Processing Service standard of the Open Geospatial Consortium (WPS, Schut and Whiteside, 2007). This maximises re-usability, because WPS standardises the representation of input, parameters, and output through XML descriptions of their types and expected contents.

DataMiner parallelises models training on a network of 100 machines, choosing the best computational configuration among a range of powerful multi-core virtual machines (Ubuntu 14.04.5 LTS ×86 64 with 16 virtual CPUs, 16 GB of random access memory, 100 GB of disk) and more "lightweight" virtual machines (Ubuntu 14.04.5 LTS ×86 64 with 2 virtual CPUs, 2 GB of random access memory, 10 GB of disk). Apart from its high performance (Coro et al., 2017), this platform was selected because we wanted our method to be used by other scientists, possibly on other invasive species, through the repeatability, reproducibility, and re-usability of each step of our method.

2.1.2. Occurrence records

Occurrence records of *L. sceleratus* were retrieved and harmonised through the Species Product Discovery (SPD) service of the D4Science e-Infrastructure[2] (Candela et al., 2015) from biodiversity data collections such as OBIS (Vanden Berghe et al., 2010), GBIF (Lane and Edwards, 2007), and the Catalogue of Life (Wilson, 2003). The SPD attaches additional information to presence coordinates, i.e: the ownership of the observation, its source (e.g. human observation, specimen etc.), and a flag stating if the record underwent expert review. This allows retrieving only data having "good" quality because they were checked by an expert. SPD produced a set of 284 expert-reviewed records for *L. sceleratus*, from its native habitat environment (Fig. 1a). Locations in the Mediterranean Sea were excluded, because this data set was used in model training and could not contain locations in the projection/testing area.

The SPD records for the Mediterranean Sea were combined with records collected from a manual literature review of published articles and grey literature (Fig. 1b). A total of 263 Mediterranean records, sometimes referring to several individuals, was obtained, where the first reliable records dated back to 2003. In order to build up our definitive distribution of *L. sceleratus*, 20% of these data were used, whereas the remaining 80% were used to validate the model (Sections 2.2 and 3). This choice was due to the fact that we wanted to use the Mediterranean observations to evaluate the performance of our models rather than to train the models. In fact, this approach makes the models less sensitive to prior information about the distribution of the species in the invaded region, and it also makes our approach applicable to other species for which few observations are available in the invaded area.

2.1.3. Estimating absence locations

Some models used by our method require to estimate locations of habitat unsuitability of *L. sceleratus*. To this aim, a DataMiner process to estimate pseudo-absence locations was used[3] (Coro et al., 2016a). This process performs a statistical analysis on scientific survey data in the OBIS data collection (Grassle, 2000). It accesses OBIS through REST APIs and uses four input parameters: (i) a time frame for the observations, (ii) a "spatial resolution" for the estimated absence locations, (iii) the geographical area where to estimate the absence locations, and (iv) an "observation frequency threshold". This last parameter is used to retrieve all the surveys hosting species' experts on the vessels, who observed *L. sceleratus* in specific locations with higher monthly frequency than this value. In these surveys, locations where the experts reported only other species than the target one are recorded. By intersecting all the surveys' trails, pseudo-absences are estimated as those locations (i) where no expert ever reported the target species presence, (ii) whose distance is higher or equal to the "spatial resolution" parameter, and (iii) that are non-overlapping with presence locations at this resolution. This process is reliable especially when a high number of survey data is available. In the case of *L. sceleratus*, it estimated 184 absence locations in its known native range (Fig. 1a).

2.1.4. Environmental data

The biotic and abiotic characteristics possibly associated to the *L. sceleratus* preferred habitat were not known *a priori*, although suggestions about some parameters were found in scientific papers (Nader et al., 2012; Yaglioglu et al., 2011). In such uncertainty scenarios, the AquaMaps Consortium advises to use a set of 18 eco-geographical environmental parameters (Table 1), whose combination is likely to be associated to species environmental preferences and thus to suitable habitat (Corsi et al., 2000). These parameters include those indicated by literature studies for *L. sceleratus*, e.g. depth, distance from land, and water temperature etc. Further, the AquaMaps ecological niche model is constrained to this set of 18 features and is part of our method. For these reasons, all the environmental parameters listed in Table 1 were taken into account.

The values of these parameters in the Mediterranean Sea were retrieved through the D4Science geospatial catalogue[4] (Assante et al., 2016), and belong to (i) the AquaMaps Consortium, (ii) the Copernicus Marine Environment monitoring service, and (iii) the National Oceanic and Atmospheric Administration (NOAA). Estimates for the same 18 parameters in 2050 were available too, as produced by the AquaMaps Consortium using the ECHAM model (Roeckner et al., 1992).

Environmental data were attached to both the retrieved presence and pseudo-absence locations of *L. sceleratus* at 0.5° resolution, which is the minimal resolution of all the models used by our method. This operation was realised by means of the DataMiner "occurrence-enrichment" process[5] that associates environmental information to a number of locations, producing a CSV file containing an "enriched" data set.

2.2. Modelling

2.2.1. AquaMaps

The AquaMaps ecological niche models (Kaschner et al., 2006) estimate species habitat suitability under different environmental scenarios. These presence-only models incorporate scientific expert knowledge to account for known biases and limitations of marine species occurrence record data sets (Ready et al., 2010). AquaMaps includes two models to estimate the actual (named *native*) distribution of a species today (2017) and in 2050, and other two models to estimate *potential* habitat suitability in locations where the species has never been observed. These models estimate species' habitat at global scale with 0.5° resolution, calculating the association between the observed locations and a predefined number of environmental variables (Table 1). This association is estimated by multiplying 18 envelope functions, each traced on one environmental variable, and by successively applying mechanistic assumptions (in the form of rule-based algorithms) to produce a species-presence probability distribution. AquaMaps is reliable if compared to other more complex approaches, although its accuracy decreases when expert knowledge is missing (Ready et al., 2010). The AquaMaps native and potential algorithms for the 2050 scenarios use information about expected modifications in the

[2] https://services.d4science.org/group/biodiversitylab/species-discovery.

[3] https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.ABSENCE_GENERATION_FROM_OBIS.

[4] https://services.d4science.org/group/biodiversitylab/geo-visualisation.

[5] https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.OCCURRENCE_ENRICHMENT.

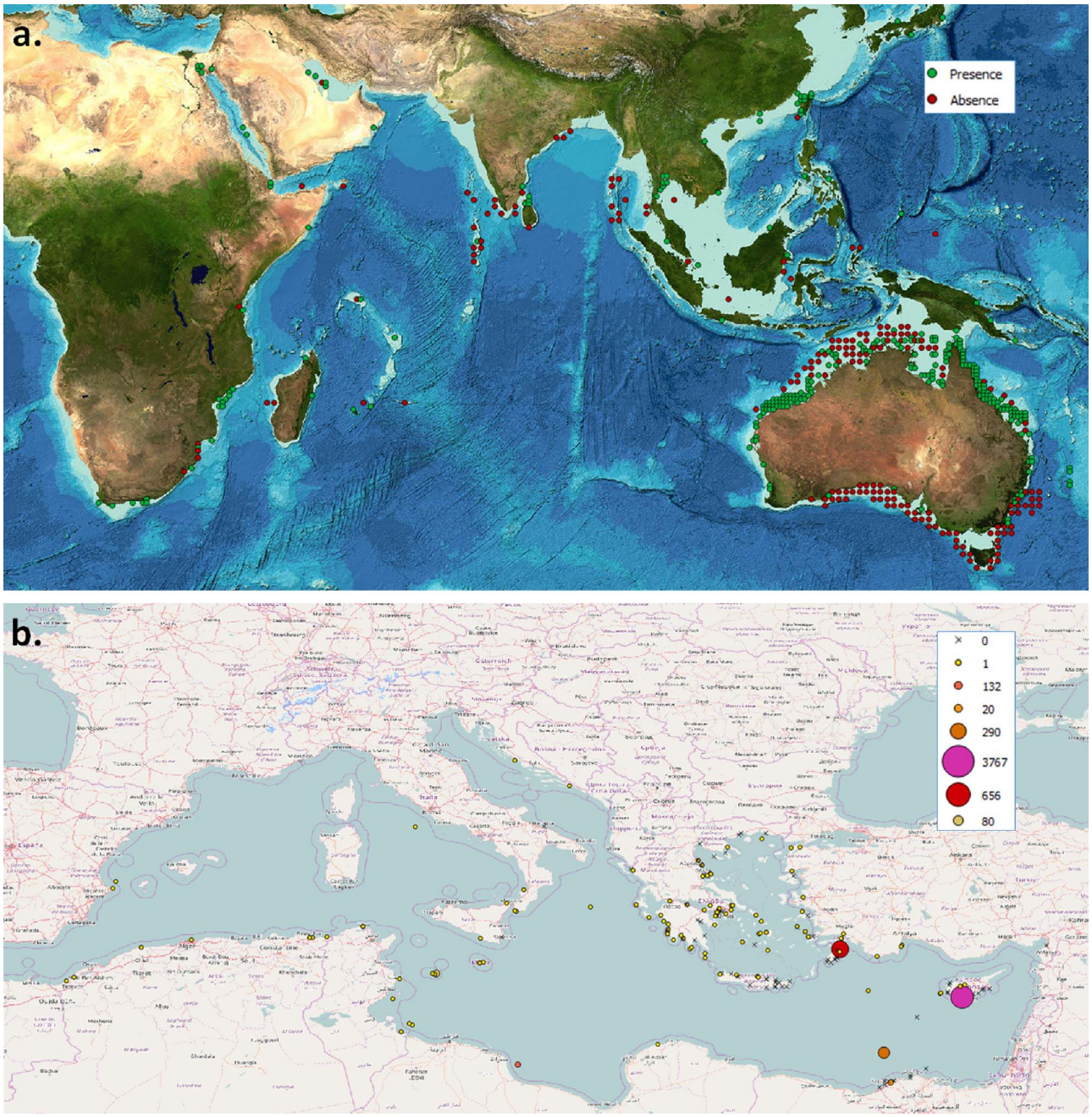

Fig. 1. (a) Presence (brighter points) and pseudo-absence (darker points) locations of *L. sceleratus* in its known inhabited/native environment, used to train our models. (b) Reported observation records in the Mediterranean Sea collected from published articles, grey literature, and from the GBIF data and OBIS providers, with indication of abundance and extension.

global FAO major fishing areas and in the overall change of the oceans water surface level. These two processes rely on environmental variables estimations for 2050 under the IPCC SRES A1B scenario (Nakicenovic and Swart, 2000) of a future of rapid global economic, population, and technological growth, where the average surface temperature increases, the ice concentration decreases and the salinity increases globally but decreases in some locations (Reyes, 2015). An AquaMaps potential habitat suitability model for *L. sceleratus* in 2017[6] was trained using presence data in its native environment and the 18 environmental variables mentioned above, and was projected on the Mediterranean Sea.

[6] https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.generators.AQUAMAPS_SUITABLE.

2.2.2. Artificial Neural Networks

Artificial Neural Networks (ANNs) are machine learning models made up of interconnected digital representations of neurons (Minsky, 1963). These models have been used for long time in many domains (Patterson, 1998), including ecological modelling (Lek and Guégan, 1999; Olden et al., 2004; Lek et al., 1996; Gevrey et al., 2003), because they allow to model non-linear functions between an input vector of Real numbers and an output vector of Real numbers. ANNs can also simulate automatic classifiers (Bishop, 1995) that associate an input vector to one category among several. In Feed-Forward Neural Networks (Bebis and Georgiopoulos, 1994), the digital neurons of an ANN are organised into "layers", where the first layer receives and processes the input vector directly and the last layer produces the output vector, and intermediate layers are named "hidden layers". One layer is fully connected only to the next layer by means of weighted edges, i.e. each

Table 1
The complete set of environmental features used in our method. The values are calculated on 0.5° cells and annual averages are used for time-dependent variables. The two columns on the right side report the features selected by MaxEnt and SVM respectively, as carrying the most important information to assess the habitat of *L. sceleratus*.

Environmental parameters	MaxEnt-selected	SVM-LOF
Mean Depth (m)	×	✓
Maximum Depth (m)	×	×
Minimum Depth (m)	✓	×
Depth Standard Deviation	×	✓
Distance from Land (m)	✓	✓
Ocean Area (m^2)	×	×
Annual Mean Ice Concentration (percentage)	✓	×
Annual Mean Primary Production ($g\,C\,m^{-2}\,yr^{-1}$)	×	×
Annual Mean Sea Surface Temperature (° C)	×	✓
Maximum Annual Sea Surface Temperature (°C)	×	×
Minimum Annual Sea Surface Temperature (°C)	✓	×
Sea Surface Temperature Standard Deviation	×	×
Sea Surface Temperature Range (°C)	×	×
Annual Mean Sea Bottom Temperature (°C)	✓	×
Annual Mean Salinity (PSU)	×	✓
Minimum Salinity (PSU)	×	×
Maximum Salinity (PSU)	✓	×
Annual Mean Bottom Salinity (PSU)	×	×

neuron in one layer has edges only towards neurons in the next layer. An ANN can be trained to simulate a function on known data by means of a learning algorithm (e.g. the "backpropagation", Rumelhart et al., 1986). The training algorithm adjusts the weights of the networks edges to produce expected output on the training data. Thereafter, the ANN is used with known input data that had not been included in the training set, and the performance of the trained model is evaluated (*test* session). The number of hidden layers and neurons of the ANN with the highest performance on the test set (i.e the *best* topology) can be found by running the learning algorithm multiple times and by testing every topology more than once in order to avoid local minima issues (Özesmi et al., 2006). One approach to find the best topology is the "growing" strategy (Bishop, 1995), where neurons and layers are added as far as the error on the training set decreases down to a certain threshold, which is empirically set to avoid overfitting of the ANN to the training data. Although ANNs are powerful models, one disadvantage in using them is that they do not provide the analytical form of the simulated function, thus it is not possible to understand how the input variables are really combined within the network.

For the scopes of this paper, an ANN was trained[7] on the same environmental features used by AquaMaps, extracted at the collected presence and absence records sets of the pufferfish in its native environment. Thus, the ANN had 18 input features, one for each environmental parameter, and one output neuron conceptually associated to a habitat suitability score, ranging between 0 (absence) and 1 (presence). In the training phase of the ANN, the environmental features associated to the presence locations were used as positive cases on which the ANN output was forced to output 1, whereas those associated to absence locations were used as negative cases with ANN output forced to 0. Topologies ranging between one and three hidden layers containing a variable number of neurons were explored using a "growing" strategy, which required training ~100,000 models. The training and testing phases were based on a 80% (train) – 20% (test) cross-validation, since data outside the Mediterranean Sea were used in this phase. Eventually, the best model was identified as an ANN containing one hidden layer with 100 neurons.

[7] https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.modellers.FEED_FORWARD_ANN.

2.2.3. Maximum Entropy

Maximum Entropy (MaxEnt) is a presence-only machine learning model commonly used in ecological modelling (Phillips et al., 2006, 2004; Phillips and Dudik, 2008; Baldwin, 2009; Coro et al., 2015). MaxEnt approximates a probability density function defined on a vector space of environmental features, with the constraint that this function is compliant with predefined mean values at presence locations and that the overall entropy of the probability distribution is maximum (Elith et al., 2011). During the model's training phase, MaxEnt performs a relative maximisation of the entropy function $H = -\sum \pi(\overline{x}) \ln(\pi(\overline{x}))$, defined on the environmental features $\overline{x}$ at the provided presence locations, with respect to the entropy function applied to the features of random points taken all over the area under study (Phillips et al., 2006). Presence points are taken as constraints during this maximisation and the model uses a linear combination of the features as π function, where the coefficients of the combination are changed to reflect the influence of each variable in predicting the distribution of the species. After the training phase, these coefficients can be used to select the most influential environmental parameters given the known presence locations. Thus, MaxEnt can also be used to select the features that carry the highest quantity of information about the species according to the entropy maximisation process. These features can be possibly used into other models to obtain new habitat projections (Coro et al., 2015). One drawback of MaxEnt, is that it is very sensible to bias in the data, thus its performance increases if the presence records are reliable (Elith and Leathwick, 2009).

MaxEnt was trained on the environmental features associated to the occurrence records of *L. sceleratus* in its native environment[8] and was projected on the Mediterranean Sea. Afterwards, the variables having highest association with the presence data of *L. sceleratus* were identified and another MaxEnt model was trained using only these variables. In particular, only variables with a coefficient value in the estimated π function higher than 5% of the maximum coefficient value were used in this model.

2.2.4. Support Vector Machines

Support Vector Machines (SVM, Boser et al., 1992), a machine learning method also used in ecological modelling (Brown et al., 1999; Guo et al., 2005; Drake et al., 2006), can be used to build a binary classifier (Vapnik, 2013; Schölkopf et al., 1999) by projecting the input data onto a higher dimensional, "simpler", features space through a *kernel* function, and then by searching for a linear separation of this space. In most of the applications, this process consists in finding an optimal separation hyperplane that maximises the distance (*margin*) from the closest training instances (*support vectors*). Thus, training a SVM usually requires maximising the *margin* by solving an optimisation problem constrained by linear relations. These constraints may be relaxed allowing some classification error in order to avoid overfitting (Cristianini and Shawe-Taylor, 2000). SVM can also be used to select the input features that carry the highest quantity of information, for example through a leave-one-out (LOF) process that records when the SVM performance decreases during a cross-validation assessment after one of the features is removed in turn (Chang and Lin, 2011; Vilas et al., 2014).

Three binary (presence-absence) SVM training processes, based on a Sequential Minimal Optimisation (SMO) algorithm[9] (Chang and Lin, 2011), were executed on three different environmental features subsets associated to presence and absence locations of *L. sceleratus* in its native range. The first model used the complete set of 18 environmental

[8] https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.MAX_ENT_NICHE_MODELLING.

[9] https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.SUPPORT_VECTOR_MACHINES_MODELLING.

features; the second model used only the environmental variables selected by a leave-one-out process; and the third model used the environmental features selected by the MaxEnt process presented in the previous section. The SVM-selected variables were not used to feed the other models because a LOF selection process is usually beneficial only for a linear binary classifier. All models were configured (through "pairwise coupling", Wu et al., 2004) to give a continuous value between 0 (absence) and 1 (presence) in order to simulate a probability distribution. A weighted SVM training process (Suykens et al., 2002) was used for each SVM to account for imbalance between the number of absence and presence data, and the environmental features were normalised between 0 and 1. As *kernel* function, the Gaussian Radial Basis function (RBF) was adopted because it is often used when the features space is large (Camps-Valls and Bruzzone, 2005) and uses a lower number of initialisation parameters (*gamma* and *C*) with respect to other kernels (Hsu et al., 2003). Using a 10-fold cross-validation on our training set (Chang and Lin, 2011), $gamma = 22$ and $C = 40$ were found optimal values for these parameters. The three binary SVMs were trained and consequently projected[10] on the Mediterranean Sea.

2.3. Merging the habitat models

The baseline ENM models were merged together in order to take advantage of possible complementary indications about the species' habitat distribution. Complementarity between these models was expected because they are complex by construction and their analytical forms are likely to be very different from each other, due to the different training processes used and the involvement of random variables.

In order to obtain one overall merged probability distribution for the Mediterranean Sea at 0.5° resolution, the normalised sum of the distributions was used, i.e.:

$$\begin{aligned} P_{\text{Overall Habitat}}(\bar{x}) \propto\ & P_{\text{AquaMaps}}(\bar{x}) + P_{\text{ANN}}(\bar{x}) + P_{\text{MaxEnt}}(\bar{x}) \\ & + P_{\text{MaxEnt with ME-selected variables}}(\bar{x}) + P_{\text{SVM}}(\bar{x}) \\ & + P_{\text{SVM with ME-selected variables}}(\bar{x}) \\ & + P_{\text{SVM with SVM-LOF variables}}(\bar{x}) \end{aligned}$$

where $P_{\text{Overall Habitat}}(\bar{x})$ is the overall merged probability function of habitat suitability defined on the features domain $X = \{\bar{x}\}$, normalised between 0 and 1. The involved distributions are those from the previous sections, i.e. respectively AquaMaps, ANN, MaxEnt, MaxEnt using MaxEnt-selected variables, SVM using all the environmental variables, SVM using MaxEnt-selected variables, and SVM using variables selected with the leave-one-out process. The formula above gives the same weight to all models, using the rationale that high suitability in a certain location could be due even just to one distribution estimating high probability there. On the contrary, using a multiplication between the functions would have led to very low or zero probability in most of the locations, since there are only few locations where all models estimate non-zero probability at the same time. The resulting function represents the overall estimated habitat suitability of *L. sceleratus* at 0.5° resolution in the Mediterranean Sea. However, a suitable habitat does not necessarily indicate a *reachable* habitat, because of possible geographical, survival, and reproduction barriers. Thus, the merged distribution needs to be converted into a more realistic *geographical reachability* map.

2.4. Estimating the actual geographical reachability distribution

Explicitly modelling all constraints that prevent *L. sceleratus* to colonise suitable habitat is challenging. Our model starts by using real observations in the Mediterranean Sea as references. It selects 20% of the species records in the Mediterranean Sea (Fig. 1b) as anchor points to trace several Gaussian functions, each centred on one recording site, having maximum height equal to 1 and value decreasing with distance from the point. The decrease rate of the functions depends on their widths and thus on their standard deviation. One standard deviation was used for all the function and was set to the maximum distance between a 0.5° cell and all recording sites, which was calculated to be ~17°. This means that also the farthest points in the Mediterranean Sea could be reached with a certain probability, because these would have a Gaussian value higher than 0.5 by construction. In summary, the Gaussian functions represent a potential movement of the fish from one record site to another location and their decreasing trends simulate inertia to reach far locations, possibly due to heterogeneous barriers.

Finally, the overall *geographical reachability* distribution was estimated by combining the Gaussian functions and the overall habitat suitability distribution. In particular, each habitat suitability value at the 0.5° locations was multiplied by the value of a Gaussian function centred on the closest record location, providing the distance from this point as an argument to the function (Fig. 3):

$$\begin{aligned} P_{\text{Geographical Distribution}}(\bar{x}) = & P_{\text{Overall Habitat}}(\bar{x}) * \text{Gaussian}_{\mu=0,\sigma=17}(d) \\ & \text{where } d = \min(\text{distance}(\text{location}(\bar{x}), \text{observations})) \end{aligned}$$

This distribution combines the species' habitat suitability with the inertia in moving from real records. It indicates high presence probability for a real observation only if also habitat suitability is high in that location. Indeed, equal weight was given to Gaussian functions and habitat suitability because they carry equally important information. In fact, a real observation refers to a sure event at certain time instant that could be even temporary or occurred by chance, whereas habitat suitability indicates environmental conditions suited for the species to persist over time in that location. Further, also locations far from these observations are allowed to have high presence probability if habitat suitability is high, because locations as distant as the full width at half maximum of the Gaussian function 1.177σ (i.e. ~20°) have 0.5 value. The resulting half-degree distribution is referred to as the *geographical reachability* distribution.

2.5. Benchmark dynamic model

The stability and the performance of the *geographical reachability* distribution was checked against an alternative model (referred to as the *dynamic* distribution). This model is more compliant with common approaches to invasive species modelling (Section 1), where the native habitat of a species is allowed to evolve in time up to a convergence status.

In particular, an iterative process was created which goes through the following steps:

1. Produce a "native" habitat map according to a niche modelling algorithm;
2. Apply Gaussian distance function weighting, using a σ value representing the geographical extent that can be reached at each step of the dynamic evolution;
3. Apply a 0.5 cut-off threshold to the *geographical reachability* distribution to produce new pseudo-presence locations;
4. Produce and project a new niche model after adding the new pseudo-presence locations to the training set;
5. Start a new cycle from point 2;
6. End the process when the distribution does not change after four cycles.

This loop requires two parameters to be fixed before the start: the niche modelling algorithm to use and the σ of the Gaussian function. To find the best combination of these parameters, the accuracy of all the baseline models was calculated at the variation of σ on the 20% of the observation records in the Mediterranean Sea. The combination of

[10] https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.transducerers.SUPPORT_VECTOR_MACHINES_PROJECTOR.

$\sigma = 2$ and the SVM using variables selected with the leave-one-out process provided the best parametrisation of the loop. This σ value indicates that locations up to 2.35° (equal to 1.177σ, i.e. ~200 km) from the observations at one step can become new pseudo-presence locations in the next step. The *dynamic* model converges to a stable distribution after 26 steps[11] (Fig. 4d).

Each step of the loop could be interpreted as a time interval, but it is not easy to establish how much time corresponds to a maximum movement of about 2° of *L. sceleratus*. However, based on the average yearly distances between first observations in the Mediterranean Sea, a ~2° distance may correspond to the movement of the pufferfish in one year.

2.6. Agreement, performance, and impact measurements

In order to numerically estimate the *agreement* between the trained models, an automatic maps comparison process was used[12] (Coro et al., 2014, 2016b), which calculates the agreement between two maps at a time. A 0.2 probability threshold was used on the maps to indicate when two models assessed together that a species was present (or absent) in a certain comparison location. This threshold was selected as the one resulting in the highest coverage by all models of the 20% of the complete set of expert-reviewed pufferfish Mediterranean records (Section 2.1.2). Thus, the threshold represents a sensibly non-zero probability of habitat suitability or presence. The result of this maps comparison process is a matrix reporting pairwise agreement percentages.

In order to estimate the *accuracies* of the models, 80% of the expert-reviewed pufferfish Mediterranean records were used. In particular, a model's accuracy was calculated as the percentage of records on which it reported sensibly non-zero probability (> 0.2).

Finally, an *impact* (or *risk*) indicator for a Mediterranean area was calculated, similar to other studies (McGeoch et al., 2006), as the normalised density of sensibly non-zero probability locations falling into the area.

2.7. Applicability to other species

The presented method is in principle applicable to other invasive species, in order to monitor their spread in any given region. In fact, our general process can be summarised in a number of steps that are independent of the selected species:

1. Retrieve presence data for the species in its native habitat;
2. Estimate absence locations for the species in its native habitat;
3. Enrich the presence/absence data set with environmental variables information;
4. Train different habitat suitability models, using the enriched data set as a training set;
5. Merge the models using a normalised sum of their projections on the invaded area;
6. Retrieve observation records in the invaded area;
7. Produce a *geographical reachability* distribution by multiplying the merged habitat model for a set of distance-based Gaussian functions, each centred on the observations in the invaded area;
8. Assess the resulting model's performance with independent observations in the invaded area;
9. Project the model onto official subdivisions of the invaded area, in order to estimate impact indicators.

This process is complex and requires access to species data, computational facilities, and storage of intermediate and final results. Thus, it requires an e-Infrastructure oriented to Open Science that guarantees fast data retrieval, direct feeding of models with data, and fast estimation of the best models.

3. Results and discussion

In this section, the projections of the models developed for the pufferfish are presented. The models' performance is quantitatively assessed on known Mediterranean records of the species, and the discrepancies between the *geographical reachability* and the *dynamic* distributions are highlighted and checked against previous studies. Finally, an impact indicator is reported for different subdivisions of the Mediterranean Sea.

3.1. Models projections

The projection of the AquaMaps model on the Mediterranean Sea predicts suitable habitat especially in coastal areas (Fig. 2a). Likewise, in the ANN model coastal proximity influences high habitat suitability but the resulting distribution is different from the AquaMaps one due to a more complex combination of environmental variables in the Neural Network (Fig. 2b). The MaxEnt model using all environmental variables estimates high suitability for the eastern Mediterranean basin (Fig. 2c). The variables having highest association with the presence data according to MaxEnt are reported in Table 1. The MaxEnt model trained only with these variables agrees with the previous MaxEnt model about the habitat suitability of the eastern basin, and additionally reports westwards suitability especially around the southern Italian coasts (Fig. 2d). Discrepancies between these two MaxEnt models are visible off the Libyan and Turkey coasts and in the centre of the Mediterranean Sea. The three SVM models are different from each other. The model trained with all environmental variable reports high suitability in the Black Sea (Fig. 2e); the model using environmental variables selected by a leave-one-out process (Table 1), estimates high suitability in a large part of the Mediterranean Sea except in the northern Adriatic Sea and in the southern France coasts (Fig. 2f); the model using the environmental variables selected by the MaxEnt process, estimates high suitability in Greece and Turkey coasts and in the western Mediterranean coasts around the strait of Gibraltar (Fig. 2g).

It can be visually recognised that the trained baseline models predict high probability in complementary locations (Fig. 2). This is confirmed by the calculation of the agreements between the models, which never reaches 100% (Table 2a). Since all the compared maps have complementary aspects, it is not possible to select *a priori* one map instead of the other as definitive habitat estimation. Thus, the production of a merged distribution representing the overall habitat suitability of *L. sceleratus* in the Mediterranean Sea is justified. This merged distribution indicates suitable habitat in most of the Mediterranean Sea except in some areas in the centre and in the western basin (Fig. 4a).

The *geographical reachability* distribution presents complementary aspects with respect to the baseline habitat distributions (Fig. 4b and Table 2b). Another *geographical reachability* map was produced by training the model with environmental data projected in 2050 (Section 2.1.4) and by using currently known locations in the Mediterranean Sea to build Gaussian functions. The resulting distribution is very similar to the previous one (Fig. 4c). Thus, the first produced *geographical reachability* distribution represents a stable average scenario for *L. sceleratus*.

The techniques used in our experiment mostly belong to the class of *correlative approaches* to niche modelling (Pearson, 2012), and thus do not explicitly model the bio-physiological characteristics of *L. sceleratus*. Nevertheless, an *a posteriori* statistical analysis was made across all models, focussing on the 18 environmental variables values in the highest probability locations (> 0.8). This analysis helps inferring the environmental conditions the species prefers. In particular, it reveals

[11] An animation representing this convergence process is available at http://data.d4science.org/WG1QTFhsS1k2Qy90WXE5NVNaZnRoRUQ4bk44Y05NVWdHbWJQNStIS0N6Yz0.

[12] https://services.d4science.org/group/biodiversitylab/data-miner?OperatorId=org.gcube.dataanalysis.wps.statisticalmanager.synchserver.mappedclasses.evaluators.MAPS_COMPARISON.

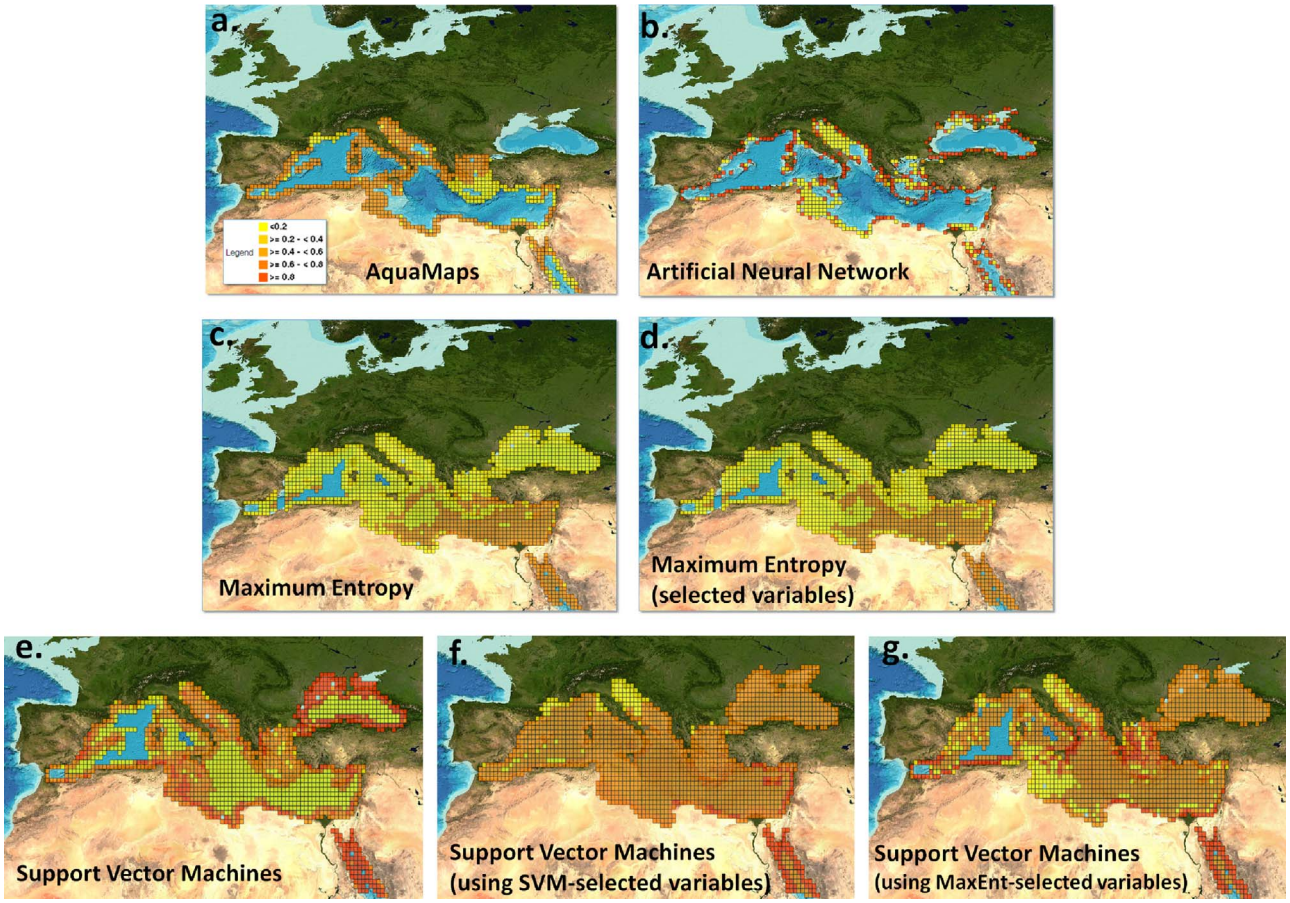

Fig. 2. Projections on the Mediterranean Sea of all the baseline ecological niche models involved in our method: (a) AquaMaps, (b) Artificial Neural Networks, (c) Maximum Entropy (MaxEnt), (d) MaxEnt trained with influential environmental variables selected by the previous MaxEnt model, (e) Support Vector Machines (SVM), (f) SVM trained on influential environmental variables selected using a leave-one-out process, and (g) SVM trained on the environmental variables selected by MaxEnt. The legend is the same for all the maps.

that for all models the average surface temperature is around (19 ± 0.2)° and the average difference with sea bottom temperature is 5°. Average sea surface salinity is 34.5 ± 0.3 PSU except for AquaMaps (38 ± 0.03 PSU), and the average difference between sea surface salinity and sea bottom salinity is always around 2%. Primary production is averagely 530 g C m^{-2} yr^{-1} with high values (~1100 g C m^{-2} yr^{-1}) reached by few locations. Distance from land is always below 200 km, with 80 km average for all models except for ANN (32 km) and AquaMaps (31 km). Finally, average water column depth is 1150 m except for ANN (207 m) and AquaMaps (407 m). However, all models also include high probability locations in shallow waters (down to 1 m) and averagely deep waters (up to 3000 m).

3.2. Performance evaluation

Accuracy was calculated for all models as described in Section 2.6 (Table 3). The *geographical reachability* distribution gains the highest performance in predicting the Mediterranean records (98%), whereas the *dynamic* model reaches lower performance (83%). All the baseline models have high performance on predicting observation records. However, a visual comparison (Fig. 2) suggests that most of these models possibly overestimate the presence locations. Nevertheless, there is useful information in this overestimation; for example, the SVM distribution in Fig. 2f assigns non-zero values to most of the Mediterranean Sea half-degree locations and may seem uninformative, but this distribution also includes a westward decreasing gradient, which is crucial information when introduced in the merged model.

Overall, the percentages in Tables 2 and 3 suggest that the *geographical reachability* distribution (i) is complementary to the other maps, (ii) presents stability with respect to a 2050 scenario, and (iii) has high performance at predicting real observations in the Mediterranean Sea.

3.3. Discrepancies evaluation

The *dynamic* distribution has overall 83% agreement with the *geographical reachability* distribution and is generally similar to this distribution (Fig. 4b and d). However, the overall performance of the *dynamic* model on known records in the Mediterranean Sea is lower and there are specific discrepancy areas (highlighted in Fig. 4d) that need further analysis. Overall, unlike the *geographical reachability* distribution, the *dynamic* distribution predicts species invasion also up to the continental edge, where currently there are very few records (Çinar et al., 2014).

The *geographical reachability* distribution predicts medium-to-high occurrence probability (0.4–0.6) in eastern Mediterranean (e.g. Aegean Sea and Cyprus Sea), southern Ionian Sea (Albanian coasts), and southern Tyrrhenian Sea (Sicily, Tunisian and Libyan coasts), whereas it predicts the highest probability (> 0.8) in several coastal areas (e.g. south Turkey, south Greece and east Libya). Indeed, the highest abundance of *L. sceleratus* has been recorded in the eastern Mediterranean (Nader et al., 2012; Michailidis, 2010). Also, in the southern Tyrrhenian Sea, between Malta and the Tunisian coast, more than 80 individuals have been reported in the last 5 years (Azzurro et al., 2014b). In the Bosporus, few records have been officially reported (Vacchi et al., 2007), which agrees with the *geographical reachability* distribution on a possible future presence in this area, despite the opposite indication by the *dynamic* distribution. As for the Italian coasts, observations have been increasingly reported around Sicily (Azzurro et al., 2014a, 2016),

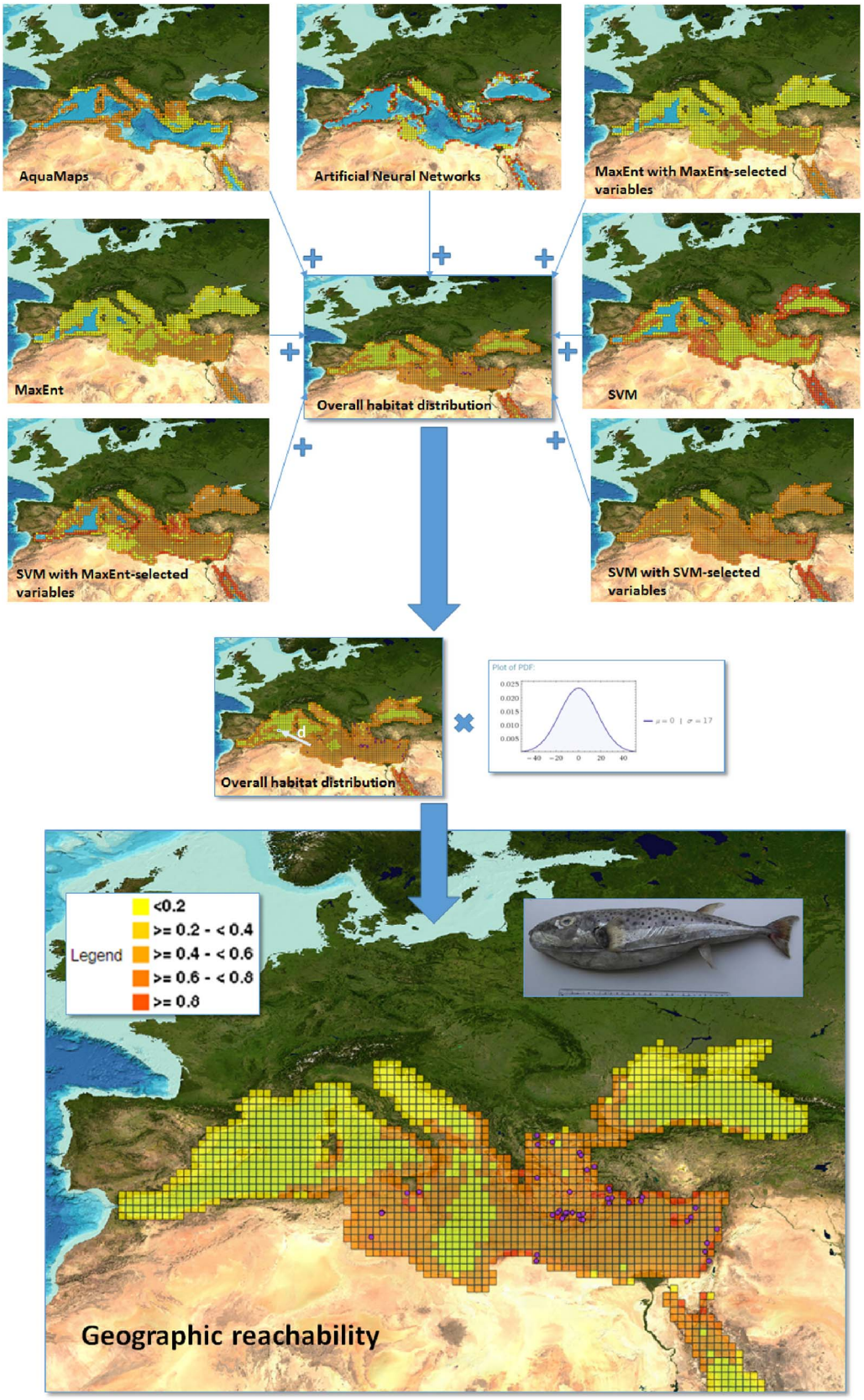

Fig. 3. Complete work flow of our method: the baseline models are projected on the Mediterranean Sea and are merged together through a normalised sum of the distributions (overall habitat distribution). A *geographical reachability* map is estimated by multiplying habitat suitability for a decreasing *geographical reachability* Gaussian function, based on the minimum distance from known observations (image of *L. scleratus* retrieved from fishbase.org). The overlaid dark dots represent the observations in the Mediterranean Sea used to train the *geographical reachability* model. The legend is the same for all the maps.

in the central Tyrrhenian Sea (Jribi et al., 2012), and in the southern Adriatic Sea (Dulčić and Dragičević, 2014; Nader et al., 2012), still in agreement with the *geographical reachability* distribution. On the contrary, although both the compared distributions indicate lower invasion probability in the western Mediterranean Sea, the fish has been observed very recently along the Algerian and southern Spanish coasts (Dailianis et al., 2016; Katsanevakis et al., 2014), but it still early to assess a population establishment.

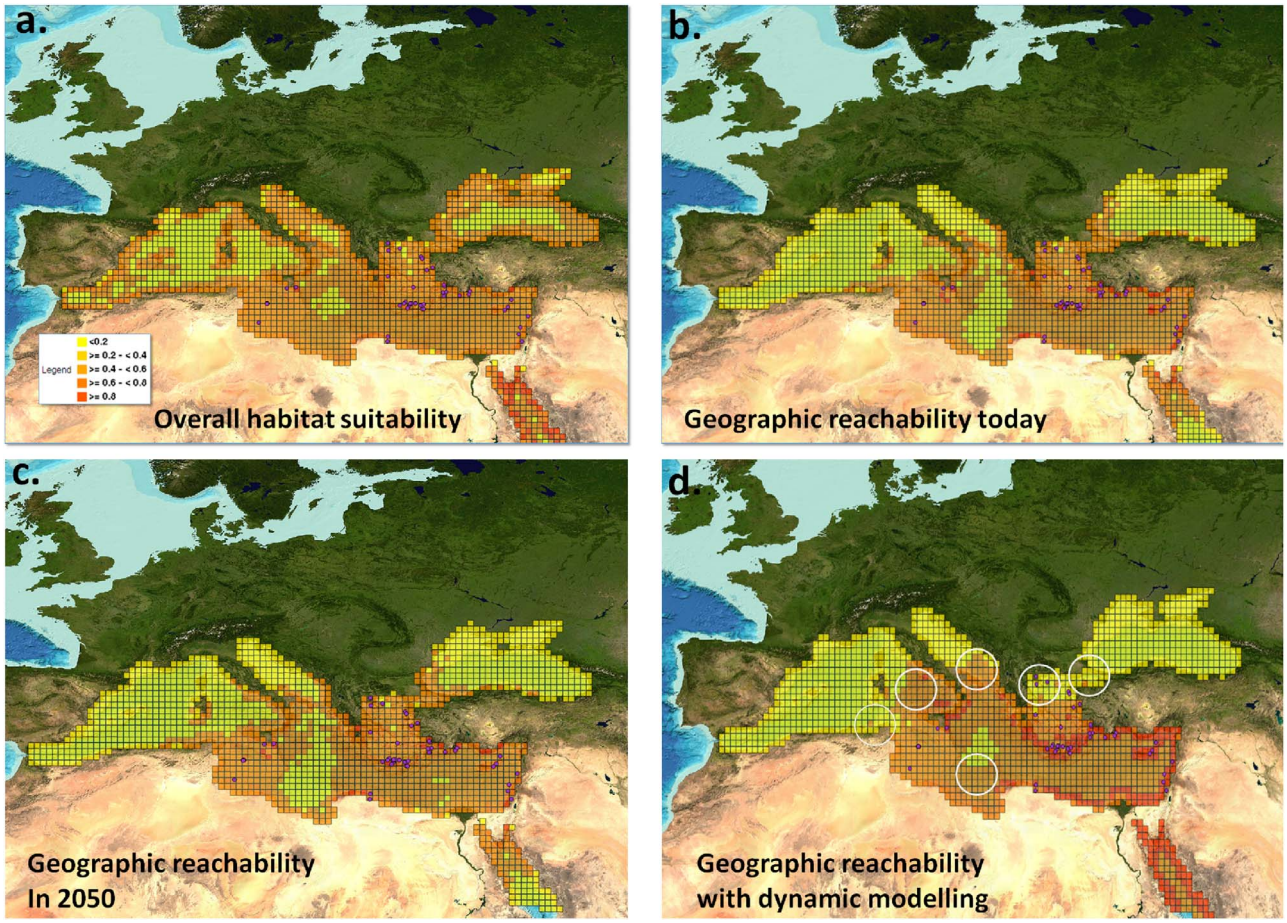

Fig. 4. Comparison between several distributions estimated for *L. sceleratus* in the Mediterranean Sea: (a) overall potential habitat map, obtained by merging the baseline models, (b) *geographical reachability* distribution calculated by applying a Gaussian spatial weighting operation to the overall potential habitat map, (c) *geographical reachability* distribution estimated for 2050, and (d) distribution estimated by means of a *dynamic* model. The overlapping points indicate the real observations of the pufferfish in the Mediterranean Sea that were used to tune the spatial models. The circles in figure (d) indicate the highest discrepancy areas with respect to the distribution in figure (b). The overlaid dark dots represent the observations in the Mediterranean Sea used to train the *geographical reachability* model. The legend is the same for all the maps.

Table 2
Agreements between all the estimated distributions of *L. sceleratus* in the Mediterranean Sea. AquaMaps – AquaMaps ecological niche model; ANN – Artificial Neural Networks; MaxEnt – Maximum Entropy; MaxEnt-sel – Maximum Entropy model trained with features selected as carrying most of the information according to MaxEnt; SVM – Support Vector Machines; SVM-MaxEnt – SVM trained using MaxEnt-selected features; SVM-LOF – SVM trained with features selected as carrying most of the information according to SVM.

	AquaMaps	ANN	MaxEnt	MaxEnt-sel	SVM	SVM-MaxEnt
(a) Agreement between all the trained baseline habitat suitability distribution models of *L. sceleratus* in the Mediterranean Sea						
ANN	80%					
MaxEnt	57%	42%				
MaxEnt-sel	58%	43%	90%			
SVM	70%	93%	46%	47%		
SVM-MaxEnt	48%	47%	51%	52%	52%	
SVM-LOF	49%	52%	55%	56%	61%	83%

	AquaMaps	ANN	MaxEnt	MaxEnt-sel	SVM	SVM-MaxEnt	SVM-LOF
(b) Agreement between the *geographical reachability* distribution of *L. sceleratus* and the baseline models							
Geographical Reachability Distribution	69%	58%	77%	76%	60%	53%	55%

3.4. Impact indicators

The impact indicator described in Section 2.6, was calculated for the *geographical reachability* distribution after its projection onto three reference subdivision sets of the Mediterranean Sea corresponding to different ecological and economical criteria. A coloured gradient was used to give direct visual understanding of the impact of the pufferfish on these areas.

A first indicator was calculated on the official major fishing areas in the Mediterranean Sea of the Food and Agriculture Organisation of the United Nations (FAO). This indicator (Fig. 5a) has a westward decreasing gradient and estimates low-medium impact (0.2–0.6) in the Tyrrhenian and Ionian Seas (Italian, Tunisian, and Libyan coasts) and high impact (0.8–1) around the coasts of Greece, Turkey, Lebanon, and Egypt.

A second indicator was calculated on the Exclusive Economic Zones associated to the United Nations Convention on the Law of the Sea (Lowe, 1990), which define the regions where a state has special rights

Table 3
Accuracy of all the models involved in our method to predict real observations of *L. sceleratus* in the Mediterranean Sea. Geographical Reachability Distribution – estimated future distribution of the pufferfish in the Mediterranean Sea; Dynamic model – benchmark model that simulates the spread of the pufferfish over time; AquaMaps – AquaMaps ecological niche model; ANN – Artificial Neural Networks; MaxEnt – Maximum Entropy; MaxEnt-sel – Maximum Entropy model trained with features selected as carrying most of the information according to MaxEnt; SVM – Support Vector Machines; SVM-MaxEnt – SVM trained using MaxEnt-selected features; SVM-LOF – SVM trained with features selected as carrying most of the information according to SVM.

	Geographical Reachability Distribution	Dynamic model	AquaMaps	ANN	MaxEnt	MaxEnt-sel	SVM	SVM-MaxEnt	SVM-LOF
Accuracy	**98%**	83%	95%	96%	79%	81%	96%	92%	96%

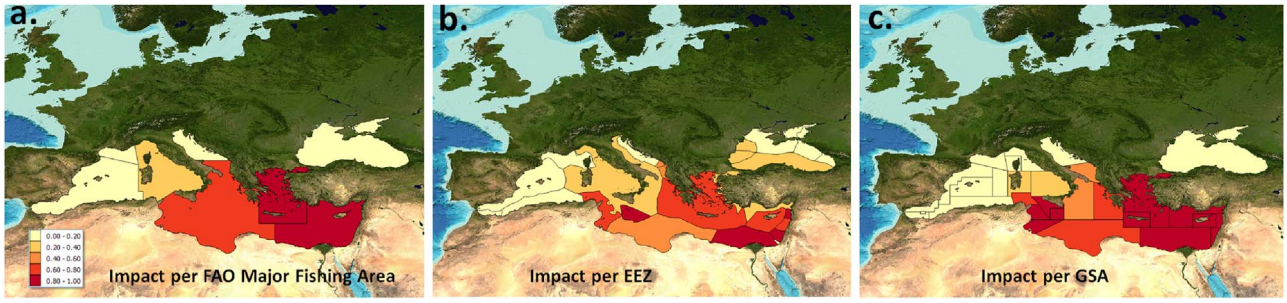

Fig. 5. Impact indicator calculated as the density of sensibly non-zero probability locations of the *geographical reachability* distribution that fall in different subdivisions of the Mediterranean Sea: (a) FAO major fishing areas, (b) Exclusive Economic Zones, and (c) general subdivisions of the General Fisheries Commission for the Mediterranean Sea (GSAs). The legend is the same for all the maps.

for the exploration and use of marine resources. Although EEZs are not accepted by several international organisations (including FAO), they may give insight on how the Mediterranean countries could be economically impacted by the *L. sceleratus* invasion. In particular (Fig. 5b), highly impacted areas are highlighted in the eastern Mediterranean Sea (e.g. in Greece and Cyprus, whose situation is already known to be alarming) and in the south (i.e. Malta and Tunisia).

Finally, a third indicator, with higher resolution, was calculated on the general subdivisions of the General Fisheries Commission for the Mediterranean Sea (GSAs). These areas are commonly used to monitor and manage marine fishery resources in the Mediterranean Sea and can better highlight the impact of *L. sceleratus* on smaller areas (Fig. 5c). Again, a westward decreasing gradient is observed and high impact zones are now highlighted in southern Sicily, in the Ionian Sea, and in southern Adriatic Sea. Further, impact in Greece is reported to be higher in the Aegean Sea than in the Ionic side.

4. Conclusions

In this paper, a method to estimate the spread of the silver-cheeked toad-fish *L. sceleratus* in the Mediterranean Sea has been presented. Seven niche models based on machine learning algorithms were trained and merged together. The method generates an overall *geographical reachability* distribution by means of a distance weighting process that takes real observations into account. The reliability of this method has been assessed with respect to a reference *dynamic* model and Mediterranean observation records of the pufferfish.

A risk estimate of the invasion has been reported as the density of non-zero probability locations falling in different subdivisions of the Mediterranean Sea related to marine resources exploitation. A general westward decreasing impact pattern has been highlighted, but high risk zones have been predicted also in the middle and in the south of the Mediterranean Sea (e.g. Sicily, Malta, and Tunisia). The overall depicted scenario is that *L. sceleratus* is a great risk for fisheries (and consequently on health security) of many Mediterranean countries in the near future. Further, our estimated distribution foresees the invasion by the pufferfish of the Bosporus, which could enable it to spread in the Black Sea. Thus, strategies such as selective fishing to decrease its population should be considered, especially in the most likely future impacted areas, in order to stem this spread and prevent severe economic damages. In this context, our maps can support these strategies and can also help fisheries researchers to advise managers and decision makers.

This experiment used an e-Infrastructure for every step of the method, from data retrieval to models' training and projection. The e-Infrastructure enabled the authors in their collaboration and every step of the experiment has been made repeatable, because it is published as-a-Service under a standard representation (WPS) and all the data, processes, and results are freely accessible on-line.[13] Finally, the presented approach is general enough to be applied to other invasive fish.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the BlueBRIDGE project (grant agreement no. 675680). Thanks to A. Zenetos for sharing the occurrence records of *L. sceleratus* from the ELNAIS project.

References

Akyol, O., Ünal, V., Ceyhan, T., Bilecenoglu, M., 2005. First confirmed record of *Lagocephalus sceleratus* (Gmelin, 1789) in the Mediterranean Sea. J. Fish Biol. 66 (4), 1183–1186.

Assante, M., Candela, L., Castelli, D., Coro, G., Lelii, L., Pagano, P., 2016. Virtual research environments as-a-service by gCube. PeerJ Prepr. 4, e2511v1.

Attard, D.J., 1987. The Exclusive Economic Zone in International Law, vol. 1 Clarendon Press, Oxford, UK.

Azzurro, E., Ben Souissi, J., Boughedir, W., Castriota, L., Deidun, A., Falautano, M., Ghanem, R., Zammit-Mangion, M., Andaloro, F., 2014a. The Sicily strait: a transnational observatory for monitoring the advance of non indigenous species. Biol. Mar. Mediterr. 21 (1), 105–106.

Azzurro, E., Castriota, L., Falautano, M., Bariche, M., Broglio, E., Andaloro, F., et al., 2016. New records of the silver-cheeked toadfish *Lagocephalus sceleratus* (Gmelin, 1789) in the Tyrrhenian and Ionian Seas: early detection and participatory monitoring in practice. Bioinvasions Rec. 5 (4), 295–299.

Azzurro, E., Castriota, L., Falautano, M., Giardina, F., Andaloro, F., 2014b. The silver-cheeked toadfish *Lagocephalus sceleratus* (Gmelin, 1789) reaches Italian waters. J.

[13] An ensemble is available at http://data.d4science.org/workspace-explorer-app/?folderId=Nk0yMHVVdWxCWVZHM00xck5qRGd2N3Q1dnN6Zk1YUDNhRjJsMnAvcHZhSHg1UXNocUZOMEliN0ZuTktycDFUQw.

Appl. Ichthyol. 30 (5), 1050–1052.
Baldwin, R.A., 2009. Use of maximum entropy modeling in wildlife research. Entropy 11 (4), 854–866.
Barbosa, F.G., Schneck, F., Melo, A.S., 2012. Use of ecological niche models to predict the distribution of invasive species: a scientometric analysis. Braz. J. Biol. 72 (4), 821–829.
Bebis, G., Georgiopoulos, M., 1994. Feed-forward neural networks. IEEE Potentials 13 (4), 27–31.
Bentur, Y., Ashkar, J., Lurie, Y., Levy, Y., Azzam, Z.S., Litmanovich, M., Golik, M., Gurevych, B., Golani, D., Eisenman, A., 2008. Lessepsian migration and tetrodotoxin poisoning due to *Lagocephalus sceleratus* in the eastern Mediterranean. Toxicon 52 (8), 964–968.
Bidegain, G., Bárcena, J.F., García, A., Juanes, J.A., 2015. Predicting coexistence and predominance patterns between the introduced manila clam (*Ruditapes philippinarum*) and the European native clam (*ruditapes decussatus*). Estuar. Coast. Shelf Sci. 152, 162–172.
Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.
Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. ACM. pp. 144–152.
Broennimann, O., Treier, U.A., Müller-Schärer, H., Thuiller, W., Peterson, A., Guisan, A., 2007. Evidence of climatic niche shift during biological invasion. Ecol. Lett. 10 (8), 701–709.
Brown, M., Gunn, S.R., Lewis, H.G., 1999. Support vector machines for optimal classification and spectral unmixing. Ecol. Model. 120 (2), 167–179.
Camps-Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 43 (6), 1351–1362.
Candela, L., Castelli, D., Coro, G., Lelii, L., Mangiacrapa, F., Marioli, V., Pagano, P., 2015. An infrastructure-oriented approach for supporting biodiversity research. Ecol. Inform. 26, 162–172.
Carlos-Júnior, L., Barbosa, N., Moulton, T., Creed, J., 2015. Ecological niche model used to examine the distribution of an invasive, non-indigenous coral. Mar. Environ. Res. 103, 115–124.
Castelar, B., de Siqueira, M.F., Sánchez-Tapia, A., Reis, R.P., 2015. Risk analysis using species distribution modeling to support public policies for the alien alga *Kappaphycus alvarezii* aquaculture in Brazil. Aquaculture 446, 217–226.
Chang, C.-C., Lin, C.-J., 2011. Libsvm: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2 (3), 27.
Ç inar, M.E., Arianoutsou, M., Zenetos, A., Golani, D., 2014. Impacts of invasive alien marine species on ecosystem services and biodiversity: a pan-European review. Aquat. Invasions 9 (4), 391–423.
Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Lasram, F.B.R., Aguzzi, J., Ballesteros, E., Bianchi, C.N., Corbera, J., Dailianis, T., et al., 2010. The biodiversity of the Mediterranean Sea: estimates, patterns, and threats. PLoS ONE 5 (8), e11842.
Coro, G., Magliozzi, C., Berghe, E.V., Bailly, N., Ellenbroek, A., Pagano, P., 2016a. Estimating absence locations of marine species from data of scientific surveys in OBIS. Ecol. Model. 323, 61–76.
Coro, G., Magliozzi, C., Ellenbroek, A., Kaschner, K., Pagano, P., 2016b. Automatic classification of climate change effects on marine species distributions in 2050 using the AquaMaps model. Environ. Ecol. Stat. 23 (1), 155–180.
Coro, G., Magliozzi, C., Ellenbroek, A., Pagano, P., 2015. Improving data quality to build a robust distribution model for *Architeuthis dux*. Ecol. Model. 305, 29–39.
Coro, G., Pagano, P., Ellenbroek, A., 2014. Comparing heterogeneous distribution maps for marine species. GISci. Remote Sens. 51 (5), 593–611.
Coro, G., Panichi, G., Scarponi, P., Pagano, P., 2017. Cloud computing in a distributed e-infrastructure using the web processing service standard. Concurrency and Computation: Practice and Experience.
Corsi, F., de Leeuw, J., Skidmore, A., 2000. Modeling species distribution with GIS. Research Techniques in Animal Ecology. Columbia University Press, New York, pp. 389–434.
Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press.
Dailianis, T., Akyol, O., Babali, N., Bariche, M., Crocetta, F., Gerovasileiou, V., Chanem, R., Gökoğlu, M., Hasiotis, T., Izquierdo Muñoz, A., et al., 2016. New Mediterranean Biodiversity Records (July 2016). https://ejournals.epublishing.ekt.gr/index.php/hcmr-med-mar-sc/article/view/13484.
Drake, J.M., Randin, C., Guisan, A., 2006. Modelling ecological niches with support vector machines. J. Appl. Ecol. 43 (3), 424–432.
Dulčić, J., Dragičević, B., 2014. Occurrence of Lessepsian migrant *Lagocephalus sceleratus* (Tetraodontidae) in the Adriatic Sea. Cybium 38 (3), 238–240.
Elith, J., Graham, C.H., 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. Ecography 32 (1), 66–77.
Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40 (1), 677–697.
Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J., 2011 Jan. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17 (1), 43–57.
FAO, 2007. FAO Fisheries and Aquaculture Technical Paper. No. 530. Climate Change Implications for Fisheries and Aquaculture. http://www.fao.org/docrep/012/i0994e/i0994e00.htm.
Farashi, A., Najafabadi, M.S., 2015. Modeling the spread of invasive nutrias (*Myocastor coypus*) over Iran. Ecol. Complex. 22, 59–64.
Ficetola, G.F., Thuiller, W., Miaud, C., 2007. Prediction and validation of the potential global distribution of a problematic alien invasive species – the American bullfrog. Divers. Distrib. 13 (4), 476–485.
Fricke, R., Golani, D., Appelbaum-Golani, B., et al., 2015. First record of the indian anchovy *Stolephorus indicus* (Van Hasselt, 1823) (Clupeiformes: Engraulidae) in the Mediterranean Sea. Bioinvasions Rec. 4 (4), 293–297.
Galil, B., 2008. Alien species in the Mediterranean Sea – which, when, where, why? Hydrobiologia 606 (1), 105–116.
Galil, B., 2009. Taking stock: inventory of alien species in the Mediterranean Sea. Biol. Invasions 11 (2), 359–372.
Galil, B., Marchini, A., Occhipinti-Ambrogi, A., Minchin, D., Narščius, A., Ojaveer, H., Olenin, S., 2014. International arrivals: widespread bioinvasions in European seas. Ethol. Ecol. Evol. 26 (2-3), 152–171.
Galil, B.S., Boero, F., Campbell, M.L., Carlton, J.T., Cook, E., Fraschetti, S., Gollasch, S., Hewitt, C.L., Jelmert, A., Macpherson, E., et al., 2015. 'Double trouble': the expansion of the Suez Canal and marine bioinvasions in the Mediterranean Sea. Biol. Invasions 17 (4), 973–976.
Ganeshaiah, K., Barve, N., Nath, N., Chandrashekara, K., Swamy, M., Uma Shaanker, R., 2003. Predicting the potential geographical distribution of the sugarcane woolly aphid using GARP and DIVA-GIS. Curr. Sci. 85 (11), 1526–1528.
Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 160 (3), 249–264.
Golani, D., 2010. Colonization of the Mediterranean by Red Sea fishes via the Suez Canal – Lessepsian migration. Fish Invasions of the Mediterranean Sea: Change and Renewal. Pensoft Publishers, Sofia-Moscow, pp. 145–188.
Grassle, J.F., 2000. The Ocean Biogeographic Information System (OBIS): an on-line, worldwide atlas for accessing, modeling and mapping marine biological data in a multidimensional geographic context. Oceanography 13 (3), 5–7.
Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., Kueffer, C., 2014. Unifying niche shift studies: insights from biological invasions. Trends Ecol. Evol. 29 (5), 260–269.
Guo, Q., Kelly, M., Graham, C.H., 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. Ecol. Model. 182 (1), 75–90.
Hey, T., Tansley, S., Tolle, K.M., 2009. The Fourth Paradigm: Data-Intensive Scientific Discovery, vol. 1 Microsoft Research, Redmond, WA.
Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al., 2003. A Practical Guide to Support Vector Classification. https://www.google.it/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjTpajeh_DVAhVBtRQKHcGqBd0QFggsMAA&url=http%3A%2F%2Fwww.csie.ntu.edu.tw%2F~cjlin%2Fpapers%2Fguide%2Fguide.pdf&usg=AFQjCNFol0McRktHC6gsBxKXqQMvmQUFeg.
ICES, 2007. Report of the Working Group on Introductions and Transfers of Marine Organisms. http://icesdk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2007/WGITMO/WGITMO07.pdf.
Jeschke, J.M., Strayer, D.L., 2008. Usefulness of bioclimatic models for studying climate change and invasive species. Ann. N. Y. Acad. Sci. 1134 (1), 1–24.
Jribi, I., Bradai, M.N., et al., 2012. First record of the Lessepsian migrant species *Lagocephalus sceleratus* (Gmelin, 1789) (Actinopterygii: Tetraodontidae) in the Central Mediterranean. Bioinvasions Rec. 1 (1), 49–52.
Kalogirou, S., Corsini-Foka, M., Sioulas, A., Wennhage, H., Pihl, L., 2010. Diversity, structure and function of fish assemblages associated with *Posidonia oceanica* beds in an area of the eastern Mediterranean Sea and the role of non-indigenous species. J. Fish Biol. 77 (10), 2338–2357.
Kaschner, K., Watson, R., Trites, A., Pauly, D., 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. Mar. Ecol. Prog. Ser. 316, 285–310.
Katsanevakis, S., Acar, U., Ammar, I., Balci, B., Bekas, P., Belmonte, M., Chintiroglou, C., Consoli, P., Dimiza, M., Fryganiotis, K., et al., 2014. New Mediterranean biodiversity records (october, 2014). Mediterr. Mar. Sci. 15 (3).
Kheifets, J., Rozhavsky, B., Girsh Solomonovich, Z., Marianna, R., Soroksky, A., 2012. Severe tetrodotoxin poisoning after consumption of *Lagocephalus sceleratus* (pufferfish, fugu) fished in Mediterranean Sea, treated with cholinesterase inhibitor. Case Rep. Crit. Care 2012.
Kulhanek, S.A., Leung, B., Ricciardi, A., 2011. Using ecological niche models to predict the abundance and impact of invasive species: application to the common carp. Ecol. Appl. 21 (1), 203–213.
Lane, M., Edwards, J., 2007. The global biodiversity information facility (GBIF). Systematics Association Special Volume, 76. pp. 1.
Lauzeral, C., Leprieur, F., Beauchard, O., Duron, Q., Oberdorff, T., Brosse, S., 2011. Identifying climatic niche shifts using coarse-grained occurrence data: a test with non-native freshwater fish. Glob. Ecol. Biogeogr. 20 (3), 407–414.
Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., Zhao, J., 2013. PROV-O: The PROV Ontology, vol. 30 W3C Recommendation.
Leidenberger, S., Obst, M., Kulawik, R., Stelzer, K., Heyer, K., Hardisty, A., Bourlat, S.J., 2015. Evaluating the potential of ecological niche modelling as a component in marine non-indigenous species risk assessments. Mar. Pollut. Bull. 97 (1), 470–487.
Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90 (1), 39–52.
Lek, S., Guégan, J.-F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Model. 120 (2), 65–73.
Lowe, V., 1990. Exclusive economic zones. Int. J. Estuar. Coast. Law 5, 409.
McGeoch, M.A., Chown, S.L., Kalwij, J.M., 2006. A global indicator for biological invasion. Conserv. Biol. 20 (6), 1635–1646.
Mellin, C., Lurgi, M., Matthews, S., MacNeil, M., Caley, M., Bax, N., Przeslawski, R., Fordham, D., 2016. Forecasting marine invasions under climate change: biotic interactions and demographic processes matter. Biol. Conserv. 204, 459–467.
Michailidis, N., 2010. Study on the Lessepsian Migrant *Lagocephalus sceleratus* in Cyprus,

vol. 4. EastMEd Technical Documents, pp. 74–87.
Minsky, M., 1963. Steps toward artificial intelligence. Computers and Thought, vol. 406. pp. 450.
Nader, M., Indary, S., Boustany, L., 2012. The Puffer Fish *Lagocephalus sceleratus* (Gmelin, 1789) in the Eastern Mediterranean. EastMed Technical Documents (FAO).
Nakicenovic, N., Swart, R., July 2000. Special report on emissions scenarios. In: Nakicenovic, N., Swart, R. (Eds.), Special Report on Emissions Scenarios. Cambridge University Press, Cambridge, UK, pp. 612 ISBN: 0521804930.
National Research Council of Italy, 2016. The D4Science Distributed e-Infrastructure. http://www.d4science.org.
Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Model. 178 (3), 389–397.
Özesmi, S.L., Tan, C.O., Özesmi, U., 2006. Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecol. Model. 195 (1), 83–93.
Padalia, H., Srivastava, V., Kushwaha, S., 2014. Modeling potential invasion range of alien invasive species, *Hyptis suaveolens* (L.) Poit. in India: comparison of MaxEnt and GARP. Ecol. Inform. 22, 36–43.
Patterson, D.W., 1998. Artificial Neural Networks: Theory and Applications. Prentice Hall PTR.
Pearman, P.B., Guisan, A., Broennimann, O., Randin, C.F., 2008. Niche dynamics in space and time. Trends Ecol. Evol. 23 (3), 149–158.
Pearson, R.G., 2012. Species Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.
Peristeraki, P., Lazarakis, G., Skarvelis, C., Georgiadis, M., Tserpes, G., 2006. Additional records on the occurrence of alien fish species in the eastern Mediterranean Sea. Mediterr. Mar. Sci. 7 (2), 61–66.
Peterson, A.T., 2003. Predicting the geography of species' invasions via ecological niche modeling. Q. Rev. Biol. 78 (4), 419–433.
Peterson, A.T., Robins, C.R., 2003. Using ecological-niche modeling to predict barred owl invasions with implications for spotted owl conservation. Conserv. Biol. 17 (4), 1161–1165.
Peterson, A.T., Vieglais, D.A., 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem: a new approach to ecological niche modeling, based on new tools drawn from biodiversity informatics, is applied to the challenge of predicting potential species' invasions. BioScience 51 (5), 363–371.
Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., Guisan, A., 2012. Climatic niche shifts are rare among terrestrial plant invaders. Science 335 (6074), 1344–1348.
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190 (3–4), 231–259.
Phillips, S.J., Dudik, M., 2008. Modeling of species distributions with MAXENT: new extensions and a comprehensive evaluation. Ecography 31, 161–175.
Phillips, S.J., Dudík, M., Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. In: Proceedings of the Twenty-First International Conference on Machine Learning. ACM. pp. 83.
Pouteau, R., Meyer, J.-Y., Stoll, B., 2011. A SVM-based model for predicting distribution of the invasive tree *Miconia calvescens* in tropical rainforests. Ecol. Model. 222 (15), 2631–2641.
Ready, J., Kaschner, K., South, A.B., Eastwood, P.D., Rees, T., Rius, J., Agbayani, E., Kullander, S., Froese, R., 2010. Predicting the distributions of marine organisms at the global scale. Ecol. Model. 221 (3), 467–478.
Reyes, K., 2015. AquaMaps: Algorithm and Data Sources for Aquatic Organisms. Available at http://www.aquamaps.org/main/FB_Book_KReyes_AquaMaps_JG.pdf.
Roeckner, E., Arpe, K., Bengtsson, L., Brinkop, S., Dümenil, L., Esch, M., Kirk, E., Lunkeit, F., Ponater, M., Rockel, B., et al., 1992. Simulation of the Present-Day Climate with the ECHAM Model: Impact of Model Physics and Resolution. Max-Planck-Institut für Meteorologie, Hamburg, Germany.
Rumelhart, D., Hinton, G., Williams, R., 1986. Steps toward artificial intelligence. Nature 6088 (323), 533–536.
Sadeghi, R., Zarkami, R., Sabetraftar, K., Van Damme, P., 2012. Use of support vector machines (SVMs) to predict distribution of an invasive water fern *Azolla filiculoides* (Lam.) in Anzali wetland, southern Caspian Sea, Iran. Ecol. Model. 244, 117–126.
Sanchez-Flores, E., Rodriguez-Gallegos, H., Yool, S., 2008. Plant invasions in dynamic desert landscapes. A field and remote sensing assessment of predictive and change modeling. J. Arid Environ. 72 (3), 189–206.
Sax, D.F., Stachowicz, J.J., Brown, J.H., Bruno, J.F., Dawson, M.N., Gaines, S.D., Grosberg, R.K., Hastings, A., Holt, R.D., Mayfield, M.M., et al., 2007. Ecological and evolutionary insights from species invasions. Trends Ecol. Evol. 22 (9), 465–471.
Schölkopf, B., Burges, C.J., Smola, A.J., 1999. Advances in Kernel Methods: Support Vector Learning. MIT Press.
Schut, P., Whiteside, A., 2007. OpenGIS Web Processing Service. OGC Project Document. http://www.opengeospatial.org/standards/wps.
Searight, S., 2016. "A dismal but profitable ditch": the Suez Canal then and now. Asian Aff. 47 (1), 93–100.
Shabani, F., Kumar, L., 2015. Should species distribution models use only native or exotic records of existence or both? Ecol. Inform. 29, 57–65.
Shine, C., Williams, N., Gündling, L., 2000. A Guide to Designing Legal and Institutional Frameworks on Alien Invasive Species, vol. 40 IUCN.
Sobek-Swant, S., Kluza, D.A., Cuddington, K., Lyons, D.B., 2012. Potential distribution of emerald ash borer: what can we learn from ecological niche models using Maxent and GARP? Forest Ecol. Manag. 281, 23–31.
Stockwell, D., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. Int. J. Geogr. Inf. Sci. 13 (2), 143–158.
Strubbe, D., Broennimann, O., Chiron, F., Matthysen, E., 2013. Niche conservatism in non-native birds in Europe: niche unfilling rather than niche expansion. Glob. Ecol. Biogeogr. 22 (8), 962–970.
Sutherst, R.W., 2000. Climate change and invasive species: a conceptual framework. Invasive Species in a Changing World. pp. 211–240.
Suykens, J.A., De Brabanter, J., Lukas, L., Vandewalle, J., 2002. Weighted least squares support vector machines: robustness and sparse approximation. Neurocomputing 48 (1), 85–105.
Thuiller, W., Richardson, D.M., PYŠEK, P., Midgley, G.F., Hughes, G.O., Rouget, M., 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Glob. Change Biol. 11 (12), 2234–2250.
Ünal, V., Bodur, H., et al., 2017. The socio-economic impacts of the silver-cheeked toadfish on small-scale fishers: a comparative study from the Turkish coast. J. Fish. Aquat. Sci. (Su Ürünleri Dergisi) 34 (2), 119–127.
Ünal, V., Göncüoglu, H., Durgun, D., Tosunoglu, Z., Deval, M.C., Turan, C., 2015. Silver-cheeked toadfish, *Lagocephalus sceleratus* (Actinopterygii: Tetraodontiformes: Tetraodontidae), causes a substantial economic losses in the Turkish Mediterranean coast: a call for decision makers. Acta Ichthyol. Piscat. 45 (3), 231.
Underwood, E.C., Klinger, R., Moore, P.E., 2004. Predicting patterns of non-native plant invasions in Yosemite National Park, California, USA. Divers. Distrib. 10 (5-6), 447–459.
Vacchi, M., Bussotti, S., Miglietta, A., Guidetti, P., 2007. Presence of the Guinean puffer *Sphoeroides marmoratus* (Lowe, 1838) in the Mediterranean Sea. J. Fish Biol. 71 (4), 1215–1219.
Vanden Berghe, E., Stocks, K.I., Grassle, J.F., 2010. Data integration: the ocean biogeographic information system. Life in the World's Oceans: Diversity, Distribution, and Abundance. pp. 333.
Vapnik, V., 2013. The Nature of Statistical Learning Theory, vol. 1 Springer Science & Business Media.
Vilas, L.G., Spyrakos, E., Palenzuela, J.M.T., Pazos, Y., 2014. Support vector machine-based method for predicting *Pseudo-nitzschia* spp. blooms in coastal waters (Galician rias, NW Spain). Prog. Oceanogr. 124, 66–77.
West, A.M., Kumar, S., Brown, C.S., Stohlgren, T.J., Bromberg, J., 2016. Field validation of an invasive species Maxent model. Ecol. Inform. 36, 126–134.
Wilson, E.O., 2003. The encyclopedia of life. Trends Ecol. Evol. 18 (2), 77–80.
Wu, T.-F., Lin, C.-J., Weng, R.C., 2004. Probability estimates for multi-class classification by pairwise coupling. J. Mach. Learn. Res. 5 (August), 975–1005.
Yaglioglu, D., Turan, C., Erguden, D., Gurlek, M., 2011. Range expansion of silverstripe blaasop, *Lagocephalus sceleratus* (Gmelin, 1789), to the northeastern Mediterranean Sea. Biharean Biol. 5 (2), 159–161.
Zenetos, A., 2010. Trend in aliens species in the Mediterranean. An answer to Galil, 2009 «taking stock: inventory of alien species in the Mediterranean Sea». Biol. Invasions 12 (9), 3379–3381.
Zenetos, A., Arianoutsou, M., Bazos, I., Balopoulou, S., Corsini-Foka, M., Dimiza, M., Drakopoulou, P., Katsanevakis, S., Kondylatos, G., Koutsikos, N., et al., 2015. ELNAIS: a collaborative network on Aquatic Alien Species in Hellas (Greece). Manag. Biol. Invasions 6 (2), 185–196.
Zenetos, A., Gofas, S., Morri, C., Rosso, A., Violanti, D., Garcia Raso, J., Cinar, M., Almogi-Labin, A., Ates, A., Azzurro, E., et al., 2016. Alien species in the Mediterranean Sea by 2012. A contribution to the application of European Union's Marine Strategy Framework Directive (MSFD). Part 2. Introduction trends and pathways. Mediterr. Mar. Sci. 13, 328–352.