The distribution of an invasive fish species is highly affected by the presence of native fish species: evidence based

on species distribution modelling

Pieterjan Verhelst . Pieter Boets . Gerlinde Van Thuyne . Hugo Verreycken .

Peter L. M. Goethals . Ans M. Mouton

Abstract Topmouth gudgeon (*Pseudorasbora parva*) is one of the most invasive aquatic fish species in Europe and causes adverse effects to ecosystem structure and functioning. Knowledge and understanding of the species’ interactions with the environment and with native fish are important to stop and prevent the further spread of the species. Creating species distribution models is a useful technique to determine which factors influence the occurrence and abundance of a species. We applied three different modelling techniques: gen- eral additive models, random forests and fuzzy habitat suitability modelling (FHSM) to assess the habitat suitability of topmouth gudgeon. The former two techniques indicated that the abundance of native fish (i.e. biotic variables) was more important than environ- mental variables when determining the abundance of

.

P. Verhelst (&) · P. Boets · P. L. M. Goethals Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, Jozef Plateaustraat 22, 9000 Ghent, Belgium

e-mail: Pieterjan.Verhelst@UGent.be

P. Verhelst

Research Group Marine Biology, Ghent University, Krijgslaan 281, 9000 Ghent, Belgium

G. Van Thuyne · H. Verreycken

Research Institute for Nature and Forest (INBO), Duboislaan 14, 1560 Hoeilaart, Belgium

topmouth gudgeon in Flanders (Belgium). Bitterling (*Rhodeus amarus*), stone loach (*Barbatula barbatula*), three-spined stickleback (*Gasterosteus aculeatus*) and predator abundance were selected as the most important biotic variables and implemented in the FHSM to investigate species interactions. Depending on the preferred food source and spawning behaviour, either coexistence or interspecific competition can occur with bitterling, stone loach and three-spined stickleback. In contrast, the presence of predators clearly had a top down effect on topmouth gudgeon abundance. These findings could be applied as a biological control measure and implemented in conservation strategies in order to reduce the abundance of earlier established populations of topmouth gudgeon.

Keywords Topmouth gudgeon · Non-native · Belgium · Biotic resistance · Species distribution modelling · Invasive fish species

P. Verhelst · A. M. Mouton

Research Institute for Nature and Forest (INBO), Kliniekstraat 25, 1070 Brussels, Belgium

P. Boets

Provincial Centre of Environmental Research, Godshuizenlaan 95, 9000 Ghent, Belgium

Introduction

Globalization enables organisms to extend their dis- tribution area to regions, which would be impossible to colonize without human help (McKinney and Lock- wood [1999](#_bookmark53)). Newly introduced species (i.e. alien species) that become invasive can pose serious economic and ecological consequences (Kolar and Lodge [2001](#_bookmark72); Andersen et al. [2004](#_bookmark16)). They can cause shifts in ecological communities (Andersen et al. [2004](#_bookmark16); McKinney and Lockwood [1999](#_bookmark53)) thereby affect- ing ecosystem structure and functioning. By compet- ing with (economically important) native species, they can severely affect the density of different species within a certain ecosystem, resulting in environmental and economic costs (Pimentel et al. [2005](#_bookmark83)).

Topmouth gudgeon (*Pseudorasbora parva* (Tem- minck and Schlegel, 1842)) is one of the most widespread aquatic invasive species in Europe. Native to East Asia, the fish was introduced in Europe in the 1960s and has since then been spreading to many European countries including Belgium (since 1992) (Banarescu [1990](#_bookmark21); Pinder et al. [2005](#_bookmark81); Verreycken et al. [2007](#_bookmark85)). Nowadays, when considering all European countries the species is most widespread in Flanders (Verreycken et al. [2007](#_bookmark85)). Topmouth gudgeon has a very broad habitat range and is therefore a successful invasive species with a high colonization speed (Arnold [1990](#_bookmark17); Beyer et al. [2007](#_bookmark23); Britton et al. [2007](#_bookmark33)). However, the species seems to have a preference for shallow, stagnant waters with vegetation (Pollux and Korosi [2006](#_bookmark86); Beyer et al. [2007](#_bookmark23); Kapusta et al. [2008](#_bookmark69); Gozlan et al. [2010](#_bookmark46); Onikura and Nakajima [2013](#_bookmark70)).

Being an invasive species, topmouth gudgeon com- petes with native species (Beyer et al. [2007](#_bookmark23)). Pollux and Korosi ([2006](#_bookmark86)) suggest topmouth gudgeon might compete with limnophillic species like bitterling (*Rhodeus amarus* (Bloch, 1782)), rudd (*Scardinius erythrophthalmus* (Linnaeus, 1758)), sunbleak (*Leucaspius delineates* (Heckel, 1843)), nine-spined stickleback (*Pungitius pungitius* (Linnaeus, 1758)) and tench (*Tinca tinca* (Linnaeus, 1758)) as well as some eurytopic species, which share for at least part of their lifecycle the same habitat as topmouth gudgeon: perch (*Perca fl atilis* (Linnaeus, 1758)), roach (*Rutilus rutilus* (Linnaeus, 1758)), bream (*Abramis brama* (Linnaeus, 1758)), three-spined stickleback (*Gasterosteus aculeatus* (Lin- naeus, 1758)), common carp (*Cyprinus carpio* (Lin- naeus, 1758)) and pike (*Esox lucius* (Linnaeus, 1758)).

In general, topmouth gudgeon can have a negative impact on native species in three ways. Firstly, being an opportunistic feeder, interspecific competition for food with native species has been observed (Rosecchi et al. [1993](#_bookmark91); Declerck et al. [2002](#_bookmark24); Britton et al. [2007](#_bookmark33); Quyen Tran et al. [2015](#_bookmark88)). Secondly, the species preys upon juveniles and eggs of native fish (Xie et al. [2000](#_bookmark92)). Thirdly, given the colonizing capabilities of topmouth gudgeon and its opportunistic feeding behaviour, this species could play a major role in the transmission of parasites and diseases, especially to larger predators (Rosecchi et al. [1993](#_bookmark91); Gozlan et al. [2005](#_bookmark44), [2010](#_bookmark46); Pinder et al. [2005](#_bookmark81)). All of these negative impacts can result in a disruption of the ecosystem balance (Britton et al. [2010b](#_bookmark38); Gozlan et al. [2010](#_bookmark46)).

Because of the invasive character and potential adverse effects to native fish species (Pinder et al. [2005](#_bookmark81); Pollux and Korosi [2006](#_bookmark86); Verreycken et al. [2007](#_bookmark85)) it is important to assess and predict the habitats suitable to invasion by topmouth gudgeon. Species distribution models (SDMs) are useful tools to analyse species environment relationships (Mouton et al. [2008](#_bookmark59); Boets et al. [2014](#_bookmark25)). The current state of the art in SDMs promotes to apply different modelling techniques and compare their outputs in order to make robust predic- tions about the distribution of the species (Elith et al. [2006](#_bookmark29); Elith and Graham [2009](#_bookmark27); Fukuda et al. [2013](#_bookmark39)). Several data-driven techniques have been previously used to model the habitat suitability of invasive fish species. Random forests (RF) has proven to be an efficient technique to model a species’ distribution, as it does not overfit the data and it assess the importance of each variable (Breiman [2001](#_bookmark30); Fukuda et al. [2013](#_bookmark39)). Another technique, that is commonly used in species distribution studies (Guisan et al. [2002](#_bookmark54)), are general additive models (GAMs), which are an extension of General Linear Models and provide for a flexible curve- fitting approach (Hastie and Tibshirani [1990](#_bookmark55)). This technique uses non-parametric, data-defined smoothers to fit non-linear functions and is particularly popular in aquatic species distribution modelling as it allows to generate the habitat preference curves commonly applied in physical habitat modelling (Bovee [1982](#_bookmark28)). Besides these more classic data-driven modelling techniques, fuzzy habitat suitability models (FHSMs) have gained a lot in popularity lately. These models take into account the inherent uncertainty of ecological variables and enable to describe non-linear relations

between ecological variables in a transparent way. Fuzzy models use linguistic descriptions such as ‘low’, ‘moderate’ or ‘high’ for the quantification of variables together with ecological expert knowledge to transform these descriptions into a mathematical framework in which data processing can be performed (Kampichler et al. [2000](#_bookmark67)). A big advantage of FHSMs is their transparency: the fuzzy rules clearly show the interac- tion between different input variables and the output variable (Mouton et al. [2011](#_bookmark65), [2012](#_bookmark68)). This aspect may be important if optimal models are to be applied within a management and conservation context.

Studies about the invasive character and distribu- tion of topmouth gudgeon have mainly focussed on the importance of environmental and hydromorpho- logical variables (Pollux and Korosi [2006](#_bookmark86); Kapusta et al. [2008](#_bookmark69); Gozlan et al. [2010](#_bookmark46); Fukuda et al. [2011](#_bookmark36), [2013](#_bookmark39); Onikura and Nakajima [2013](#_bookmark70)). Only a few papers take biotic variables into account (Declerck et al. [2002](#_bookmark24); Beyer et al. [2007](#_bookmark23); Britton et al. [2010b](#_bookmark38)). However, these were merely descriptive and used to ascribe the success of the species and were not implemented in SDMs. Biotic variables such as species abundance and interactions have recently proven to be valuable predictors in species distribu- tion modelling (Guisan and Thuiller [2005](#_bookmark49); Heikkinen et al. [2007](#_bookmark58); Elith and Leathwick [2009](#_bookmark31); Howard et al. [2014](#_bookmark60)). These variables may even be more important for invasive species, as these species are very opportunistic and generally have low physico-chem- ical habitat requirements.

This study aims to analyse to what extent envi- ronmental and biotic variables affect the distribution of topmouth gudgeon in lowland waterways in Flanders, Belgium, which we used as a case study for lowland waterways in Western Europe. Three data-driven methods were applied to assess the distribution of topmouth gudgeon: GAMs, RF and FHSMs. We compared the importance of including environmental and biotic variables to improve model reliability and investigated the interactions between topmouth gudgeon abundance and key environmen- tal and biotic variables. Based on our results, biotic variables show an important relationship with top- mouth gudgeon abundance and could help to better underpin management plans to avoid new introduc- tions and to confi e the future spread of this highly invasive fi species.

Materials and methods

Study area

Flanders, the northern part of Belgium, is highly urbanised and characterised by slow flowing lowland rivers and channels, which are part of the drainage basins of the rivers Yser, Scheldt and Meuse. The latter three drainage basins are divided into 11 Flemish river catchments, which in turn are divided into 102 sub- basins. Water quality is very variable within the different catchments, as indicated by a wide range in conductivity (102–1893 lS cm-1), turbidity (0.82–750

FTU) and dissolved oxygen (0.80–18 mg l-1). The

sampled waterways in this study varied in average depth between 0.05 and 1.84 m. Notably, only for waters where wading was possible average depth could be measured, so that average depth was not measured when a boat was used for sampling. Width varied between 0.25 and 337 m and flow velocity ranged from 0 to 11.7 m s-1 (Table [1](#_bookmark0)). Every year, fish are stocked on a selection of waterbodies for sport fishing purposes, which facilitates the distribution of invasive aquatic species (Banarescu [1990](#_bookmark21); Copp et al. [2005](#_bookmark13)).

Data collection

Between 1993 and 2013, 563 aquatic habitats were sampled by the Research Institute for Nature and Forest (INBO) in a standardised way during fish stock assess- ments as part of the Flemish Freshwater Fish Monitoring Network (Brosens et al. [2015](#_bookmark40)) (Fig. [1](#_bookmark1)). Fish were sampled by one-pass electrofi g with a 5 kW gener- ator using a control box to convert 230 V alternating current into an adjustable output voltage of 300–500 V direct current and a pulse frequency of 480 Hz. Depending on river conductivity, an electric current between 2 and 9 A was generated. Habitats were discretely approached and subjected to sampling from a small boat or by wading in an upstream direction. Depending on the river width, the number of electrofi ing devices and the number of hand-held anodes (0.3 m diameter) varied between 1 and 3. All fish were identified to species level, measured [total length (TL), 0.1 cm accuracy], weighed [wet weight (W), 0.1 g accuracy] and catch per unit effort (CPUE) estimates were made to determine species abundance. After sampling, all fish were returned to the water at the same site of collection.

Table 1 Overview of the 14 selected variables after collinearity analysis, to be included in the model development: eight envi- ronmental and six biotic variables

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Type | Minimum | Maximum |
| Muddy river bed | Environmental | NA | NA |
| Riparian habitat status | Environmental | NA | NA |
| Aquatic vegetation Conductivity (lS m-1) | Environmental Environmental | NA 102 | NA 1893 |
| Average depth (m) | Environmental | 0.05 | 1.84 |
| Flow velocity (m s-1) | Environmental | 0.00 | 11.70 |
| Turbidity (FTU) | Environmental | 0.82 | 750 |
| Oxygen concentration (mg l-1) | Environmental | 0.80 | 18.13 |
| Stone loach abundance | Biotic | 0 | 427 |
| Bitterling abundance | Biotic | 0 | 124 |
| Three-spined stickleback abundance | Biotic | 0 | 1213 |
| Spined loach abundance | Biotic | 0 | 33 |
| Sunbleak abundance | Biotic | 0 | 8 |
| Predator abundance | Biotic | 0 | 90 |

Muddy river bed, riparian habitat status and aquatic vegetation are categorical variables. As such, no minimum and maximum value is given. Muddy river bed and aquatic vegetation have two categories present and absent, whereas riparian habitat status is divided into the categories absent, medium developed and well developed. The species abundance is expressed as CPUE

At each sampling site, environmental water quality variables were recorded before the start of fi sampling and habitat characteristics were recorded. In total, 49 biotic and 31 environmental variables were measured (see Table [6](#_bookmark10) in ‘‘[Appendix](#_bookmark9)’’). The environmental vari- ables included: conductivity, average depth, pH, flow velocity, water temperature, turbidity, transparency, oxygen concentration and oxygen saturation, but also habitat characteristics such as river bed substrate, riparian land use, connectivity (i.e. presence of migration barri- ers), presence of cover habitat, presence of aquatic vegetation, sinuosity, habitat heterogeneity and riparian habitat status. Although mainly environmental variables are used to build habitat suitability models and determine a species’ distribution, including biotic variables (e.g. fi abundance) can lead to better modelling results (Heikki- nen et al. [2007](#_bookmark58); Gonza´lez-Salazar et al. [2013](#_bookmark41)).

The 49 collected biotic variables comprised of 46 fish species expressed in catch per unit effort and three variables identifying biotic guilds: predators, pelagic fish and benthic fish.

Data preparation

In total, a database of 2671 records was collected during the period 1993–2013. However, many records were incomplete regarding environmental data. Therefore,

records with more than 60 % of variable information missing were excluded (Zuur et al. [2010](#_bookmark96)). This resulted in a dataset with 554 records (2117 records were removed) of which 163 contained topmouth gudgeon. By this data process, habitats deeper than two metres were excluded. Collinearity analysis was conducted on the 80 variables to avoid collinearity (Zuur et al. [2009](#_bookmark93), [2010](#_bookmark96)) (Table [1](#_bookmark0)).

To investigate which of the selected variables obtained by the collinearity analysis had an impact on the distribution of topmouth gudgeon, generalized additive modelling (GAM) and random forests (RF) were applied. In addition to these two classic data- driven modelling techniques, a fuzzy habitat suitabil- ity model (FHSM) was applied on the four most important biotic variables resulting from the GAM and RF. FHSM is able to depict interactions between the input variables and the output variable (Mouton et al. [2011](#_bookmark65), [2012](#_bookmark68)) (Fig. [2](#_bookmark2)). The models were evaluated based on the determination coefficient (R2), the percentage of correctly classified instances (CCI) and kappa statistic (j). The kappa index ranges between -1 and 1. Positive kappa values mean that the model is better than a random model whereas zero or negative values mean that the model is equal to or not better than a random model. Cohen’s kappa is a reliable performance measure, since the effect of



Fig. 1 Map of Flanders with indication of the main rivers and

563 sampling locations where fish stock assessments were undertaken between 1990 and 2013 by the Research Institute for

Nature and Forest (INBO). *Black dots* represent locations where topmouth gudgeon was present, whereas *white dots* are locations where the fish is absent

prevalence on Cohens’s kappa appeared to be negli- gible (Dedecker et al. [2004](#_bookmark26); D’heygere et al. [2006](#_bookmark18); Goethals et al. [2007](#_bookmark42)). Therefore, kappa provides a more reliable representation of model performance (Cohen [1960](#_bookmark14)). It should be noted that invasive species have a very broad habitat range and are therefore hard to model. As such, model accuracy might be below good model estimates (CCI \ 70 %; j \ 0.75;

R2 « 1). In this paper we focussed on the variable

importance output of two different modelling tech- niques, GAM and RF, to find out which variables contribute the most to the habitat suitability of topmouth gudgeon. In a next step we studied how these variables were related to habitat suitability of topmouth gudgeon by FHSM. In the scope of this paper, we tried to find a trade-off between model accuracy and expert knowledge.

Model development

*Generalized additive modelling*

Generalised additive modelling (GAM) was applied to find which variables determined the abundance of topmouth gudgeon. GAM is an extension of gener- alised linear modelling (GLM), which is itself an extension of ordinary linear regression. Like in GLM, each class of GAM is defined in terms of a link function that specifies the relationship between the mean of the output variable and the predictor (the sum of the effects of the individual input variables), and a variance function which relates the variance of the output variable to its mean (Zuur et al. [2009](#_bookmark93)). Due to the linear nature of the predictor in the link function, the flexibility of GLMs for modelling species



Fig. 2 Schematic presentation of the workflow. Collinearity analysis was conducted to avoid collinearity. By means of GAM and RF, four biotic variables were selected. These four variables were implemented in a FHSM by fuzzy sets

responses to environmental variables is limited. GAM provides a flexible solution to this problem (Yee and Mitchell [1991](#_bookmark94); Leathwick [1995](#_bookmark76)). In models fitted using GAM, the effect of each input variable is specified as a non-parametric smooth function, esti- mated from the data using techniques originally developed for smoothing scatter plots (most com- monly cubic splines). GAM determines the shape of the response curve based on the observed data and thus fewer assumptions are made about how species respond to their environment.

GAM selects variables based on the Akaike weight procedure (Burnham and Anderson [2002](#_bookmark43)). The models were optimised by means of a stepwise backward elimination. When plotting the residuals, no patterns

were found. As such, the model approved to be valid (Zuur et al. [2009](#_bookmark93)). We used the ‘mgcv’ package (Wood [2011](#_bookmark90)) of the R environment (R Development Core Team [2014](#_bookmark90)), in which the default settings were used (i.e. logit link function under a binomial error distribu- tion) and the restricted maximum likelihood (REML) method was applied for smoothness selection.

*Random forests*

Random forests is a widely applied machine learning technique (Breiman [2001](#_bookmark30); Cutler et al. [2007](#_bookmark19); Peters et al. [2007](#_bookmark77)). This approach creates *k* classification trees that are aggregated, based on majority voting. Each tree is grown based on another bootstrap subset of the training dataset and considers the best split predictive variable among a subset of *m* randomly chosen variables to split each node. Both *k* and *m* are user- defined parameters that should be optimised to ensure convergence of the developed models (Peters et al. [2007](#_bookmark77)). One important feature of RF is the availability of some measures to assess the importance of each variable and to detect outliers. Breiman ([2001](#_bookmark30)) showed that Bayes consistency is achieved with a simple version of RF and claimed that RF rarely overfits the data. Several studies showed that RF often outperform other state-of-the-art modelling techniques such as decision trees or support vector machines (Cutler et al. [2007](#_bookmark19); Fukuda et al. [2013](#_bookmark39)). We used the ‘ran- domForest’ package (Liaw and Wiener [2002](#_bookmark47)) of the R environment (R Development Core Team [2014](#_bookmark90)), in which the default settings were applied. To know the individual variable importance, the mean decrease in accuracy was calculated.

*Fuzzy logic*

Based on the results of the GAM and the mean decrease in accuracy of the RF, four biotic variables were selected as input for the FHSM to analyse interactions (Fig. [2](#_bookmark2)): stone loach (*Barbatula barbatula* (Linnaeus, 1758)), bitterling, three-spined stickleback and preda- tor abundance. The linguistic values ‘low’ and ‘high’ were assigned to the four input variables and output variable (habitat suitability). These linguistic values were defined by fuzzy sets (Zadeh [1965](#_bookmark95)) and not by conventional sets with crisp boundaries (hereafter called crisp sets). A membership function of a partic- ular fuzzy set indicates the degree to which an element

belongs to this fuzzy set, with membership ranging from zero to one. This fuzzy approach with overlap- ping set boundaries is in line with the ecological

Table 2 The selected measured variables at each sampling campaign, the linguistic values assigned to the input variables of the habitat suitability models and the fuzzy sets describing

these linguistic values

gradient theory (Strayer et al. [2003](#_bookmark78)), because an

element can partially belong to a fuzzy set and thus have a membership degree to this fuzzy set ranging

Variable Linguistic

value

Fuzzy set

parameters

from zero to one. Consequently, the linguistic state- ment ‘the depth is quite low but tending to be moderate’ can for instance be translated into a depth which has a membership degree of 0.4 to the fuzzy set ‘low’ and of

Stone loach L (0, 0, 15, 25)

H (15, 25, 430, 430)

Bitterling L (0, 0, 1, 3)

H (1, 3, 125, 125)

0.6 to the fuzzy set ‘moderate’. In this study, all membership functions had trapezoidal shapes and were defined by four parameters (am, bm, cm and dm): the

Three-spined stickleback

L (0, 0, 20, 30)

H (20, 30, 1220,

1220)

membership degree linearly increases between am and

bm from 0 to 1, is equal to 1 between bm and cm and linearly decreases from 1 to 0 between cm and dm. A triangular membership function is obtained when bm equals cm. The parameters of the membership func-

Predators L (0, 0, 2, 4)

H (2, 4, 91, 91)

Habitat suitability L (0, 0, 3, 5)

H (3, 5, 130, 130)

Each fuzzy set is defined by four parameters (a , b , c

and

m m m

tions corresponding to the fuzzy sets used in this work

are given in Table [2](#_bookmark3) and were optimised to create a uniform distribution of the input variables over the fuzzy sets. The Shannon–Weaver entropy (Shannon and Weaver [1963](#_bookmark97)), of which a threshold of 0.85 was applied, quantified this uniformity and was used as an optimisation criterion to increase the quality of the fuzzy sets (Mouton et al. [2009b](#_bookmark63)). The parameters of the membership functions corresponding to the fuzzy sets of the input variables have often been derived from expert knowledge (Mouton et al. [2007](#_bookmark56)). However, this may result in an input variable described by fuzzy sets that contains very few data points.

The fuzzy rule base relates the input variables to the habitat suitability for topmouth gudgeon and consists of ‘if–then’ rules, such as ‘IF bitterling abundance IS high AND three-spined stickleback abundance IS high AND predator abundance IS low THEN habitat suitability IS high’. The if-part of the rule (the antecedent) describes in which situation this rule applies, while the then-part (the consequent) indicates whether the habitat in this situation is suitable or not for topmouth gudgeon. Given crisp values of the four input variables, the output of the fuzzy model is calculated as described by Van Broekhoven et al. ([2006](#_bookmark82)). For each instance, we calculated the member- ship degrees to the fuzzy sets of the input variables. The degree of fulfilment of each rule was then calculated as the minimum of the membership degrees in its antecedent. Finally, to each linguistic output value a fulfilment degree was assigned equal to the

dm): the membership degree linearly increases between am and bm from 0 to 1, is equal to 1 between bm and cm and linearly decreases from 1 to 0 between cm and dm

maximum of the fulfilment degrees of all rules with the output value under consideration in their consequent. The approach is similar to the Mamdani–Assilian procedure (Assilian [1974](#_bookmark20); Mamdani [1974](#_bookmark50)) in which the fuzzy output is defuzzified in a crisp one based on the fuzzy sets of the output variables. However, in this work a different type of model was applied: a fuzzy classifier. The modelled values of the output variable were assigned to the fuzzy set with the highest fulfilment degree, which allowed comparison of the modelled output with the observed output and calcu- lation of performance measures (Van Broekhoven et al. [2006](#_bookmark82)). Specifically, if a value of habitat suitability had 0.6 membership in fuzzy set ‘absent’ and 0.4 membership in fuzzy set ‘present’, then the value was assigned to ‘absent’.

The adjusted average deviation (Van Broekhoven et al. [2006](#_bookmark82)) was applied because it incorporates the specific characteristics of fuzzy classifiers with an ordered set of output classes and can deal with the fuzzy outputs of these models. Specifically, several performance criteria have been developed to evaluate and train presence/absence models, but most of these criteria are based on the confusion matrix which requires a threshold to distinguish between present and absent predictions. Since these criteria cannot deal with the fuzzy output of a fuzzy ordered classifier,

valuable information may be lost by transferring this fuzzy output to the crisp output which is needed to generate the confusion matrix. Performance measures which are derived from the confusion matrix, for instance, are not sensitive to the position of the classes where the wrong classification occurs (Van Broekho- ven et al. [2006](#_bookmark82)). For further details about fuzzy rule- based modelling and rule base optimisation we refer to Mouton et al. ([2009a](#_bookmark61), [2011](#_bookmark65), [2012](#_bookmark68)).

To generate a reliable habitat suitability model, the consequents of the fuzzy rules were optimised using a nearest ascent hill-climbing algorithm (Michalewicz and Fogel [2000](#_bookmark57)). Starting from fixed fuzzy sets (Table [2](#_bookmark3)) and a randomly selected rule base, the consequent of one rule is changed into the linguistic value of one of its neighbouring fuzzy sets (e.g. the replacement of ‘moderate’ by ‘low’ or ‘high’) and the impact on model performance is calculated. If model performance increases, the algorithm continues with the adjusted rule base, if not, it continues with the original one. Each training iteration was stopped when no further increase of the performance criterion on the test fold was observed. Each training iteration was repeated and the new rule base was compared to the rule bases obtained from each of the previous itera- tions. The resulting rule base similarity was the percentage of rule consequents that were identical for two rule bases. If the rule base with the highest performance on the test fold was obtained 3 times, this rule base was selected as the final rule base and training continued on another fold (Mouton et al. [2009a](#_bookmark61)). Ten- fold cross-validation was applied to indicate the robustness of the optimisation results. The folds were constructed by randomising the original data set and assigning each data point to one fold without replace- ment. The species prevalence (i.e. the frequency of occurrence) was constant for all tenfolds and equal to the prevalence of the original dataset.

Models were trained based on Cohen’s kappa (Cohen [1960](#_bookmark14)), which ranges from -1 to 1 and is derived from the confusion matrix (Fielding and Bell [1997](#_bookmark34); Manel et al. [2001](#_bookmark51)). This measure was selected as a training performance criterion because model training based on kappa showed good results in a previous study using fuzzy logic (Mouton et al. [2008](#_bookmark59)). The training algorithm was described into detail in Mouton et al. ([2008](#_bookmark59)). All fuzzy algorithms were implemented in a C# toolbox (FISH 3.7, freely available upon request).

Results

Collinearity analysis

The collinearity analysis resulted in a selection of eight out of 31 environmental variables used for model development: muddy river bed, riparian habitat status, presence of aquatic vegetation, conductivity, average depth, flow velocity, turbidity and oxygen concentra- tion. Considering the biotic variables, six variables were found not to be collinear: stone loach, bitterling, three-spined stickleback, spined loach (*Cobitis taenia* (Linnaeus, 1758)), sunbleak and predator abundance (see Tables [7](#_bookmark11), [8](#_bookmark12) in ‘‘[Appendix](#_bookmark9)’’). The collinearity analysis showed that bitterling abundance was strongly correlated with pelagic fish species abundance. In contrast, stone loach abundance was correlated with benthic fish abundance. As such, different habitats (i.e. pelagic and benthic habitat) were taken into account in this study.

General additive model

The selection of 14 variables resulting from the collinearity analysis was used as input variables for the GAM. Variable importance could be assessed based on a purely data-driven method. The five most important variables of the full model (R2 = 0.135, AIC = 4102.599) in order of importance were bitter- ling abundance (*p* \ 0.001), three-spined stickleback abundance (*p* \ 0.001), stone loach abundance (*p* = 0.005), conductivity (*p* = 0.34) and flow veloc- ity (*p* = 0.30). After a stepwise backward elimination, a model (R2 = 0.133, AIC = 4089.107) with three significant biotic variables (*p* \ 0.05) was found: bitterling abundance (*p* \ 0.001) was evaluated as the most important variable, followed by three-spined stickleback (*p* \ 0.001) and stone loach (*p* = 0.0031). However, the model with the highest performance (AIC = 4087.729) contained the additional variables predator abundance (*p* = 0.40) and flow velocity (*p* = 0.27). Nonetheless, these were not significant. Although the accuracy of the model is rather low, only biotic variables were selected. Based on expert knowledge and literature (Pollux and Korosi [2006](#_bookmark86); Beyer et al. [2007](#_bookmark23)), the output of the model will be used in the further steps of this study (Fig. [2](#_bookmark2)).

10

8

Mean decrease in accuracy

6

4

2

0

-2

-4



Variable

Fig. 3 The mean decrease in accuracy based on RF for biotic and abiotic variables

Random forests

Similar to the GAM, the 14 selected variables resulting from the collinearity analysis were used as input variables for the RF. The mean decrease in accuracy was applied as a measure of variable importance. The biotic variables bitterling, stone loach, three-spined stickleback and predator abundance were by far selected as most important variables to determine topmouth gudgeon abundance (Fig. [3](#_bookmark4)). They had a mean decrease in accuracy higher than three, whereas the other 11 variables had a value below two. It should be noted that the accuracy of the model with these four variables is rather low (R2 = 0.387), but based on expert knowl- edge, literature and the output of the GAM, the four most contributing variables were used as an input for the fuzzy habitat suitability model. Bitterling abundance explained most of the variation. Stone loach abundance was more important than three-spined stickleback abundance. RF confirmed the importance of predator abundance, which was not selected by the GAM. Amongst the environmental variables, turbidity, depth and flow velocity were the most important. However,

their standard error was similar, so no ranking could be performed.

Fuzzy habitat suitability model

According to the GAM and RF, bitterling, stone loach, three-spined stickleback and predator abundance were the most important variables explaining topmouth

gudgeon abundance (Table [3](#_bookmark5)). As such, these four variables were used in the FHSM to examine species interactions. A stepwise forward selection approved bitterling abundance to be the most important variable (j = 0.27), followed by predator abundance (j = 0.25), stone loach abundance (j = 0.24) and three-spined stickleback abundance (j = 0.15).

The fuzzy rule base of the optimal fuzzy model (j = 0.50) illustrated that if three out of four biotic variables were high, topmouth gudgeon abundance was high as well. However, this was not the case for two fuzzy rules. When one or two biotic variables were high and one of them involved bitterling abundance, this resulted in high topmouth gudgeon abundance. In contrast, one fuzzy rule described a high abundance of topmouth gudgeon, whereas bitterling abundance was low. In this case, stone loach and three-spined stickle- back had a high abundance and predator abundance was found to be low (Table [4](#_bookmark6)). This illustrates that multiple species interactions were not straightforward and difficult to interpret. As such, confusion matrices were calculated to clarify the single species interactions (Table [5](#_bookmark7)). Bitterling abundance reflected topmouth gudgeon abundance for 87.5 % of the rules. For the other 12.5 % topmouth gudgeon and bitterling had an opposite trend in abundance. This was found to a lesser extent for stone loach: in 62.5 % of the rules both had the same abundance value and in 37.5 % of the cases an opposite trend was observed. The interaction between three-spined stickleback and topmouth gudgeon was less straightforward as only 50 % of the species’

Table 3 Overview of the variables that best explained the distribution of topmouth gudgeon based on three different

 modeling techniques RF, GAM and FHSM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Stone | Bitterling | Three-spined | Predators |
| loach |  | stickleback |  |
| RF | 2 | 1 | 3 | 4 |
| GAM | 3 | 1 | 2 | 4 |
| FHSM | 3 | 1 | 4 | 2 |

A ranking was given to the different variables, with the lowest value (1) being the most important and the highest value (4) the least important

abundance reflected topmouth gudgeon abundance. Considering predator abundance, an inverse relation- ship was found: when predator abundance was low, habitat suitability was high and vice versa (62.5 % of the rules). Most of the observations were in line with our results (Fig. [4](#_bookmark8)). However, a few extreme values

suggested a trend towards dominance of one of both species. FHSM considered these as outliers, hence, they did not contribute to the output.

Discussion

Variable importance

In the current study, we assessed the importance of environmental and biotic variables (i.e. native species abundance) to determine the distribution of topmouth gudgeon based on habitat suitability mod- els. We found that, based on different modelling techniques, the distribution of the species was mainly determined by biotic variables. Previous research on the distribution of topmouth gudgeon solely focused on the importance of environmental variables. According to literature, the species prefers shallow

Table 4 The fuzzy rule base of the optimal fuzzy model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Stone loach | Bitterling | Three-spined stickleback | Predators | Habitat suitability | Rule coverage |
| L | L | L | L | L | 278 |
| H | L | L | L | L | 45 |
| L | H | L | L | H | 4 |
| H | H | L | L | H | 1 |
| L | L | H | L | L | 85 |
| H | L | H | L | H | 29 |
| L | H | H | L | H | 8 |
| H | H | H | L | H | 4 |
| L | L | L | H | L | 44 |
| H | L | L | H | L | 26 |
| L | H | L | H | H | 1 |
| H | H | L | H | H | 4 |
| L | L | H | H | L | 7 |
| H | L | H | H | L | 12 |
| L | H | H | H | L | 3 |
| H | H | H | H | H | 3 |

The number of habitats that covered each environmental condition is shown in the last column (rule coverage)

*L* low, *H* high

Table 5 Confusion

Bitterling Three-spined stickleback Stone loach Predators

matrices derived from the

FHSM output

L H L H L H L H

*Topmouth gudgeon*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| L | 7 | 1 | 4 | 4 | 5 | 3 | 3 | 5 |
| H | 1 | 7 | 4 | 4 | 3 | 5 | 5 | 3 |

Fig. 4 Topmouth gudgeon abundance in function of the four most important biotic variables [stone loach, bitterling, three-spined stickleback (3S) and predator abundance].

Species abundance is expressed as CPUE

lentic and vegetated waters (Arnold [1990](#_bookmark17); Pollux and Korosi [2006](#_bookmark86); Britton et al. [2007](#_bookmark33); Beyer et al. [2007](#_bookmark23); Kapusta et al. [2008](#_bookmark69); Gozlan et al. [2010](#_bookmark46); Onikura and Nakajima [2013](#_bookmark70)). In Flanders, it has already been shown that topmouth gudgeon reaches high densities in these optimal habitats (Verreycken et al. [2007](#_bookmark85)). However, we tested for topmouth gudgeon abun- dance in lotic water, a suboptimal habitat, which is important as distribution corridor. Based on our models, environmental factors did not seem to affect the abundance of topmouth gudgeon. Nonetheless, GAM and RF did show that next to biotic variables, flow velocity could affect the abundance of topmouth gudgeon as the variable still contributed to explain the abundance of the species. In addition, water depth and turbidity were of similar importance when using RF. Related to the optimal habitat of topmouth gudgeon, shallow water and low turbidity promote vegetation growth. The absence of a significant relationship between environmental variables and the abundance of topmouth gudgeon can be attributed to

different reasons. Firstly, alien invasive species have a high phenotypic plasticity and therefore can withstand a variety of environmental conditions (Arnold [1990](#_bookmark17); Beyer et al. [2007](#_bookmark23); Davidson et al. [2011](#_bookmark22)). In this way, environmental variables may play an inferior role in the species’ abundance, whereas biotic interactions might be more important and determine the species’ abundance (Heikkinen et al. [2007](#_bookmark58); Howard et al. [2014](#_bookmark60)). Secondly, topmouth gudgeon recently invaded Belgium and probably did not reach its maximum expansion yet in lotic waters. As such, a delineated relationship with environmen- tal variables is still lacking (Sutherst and Bourne [2009](#_bookmark79)). Thirdly, studies reporting on the optimal habitat where topmouth gudgeon can reach high abundances comprise lentic waters such as pools and shallow lakes (Arnold [1990](#_bookmark17); Britton et al. [2007](#_bookmark33); Kapusta et al. [2008](#_bookmark69); Pollux and Korosi [2006](#_bookmark86)). However, in this study, slow flowing lowland rivers were investigated, which are considered a suboptimal habitat.

Both GAM and RF showed the importance of biotic variables to model the species distribution of topmouth gudgeon. The selected variables affecting topmouth gudgeons’ distribution were bitterling, three-spined stickleback, stone loach and predator abundance. The habitat of these fish species is similar to the habitat of topmouth gudgeon. Bitterling and three-spined stickle- back prefer shallow (up to one meter) lentic waters with vegetation (Hynes [1950](#_bookmark62); Kottelat and Freyhof [2007](#_bookmark74)). Stone loach prefers lotic waters but can be found in a variety of habitats, including lake shores (Kottelat and Freyhof [2007](#_bookmark74); Smyly et al. [1955](#_bookmark77)). Considering native predators, pike (*E. lucius*) and perch (*P. fluviatilis*) are the most important species. Pike is a top predator in Belgian aquatic habitats. It prefers stagnant to slow flowing vegetated waters (Pauwels [2014](#_bookmark72)). A smaller, but more common predator is perch. Perch is a eurytopic species and can manage stable populations in a variety of habitats. As such, both predators can reside in optimal and suboptimal topmouth gudgeon habitat. Although environmental variables were not the most important variables determining the abundance of topmouth gudgeon, they were indirectly taken into account by including fish species (biotic variables), which have similar habitat preferences.

Biotic interactions

Invasive species have a high phenotypic plasticity and can withstand numerous environmental conditions (Davidson et al. [2011](#_bookmark22)), which makes these species difficult to model. Hence, biotic variables and their interactions are important processes in relation to the distribution and abundance of invasive species (Guisan and Thuiller [2005](#_bookmark49); Heikkinen et al. [2007](#_bookmark58); Gonza´lez- Salazar et al. [2013](#_bookmark41); Howard et al. [2014](#_bookmark60)). However, very few SDMs explicitly include biotic factors, which can be used to infer or provide clues about interspecific interactions (Elith and Leathwick [2009](#_bookmark31)). The focus on environmental variables is attributed to the ability of control. Nonetheless, biotic variables can be controlled as well, for example due to fish stock assessments.

The result of the fuzzy model indicates interactions between the abundance of topmouth gudgeon and the four selected biotic variables. As topmouth gudgeon occupies a similar habitat compared to bitterling, stone loach and three-spined stickelback, either coexistence (i.e. niche divergence) or interspecific competition (i.e. niche overlap) can exist. Our results suggest the

former applies to bitterling, since there is a low competition for food due to different feeding habits. Topmouth gudgeon is an omnivorous species (Rosec- chi et al. [2001](#_bookmark95); Gozlan et al. [2010](#_bookmark46)) with a wide dietary plasticity (Jackson and Britton [2014](#_bookmark64)), whereas bitter- ling is considered to be a herbivorous fish (Schmidt and McGurk [1982](#_bookmark98)). Another reason to expect coex- istence is the spawning of bitterling in freshwater mussels as it becomes less susceptible to egg predation by topmouth gudgeon (Aldridge [1999](#_bookmark15)).

The same holds for stone loach and three-spined stickleback (Quyen Tran et al. [2015](#_bookmark88)), nonetheless, the trend is less strong than for bitterling, suggesting a certain level of interspecific competition. This might be explained by the fact that all three species have a preference for chironomid larvae as a food source (Declerck et al. [2002](#_bookmark24); Hynes [1950](#_bookmark62); Smyly [1955](#_bookmark77); Wolfram-Wais et al. [1999](#_bookmark89)). This can force the compet- itively weaker species to rely upon suboptimal food resources (Persson and Greenberg [1990](#_bookmark75)). If the food quality is still substantial or food is not a limiting resource, coexistence can take place. Concerning the habitat, topmouth gudgeon is a bentho-pelagic fish species (Pinder and Gozlan [2003](#_bookmark80)). Stone loach and three-spined stickleback are benthic and pelagic species, respectively (Jurvelius et al. [1996](#_bookmark66); Smyly [1955](#_bookmark77)). As such, an incomplete overlap in habitat exists, allowing some niche divergence and coexistence. The same is true for the spawning habitat. Sufficient spawning habitat is necessary to sustain stable populations. Top- mouth gudgeon shows a wide variability in spawning substrate, including rocks, surfaces of shells and plants and artificial substrates (Gozlan et al. [2010](#_bookmark46); Pinder and Gozlan [2003](#_bookmark80)). This flexible spawning strategy, might allow coexistence. However, the spawning habitats are often aggressively defended by topmouth gudgeon males (Pinder and Gozlan [2003](#_bookmark80)) leading to higher interspecific competition. When topmouth gudgeon increases in numbers, spawning habitat may become scarce and unavailable for other species because of the territorial drift. Yet, three-spined stickleback is a nest guarding species as well. They defend their eggs and young fry aggressively towards intruders (Pressley [1980](#_bookmark87)). As such, they are probably more robust against topmouth gudgeon displacement. This could explain the ambiguous interaction observed in the FHSM as both coexistence and interspecific competition can exist between topmouth gudgeon and three-spined stickle- back, depending on the availability of food and habitat.

If food and habitat resources are scarce, they might be maximally exploited by the most competitive species. Nonetheless, highly productive systems with sufficient food and habitat could allow coexistence between different fish species.

It should be noted that only single species interac- tions were taken into account from the FHSM output. Multiple species interactions are difficult to interpret. Nonetheless, these could influence species abundance. For example, coexistence between topmouth gudgeon and bitterling can result in a high population density of both species. This could reduce predation impact on topmouth gudgeon, as predators of temperate waters are opportunistic. In addition, it is hard to model a few extreme values as these are seen as outliers (Fig. [4](#_bookmark8)). Therefore, the rare situations where one of the biotic variables became abundant and seemed to supress topmouth gudgeon abundance were not brought forward in the FHSM. At the same time, this is valuable information as it might indicate a bottom up effect of native species. Clearly, mechanisms account- ing for biotic interactions within species distribution studies merit more exploration.

Management

Our analysis indicated that predators negatively affect the abundance of topmouth gudgeon (Table [5](#_bookmark7)), sug- gesting a top down effect. As such, stocking native predators to control invasive species may be a cost- effective and sustainable measure. Predatory fish of cold-temperate waters tend to be opportunistic feed- ers, as productivity in these areas is lower and prey abundance depends on season and temperature (Keast [1979](#_bookmark71); MacLean and Magnuson [1977](#_bookmark48); Margenau et al. [1998](#_bookmark52)). This means that the prey species with the highest abundance (often the alien invasive species) will have the highest chance to be predated. Topmouth gudgeon might be highly susceptible to this oppor- tunistic predation behaviour due to its fast recruitment and establishment of a large population within a short time span leading to extremely high abundances. This was confirmed by the study of Musil and Ada´mek ([2007](#_bookmark69)) who found that topmouth gudgeon was the most important prey species in the diet of predators.

In the UK, studies report on the eradication of alien invasive fish species with rotenone, a chemical agent. The disadvantage of this management measure is that the product is not host-specific, which likely causes

adverse effects on native fish (Britton and Brazier [2006](#_bookmark32); Britton et al. [2008](#_bookmark35), [2010a](#_bookmark37)). In addition, the use of rotenone is not robust and re-colonization can appear; even immunity can arise (Orciari [1979](#_bookmark73)). Therefore, another approach is required with a focus on control of abundance, rather than on presence-absence. Moreover, instead of applying chemical agents, a biological approach may have less severe effects on the native flora and fauna. Based on our results, we propose biomanipulation by stocking native predatory fish, which can result in a top-down control. Only a few studies demonstrated the importance of predators and the role of topmouth gudgeon as a prey fish (Musil and Ada´mek [2007](#_bookmark69); Lemmens et al. [2014](#_bookmark45)). In the study of Lemmens et al. (2014) juvenile pike were introduced in ponds colonized by topmouth gudgeon to investigate the predatory effect on the invasive species. This technique appeared to be effective with complete eradication of topmouth gudgeon and almost no reduction in native species. The effect on native fish might be minor as these species have evolved strategies to cope with this predator. Moreover, as Lemmens et al. ([2014](#_bookmark45)) stated, the presence of a native predator in the ecosystem can enhance the biotic resistance of this system. However, juvenile pike reside in the littoral zone (Pauwels [2014](#_bookmark72)), forcing prey fish into the pelagic. A combination with a pelagic predator such as perch might give better results. It should be noted that the studies of Musil and Ada´mek ([2007](#_bookmark69)) and Lemmens et al. (2014) were conducted in the optimal habitat of topmouth gudgeon, so extrapolation to connected running waters should be done with caution. In these situations, fish can move to other habitats and avoid predators more easily. This management measurement might also be applicable to other small invasive species. Recent research of van Kleef and Jongejans ([2014](#_bookmark84)) showed a negative relationship between another inva- sive fish species, pumpkinseed (*Lepomis gibbosus* (Linnaeus, 1758)), and predatory fish. Moreover, information about top-down control to eradicate inva- sive species is scarce, especially in lotic systems.

Conclusion

Our study revealed important insights regarding inter- actions of topmouth gudgeon with its surroundings. Topmouth gudgeon appear to coexist with native species such as bitterling, stone loach and three-spined

stickleback. However, a certain level of interspecific competition might occur with the latter two. Predators showed a negative relation with topmouth gudgeon abundance. This top down effect could be used as a management approach in order to reduce the abundance of earlier established populations of topmouth gudgeon. Hence, such information can be implemented in conservation strategies to reduce the impact caused by alien invasive fish species.

Acknowledgments Pieterjan Verhelst is a recipient of a Ph.D. Grant financed by the Agency for Innovation by Science and Technology in Flanders (IWT) and is affiliated with Ghent University and the Research Institute for Nature and Forest (INBO). Pieter Boets was supported by a postdoctoral fellowship from the Research Foundation Flanders (FWO- Vlaanderen). This work was supported by the Interreg IIa Two Seas project RINSE. The author would like to thank the fishing team of INBO Groenendaal and Tom De Boeck for data extraction.

Appendix

See Tables [6](#_bookmark10), [7](#_bookmark11) and [8](#_bookmark12).

Table 6 A total of 80 variables were sampled, of which 31 environmental and 49 biotic variables

Variable Type

Algae Environmental

Average depth Environmental

Bank Environmental

Bank slope Environmental

Brakisch Environmental

Conductivity Environmental

Curves Environmental

Date Environmental

Disolved oxygen Environmental

Disturbance Environmental

Floating plants Environmental

Flow velocity Environmental

Table 6 continued

Variable Type

Landuse Environmental

Method Environmental

Migration barrier Environmental

Natural hiding Environmental

pH Environmental

Pool Environmental

Riffle Environmental

Slope Environmental

Submersed plants Environmental

Substrate Environmental

Temp Environmental

Tide Environmental

Total mass Environmental

Transact length Environmental

Transact width Environmental

Transparency Environmental

Turbidity Environmental

Waterdepth Environmental

Width Environmental

African catfish (*Clarias gariepinus*) Biotic

Asp (*Aspius aspius*) Biotic

Barbel (*Barbus barbus*) Biotic

Bitterling (*Rodeus amarus*) Biotic

Bleak (*Alburnus alburnus*) Biotic

Bream (*Abramis brama*) Biotic

Brook lamprey (*Lampetra planeri*) Biotic Brown bullhead (*Ameiurus nebulosus*) Biotic Brown trout (*Salmo trutta*) Biotic

Bullhead (*Cottus gobio*) Biotic

Burbot (*Lota lota*) Biotic

Chub (*Squalius cephalus*) Biotic

Common carp (*Cyprinus carpio*) Biotic

Crucian carp (*Carassius carassius*) Biotic

Dace (*Leuciscus leuciscus*) Biotic

European eel (*Anguilla anguilla*) Biotic European weather loach (*Misgurnus fossilis*) Biotic Fathead minnow (*Pimephales promelas*) Biotic Grass carp (*Ctenopharyngodon idella*) Biotic Gudgeon (*Gobio gobio*) Biotic

Table 6 continued

Variable

Muddy river bed

Muddy river bed

Aquatic vegetation

Bank status

0.00E?00

6.75E-01

3.54E-06

Table 8 The correlation matrix of the 11 selected continuous variables as output from the correlation analysis

Type

Ide (*Leuciscus idus*)

Kessler’s goby (*Neogobius kessleri*) Minnow (*Phoxinus phoxinus*)

Nase (*Chondrostoma nasus*)

Nine-spined stickleback (*Pungitius pungitius*) Perch (*Perca fluviatilis*)

Pike (*Esox lucius*)

Prussian carp (*Carassius gibelio*) Pumpkinseed (*Lepomis gibbosus*) Rainbow trout (*Oncorhynchus mykiss*) Roach (*Rutilus rutilus*)

Round goby (*Neogobius melanostomus*) Rud (*Scardinius erythrophthalmus*) Ruffe (*Gymnocephalus cernua*) Schneider (*Alburnoides bipunctatus*)

Silver carp (*Hypophthalmichthys molitrix*) Spined loach (*Cobitis taenia*)

Stone loach (*Barbatula barbatula*) Striped mudminnow (*Umbra pygmaea*) Sunbleak (*Leucaspius delineatus*) Tench (*Tinca tinca*)

Three-spined stickleback (*Gasterosteus aculeatus*)

Topmouth gudgeon (*Pseudorasbora parva*) European catfish (*Silurus glanis*)

Western tubenose goby (*Proterorhinus semilunaris*)

White bream *(Blicca bjoerkna*) Benthic fish

Pelagic fish

Predatory fish

Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic Biotic

Biotic

Biotic Biotic

Biotic

Biotic Biotic Biotic

Biotic

Table 7 The correlation matrix of the three selected categor- ical environmental variables as output from the association analysis

Aquatic vegetation

Bank status

6.75E-01

3.54E-06

0.00E?00

1.29E-01

1.29E-01

0.00E?00

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Conductivity | Average depth | Flow velocity | Turbidity | Oxygen concentration | Stone loach | Bitterling | Three-spined stickleback | Spined loach | Sunbleak | Predators |
| Conductivity | 0.00E?00 | 0.6028206 | 0.2044197 | 7.39E-01 | 2.05E-05 | 1.53E-01 | 8.90E-01 | 0.6610353 | 0.2265934 | 4.13E-02 | 6.86E-01 |
| Average depth | 6.03E-01 | 0 | 0.1624192 | 6.19E-01 | 1.20E-01 | 6.34E-03 | 1.31E-01 | 0.3917137 | 0.5932949 | 1.05E-03 | 1.84E-02 |
| Flow velocity | 2.04E-01 | 0.1624192 | 0 | 6.43E-01 | 9.11E-01 | 4.98E-01 | 6.37E-01 | 0.313881 | 0.6099453 | 5.02E-01 | 6.07E-01 |
| Turbidity | 7.39E-01 | 0.619164 | 0.6426673 | 0.00E?00 | 4.67E-08 | 8.46E-02 | 3.33E-01 | 0.1497487 | 0.4992215 | 8.49E-01 | 9.75E-01 |
| Oxygen concentration | 2.05E-05 | 0.1199582 | 0.9110298 | 4.67E-08 | 0.00E?00 | 1.66E-05 | 9.66E-01 | 0.6675519 | 0.1088053 | 4.75E-01 | 1.17E-02 |
| Stone loach | 1.53E-01 | 0.0063394 | 0.4981108 | 8.46E-02 | 1.66E-05 | 0.00E?00 | 9.83E-01 | 0.011572 | 0.1915946 | 6.80E-02 | 4.45E-05 |
| Bitterling | 8.90E-01 | 0.1314631 | 0.637408 | 3.33E-01 | 9.66E-01 | 9.83E-01 | 0.00E?00 | 0.0837119 | 0.7562474 | 2.43E-09 | 1.26E-01 |
| Three-spined stickleback | 6.61E-01 | 0.3917137 | 0.313881 | 1.50E-01 | 6.68E-01 | 1.16E-02 | 8.37E-02 | 0 | 0.6797438 | 3.23E-01 | 6.12E-01 |
| Spined loach | 2.27E-01 | 0.5932949 | 0.6099453 | 4.99E-01 | 1.09E-01 | 1.92E-01 | 7.56E-01 | 0.6797438 | 0 | 6.96E-02 | 7.79E-01 |
| Sunbleak | 4.13E-02 | 0.0010516 | 0.5023189 | 8.49E-01 | 4.75E-01 | 6.80E-02 | 2.43E-09 | 0.3226018 | 0.0695807 | 0.00E?00 | 2.61E-03 |
| Predators | 6.86E-01 | 0.0183858 | 0.6069418 | 9.75E-01 | 1.17E-02 | 4.45E-05 | 1.26E-01 | 0.6124943 | 0.7794211 | 2.61E-03 | 0.00E?00 |

References

Aldridge DC (1999) Development of European bitterling in the gills of freshwater mussels. J Fish Biol 54:138–151

Andersen MC, Adams H, Hope B, Powell M (2004) Risk assessment for invasive species. Risk Anal 24:787–793

Arnold A (1990) Eingebuergerte Fischarten: Zur Biologie und Verbreitung allochthoner Wildfische in Europa. In: Beyer K, Copp GH, Gozlan RE (2007) Microhabitat use and interspecific associations of introduced topmouth gudgeon *Pseudorasbora parva* and native fishes in a small stream. J Fish Biol 71:224–238

Assilian S (1974) Artificial intelligence in the control of real dynamical systems. London University, London

Banarescu PM (1990) On the history of the dispersal of *Pseu- dorasbora parva* in South-Eastern Europe (Pisces, Cypri- nidae). In: Verreycken H, Anseeuw D, Van Thuyne G, Quataert P, Belpaire C (2007) The non-indigenous fresh- water fishes of Flanders (Belgium): review, status and trends over the last decade. J Fish Biol 71:160–170

Beyer K, Copp GH, Gozlan RE (2007) Microhabitat use and interspecific associations of introduced topmouth gudgeon *Pseudorasbora parva* and native fishes in a small stream. J Fish Biol 71:224–238

Boets P, Pauwels IS, Lock K, Goethals PLM (2014) Using an integrated modelling approach for risk assessment of the ‘killer shrimp’ *Dikerogammarus villosus*. River Res Appl 4:403–412

Bovee KD (1982) A guide to stream habitat analysis using the instream flow incremental methodology. Instream Flow Information Paper 12, US Fish and Wildlife Service, Fort Collins, Colorado

Breiman L (2001) Random forests. Mach Learn 45:5–32 Britton JR, Brazier M (2006) Eradicating the invasive topmouth

gudgeon, *Pseudorasbora parva*, from a recreational fishery in northern England. Fish Manag Ecol 13:329–335

Britton JR, Davies GD, Brazier M, Pinder AC (2007) A case study on the population ecology of a topmouth gudgeon (*Pseudorasbora parva*) population in the UK and the implications for native fish communities. Aquat Conserv Mar Freshw Ecosyst 17:749–759

Britton JR, Brazier M, Davies GD, Chare SI (2008) Case studies on eradicating the Asiatic cyprinid *Pseudorasbora parva* from fishing lakes in England to prevent their riverine dispersal. Aquat Conserv Mar Freshw Ecosyst 18:867–876

Britton JR, Davies GD, Brazier M (2010a) Towards the suc- cessful control of the invasive *Pseudorasbora parva* in the UK. Biol Invasions 12:125–131

Britton JR, Davies GD, Harrod C (2010b) Trophic interactions and consequent impacts of the invasive fish, *Pseudorasb-* *ora parva*, in a native aquatic foodweb: a field investigation in the UK. Biol Invasions 12:1533–1542

Brosens D, Breine J, Van Thuyne G, Belpaire C, Desmet P, Verreycken H (2015) VIS—a database on the distribution of fishes in inland and estuarine waters in Flanders, Bel- gium. ZooKeys 475:119–145

Burnham KP, Anderson DR (2002) Model selection and mul- timodel inference: a practical information-theoretic approach. Springer, New York

Cohen J (1960) A coefficient of agreement for nominal scales.

Educ Psychol Meas 20:37–46

Copp GH, Bianco PG, Bogutskaya NG, Ero¨s T, Falka I, Ferreira MT, Fox MG, Freyhof J, Gozlan RE, Grabowska J, Kova´cˇ V, Moreno-Amich R, Naseka AM, Penˇa´z M, Povzˇ M, Przybylski M, Robillard M, Russell IC, Stake\_ nas S, Sˇ umer

S, Vila-Gispert A, Wiesner C (2005) To be, or not to be, a non-native freshwater fish? J Appl Ichthyol 21:242–262

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT (2007) Random forests for classification in ecology. Ecology 88:2783–2792

D’heygere T, Goethals PLM, De Pauw N (2006) Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks. Ecol Model 195:20–29

Davidson AM, Jennions M, Nicotra AB (2011) Do invasive species show higher phenotypic plasticity than native species and if so, is it adaptive? A meta-analysis. Ecol Lett 14:419–431

Declerck S, Louette G, De Bie T, De Meester L (2002) Patterns of diet overlap between populations of non-indigenous and native fishes in shallow ponds. J Fish Biol 61:1182–1197

Dedecker A, Goethals PLM, Gabriels W, De Pauw N (2004) Optimisation of Artificial Neural Network (ANN) model design for prediction of macroinvertebrate communities in the Zwalm river basin (Flanders, Belgium). Ecol Model 174:161–173

Elith J, Graham CH (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography 32:66–77

Elith J, Leathwick JR (2009) Species distribution models: eco- logical explanation and prediction across space and time. Annu Rev Ecol Evol Syst 40:677–697

Elith J, Graham CH, Anderson RP, Dud´ık M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Sobero´n J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species’ distributions from occurrence data. Ecography 29:129–151

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/ absence models. Environ Conserv 24:38–49

Fukuda S, De Baets B, Mouton AM, Waegeman W, Nakajima J, Mukai T, Hiramatsu K, Onikura N (2011) Effects of model formulation on the optimization of a genetic Takagi– Sugeno fuzzy system for fish habitat suitability evaluation. Ecol Model 222:1401–1413

Fukuda S, De Baets B, Waegeman W, Verwaeren J, Mouton AM (2013) Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. Environ Model Softw 47:1–6

Goethals PLM, Dedecker AP, Gabriels W, Lek S, De Pauw N (2007) Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat Ecol 41:491–508 Gonza´lez-Salazar C, Stephens CR, Marquet PA (2013) Com- paring the relative contributions of biotic and abiotic

factors as mediators of species’ distributions. Ecol Model 248:57–70

Gozlan RE, St-Hilaire S, Feist SW, Martin P, Kent ML (2005) Disease threat to European fish. Nature 435:1046. doi:[10.](http://dx.doi.org/10.1038/4351046a)  [1038/4351046a](http://dx.doi.org/10.1038/4351046a)

Gozlan RE, Andreou D, Asaeda T, Beyer K, Bouhadad R, Burnard D, Calola N, Cakic P, Djikanovic V, Esmaeili HR, Falka I, Golicher D, Harka A, Jeney G, Kova´c V, Musil J, Nocita A, Povz M, Poulet N, Virbickas T, Wolter C, Tarkan AS, Tricarico E, Trichkova T, Verreycken H, Witkowski A, Zhang C, Zweimueller I, Brriton JR (2010) Pan-conti- nental invasion of *Pseudorasbora parva*: towards a better understanding of freshwater fish invasions. Fish Fish 11:315–340

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecol Lett 8:993–1009

Guisan A, Edwards TC, Hastie T (2002) Generalized linear and generalized additive models in studies of species distribu- tions: setting the scene. Ecol Model 157:89–100

Hastie T, Tibshirani R (1990) Generalized additive models.

Chapman and Hall, London

Heikkinen RK, Luoto M, Virkkala R, Pearson RG, Ko¨rber J-H (2007) Biotic interactions improve prediction of boreal bird distributions at macro-scales. Glob Ecol Biogeogr 16:754–763

Howard C, Stephens PA, Pearce-Higgins JW, Gregory RD, Willis SG (2014) Improving species distribution models: the value of data on abundance. Methods Ecol Evol 5:506–513

Hynes HBN (1950) The food of fresh-water sticklebacks (*Gasterosteus aculeatus* and *Pygoteus pungitius*), with a review of methods used in studies of the food of fishes. J Anim Ecol 1:36–58

Jackson MC, Britton JR (2014) Divergence in the trophic niche of sympatric freshwater invaders. Biol Invasions 16:1095–1103

Jurvelius J, Leinikki J, Mamylov V, Pushkin S (1996) Stock assessment of pelagic three-spined stickleback (*Gasteros-* *teus aculeatus*): a simultaneous up- and down-looking echo-sounding study. Fish Res 27:227–241

Kampichler C, Barthel J, Wieland R (2000) Species density of foliage-dwelling spiders in field margins: a simple, fuzzy rule-based model. Ecol Model 129:87–99

Kapusta A, Bogacka-Kapusta E, Czarnecki B (2008) The sig- nificance of stone moroko *Pseudorasbora parva* (Tem- minck and Schlegel) in the small-sized fish assemblages in the littoral zone of the heated Lake Lichenskie. Arch Pol Fish 16:49–62

Keast A (1979) Patterns of predation in generalist feeders. In: Beaudoin CP, Tonn WM, Prepas EE, Wassenaar LI (1999) Individual specialization and trophic adaptability of northern pike (*Esox lucius*): an isotope and dietary analysis. Oecologia 120:386–396

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. Trends Ecol Evol 16:199–204

Kottelat M, Freyhof J (2007) Handbook of European freshwater fishes. Kottelat, Freyhof, Cornol, Berlin

Leathwick JR (1995) Climatic relationships of some New- Zealand forest tree species. J Veg Sci 6:237–248

Lemmens P, Mergeay J, Vanhove T, De Meester L, Declerck SAJ (2014) Suppression of invasive topmouth gudgeon *Pseudorasbora parva* by native pike *Esox lucius* in ponds. Aquat Conserv Mar Freshw Ecosyst 25:41–48

Liaw A, Wiener M (2002) Classification and regression by random forest. R News 2:18–22

MacLean J, Magnuson JJ (1977) Species interactions in percid communities. J Fish Res Board Can 34:1941–1951

Mamdani EH (1974) Application of fuzzy algorithms for control of simple dynamic plant. Proc Inst Electr Eng 121:1585–1588

Manel S, Williams HC, Ormerod SJ (2001) Evaluating pres- ence–absence models in ecology: the need to account for prevalence. J Appl Ecol 38:921–931

Margenau TL, Rasmussen PW, Kampa JM (1998) Factors affecting growth of northern pike in small northern Wis- consin lakes. N Am J Fish Manag 18:1998

McKinney ML, Lockwood JL (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. Trends Ecol Evol 14:450–453

Michalewicz Z, Fogel DB (2000) How to solve it: modern heuristics. Springer, Berlin

Mouton AM, Schneider M, Depestele J, Goethals PLM, De Pauw N (2007) Fish habitat modelling as a tool for river management. Ecol Eng 29:305–315

Mouton AM, Schneider M, Peter A, Holzer G, Mu¨ller R, Goe- thals PLM, De Pauw N (2008) Optimisation of a fuzzy habitat model for spawning European grayling (*Thymallus thymallus* L.) in the Aare river (Thun, Switzerland). Ecol Model 215:122–132

Mouton AM, De Baets B, Goethals PLM (2009a) Knowledge- based versus datadriven fuzzy habitat suitability models for river management. Environ Model Softw 24:982–993

Mouton AM, Jowett I, Goethals PLM, De Baets B (2009b) Prevalence-adjusted optimisation of fuzzy habitat suit- ability models for aquatic invertebrate and fish species in New Zealand. Ecol Inform 4:215–225

Mouton AM, Alcaraz-Herna´ndez JD, De Baets B, Goethals PLM, Mart´ınez-Capel F (2011) Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediter- ranean rivers. Environ Model Softw 26:615–622

Mouton AM, Dillen A, Van den Neucker T, Buysse D, Stevens M, Coeck J (2012) Impact of sampling efficiency on the performance of data-driven fish habitat models. Ecol Model 245:94–102

Musil J, Ada´mek Z (2007) Piscivorous fishes diet dominated by the Asian cyprinid invader, topmouth gudgeon (*Pseudo- rasbora parva*). Biologia 62:488–490

Onikura N, Nakajima J (2013) Age, growth and habitat use of the topmouth gudgeon, *Pseudorasbora parva* in irrigation ditches on northwestern Kyushu Island, Japan. J Appl Ichthyol 29:186–192

Orciari RD (1979) Rotenone resistance of golden shiners from a periodically reclaimed pond. Trans Am Fish Soc 108:641–645

Pauwels I (2014) Spatially explicit migration models of pike to support river management Doctoral dissertation, Ghent University

Persson L, Greenberg L (1990) Juvenile competitive bottle- necks: the perch (*Perca fluviatilis*)–roach (*Rutilus rutilus*) interaction. Ecology 71:44–56

Peters J, De Baets B, Verhoest NEC, Samson R, Degroeve S, De Becker P, Huybrechts W (2007) Random forests as a tool for ecohydrological distribution modelling. Ecol Model 207:304–318

Pinder AC, Gozlan RE (2003) Sunbleak and topmouth gudgeon: two new additions to Britain’s freshwater fishes. Br Wildl 15:77–83

Pinder AC, Gozlan RE, Britton JR (2005) Dispersal of the invasive topmouth gudgeon, *Pseudorasbora parva* in the UK: a vector for an emergent infectious disease. Fish Manag Ecol 12:411–414

Pimentel D, Zuniga R, Morrison D (2005) Update on the envi- ronmen tal and economic costs associated with alien-in- vasive species in the United States. Ecol Econ 52:273–288 Pollux BJA, Korosi A (2006) On the occurrence of the Asiatic cyprinid *Pseudorasbora parva* in the Netherlands. J Fish

Biol 69:1575–1580

Pressley PH (1980) Parental effort and the evolution of nest- guarding tactics in the threespine stickleback, *Gasterosteus aculeatus* L. Evolution 35:282–295

Quyen Tran TN, Jackson MC, Sheath D, Verrecyken H, Britton JR (2015) Patterns of trophic niche divergence between invasive and native fishes in wild communities are predictable from mesocosm studies. J Anim Ecol 84:1071–1080

R Development Core Team (2014) R: a Language and Envi- ronment for Statistical Computing. R Foundation for Sta- tistical Computing, Vienna, Austria. [http://www.R-project.](http://www.R-project.org/) [org/](http://www.R-project.org/)

Rosecchi E, Crivelli AJ, Catsadorakis G (1993) The establish- ment and impact of *Pseudorasbora parva*, an exotic fish species introduced into Lake Mikri Prespa (north-western Greece). Aquat Conserv Mar Freshw Ecosyst 3:223–231

Rosecchi E, Thomas F, Crivelli AJ (2001) Can life-history traits predict the fate of introduced species? A case study on two cyprinid fish in southern France. Freshw Biol 46:845–853 Schmidt RE, McGurk J (1982) Biology of the European bitter- ling *Rhodeus sericeus* (Pisces: Cyprinidae) in the Bronx River, New York, USA: an apparently benign exotic spe-

cies. Biol Conserv 24:157–162

Shannon CE, Weaver W (1963) Mathematical theory of com- munication. University of Illinois Press, Urbana

Smyly WJP (1955) On the biology of the stone-loach *Nema- cheilus brabatula* (L.). J Anim Ecol 24:167–186

Strayer DL, Power ME, Fagan WF, Pickett STA, Belnap J (2003) A classification of ecological boundaries. Bio- science 53:723–729

Sutherst RW, Bourne AS (2009) Modelling non-equilibrium distributions of invasive species: a tale of two modelling paradigms. Biol Invasions 11:1231–1237

Van Broekhoven E, Adriaenssens V, De Baets B, Verdonschot PFM (2006) Fuzzy rule-based macroinvertebrate habitat suitability models for running waters. Ecol Model 198: 71–84

Van Kleef HH, Jongejans E (2014) Identifying drivers of pumpkinseed invasiveness using population models. Aquat Invasions 9:315–326

Verreycken H, Anseeuw D Van, Thuyne G, Quataert P, Belpaire C (2007) The non-indigenous freshwater fishes of Flanders (Belgium): review, status and trends over the last decade. J Fish Biol 71:160–170

Wolfram-Wais A, Wolfram G, Auer B, Mikschi E, Hain A (1999) Feeding habits of two introduced fish species (*Lepomis gib- bosus*, *Pseudorasbora parva*) in Neusiedler See (Austria), with special reference to chironomid larvae (Diptera: Chi- ronomidae). Hydrobiologia 408(409):123–129

Wood SN (2011) mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. R Package Version 1.7-6. [http://www.r-project.org](http://www.r-project.org/)

Xie S, Cui Y, Zhang T, Li Z (2000) Seasonal patterns in feeding ecology of three small fishes in the Biandantang Lake, China. J Fish Biol 57:867–880

Yee TW, Mitchell ND (1991) Generalized additive-models in plant ecology. J Veg Sci 2:587–602

Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith G (2009) Mixed effects models and extensions in ecology with R. Springer, New York

Zuur AF, Ieno EN, Elphick CS (2010) A protocol for data exploration to avoid common statistical problems. Meth- ods Ecol Evol 1:3–14