Muñoz Mas, R.; Vezza, P.; Alcaraz-Hernández, JD.; Martinez-Capel, F. (2016). Risk of invasion predicted with support vector machines: A case study on northern pike (Esox Lucius, L.) and bleak (Alburnus alburnus, L.). Ecological Modelling. 342:123-134. doi:10.1016/j.ecolmodel.2016.10.006.

Risk of invasion predicted with support vector machines: a case study on northern pike (*Esox Lucius*, L.) and bleak (*Alburnus alburnus*, L.)

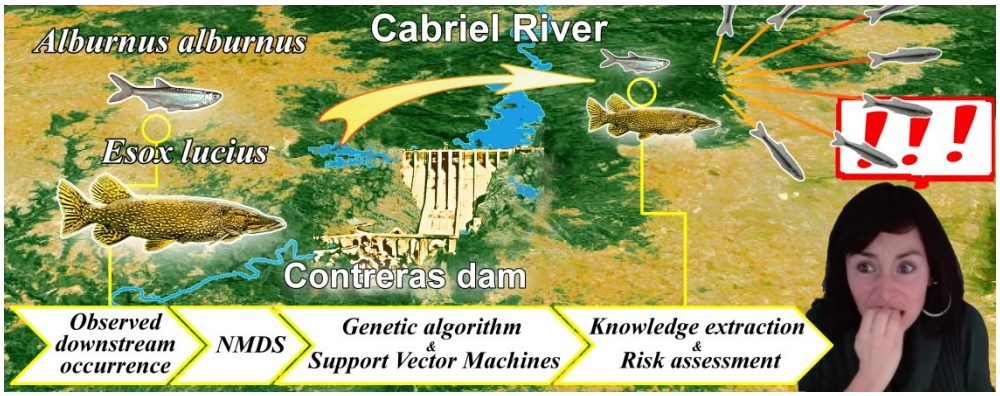
Rafael Muñoz-Mas[1]*, Paolo Vezza[2], Juan Diego Alcaraz-Hernández[1], Francisco Martínez-Capel[1]

[1] Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC) - Universitat Politècnica de València, C/ Paranimf 1, 46730 Grau de Gandia, València (Spain).

[2] International Centre for Ecohydraulics Research (ICER) - University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom.

*Corresponding author: Rafael Muñoz-Mas, e-mail: pitifleiter@hotmail.com, voice: +34622098521

Graphical abstract



Abstract

The impacts of invasive species are recognised as a major threat to global freshwater biodiversity. The risk of invasion (probability of presence) of two avowed invasive species, the northern pike (*Esox Lucius*, L.) and bleak (*Alburnus alburnus*, L.), was evaluated in the upper part of the Cabriel River (eastern Iberian Peninsula). Habitat suitability models for these invasive species were developed with Support Vector Machines (SVMs), which were trained with data collected downstream the Contreras dam (the last barrier impeding the invasion of the upper river segment). Although SVMs gained visibility in habitat suitability modelling, they cannot be considered

widespread in ecology. Thus, with this technique, there is certain controversy about the necessity of performing variable selection procedures. In this study, the parameters tuning and the variable selection for the SVMs was simultaneously performed with a genetic algorithm and, contradicting previous studies in freshwater ecology, the variable selection proved necessary to achieve almost perfect accuracy. Further, the development of partial dependence plots allowed unveiling the relationship between the selected input variables and the probability of presence. Results revealed the preference of northern pike for large and wide mesohabitats with vegetated shores and abundant prey whereas bleak preferred deep and slightly fast flow mesohabitats with fine substrate. Both species proved able to colonize the upper part of the Cabriel River but the habitat suitability for bleak indicated a slightly higher risk of invasion. Altogether may threaten the endemic species that actually inhabit that stretch, especially the Júcar nase (*Parachondrostoma arrigonis*; Steindachner), which is one of the most critically endangered Iberian freshwater fish species.

Key words

Genetic algorithm; habitat suitability model; Iberian Peninsula; Mediterranean river; mesohabitat scale; variable selection.

1 Introduction

The impacts of invasive species are recognised as a major threat to global freshwater biodiversity via a variety of adverse impacts, such as predation, hybridisation, vectoring diseases, food web alteration and interspecific competition (Almeida and Grossman, 2012). Therefore, several authors highlighted the importance of risk assessment and management for controlling these invasive species (Almeida et al., 2013). In Iberian rivers, native fish have suffered from multiple and recurrent introductions during the last century, which has been stressed as one of the main negative factors affecting the survival of these native, mostly endemic, species (Elvira and Almodóvar, 2009). It is

consequently the responsibility of conservationists to elucidate the link between the level and nature of propagule pressure and its potential impact on native species (Ribeiro et al., 2008).

In the Iberian Peninsula, most of the conducted research quantified the invasiveness degree of several fish species at the basin scale identifying key biological traits that would facilitate successful establishments (Almeida et al., 2013; Clavero, 2011; Ribeiro et al., 2008). Although ecological impacts such as changes in species survival, microhabitat selection or competition for spawning areas have been reported (Ribeiro and Leunda, 2012), very few studies have been performed at detailed scales (*i.e.* micro or mesohabitat scales) (Almeida et al., 2014a; Elkins and Grossman, 2014). In the Iberian Peninsula, introduced species are widespread and they are still expanding their distribution ranges (Ribeiro and Leunda, 2012). Furthermore, many of them are piscivorous species, which form a trophic group almost absent in the original ichthyofauna (Clavero et al., 2004; Ribeiro and Leunda, 2012). Damming of rivers has favoured the establishment of these typically lentic species by reducing the natural intra- and inter-annual flow variations (Clavero et al., 2004; Muñoz-Mas et al., 2016). Thus the native fish communities are increasingly being cornered to the upper part of the stream networks, being isolated from one to another by these artificial barriers (Alcaraz et al., 2014; Aparicio et al., 2000). In this context, the basin scale can be too coarse resolution to render effective tools with management purposes.

The benefits of the mesohabitat scale have been highlighted among the other spatial scales to analyse fish habitat requirements (Costa et al., 2012; Vezza et al., 2015) because using this scale is possible to describe the environmental conditions around an aquatic organism not only limiting the analysis to the point where it is observed (Vezza et al., 2015). Furthermore, mesohabitats – generally corresponding in size and location to Hydro-Morphological Units (HMU) such as, pool, riffle or rapid – can be used to describe fish ecology with a broader range of variables even including biotic predictors (Muñoz-Mas et al., 2015; Vezza et al., 2015).

In numerous occasions, machine learning habitat suitability models proved to be adequate tools to understand the habitat requirements of fish species (Mouton et al., 2007; Olden et al., 2008). Thus, they can be considered adequate tools to characterize the suitability of the recipient mesohabitats allowing the evaluation of the risk of invasion. To date, the mesohabitat scale has been used in

combination with several modelling approaches to predict the presence or abundance of fish, for instance, with logistic regression (*e.g.* Vezza et al., 2014), Support Vector Machines (SVMs) (*e.g.* Tirelli et al., 2012), or random forests (*e.g.* Vezza et al., 2014). The strengths and weaknesses of every machine learning technique must be considered; otherwise, the development of an inappropriate habitat suitability model may result in erroneous predictions (Lin et al., 2015). In this regard, SVMs produced very competitive results when compared with the best accessible classification methods (Fukuda et al., 2013; Sadeghi et al., 2014; Tirelli et al., 2012) and, in addition, they only need the optimization of very few parameters (Hoang et al., 2010; Sadeghi et al., 2014). Furthermore, SVMs rely on convex quadratic programming; thus, no local optima exist and efficient optimization procedures can be used to find the unique global optimum (Fukuda and De Baets, 2016). In accordance to previous statements, SVMs gained visibility in habitat suitability modelling in the last few years (Fukuda and De Baets, 2016; Fukuda et al., 2013; Hoang et al., 2010; Sadeghi et al., 2014) but, despite the existence of the aforementioned references, they still cannot be considered widespread in ecology (Hoang et al., 2010). Thus, within the ecological modelling discipline, it does exist certain controversy about the necessity of performing variable selection. Thereby, in ecology, some authors suggested it unnecessary (Hoang et al., 2010; Sadeghi et al., 2014; Tirelli et al., 2012) or explained it very cryptically (Poulos et al., 2012) whereas in machine learning or medical studies it has been stated as a fundamental step to improve generalization capability (Fröhlich et al., 2003; Guyon et al., 2002; Huang and Wang, 2006).

This study focuses on the Cabriel River, the main tributary of the Júcar River (eastern Iberian Peninsula). The Cabriel River has 220 km in length, 4,754 km^2 of drainage area and 10.8 m^3/s of mean flow. It harbours the most important populations, in terms of presence and fish density, of the Júcar nase (*Parachondrostoma arrigonis*; Steindachner, 1866) a fish species in imminent danger of extinction (Alcaraz et al., 2014). The Cabriel River is actually split into two main stretches of similar length (upper Cabriel and lower Cabriel) by a sequence of weirs and dams – the most noticeable the Contreras dam (Costa et al., 2012) – conforming a complex of storage and hydropower facilities. Both stretches harbour invasive species, but the larger amount of invasive species is hosted in the lower part (Alcaraz et al., 2014). The most remarkable ones are the northern pike (*Esox lucius*;

Linnaeus, 1758) and the bleak (*Alburnus alburnus*; Linnaeus, 1758) (Alcaraz et al., 2014; Costa et al., 2012) both categorized as highly invasive species (Almeida et al., 2013). Northern pike has been introduced in freshwater systems across the globe, and, in the Iberian Peninsula, it has been linked to the decline or extirpation of multiple fish species (Elvira et al., 1996; Ribeiro and Leunda, 2012; Rincón et al., 1990). Conversely, the impacts of bleak introduction have been poorly documented; especially taking into account that the species has shown an incredible high fecundity allowing bleak to outcompete other species (Vinyoles et al., 2007). Thus, it has been only confirmed that the species is able to easily hybridize with the calandino (*Squalius alburnoides* complex; Steindachner, 1866) and the southern Iberian chub (*Squalius pyrenaicus*; Günther, 1868) (Almodóvar et al., 2012) and to proficiently compete for feeding resources (Almeida et al., 2014b). In the Iberian Peninsula, the northern pike was introduced in 1949 with recreational purposes, whereas the introduction of bleak took place in 1992, principally to provide forage for large predator fish (*e.g.* northern pike) (Elvira and Almodóvar, 2001). Reservoirs favour the establishment and rearing of these invasive species (Ribeiro and Leunda, 2012), and, in the current situation, the Contreras complex can be considered the bridgehead in their invasion of the upper part of the Cabriel River.

As a consequence, the study aim was (i) to infer the habitat preferences of the northern pike and bleak at the mesohabitat scale (based on data collected in the lower part of the Cabriel River), and (ii) to predict the risk of invasion of these species in the upper part of the Cabriel River (upstream of the Contreras complex of storage and hydropower facilities) whereas (iii) the triviality of the variable selection procedures was ruled out in a subsidiary way. To achieve these aims, habitat suitability models (*i.e.* probability of presence estimation) were developed by means of SVMs optimized simultaneously performing the variable selection and the parameter tuning with a genetic algorithm. Then, the optimal SVMs (*i.e.* the aforementioned SVMs trained with data collected throughout the downstream river segment) were used to assess the risk of invasion (potential suitable habitat) in the upper part of the Cabriel River.

2 Methods

2.1 Previous knowledge on northern pike and bleak ecology

The northern pike is a large ambushing predator (maximum body length ≈ 150 cm) with circumpolar origins (Harvey, 2009). Northern pike has shown an opportunistic diet (Sepulveda et al., 2013). Consequently, it can become a keystone predator able to control fish community composition (Kobler et al., 2008). In accordance with the interest and the aftermath of its introduction northern pike has profusely been the subject of ecological modelling from the plain univariate habitat suitability criteria (Inskip, 1982) to the more complex cellular automata (Pauwels et al., 2013) or individual-based models (Baetens et al., 2013). However studies on habitat requirements have usually focused on lentic environments (Casselman and Lewis, 1996; Kobler et al., 2008 and references therein) and only few studies have been carried out in lotic ecosystems (Inskip, 1982; Kerle et al., 2001; Zarkami, 2008). Although results partially differ between sites, it has been stated the imperative necessity for aquatic vegetation, submerged or emerged (*i.e.* reeds) (Harvey, 2009; Inskip, 1982; Kerle et al., 2001), and the preference for large depths (up to 5 m) (Kerle et al., 2001; Stojkovic et al., 2014).

On the other hand, bleak is a small cyprinid (maximum body length ≈ 30 cm) with a wide natural distribution in Europe, from the north-eastern slopes of the Pyrenees to the Urals (Vinyoles et al., 2007). It inhabits open waters of lakes and medium-to-large rivers conforming large aggregations in backwaters and other still waters (Kottelat and Freyhof, 2007). Larvae live in the littoral zone of rivers and lakes with preference for vegetated shorelines (Mouton et al., 2009) while elder individuals tend to leave shores, occupying pelagic habitats and feeding on plankton, drifting insects or invertebrates (Kottelat and Freyhof, 2007). Despite the extremely rapid expansion of this exotic cyprinid (Vinyoles et al., 2007), there are no dedicated studies specifically studying its habitat preferences. Nevertheless, literature provides valuable hints; therefore, it has been suggested to inhabit a large range of environments, water depth from 0.4 to 5 m, temperature from 10.6 to 29.6 ºC and elevation from 78 to 308 m a.s.l., whereas other authors suggested narrower optimal ranges, water depth < 0.7 m and flow velocity < 0.5 m/s (Harby et al., 2007; Stojkovic et al., 2014).

2.2 Data collection

The Cabriel River was surveyed from 2006 to 2008 stratifying the river in eight different study sites (Fig. 1), four sites located upstream the Contreras complex (from U1 to U4) and four downstream it (from L1 to L4), to ensure an equal sampling effort in the upper and the lower parts (Costa et al., 2012; Vezza et al., 2015).

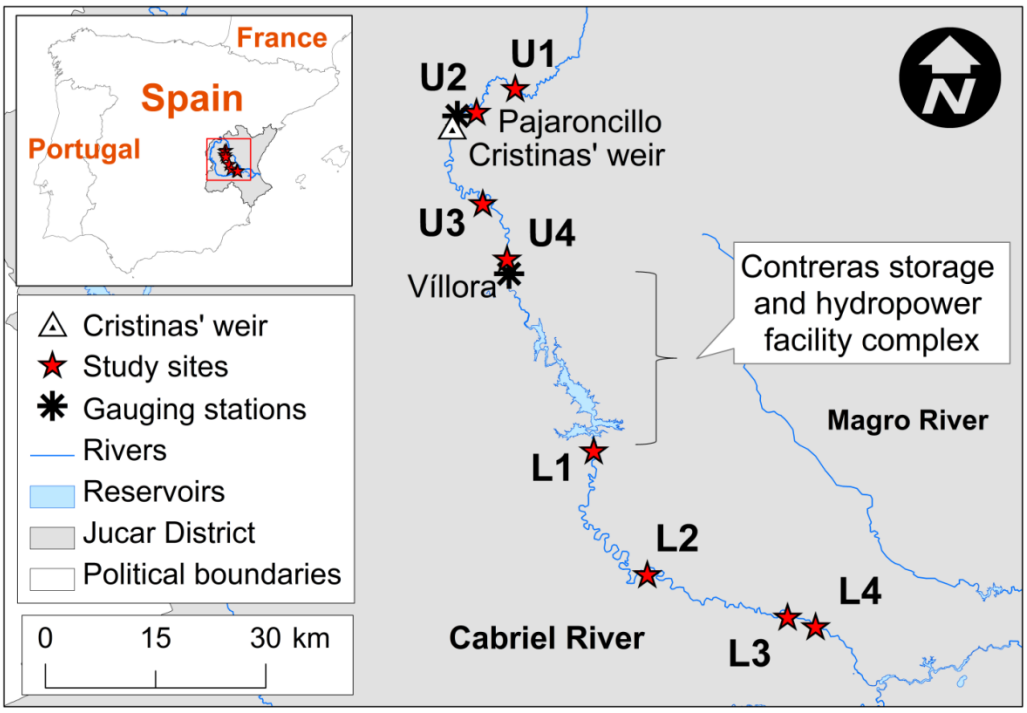


Fig. 1. Location of the study sites in the upper (U1-U4) and lower (L1-L4) segments of the Cabriel River (Júcar River tributary – eastern Iberian Peninsula).

2.2.1 Physical habitat survey

The physical habitat survey in the upper and the lower part of the Cabriel River was conducted at the mesohabitat scale (Fig. 2). Mesohabitats and Hydro-Morphological Units (HMU) were assimilated (Costa et al., 2012; Vezza et al., 2015) thus the HMU was considered the sampling unit for the study. Each year, the four sites in the upper and the four in the lower part of the river were firstly stratified by HMU, classified as pool, glide, run, riffle, and rapid (Costa et al., 2012; Vezza et al., 2015), surveying a sequence of HMUs sufficiently long to exceed one km . The HMU class was considered as an ordinal variable in the models development and, for each of the HMU, several attributes were recorded. Length was measured with CMII Hip Chain (CSP Forestry Ltd. Alford, Scotland), average width was measured with laser distancemeter DISTO A5 (Leica Geosystems, Heerbrugg, Switzerland) and obtained from four to eight cross-sections. Mean depth (hereafter depth) was measured with a wading rod and calculated from 20 to 40 point measurements at a rate of five measurements per transect whereas maximum depth was measured in the corresponding

point (Vezza et al., 2015). Then HMU area and volume were calculated by correspondingly considering length, width and depth. Additionally the backwaters area was recorded if present, considering presence if waters visibly were stagnated or backed up by obstructions. Canopy shading (as percentage of the overall HMU area), undercut banks (as percentage of the HMU length) and the presence of emerged vegetation on the shoreline (as percentage of the HMU length) were visually estimated. The percentages of substrate types following a simplified classification from the American Geophysical Union (Muñoz-Mas et al., 2012) were also visually estimated and summarized in the substrate index (Mouton et al., 2011) that typically ranges from zero (silt and vegetated soil) to eight (bedrock). The cover index (García de Jalón and Schmidt, 1995) was then determined by characterizing the available refuge due to caves, shading, substrate, submerged vegetation and water depth producing an index ranging from 0 to 10. Finally, the number of big boulders and woody debris were counted (Table 1).

2.2.2 Biological survey

The biological survey took place concomitantly with the physical habitat survey (Fig. 2). To ensure a reasonably uniform probability of detection, during the three campaigns, two divers conducted the underwater counts (snorkelling) in three independent passes (from downstream to upstream) throughout each HMU (Schill and Griffith, 1984). In each HMU, fish were counted, dividing fish species into small (body length <10 cm) and large (body length > 10 cm) size classes but the northern pike that included only large specimens (body length > 0.5 m). Divers were trained to maintain constant the fish sampling effort and to ensure that each pass was independent and not affected by previous passes, a time delay was programmed among replicate counts (Bain et al., 1985). The snorkelling technique was chosen for its effectiveness in assessing fish population at the mesohabitat scale and to avoid any damage to any endangered species (Costa et al., 2012; Vezza et al., 2015). Moreover, we considered it the most appropriate methodology for this study due to the morphological characteristics of the river (clear water and deep pools, max. depth *ca.* 4 m). Finally, in addition to northern pike and bleak, seven species where profusely observed, namely: brown

trout (*Salmo trutta*; Linnaeus, 1758), Eastern Iberian barbel (*Luciobarbus guiraonis*; Steindachner, 1866), valencia chub (*Squalius valentinus*, Doadrio & Carmona, 2006), Júcar nase (*Parachondrostoma arrigonis*; Steindachner, 1866), Iberian gudgeon, (*Gobio lozanoi*; Doadrio & Madeira, 2004) and freshwater blenny (*Salaria fluviatilis*; Asso, 1801). The Iberian nase (*Pseudochondrostoma polylepis*; Steindachner, 1865) was observed only in U3 and U4 (Fig. 1), whereas European eel (*Anguilla anguilla*; Linnaeus, 1758) and eastern mosquito fish (*Gambusia holbrooki*; Girard, 1859) were seldom observed in the lower Cabriel.

Fish species distribution is conditioned by a suite of biotic and abiotic factors (Guisan and Thuiller, 2005). Among the biotic set, interactions between freshwater species play an important role in the habitat selection (Jackson et al., 2001), although very few studies on predictive models included biotic variables explicitly (Elith and Leathwick, 2009). In accordance, to investigate the possibility of any predator-prey relationship, interspecific competition or mutualism, fish counts were included as input variables. Northern pike can switch to alternative prey species when the abundance of the preferred ones has declined (Sepulveda et al., 2013) and it has demonstrated preference for relatively small prey instead of large (Nilsson, 2001). In the Iberian Peninsula, the occurrence of shoals encompassing multiple cyprinid species has been reported (Martínez-Capel et al., 2009; Muñoz-Mas et al., 2015) likewise they were observed during the performance of the surveys. Therefore cyprinid counts were grouped in two different variables; large cyprinids, encompassing large Eastern Iberian barbel and large Iberian nase, and small cyprinids, which encompassed the remaining specimens (including bleak counts in the northern pike model) (Table 1). Bleak was removed from that aggregation to train the corresponding habitat suitability model and freshwater blenny was finally eliminated from the data analysis because this species is naturally absent from the upper Cabriel (Vezza et al., 2015) (Table 1).

Presence-absence modelling was the selected choice because it is likely to yield better performance (Fukuda et al., 2011) and its output better fits the goal of risk assessment by providing simple and interpretable outputs (*i.e.* probabilistic-like index ranging from zero to one). In the end the northern pike was observed in 59 out of 177 HMUs whereas bleak was observed in 68 resulting in a data prevalence of 0.33 and 0.38 for pike and bleak respectively.

Table 1 Summary and units of input variables collected downstream Contreras' dam (L1-L4). HMU means Hydro-Morphological Unit, N. number, N.L. number of large and N.S. number of small individuals.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Units
HMU class	1	1	3	2.627	4	5	(-)
Length	9	39.6	66.8	77.01	96	330	(m)
Width	5.563	13.86	15.78	15.37	17.31	25.82	(m)
HMU area	79.3	581.9	1054	1215	1532	4174	(m^2)
Volume	14.83	336.9	854.3	1210	1664	6724	(m^3)
Depth	0.187	0.502	0.75	0.882	1.145	3.929	(m)
Max. Depth	0.3	1.07	1.5	1.681	2.1	4.5	(m)
Velocity	0.016	0.125	0.224	0.264	0.362	0.858	(m/s)
Substrate index	0.1	3.05	4.25	4.069	5.35	7.45	(-)
N. Woody debris	0	0	0.25	1.022	1	9	(#)
Shade	0	5	10	14.9	20	95	(%)
Cover index	0.16	0.4	0.52	0.497	0.56	0.84	(-)
Backwaters area	0	5.25	17.75	36.15	33.5	880	(m^2)
N. Boulders	0	2	12	27.58	35	354	(#)
Undercut banks	0	0	5	12.8	15	90	(%)
Vegetation	0	3	10	22.1	35	93	(%)
N.L. Trout	0	0	0	0.17	0	4	(#)
N.S. Trout	0	0	0	0.181	0	6	(#)
N.L. Cyprinids	0	0	0	10.05	2	213	(#)
N.S. Cyprinids (N. Pike model)	0	8	86	164.9	249	1033	(#)
N.S. Cyprinids (Bleak model)	0	8	81	149.6	211	1033	(#)

2.3 Adequacy of data packing – Non-metric Multi-Dimensional Scaling (NMDS)

Although the surveys in the lower Cabriel were performed every year during low flow (from 0.49 m^3/s at L1 to 6.05 m^3/s at L4), the study encompassed a three year period thus it could implicate significant changes in the physical habitat or in the biotic components since fish abundance is a fed-back phenomena (Mas-Martí et al., 2010). To rule out the significant influence of study year, Non-metric Multi-Dimensional Scaling (NMDS) (Kruskal, 1964a, 1964b) was performed to summarize the physical habitat and the fish counts collected each year (Fig. 2). NMDS is an ordination method that preserves the distances between sample points in the ordination space becoming a useful approach for visualizing spatial (Garófano-Gómez et al., 2011) and temporal similitudes (Marchetti et al., 2006). NMDS uses an iterative approach that rearranges samples in the ordination space to minimize a measure of disagreement (referred to as stress) between the compositional

dissimilarities and the distance between the points; stress values below 5 correspond to a good ordination with no real prospect of a misleading interpretation (Garófano-Gómez et al., 2011; Marchetti et al., 2006). The entire analysis was carried out in *R* (R Core Team, 2015) with the package *vegan* (Dixon, 2003) setting the function with the Bray-Curtis distance and reducing the input space to two dimensions. Similarities between sites and campaigns were inspected by plotting the NMDS; overlapping samples (depicted as dots) meant similar physical habitat and biotic predictors and separated, dissimilar.

2.4 Habitat suitability models – Genetically optimized Support Vector Machines (SVMs)

The habitat preferences of the invasive species (*i.e.* probability of presence of northern pike and bleak) were modelled by means of presence-absence SVMs (Vapnik, 1995). The basic idea with SVMs is to map the training data into a higher dimensional feature space via some mapping functions $\varphi(x)$ (*e.g.*, linear or polynomial) constructing a discriminant (classificatory) hyperplane with the maximum discriminant margin (Huang and Wang, 2006), which is called Optimal Separating Hyperplane (OSH) (Vapnik, 1995).

In the very beginning the OSH was adjusted using linear functions though data might not be linearly separable (Howley and Madden, 2005). As a consequence, the use of non-linear functions was promptly popularized (Cristianini and Schölkopf, 2002). The most popular non-linear functions are polynomial, radial basis and sigmoid, each one with a small number of parameters to be optimized (Hoang et al., 2010; Howley and Madden, 2005). A common approach is to use a grid-search of these parameters, often starting from a very coarse grid covering the whole searching space and iteratively refining both grid resolution and search boundaries (Howley and Madden, 2005).

In addition of the optimization of these parameters, modellers usually had to deal with very high dimensional input spaces. Therefore, the necessity for finding out the combination of input variables which contribute most to the proper classification has been highlighted (Fröhlich et al., 2003). There are two main approaches to solve the variable selection problem, the filter approach and the

wrapper approach (Kohavi and John, 1997). In the filter approach, variable selection is performed as a pre-processing step whereas wrapper methods are based on the generalized performance derived of models' training with the considered variables' subset.

Then, two problems can be confronted during the development of optimal SVM, choosing the optimal input variables' subset and the best mapping function with its corresponding parameters. These two elements are crucial, because the choice of the variables' subset influences the appropriate function parameters and vice versa (Fröhlich et al., 2003; Huang and Wang, 2006).

Genetic algorithms (GA) (Holland, 1992), a group of heuristics and optimization algorithms based on the process of natural selection, have demonstrated proficient to simultaneously infer both, the optimal variables' subset and the parameters of the SVM, in a sort of extension of the concepts stated for the wrapper approach (Huang and Wang, 2006). Therefore, to develop optimal SVMs, we used a GA to infer these elements (Fig. 2).

The SVMs, employed to evaluate the risk of invasion, were developed in *R* (R Core Team, 2015) with the function *svm* implemented within the *e1071* package (Dimitriadou et al., 2011). The sigmoid function seems to work well in practice, but is not better than the radial basis function (Wu et al., 2012), thus we restricted the tested functions to (i) linear, (ii) polynomial and (iii) radial basis. Thereby the GA searched for the optimal function and parameters; (i) degree, (ii) C, (iii) γ and (iv) C_0. Degree corresponds to the power in the polynomial and controls the curvature of the mapping function. The parameter C corresponds to the misclassification cost and allows balancing the bias and variance. The parameter γ is the inverse of the radius of influence of samples selected by the model as support vectors for radial basis functions or the degree of symmetry in polynomial ones. Finally, C_0 is the intercept in polynomial functions. The tested ranges of the parameters were based on Huang and Wang (2006), who modelled a large range of different datasets, with degree, γ, C and C_0, ranging from 1 to 4, from 0 to 10, from 0 to 1 and from 0 to 300 respectively. In addition, to improve generalization during the training of each SVM, threefold cross-validation was internally performed. Data prevalence can have a strong effect on SVM classification capability (Osuna et al., 1997). Therefore, the training cases were weighted accordingly to their prevalence: absence with 0.33 and 0.38 and presence with 0.66 and 0.62 for pike and bleak respectively.

The selected GA was the one comprised in the *rgenoud* package (Mebane Jr and Sekhon, 2011); this function combines evolutionary algorithm methods with a derivative based (quasi-Newton) method to solve difficult optimization problems. GA optimization is based on selection, crossover and mutation (Fröhlich et al., 2003; Huang and Wang, 2006). *Rgenoud* presents nine operators driving the optimization which correspond to cloning, uniform mutation, boundary mutation, non-uniform mutation, polytope crossover, simple crossover, whole non-uniform mutation, heuristic crossover and local-minimum crossover (Mebane Jr and Sekhon, 2011). These operators were set to 0.25, 0.75, 0.15, 0.10, 0.15, 0.75, 0.15, 0.35 and 0.0 respectively and we disabled the use of derivatives in the searching process. Finally, the population size and the number of generations were both set to 1000.

We followed a modification of the procedure presented by Huang and Wang (2006) then instead of encoding the parameters of the SVM and the input variables in bit strings (*i.e.* 0 or 1), the mapping function, the corresponding parameters and the variables were encoded with real numbers and those elements encoded by integers (*i.e.* function, degree and the selected variables) were implemented within the optimized functions by rounding up the tested value (*e.g.*, $\lVert x_1 \rVert$=1 meant linear, $\lVert x_1 \rVert$=2 polynomial, etc.).

GAs effectively solve problems that are nonlinear or perhaps even discontinuous in the parameters of the function to be optimized (Mebane Jr and Sekhon, 2011). As a consequence, a modification of the multi-objective function presented by Huang and Wang (2006), which is based on the principle of parsimony, was maximized (equation 1). The GA searched the maximal classification strength with the minimum number of selected features applying a three times threefold cross-validation scheme ($3 \times 3\ \mathrm{cross} - \mathrm{validation}$). The selected performance criterion was the *Balanced accuracy* ($B.accuracy\ =\ (Sn + Sp)/2$), which corresponded to the mean value of the nine models. Sensitivity (*Sn*) corresponds to the mean of the ratio of presence classified as presence whereas *S*pecificity (*Sp*) to the ratio of absence classified as absence. Finally, n_V represented to the total amount of variables (Objective range from 1 to -0.86).

$$Objective = 0.875 \times \left(\overline{Balanced\ accuracy} + \min\{0, \overline{Sn} - \overline{Sp}\}\right) + 0.125 \times \left({}^{1}/_{n_V}\right) \quad (1)$$

We stimulated the overprediction (*i.e.* $\min\{0, \overline{Sn} - \overline{Sp}\}$) *because* it has been stated more defendable from an ecological viewpoint (Fukuda et al., 2013). In addition we favoured the classification strength over the number of inputs by slating the assigned weights to the former (0.875 in front of 0.125). Finally, an additional constraint was stated by impeding correlated (r^2 > 50 %) combinations of variables. The input database was a combination of ordinal and continuous variables; then, the function *hetcor* in the package *polycor* (Fox, 2010) was used to calculate the variables' correlation (Appendix A). Following previous studies (*i.e.* Fukuda et al., 2013), once the optimal parameters and variables were obtained, the optimal SVMs used for knowledge extraction and risk assessment were trained employing the optimal settings and input variables but considering the entire datasets (*i.e.* no external cross-validation was performed).

2.5 Knowledge extraction and risk assessment

Model reliability and transparency is of major concern in habitat suitability modelling (Muñoz-Mas et al., 2015) and is fundamental to rule out ecologically unreliable models (Austin, 2007). However, to date, the influence of the input variables in the ultimate prediction carried out with SVMs has typically remained veiled (Fukuda et al., 2013; Hoang et al., 2010; Tirelli et al., 2012). To overcome such limitation, the partial dependence plots (Friedman, 2001) implemented in the package *randomForests* (Liaw and Wiener, 2002) were developed for the optimal SVMs (*i.e.* the ones developed in the previous section employing optimal settings and input variables but considering the entire datasets) (Fig. 2). The partial dependence plots graphically characterise the relationship between the input variables and the probability of presence and were developed adjusting the original code.

To evaluate the risk of invasion, the aforementioned optimal SVMs were used to assess the data obtained during the survey performed in the upper part of the Cabriel River but employing uniquely the variables considered relevant for each target species (Fig. 1). The upper Cabriel comprises five

water bodies as defined by the water framework directive (European Parliament & Council, 2000), all of them classified with 'good ecological status' (CHJ, 2009a) and separated because some sections of the upper Cabriel are included within Sites of Community Importance (SCI) (CHJ, 2009b) or because they present small differences on the riparian habitat quality index (QBR, Munné et al., 2003), which were considered irrelevant to differentiate the risk of invasion.

Although the study had good spatial distribution, it comprises a series of snapshots corresponding to the different surveyed flows. The upper Cabriel corresponds to the unregulated segment of the Cabriel River therefore; a wide flow range was sampled (*i.e.* from Q_{26} to Q_{97}). Some of these flows could be relatively infrequent thus misleading the risk of invasion if they were evaluated. Therefore, we selected the most representative flow for each study site. Flow duration curves were developed using daily flow time series from the Víllora (1970-2010) and Pararoncillo (1949-2010) gauging stations (Fig. 1) and the selected flow was comprised within the range between the Q_{90} and Q_{75} (*i.e.* the nearest to the Q_{83}). The areas of the HMUs and their assessed suitability (*i.e.* probability of presence) were compared in four broad categories Very Low risk of invasion, Low, High and Very High and finally, results for both species were discussed (Fig. 2).

Table 2 Summary of inputs used to evaluate the risk of invasion in the part upper Cabriel River. The risk assessment was restricted to the most representative flows (from Q_{90} and Q_{75}). They corresponded to Q_{91} (2007), Q_{71} (2006), Q_{74} (2008) and Q_{80} (2007) from U1 to U4 respectively. HMU means Hydro-Morphological Unit, N. number, N.L. number of large and N.S. number of small individuals.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Units
HMU class	1	1	4	3	4	5	(-)
Length	5.3	19	32.5	46.22	63.68	157.5	(m)
Width	4.23	6.518	8.238	8.606	10.72	17.75	(m)
HMU area	35.25	133.7	329.1	1268	1142	11970	(m^2)
Volume	12.76	83.12	230.1	1244	1275	12860	(m^3)
Depth	0.296	0.55	0.706	0.854	1.017	3.384	(m)
Max. Depth	0.34	0.91	1.24	1.352	1.772	3.42	(m)
Velocity	0.06	0.144	0.207	0.273	0.326	0.826	(m/s)
Substrate index	0.95	3.875	4.4	4.391	4.962	6.75	(-)
N. Woody debris	0	0	0	0.61	0.712	6	(#)
Shade	0	20	30	35.5	51.2	90	(%)
Cover index	2.5	4	4.5	5.112	5.812	8.75	(-)
Backwaters area	0	2.385	6.085	20.21	18.25	389	(m^2)
N. Boulders	0	1	7	7.823	11	42	(#)
Undercut banks	0	0	0.2	9.4	15	95	(%)
Vegetation	0	10	32.5	38.2	60	97.5	(%)
N.L. Trout	0	3	6	8.812	11	69	(#)
N.S. Trout	0	2	6	7.292	10	29	(#)
N.L. Cyprinids	0	0	0	49.4	4	1374	(#)
N.S. Cyprinids	0	0	3.5	103.9	59.25	1614	(#)

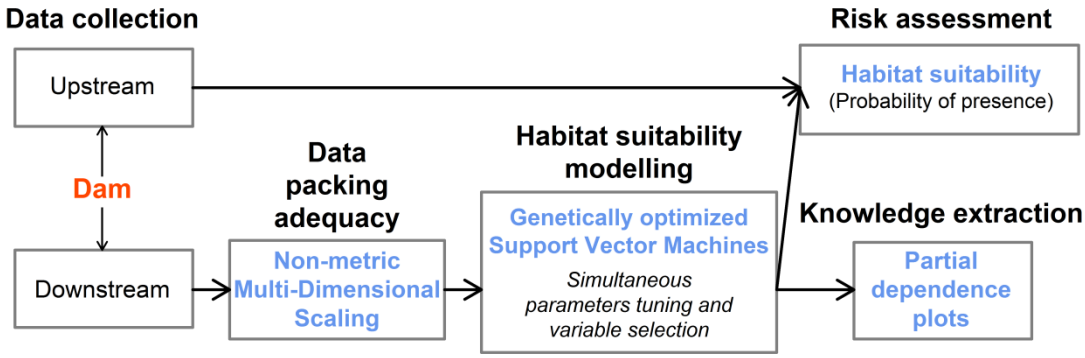


Fig. 2. Flowchart of the steps followed in the development of the habitat suitability models (Genetically optimized Support Vector Machines - SVMs) and risk assessment.

3 Results

3.1 Adequacy of data packing – Non-metric Multi-Dimensional Scaling (NMDS)

The NMDS analysis achieved the *stress* values of 0.21 for the lower Cabriel dataset (L1-L4) corresponding to a perfect ordination. The NMDS plot showed a clear overlapping between years – highlighted by the *ordiplots* depicted in Fig. 3 (*i.e.* depicted ellipses) – and between sites. The only reach that presented lesser overlap with the other sites was L1 because that reach is the most affected by the Contreras dam. However, we considered such phenomena of minor importance and we did not remove these data from the training dataset because certain imbrication between the sampled HMUs and the remaining sites was observed (dots in Fig. 3). Altogether, we considered the data packing adequate for the purpose of the study.

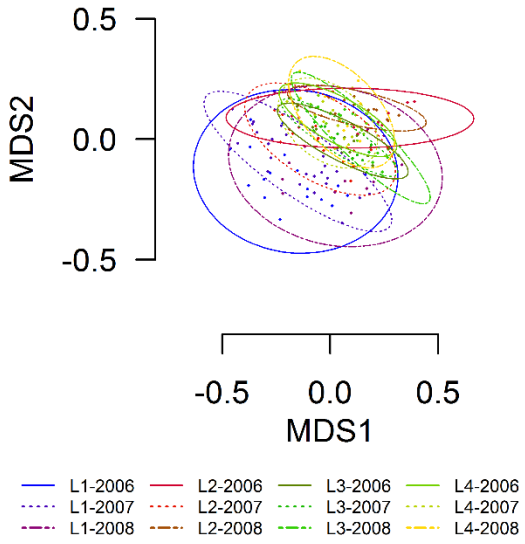


Fig. 3. Non-metric Multi-Dimensional scaling (NMDS) plot of the sites sampled in the lower Cabriel. Dots are coloured in accordance with reach and year whereas *ordiplots* (*i.e.* depicted ellipses) encompass the corresponding site and year. Its amplitude was based on the standard deviation.

3.2 Habitat suitability models - Genetically-optimized support vector machines (SVMs)

The optimal SVM for northern pike was obtained with a radial basis function (γ = 0.34 and C = 156.20) selecting four input variables; width, volume, vegetation (emergent) and number of small cyprinids. On the other hand, the optimal SVM for bleak was obtained with a polynomial function (degree = 3, γ = 0.75, C = 105.03 and C_0 = 2.54). In this case three input variables were selected: depth, velocity and substrate. The bleak model almost presented perfect accuracy (Balanced accuracy = 0.92) then it outperformed the northern pike model (Balanced accuracy = 0.77) for every performance criterion (Table 3). Although the overprediction was stimulated by penalizing *Sn* lower than *Sp* the performance of the pike model trained with the entire database yielded lower *Sn* than *Sp*.

Table 3 Summary of the performance criteria calculated by cross-validation and for the ultimate models (overall); accuracy or correctly classified instances (*CCI*), sensibility (*Sn*), specificity (*Sp*), balanced accuracy (B. Accuracy), Cohen's Kappa (*Kappa*) and True Skill Statistics (*TSS*).

Criteria	Northern pike		Bleak	
	Cross-validation	Overall	Cross-validation	Overall
CCI	0.75±0.09	0.79	0.91±0.08	0.90
Sn	0.76±0.16	0.73	0.97±0.06	0.99
Sp	0.75±0.09	0.82	0.86±0.11	0.85
B. Accuracy	0.76±0.10	0.77	0.92±0.07	0.92
TSS	0.51±0.20	0.55	0.84±0.15	0.84
Kappa	0.48±0.18	0.54	0.81±0.16	0.81

In accordance with the comparison of the performance criteria, the optimal SVM for the northern pike demonstrated poorer discrimination and did not yield either the maximum or the minimum probabilities of presence (*i.e.* zero and one) (Fig. 4). Nevertheless, the bleak model did not present such a limitation. Finally, it is remarkable that the use of case weighs displaced the discriminant threshold (*i.e.* presence or absence); consequently we readjusted the thresholds for the broad categories assigned to the risk of invasion (e.g. Very low, Low, etc.), although outputs in each of the four categories were possible for both models (Fig. 4).

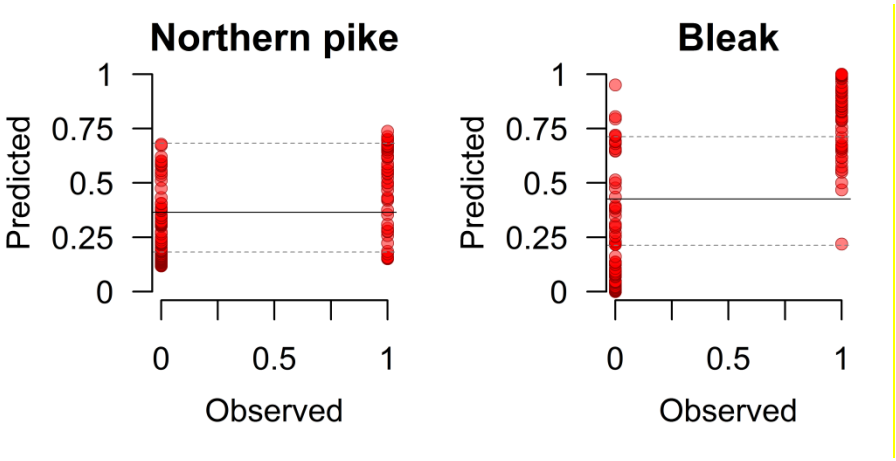


Fig. 4. Observed versus predicted plots for the northern pike and bleak models. Black solid line shows the presence and absence discriminant threshold whereas dashed ones divide ordinary risk (Low or High) from extraordinary risk (Very Low or Very High) of invasion.

3.3 Knowledge extraction and risk assessment

The partial dependence plots depicted a positive quasi linear relationship between channel width and the presence of northern pike (Fig. 5), although the curve presented a slight increment for the small widths. Volume presented similar positive correlation, although it presented a small increase around 2000 m^3. The plot for vegetation (emerged) showed a positive asymptotic pattern whereas the one for number of small cyprinids presented a unimodal curve with the peak around 9000 individuals per HMU. In accordance with the observed ranges of variation, vegetation and number of small cyprinids plaid a major role in the presence of pike, whereas width and volume plaid secondary roles.

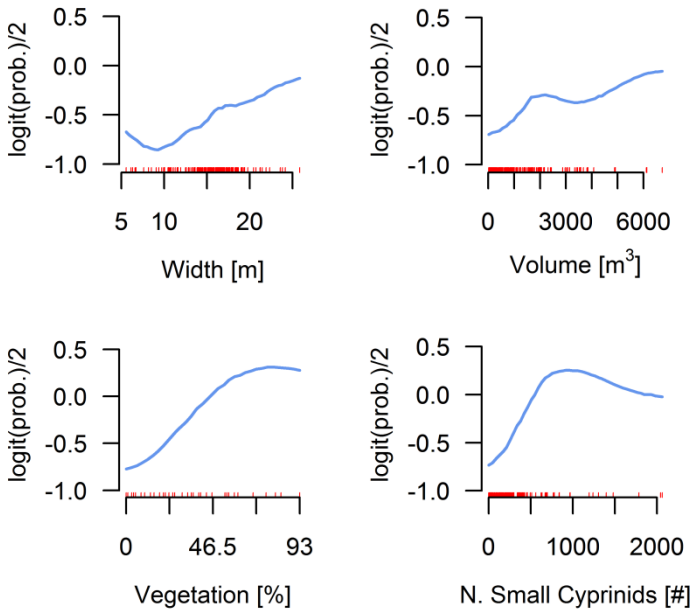


Fig. 5. Partial dependence plots of the northern pike model. Ticks close to the x-axis depict the data in the training database.

The partial dependence plots suggested a positive relationship between depth and the presence of bleak (Fig. 6). The curve presented two marked segments, a gentle one below 2.75 m and a steeped one onward that value. Velocity presented a positive quasi linear relationship with the presence of bleak whereas substrate presented a negative asymptote thus bleak would had appeared more often in HMU with significant presence of lime and vegetation. In accordance with the ranges of variation observed in the partial dependence plots, depth would play a major role in the presence of bleak followed by substrate and, although positive, velocity would play a tertiary influence.

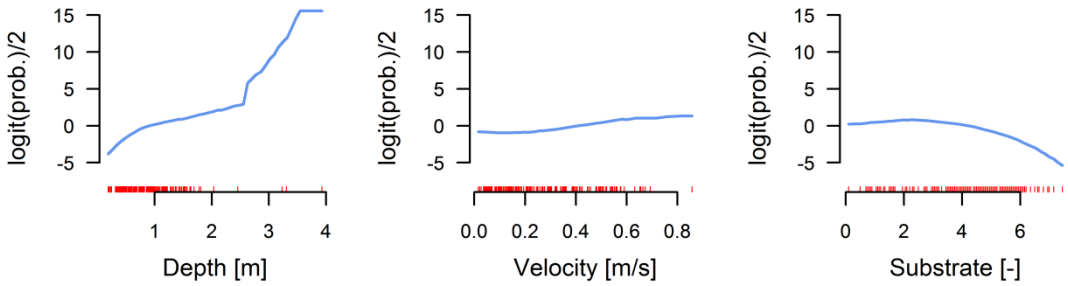


Fig. 6. Partial dependence plots of the bleak model. Ticks close to the x-axis depict the data in the training database.

Overall the northern pike demonstrated lower risk of invasion than bleak, both regarding the magnitude and the suitable area (Fig. 7). Northern pike presence (*i.e.* High and Very high risk) was predicted for 57 % the area comprised in the assessed HMUs, whereas bleak would potentially occupy the 73 % but with disparate distribution and suitability. Northern pike would be present preferably in U3 (90 %) because this stretch presents large and relatively deep HMUs with abundant cyprinids and vegetation, followed by U4 (56 %) and U2 (39 %) because they present intermediate conditions. Although some HMUs were assessed as suitable due to the presence of abundant cyprinids and vegetation, volumes and widths of U1 suggest it unfavourable for pike. In accordance with the area of the HMUs at U1 alone, it would be considered unsuitable for this species (Fig. 7).

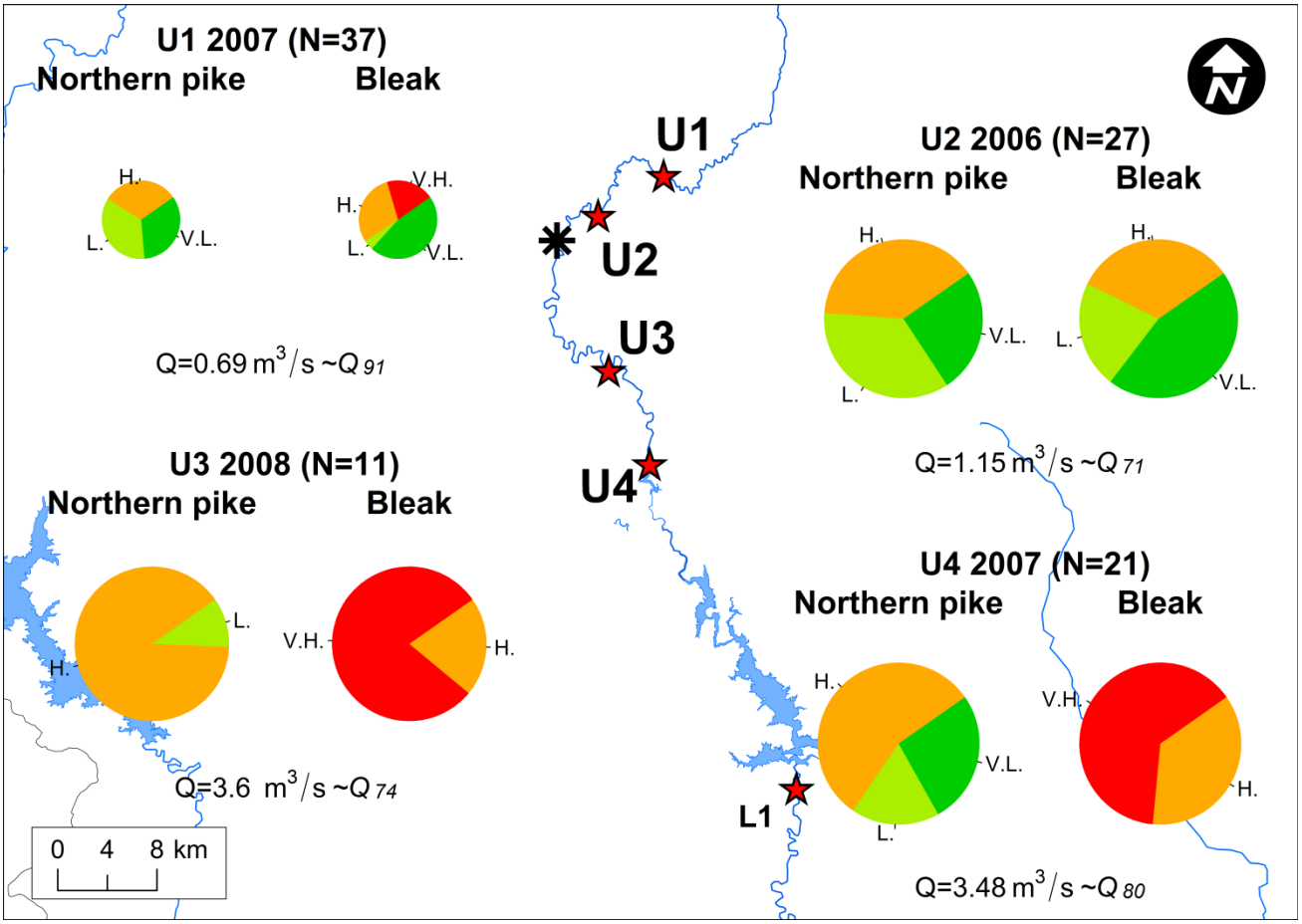


Fig. 7. Risk of invasion for the northern pike and bleak; V.L. means very low, L. means low, H. means high and V.H. very high. The size of the pie chart is proportional to the reach area corresponding to 6669 m^2 (U1-2007), the smallest, and 12214 m^2 (U4-2007), the largest.

Bleak would inhabit preferably the lower part of the upper Cabriel (U3 and U4) thus it could inhabit the 100 % of the assessed area. Likewise northern pike, bleak would be able to colonize U1 and U2 with similar proportions (49 % and 33 % respectively). The optimal SVM for bleak discriminated better the input space consequently some HMU were assessed with the maximum risk of invasion (Very high). However, the risk of invasion for bleak cannot be considered extremely higher than the calculated for northern pike as these differences on the suitability (*i.e.* probability) can be caused by mathematical drawbacks.

4 Discussion

4.1 Model optimization

The wrapper approach, based on the principles stated by Huang and Wang (2006), to simultaneously optimize the parameters of the SVM and to select the optimal set of variables has demonstrated proficient to develop optimal SVMs. They were accurate with a small number of input variables. From the accuracy point of view, the optimal SVM yielded similar or higher values than previous studies (Fukuda et al., 2013), which demonstrated that SVM performance strongly depend on the problem at hand. Bleak model almost achieved perfect accuracy with only three variables and the partial dependence plots fitted well the ecological gradient theory (Austin, 2007) by providing smooth curves without relevant irregularities (*i.e.* increases and decreases), which are hardly explicable from an ecological perspective. Only width and volume for northern pike presented small increase and decreases, although we considered them of minor importance and thus negligible. This phenomenon was caused by the radial basis function and the range of tested values of the γ parameter. Every of the selected support vectors allow an invagination or evagination of the discriminant surface as much deep and irregular as the smaller the γ parameter is. Some authors used the polynomial function by default (Hoang et al., 2010; Sadeghi et al., 2014; Tirelli et al., 2012) whereas others advocated for radial basis functions (Fukuda et al., 2013) in both cases arguably rendering accurate SVMs. However, radial basis–like approaches (*e.g.* probabilistic neural networks) are by conception more flexible in the definition of the OSH (Muñoz-Mas et al., 2014 and references therein) but such flexibility can lead to these irregular patterns. Nevertheless, in this study we performed three times threefold cross-validation. Further, during the optimizations of each of the three SVM threefold cross-validation was internally performed and none of the default parameters was held constant. Therefore, from our personal point of view, that $3\,x\,3$ cross-validation, the specific parameter settings, in addition to the low number of selected inputs and the consequent good generalization, emphasizes the credibility of the optimal SVMs regardless of the selected discriminant function.

4.2 Northern pike habitat preferences

The selected variables as well as the partial dependence plots fitted well the prior knowledge about this species. There is evidence of the northern pike preference for lake-like habitats (Harvey, 2009) thus avoiding fast waters and seeking out vegetated channels and backwaters. Such a general description matches the volume partial dependence plot; pike was present in HMUs of large area and volume, which in turn were likely to include backwater areas. Moreover, channel width presented a positive influence on pike presence because wide HMUs present more often low flow velocity and gentle slope in the river banks, which favours the settlement of vegetation (emergent reeds), critical features that are used by pike for ambushing prey (Bry, 1996; Harvey, 2009; Sepulveda et al., 2013). These results corroborated our field observations; L2 was a stretch with large water volumes and the highest incision of the river channel. The incision impeded the proliferation of vegetation and reeds and consequently, pike was seldom present in the HMUs of that stretch.

Although the presence of pike showed a relatively high correlation with the presence of bleak, the optimal SVM considered crucial the number of small cyprinids at the expense of rejecting bleak presence as a relevant input variable. The partial dependence plot showed a positive influence of number of small cyprinids up to *ca.* 9000 individuals per HMU and decreasing for the larger values of number of small cyprinids. Such a pattern is certainly plausible from a theoretical point of view and fits well with the acknowledged preference of northern pike for relatively small fish (Neill and Cullen, 1974). Cyprinid species are found in very high densities in multi-species large schools (Martínez-Capel et al., 2009; Muñoz-Mas et al., 2015), which are likely to cause the decrement of the curve because it has been reported that the frequency of success per attack decreases with increasing prey density due to the inherent protection provided by schooling. Thereby, the partial dependence plot reflected the search for relatively isolated individuals rather than the attack of those large schools (Connell, 2000; Neill and Cullen, 1974).

The positive influence of vegetation in the presence of pike either in lentic (Bry, 1996; Harvey, 2009; Sepulveda et al., 2013) and lotic (Inskip, 1982; Kerle et al., 2001; Zarkami, 2008) environments has

been profusely documented in previous literature and perfectly matches the partial dependence plots for this variable. Aquatic plants play a critical role in life cycle of pike; for spawning, provision of prey for fingerlings, and cover for adults (Bry, 1996). Only when pike are mature can occasionally be found in un-vegetated areas (Harvey, 2009). As a consequence of such behaviour, the northern pike would have lower detectability, which could be the main cause of the poorer discrimination capability of the optimal SVM in comparison with bleak's one (*i.e.* pike dataset could be much more noisy than bleak one). Although theoretically feasible, the risk assessment for northern pike did not yield the highest category (Very high); such an issue can be certainly caused by the poorer discrimination but also due to the differences between the upper and the lower Cabriel. The characteristics of the recipient ecosystem have demonstrated as relevant as the ones of the invading species (Ribeiro et al., 2008). Therefore we considered preferable the use of the data from the lower segment which are likely to present similar conditions for the unconsidered features (*e.g.* temperature, water chemistry or future potential prey) than data collected in already invaded distant river basins. In a broad sense, we considered our results were plausible for a lowland even brackish dweller species (Kobler et al., 2008; Stojkovic et al., 2014).

4.3 Bleak habitat preferences

In line with the habitat requirement suggested by Kottelat and Freyhof (2007), in the Cabriel River bleak should be categorised as a eurytopic species, preferably dwelling in run-type mesohabitats, which are characterised by relatively large depth and appreciable flow velocity. Bleak was also observed schooling in backwaters downstream of vegetated areas, which has become apparent in the partial dependence plot for substrate because these low flow areas are likely to favour the depositional substrates (*i.e.* silt) and, consequently, the maximum around the substrate index of two. These results would contradict Harby *et al.* (2007) who suggested a more limnophilic nature of the species, although the extremely rapid expansion of bleak (Vinyoles et al., 2007) suggests very general habitat requirements. These unreserved necessities are corroborated by its ability to adapt to a wide variety of Mediterranean ecosystems (Almeida et al., 2014b), which would be confirmed

by the optimal SVM since only three very general variables have been enough to almost perfectly discriminate the suitable from the unsuitable HMUs.

4.4 Pertinence of the variable selection procedure

Interestingly, Tirelli et al. (2012 and therein references), who used SVMs to develop presence-absence habitat suitability models, suggested a much more reophilic nature of one species of the same genus (*Alburnus alburnus alborella*; De Filippi, 1844). Nevertheless, from the methodological viewpoint, their approach – which was similar to the one followed by Hoang *et al.* (2010) and Sadeghi *et al.* (2014) – contradicted our findings because they suggested unnecessary the performance of any variable selection procedure, although the preference for smaller feature subsets was acknowledged. Even though we strongly favoured the model accuracy by assigning a weight of 0.875 instead of the 0.125 assigned to the number of selected variables, very few of them were selected. Furthermore, the selection of the optimal kernel has demonstrated to be a problem-dependant phenomenon (Howley and Madden, 2005); therefore, although apparently optimal for the aforementioned studies, these results cannot be extrapolated and several mapping functions should be tested in every study. Moreover, in Tirelli *et al.* (2012), and in other studies as well (Hoang et al., 2010; Sadeghi et al., 2014), the selected mapping function was polynomial, which involves the optimization of five different parameters. However, only the degree of the polynomial was optimized, although γ and C play crucial roles in the classificatory and generalization capability of SVMs (Cristianini and Schölkopf, 2002; Fukuda and De Baets, 2016; Fukuda et al., 2013; Huang and Wang, 2006). As a consequence, we would advocate at least for performing a grid-search of these parameters (Howley and Madden, 2005) to approximate the optimal values instead of using these default values, because our results indicated that optimal SVM can be also obtained with smaller input subsets. The simultaneous optimization of SVM parameters and the feature selection yielded competent models. However, dealing with noisy datasets the approach to calculate probabilities (Platt, 2000; Wu et al., 2004) has demonstrated awkward by trimming the output range. Such phenomena, apparently inherent of radial basis–like approaches (e.g. probabilistic neural networks,

Muñoz-Mas et al., 2014) should be thoroughly analysed by modelling several datasets in order to advocate for or rule out SVMs in further risk of invasion or ecological studies requiring probabilistic outputs.

4.5 Risk assessment and potential consequences

In addition to the training datasets, the selected modelling technique and the corresponding parameter settings may introduce uncertainty on models (Lin et al., 2015), potentially rendering contrasting habitat preferences (Fukuda et al., 2013; Fukuda and De Baets, 2016; Lin et al., 2015; Muñoz-Mas et al., 2016). One method to reduce the uncertainty of models' predictions is the ensemble modelling (Lin et al., 2015; Ren et al., 2016). Ensemble modelling consists of learning several models, each developed with a unique technique (Muñoz-Mas et al., 2015; Vezza et al., 2015) or with several different techniques (Muñoz-Mas et al., 2016; Thuiller et al., 2009), combining their individual forecasts into a single prediction (Ren et al., 2016). Assembling different modelling techniques shall better reproduce fish habitat preferences as it was demonstrated for the brown trout (Muñoz-Mas et al., 2016). However, ensemble modelling is not the panacea (Hannemann et al., 2016) and, thus, not only accuracy can determine the quality of additional predictions. In that case, transparency becomes a fundamental issue in the evaluation of the developed models (Austin, 2007). In our case, the developed partial dependence plots allowed us to evaluate the quality of the SVMs. Thus, the development of models with other techniques or its assembling was considered unnecessary since both models presented sound habitat preferences, especially for the better-known species (*i.e.* northern pike), and excellent accuracy with a small set of input variables, which is in agreement with the principle of parsimony. Nevertheless, in accordance with the aforementioned studies, the development of plenteous sets of models should be always advisable to obtain better insights on the habitat preferences and/or to eventually perform more accurate predictions, especially for species with no preceding references.

The habitat suitability for bleak suggested higher risk of invasion but, in accordance with the modelled habitat preferences, we considered it not extremely higher than the one posed by northern

pike. Our results confirmed in a much more detailed scale the classification of Almeida *et al.* (2013), who suggested *high* but not *very high* risk of invasion by northern pike. Conversely, the assessment for bleak would increase that score to *very high* risk of invasion. The most important impact of the northern pike introduction is the trophic alteration or the re-structuring of fish communities (Harvey, 2009) by predating the most palatable species and then readjusting its diet by predating other organisms (Sepulveda et al., 2013). The release of this piscivorous predator could then lead to a decline and/or extinction of the already threatened Cabriel's native species (Ribeiro and Leunda, 2012), as already reported in other research studies (Elvira et al., 1996). Consequently, its introduction in the upper Cabriel is likely to jeopardize the survival of the most important reservoirs of the Júcar nase, a fish species in imminent danger of extinction (Alcaraz et al., 2014), especially if pike was able to reach the river segment upstream Cristina's weir, which can be considered the last barrier impeding the invasion (Fig. 1).

The most probable impact of bleak invasion would be produced by an increase in resources competition (Almeida et al., 2014b) especially taking into account the species has shown an incredible high fecundity that allowed the bleak to outcompete other species (Vinyoles et al., 2007). In addition, its ability to exploit a widespread spectrum of prey and its temperature tolerance are not negligible. Another remarkable threat posed by bleak is its ability for hybridization, thus the hybridization of bleak and Iberian chub (*Squalius pyrenaicus*, Günther, 1868) quickly occurred after another invasion (Almodóvar et al., 2012). Likewise, it would be likely to occur with the Iberian chub specimens of the Cabriel River. Hybridisation could lead to loss of local adaptations, fitness, mating efficiency and reproductive output, as well as alteration of behaviour, migration patterns and life-cycle timing (Elvira and Almodóvar, 2001). Although the opposite may also occur thus, new invasive hybrid lineages can outcompete with native parentals through vigorous hybrids, enhancing the invasion success (Almodóvar et al., 2012). Consequently, the bleak invasion would have unpredictable consequences. Nevertheless, one aspect that may facilitate coexistence of invasive predators and native species is the spatial distribution and availability of refuges, in order to produce consistent habitat segregation between invasive and native species (Sepulveda et al., 2013). Therefore, we should strongly highlight the importance of those existent barriers, and especially the

conservation of the Cristina's weir, which is nowadays impeding the invasion of the non-indigenous Iberian nase to the last significant *stronghold* of the Júcar nase. This measure should be coupled with a close monitoring program because the positive impact of such artificial barrier would be worthless if these invasive species were able to colonize the upper stretches.

Finally, we have to acknowledge that this study is unable to foresee the ultimate impact of the simultaneous invasion by both species because the consequences of introductions tend to be negative in unpredictable ways conforming the so-called *Frankenstein effect* (Moyle et al., 1986). Further, some other studies presented a higher degree of sophistication by modelling the pattern of movements of pike (Baetens et al., 2013; Pauwels et al., 2013) or followed dynamic approaches (Vezza et al., 2012, 2014b), which would be able to render in-deep results of the habitat suitability under different scenarios in a better way than the group of the evaluated snapshots. However, the upper Cabriel corresponds to the unregulated stretch of the river thus flow management alternatives cannot be implemented. As a consequence we considered the study valuable since it analysed the habitat preferences of the invasive species in a relatively detailed scale and with high accuracy.

It has been stressed that understanding the driving mechanisms of invasions may help managers with limited resources to prioritise habitats for invasive suppression (Sepulveda et al., 2013). The optimal SVM included very few variables, which highlight the possibility of their use in studies of invasion risk in the nearby rivers (e.g. upper Júcar River). Furthermore, the ecological impacts of invasive species remain the subject of continuous debate, mainly because of a lack of indisputable evidence, which results from the scarcity of pre-invasion baseline information and specific post-invasion monitoring studies (Ribeiro and Leunda, 2012). Consequently, in addition to sealing and monitoring those existent barriers, the upper Cabriel River should be the subject of continuous monitoring programmes with special emphasis on the segments that connect the storage and hydropower facilities, either to impede the invasion of these species or to take early actions to mitigate its impact.

5 Conclusions

The simultaneous optimization of the SVM parameters and the variable selection demonstrated proficient to develop accurate SVMs with a small number of input variables, thus variable selection have demonstrated necessary for SVMs. The partial dependence plots suggested a positive relationship between width, volume and pike's probability of presence whereas vegetation showed a positive asymptotic pattern and the probability of presence decreased beyond 9000 small cyprinids per HMU. Depth presented a positive effect on bleak's probability of presence, especially above 2.75 m. Velocity presented a positive relationship with the presence of bleak whereas substrate had a negative asymptote thus bleak appeared more often in HMU with fine substrates. The habitat suitability for bleak suggested higher risk of invasion but it has been considered not extremely higher than the one posed by northern pike, due to limitations in the range of the outputs perhaps caused by mathematical limitations. The upper Cabriel River, especially the segments that connect the storage and hydropower facilities, should be regularly monitored to impede the invasion of these species or to restrict the negative impacts as soon as it took place.

Acknowledgments

The study has been partially funded by the IMPADAPT project (CGL2013-48424-C2-1-R) with Spanish MINECO (Ministerio de Economía y Competitividad) and by the Confederación Hidrográfica del Júcar (Spanish Ministry of Agriculture, Food and Environment). We also want to thank all the colleagues who worked in the field data collection, especially Rui M. S. Costa and Aina Hernández. Finally, we are especially grateful to Esther López Fernández who kindly and selflessly posed for the graphical abstract.

References

Alcaraz, C., Carmona-Catot, G., Risueño, P., Perea, S., Pérez, C., Doadrio, I., et al., 2014. Assessing population status of *Parachondrostoma arrigonis* (Steindachner, 1866), threats and

conservation perspectives. Environ. Biol. Fishes 98 (1), 443–455. http://dx.doi.org/10.1007/s10641-014-0274-3

Almeida, D., Grossman, G.D., 2012. Utility of direct observational methods for assessing competitive interactions between non-native and native freshwater fishes. Fish. Manag. Ecol. 19 (2), 157–166. http://dx.doi.org/10.1111/j.1365-2400.2012.00847.x

Almeida, D., Merino-Aguirre, R., Vilizzi, L., Copp, G.H., 2014a. Interspecific aggressive behaviour of invasive pumpkinseed *Lepomis gibbosus* in Iberian fresh waters. PLoS One 9 (2), e88038. http://dx.doi.org/10.1371/journal.pone.0088038

Almeida, D., Ribeiro, F., Leunda, P.M., Vilizzi, L., Copp, G.H., 2013. Effectiveness of FISK, an invasiveness screening tool for non-native freshwater fishes, to perform risk identification assessments in the Iberian Peninsula. Risk Anal. 33 (8), 1404–1413. http://dx.doi.org/10.1111/risa.12050

Almeida, D., Stefanoudis, P. V, Fletcher, D.H., Rangel, C., Da Silva, E., 2014b. Population traits of invasive bleak *Alburnus alburnus* between different habitats in Iberian fresh waters. Limnologica 46, 70–76. http://dx.doi.org/10.1016/j.limno.2013.12.003

Almodóvar, A., Nicola, G.G., Leal, S., Torralva, M., Elvira, B., 2012. Natural hybridization with invasive bleak *Alburnus alburnus* threatens the survival of Iberian endemic calandino *Squalius alburnoides* complex and Southern Iberian chub *Squalius pyrenaicus*. Biol. Invasions 14 (11), 2237–2242. http://dx.doi.org/10.1007/s10530-012-0241-x

Aparicio, E., Vargas, M.J., Olmo, J.M., De Sostoa, A., 2000. Decline of native freshwater fishes in a Mediterranean watershed on the Iberian Peninsula: A quantitative assessment. Environ. Biol. Fishes 59 (1), 11–19. http://dx.doi.org/10.1023/A:1007618517557

Austin, M., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecol. Modell. 200 (1–2), 1–19. http://dx.doi.org/10.1016/j.ecolmodel.2006.07.005

Baetens, J.M., Van Nieuland, S., Pauwels, I.S., De Baets, B., Mouton, A.M., Goethals, P.L.M., 2013. An individual-based model for the migration of pike (*Esox lucius*) in the river Yser, Belgium. Ecol. Modell. 258, 40–52. http://dx.doi.org/10.1016/j.ecolmodel.2013.02.030

Bain, M.B., Finn, J.T., Booke, H.E., 1985. A Quantitative Method for Sampling Riverine Microhabitats by Electrofishing. North Am. J. Fish. Manag. 5 (3), 489–493. http://dx.doi.org/10.1577/1548-8659(1985)5<489:AQMFSR>2.0.CO;2

Bry, C., 1996. Role of vegetation in the life cycle of pike. In: Craig, J.(ed.), Pike. Springer Netherlands, pp. 45-67; 3.

Casselman, J.M., Lewis, C.A., 1996. Habitat requirements of northern pike (*Esox lucius*). Can. J. Fish. Aquat. Sci. 53 (SUPPL. 1), 161–174. http://dx.doi.org/10.1139/f96-019

CHJ - Confederación Hidrográfica del Júcar (Spanish government), 2009a. Documento técnico de referencia: Evaluación del estado de las masas de agua superficial y subterránea València (Spain), pp. 55.

CHJ - Confederación Hidrográfica del Júcar (Spanish government), 2009b. Documento técnico de referencia: Identificación y delimitación de masas de agua superficial y subterránea València (Spain), pp. 55.

Clavero, M., 2011. Assessing the risk of freshwater fish introductions into the Iberian Peninsula. Freshw. Biol. 56 (10), 2145–2155. http://dx.doi.org/10.1111/j.1365-2427.2011.02642.x

Clavero, M., Blanco-Garrido, F., Prenda, J., 2004. Fish fauna in Iberian Mediterranean river basins: Biodiversity, introduced species and damming impacts. Aquat. Conserv. Mar. Freshw. Ecosyst. 14 (6), 575–585. http://dx.doi.org/10.1002/aqc.636

Connell, S.D., 2000. Is there safety-in-numbers for prey? Oikos 88 (3), 527–532. http://dx.doi.org/10.1034/j.1600-0706.2000.880308.x

Costa, R.M.S., Martínez-Capel, F., Muñoz-Mas, R., Alcaraz-Hernández, J.D., Garófano-Gómez, V., 2012. Habitat suitability modelling at mesohabitat scale and effects of dam operation on the endangered Júcar nase, *Parachondrostoma arrigonis* (River Cabriel, Spain). River Res. Appl. 28 (6), 740–752. http://dx.doi.org/10.1002/rra.1598

Cristianini, N., Schölkopf, B., 2002. Support vector machines and kernel methods: The new generation of learning machines. AI Mag. 23 (3), 31–41. http://dx.doi.org/10.1609/aimag.v23i3.1655

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, D., 2011. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: e1071), TU Wien (Austria). R Package Version 1.5-25.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14 (6), 927–930. http://dx.doi.org/10.1111/j.1654-1103.2003.tb02228.x

Elith, J., Leathwick, J.R., 2009. Species distribution models: Ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159

Elkins, D., Grossman, G.D., 2014. Invasive rainbow trout affect habitat use, feeding efficiency, and spatial organization of warpaint shiners. Biol. Invasions 16 (4), 919–933. http://dx.doi.org/10.1007/s10530-013-0548-2

Elvira, B., Almodóvar, A., 2009. Threatened fishes of the world: *Parachondrostoma turiense* (Elvira, 1987) (Cyprinidae). Environ. Biol. Fishes 86 (2), 337–338. http://dx.doi.org/10.1007/s10641-009-9516-1

Elvira, B., Almodóvar, A., 2001. Freshwater fish introductions in Spain: Facts and figures at the beginning of the 21st century. J. Fish Biol. 59 (SUPPL. A), 323–331. http://dx.doi.org/10.1006/jfbi.2001.1753

Elvira, B., Nicola, G.G., Almodovar, A., 1996. Pike and red swamp crayfish: A new case on predator-prey relationship between aliens in central Spain. J. Fish Biol. 48 (3), 437–446. http://dx.doi.org/10.1111/j.1095-8649.1996.tb01438.x

European Parliament & Council, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy.

Fox, J., 2010. polycor: Polychoric and Polyserial Correlations. R package version 0.7-8.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. http://dx.doi.org/10.1214/aos/1013203451

Fröhlich, H., Chapelle, O. and Schölkopf, B., 2003. Feature Selection for Support Vector Machines by Means of Genetic Algorithms. Proceedings: 15th IEEE International Conference on Tools with artificial Intelligence, Sacramento, CA, 142-148.

Fukuda, S., De Baets, B., 2016. Data prevalence matters when assessing species' responses using data-driven species distribution models. Ecol. Inform. 32, 69–78. http://dx.doi.org/10.1016/j.ecoinf.2016.01.005

Fukuda, S., De Baets, B., Waegeman, W., Verwaeren, J., Mouton, A.M., 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. http://dx.doi.org/10.1016/j.envsoft.2013.04.005

Fukuda, S., Mouton, A.M., De Baets, B., 2011. Abundance versus presence/absence data for modelling fish habitat preference with a genetic Takagi-Sugeno fuzzy system. Environ. Monit. Assess. 184 (10), 6159–6171. http://dx.doi.org/10.1007/s10661-011-2410-2

García de Jalón, D. and Schmidt, G., 1995. Manual práctico para la gestión sostenible de la pesca fluvial. Madrid, (Spain), pp. 169.

Garófano-Gómez, V., Martínez-Capel, F., Peredo-Parada, M., Olaya-Marín, E.J., Muñoz-Mas, R., Costa, R.M.S., et al., 2011. Assessing hydromorphological and floristic patterns along a regulated Mediterranean river: The Serpis River (Spain). Limnetica 30 (2), 307–328.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46 (1–3), 389–422. http://dx.doi.org/10.1023/A:1012487302797

Hannemann, H., Willis, K.J., Macias-Fauria, M., 2016. The devil is in the detail: unstable response functions in species distribution models challenge bulk ensemble modelling. Glob. Ecol. Biogeogr. 25 (1), 26–35. http://dx.doi.org/10.1111/geb.12381

Harby, A., Olivier, J.M., Merigoux, S., Malet, E., 2007. A mesohabitat method used to assess minimum flow changes and impacts on the invertebrate and fish fauna in the Rhône River, France. River Res. Appl. 23 (5), 525–543. http://dx.doi.org/10.1002/rra.997

Harvey, B., 2009. A biological synopsis of northern pike (*Esox lucius*). Victoria, B.C (Canada), pp 39.

Hoang, T.H., Lock, K., Mouton, A., Goethals, P.L.M., 2010. Application of classification trees and support vector machines to model the presence of macroinvertebrates in rivers in Vietnam. Ecol. Inform. 5 (2), 140–146. http://dx.doi.org/10.1016/j.ecoinf.2009.12.001

Holland, J.H., 1992. Genetic algorithms. Sci. Am. 267 (1), 66–72.

Howley, T., Madden, M.G., 2005. The genetic kernel support vector machine: Description and evaluation. Artif. Intell. Rev. 24 (3–4), 379–395. http://dx.doi.org/10.1007/s10462-005-9009-3

Huang, C.-L., Wang, C.-J., 2006. A GA-based feature selection and parameters optimization for support vector machines. Expert Syst. Appl. 31 (2), 231–240. http://dx.doi.org/10.1016/j.eswa.2005.09.024

Inskip, P.D., 1982. Habitat suitability index models: northern pike. Washington, DC (USA), pp. 50.

Jackson, D.A., Peres-Neto, P.R., Olden, J.D., 2001. What controls who is where in freshwater fish

communities - The roles of biotic, abiotic, and spatial factors. Can. J. Fish. Aquat. Sci. 58 (1), 157–170. http://dx.doi.org/10.1139/cjfas-58-1-157

Kerle, F., Zollner, F., Kappus, B., Marx, W., Giesecke, J., 2001. Fish habitats and vegetation modelling in floodplains with CASiMiR. Stuttgart (Germany), pp. 75.

Kobler, A., Klefoth, T., Wolter, C., Fredrich, F., Arlinghaus, R., 2008. Contrasting pike (*Esox lucius* L.) movement and habitat choice between summer and winter in a small lake. Hydrobiologia 601 (1), 17–27. http://dx.doi.org/10.1007/s10750-007-9263-2

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artif. Intell. 97 (1–2), 273–324. http://dx.doi.org/10.1016/S0004-3702(97)00043-X

Kottelat, M., Freyhof, J., 2007. Handbook of European Freshwater Fishes. Cornol (Switzerland) & Berlin (Germany), pp 646.

Kruskal, J.B., 1964a. Nonmetric multidimensional scaling: A numerical method. Psychometrika 29 (2), 115–129. http://dx.doi.org/10.1007/BF02289694

Kruskal, J.B., 1964b. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29 (1), 1–27. http://dx.doi.org/10.1007/BF02289565

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 3 (2), 18–22.

Lin, Y.-P., Lin, W.-C., Wu, W.-Y., 2015. Uncertainty in Various Habitat Suitability Models and Its Impact on Habitat Suitability Estimates for Fish. Water 7 (8), 4088–4107. http://dx.doi.org/10.3390/w7084088

Marchetti, M.P., Lockwood, J.L., Light, T., 2006. Effects of urbanization on California's fish diversity: Differentiation, homogenization and the influence of spatial scale. Biol. Conserv. 127 (3), 310–318. http://dx.doi.org/10.1016/j.biocon.2005.04.025

Martínez-Capel, F., García De Jalón, D., Werenitzky, D., Baeza, D., Rodilla-Alamá, M., 2009. Microhabitat use by three endemic Iberian cyprinids in Mediterranean rivers (Tagus River Basin, Spain). Fish. Manag. Ecol. 16 (1), 52–60. http://dx.doi.org/10.1111/j.1365-2400.2008.00645.x

Mas-Martí, E., García-Berthou, E., Sabater, S., Tomanova, S., Muñoz, I., 2010. Comparing fish assemblages and trophic ecology of permanent and intermittent reaches in a Mediterranean stream. Hydrobiologia 657 (1), 167–180. http://dx.doi.org/10.1007/s10750-010-0292-x

Mebane Jr, W.R., Sekhon, J.S., 2011. Genetic optimization using derivatives: The rgenoud package for R. J. Stat. Softw. 42 (11), 1–26.

Mouton, A.M., Alcaraz-Hernández, J.D., De Baets, B., Goethals, P.L.M., Martínez-Capel, F., 2011. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. Environ. Model. Softw. 26 (5), 615–622. http://dx.doi.org/10.1016/j.envsoft.2010.12.001

Mouton, A.M., Schneider, M., Depestele, J., Goethals, P.L.M., De Pauw, N., 2007. Fish habitat modelling as a tool for river management. Ecol. Eng. 29 (3), 305–315. http://dx.doi.org/10.1016/j.ecoleng.2006.11.002

Mouton, A.M., Van Der Most, H., Jeuken, A., Goethals, P.L.M., De Pauw, N., 2009. Evaluation of river basin restoration options by the application of the Water Framework Directive Explorer in the Zwalm River basin (Flanders, Belgium). River Res. Appl. 25 (1), 82–97.

http://dx.doi.org/10.1002/rra.1106

Moyle, P.B., Li, H.W. and Barton, B.A., 1986. The Frankenstein effect: Impact of introduced fishes on native fishes in North America. In: Stroud, R.H.(ed.), Fish Culture in Fisheries Management. American Fisheries Society, Bethesda, MD (USA), pp. 415-426.

Munné, A., Prat, N., Solà, C., Bonada, N., Rieradevall, M., 2003. A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR index. Aquat. Conserv. Mar. Freshw. Ecosyst. 13 (2), 147–163. http://dx.doi.org/10.1002/aqc.529

Muñoz-Mas, R., Fukuda, S., Vezza, P., Martínez-Capel, F., 2016. Comparing four methods for decision-tree induction: A case study on the invasive Iberian gudgeon (*Gobio lozanoi*; Doadrio and Madeira, 2004). Ecol. Inform. 34, 22–34. http://dx.doi.org/10.1016/j.ecoinf.2016.04.011

Muñoz-Mas, R., Lopez-Nicolas, A., Martínez-Capel, F., Pulido-Velazquez, M., 2016. Shifts in the suitable habitat available for brown trout (*Salmo trutta* L.) under short-term climate change scenarios. Sci. Total Environ. 544, 686–700. http://dx.doi.org/10.1016/j.scitotenv.2015.11.147

Muñoz-Mas, R., Martínez-Capel, F., Alcaraz-Hernández, J.D., Mouton, A.M., 2015. Can multilayer perceptron ensembles model the ecological niche of freshwater fish species? Ecol. Modell. 309–310 (0), 72–81. http://dx.doi.org/10.1016/j.ecolmodel.2015.04.025

Muñoz-Mas, R., Martínez-Capel, F., Garófano-Gómez, V., Mouton, A.M., 2014. Application of Probabilistic Neural Networks to microhabitat suitability modelling for adult brown trout (*Salmo trutta* L.) in Iberian rivers. Environ. Model. Softw. 59 (0), 30–43. http://dx.doi.org/10.1016/j.envsoft.2014.05.003

Muñoz-Mas, R., Martínez-Capel, F., Schneider, M., Mouton, A.M., 2012. Assessment of brown trout habitat suitability in the Jucar River Basin (SPAIN): Comparison of data-driven approaches with fuzzy-logic models and univariate suitability curves. Sci. Total Environ. 440, 123–131. http://dx.doi.org/10.1016/j.scitotenv.2012.07.074

Neill, S.R.J., Cullen, J.M., 1974. Experiments on whether schooling by their prey affects the hunting behaviour of cephalopods and fish predators. J. Zool. 172 (4), 549–569. http://dx.doi.org/10.1111/j.1469-7998.1974.tb04385.x

Nilsson, P.A., 2001. Predator behaviour and prey density: Evaluating density-dependent intraspecific interactions on predator functional responses. J. Anim. Ecol. 70 (1), 14–19. http://dx.doi.org/10.1046/j.1365-2656.2001.00472.x

Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: A primer for ecologists. Q. Rev. Biol. 83 (2), 171–193. http://dx.doi.org/10.1086/587826

Osuna, E., Freund, R. and Girosi, F., 1997. Training support vector machines: An application to face detection. Proceedings of the 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR (USA), 130-136.

Pauwels, I.S., Mouton, A.M., Baetens, J.M., Van Nieuland, S., De Baets, B., Goethals, P.L.M., 2013. Modelling a pike (*Esox lucius*) population in a lowland river using a cellular automaton. Ecol. Inform. 17, 46–57. http://dx.doi.org/10.1016/j.ecoinf.2012.04.003

Platt, J., 2000. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Smola, A.J. and Bartlett, P.J.(ed.), Advances in Large Margin Classifiers. MIT Press, Cambridge, MA (USA), pp. 61-74.

Poulos, H.M., Chernoff, B., Fuller, P.L., Butman, D., 2012. Ensemble forecasting of potential habitat for three invasive fishes. Aquat. Invasions 7 (1), 59–72. http://dx.doi.org/10.3391/ai.2012.7.1.007

R Core Team, 2015. R: A language and environment for statistical computing.

Ren, Y., Zhang, L., Suganthan, P.N., 2016. Ensemble Classification and Regression-Recent Developments, Applications and Future Directions. IEEE Comput. Intell. Mag. 11 (1), 41–53. http://dx.doi.org/10.1109/MCI.2015.2471235

Ribeiro, F., Elvira, B., Collares-Pereira, M.J., Moyle, P.B., 2008. Life-history traits of non-native fishes in Iberian watersheds across several invasion stages: A first approach. Biol. Invasions 10 (1), 89–102. http://dx.doi.org/10.1007/s10530-007-9112-2

Ribeiro, F., Leunda, P.M., 2012. Non-native fish impacts on Mediterranean freshwater ecosystems: Current knowledge and research needs. Fish. Manag. Ecol. 19 (2), 142–156. http://dx.doi.org/10.1111/j.1365-2400.2011.00842.x

Rincón, P.A., Velasco, J.C., González-Sánchez, N., Pollo, C., 1990. Fish assemblages in small streams in western Spain: The influence of an introduced predator. Arch. für Hydrobiol. 118 (1), 81–91. http://dx.doi.org/10.1002/aqc.679

Sadeghi, R., Zarkami, R., Van Damme, P., 2014. Modelling habitat preference of an alien aquatic fern, *Azolla filiculoides* (Lam.), in Anzali wetland (Iran) using data-driven methods. Ecol. Modell. 284, 1–9. http://dx.doi.org/10.1016/j.ecolmodel.2014.04.003

Schill, D.J., Griffith, J.S., 1984. Use of Underwater Observations to Estimate Cutthroat Trout Abundance in the Yellowstone River. North Am. J. Fish. Manag. 4 (4), 479–487. http://dx.doi.org/10.1577/1548-8659(1984)4<479:UOUOTE>2.0.CO;2

Sepulveda, A.J., Rutz, D.S., Ivey, S.S., Dunker, K.J., Gross, J.A., 2013. Introduced northern pike predation on salmonids in southcentral Alaska. Ecol. Freshw. Fish 22 (2), 268–279. http://dx.doi.org/10.1111/eff.12024

Stojkovic, M., Miloševic, D., Simic, S., Simic, V., 2014. Using a Fish-Based Model to Assess the Ecological Status of Lotic Systems in Serbia. Water Resour. Manag. 28 (13), 4615–4629. http://dx.doi.org/10.1007/s11269-014-0762-4

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD – a platform for ensemble forecasting of species distributions. Ecography (Cop.). 32 (3), 369–373. http://dx.doi.org/10.1111/j.1600-0587.2008.05742.x

Tirelli, T., Gamba, M., Pessani, D., 2012. Support vector machines to model presence/absence of *Alburnus alburnus alborella* (Teleostea, Cyprinidae) in North-Western Italy: Comparison with other machine learning techniques. Comptes Rendus - Biol. 335 (10–11), 680–686. http://dx.doi.org/10.1016/j.crvi.2012.09.001

Vapnik, V., 1995. The nature of statistical learning theory, Information Science and Statistics. New York, NY (USA), pp. 314.

Vezza, P., Muñoz-Mas, R., Martínez-Capel, F., Mouton, A., 2015. Random forests to evaluate biotic interactions in fish distribution models. Environ. Model. Softw. 67, 173–183. http://dx.doi.org/10.1016/j.envsoft.2015.01.005

Vezza, P., Parasiewicz, P., Calles, O., Spairani, M., Comoglio, C., 2014a. Modelling habitat

requirements of bullhead (*Cottus gobio*) in Alpine streams. Aquat. Sci. 76 (1), 1–15. http://dx.doi.org/10.1007/s00027-013-0306-7

Vezza, P., Parasiewicz, P., Rosso, M., Comoglio, C., 2012. Defining minimum environmental flows at regional scale: Application of mesoscale habitat models and catchments classification. River Res. Appl. 28 (6), 717–730. http://dx.doi.org/10.1002/rra.1571

Vezza, P., Parasiewicz, P., Spairani, M., Comoglio, C., 2014b. Habitat modeling in high-gradient streams: the mesoscale approach and application. Ecol. Appl. 24 (4), 844–861. http://dx.doi.org/10.1890/11-2066.1

Vinyoles, D., Robalo, J.I., de Sostoa, A., Almodóvar, A., Elvira, B., Nicola, G.G., et al., 2007. Spread of the alien bleak *Alburnus alburnus* (Linnaeus, 1758) (Actinopterygii, Cyprinidae) in the Iberian Peninsula: the role of reservoirs. Graellsia 63 (1), 101–110. http://dx.doi.org/10.3989/graellsia.2007.v63.i1.84

Wu, T.-F., Lin, C.-J., Weng, R.C., 2004. Probability estimates for multi-class classification by pairwise coupling. J. Mach. Learn. Res. 5, 975–1005.

Wu, W.-J., Lin, S.-W., Moon, W.K., 2012. Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images. Comput. Med. Imaging Graph. 36 (8), 627–633. http://dx.doi.org/10.1016/j.compmedimag.2012.07.004

Zarkami, R., 2008. Habitat suitability modelling of pike (Esox lucius) in rivers. Ghent, (Belgium), pp. 235.