

RESEARCH ARTICLE

Forecasting the future establishment of invasive alien freshwater fish species

Sam Wenaas Perrin^{1,2}  | Kim Magnus Bærum³  | Ingeborg Palm Helland³  | Anders Gravbrøt Finstad^{1,2} 

¹Department of Natural History, Norwegian University of Science and Technology, Trondheim, Norway

²Centre of Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway

³Norwegian Institute for Nature Research, Trondheim, Norway

Correspondence

Sam Wenaas Perrin
Email: sam.perrin@ntnu.no

Funding information

Norges Forskningsråd, Grant/Award Number: 243910 and 266574

Handling Editor: Steven Vamosi

Abstract

1. Invasive alien species constitute a major threat to the world's freshwater ecosystems. Human translocations as well as rising temperatures have allowed freshwater fish species to expand their distribution into novel ecosystems, often with negative effects on native biodiversity. Early intervention is key to restricting damage and further spread of invasive aliens. This makes identification of areas with high risk for the establishment of invasive alien species necessary in order to target monitoring and mitigation measures.
2. Here, we model lake-specific likelihood of establishment of five freshwater fish species which are increasing their distribution in Norway. In order to establish the likelihood of establishment resulting from human translocation, environmental factors or natural dispersal from an established population, a suite of anthropogenic and environmental covariates were included as predictors. We used these models to create a future scenario which modelled establishment risk for these species over a 50-year time period.
3. Connectivity of lakes to other extant populations and anthropogenic covariates influenced likelihood of establishment—and subsequently future establishment risk—the most across all species. The effects of temperature were variable, and for the most part had little effect on likelihood of establishment.
4. Our results indicate that human behaviour, infrastructure development and alterations of watershed connectivity are more important than climate induced range shifts on a short to medium time horizon.
5. *Synthesis and applications.* Our study demonstrates how risk assessments of invasive establishment can be synthesised based on readily available open data sources. This allows for the construction of tools to forecast invasion hotspots as a basis for designing mitigation actions, including early monitoring programs, horizon scanning initiatives and eradication measures. It also allows managers to determine where species are spreading as a result of direct human translocation, and where they are expanding as a result of increased temperatures.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. Journal of Applied Ecology published by John Wiley & Sons Ltd on behalf of British Ecological Society

KEYWORDS

connectivity, forecasting, freshwater fish, invasive alien species

1 | INTRODUCTION

Biological invasions constitute one of the world's largest threats to native biodiversity (Mack et al., 2000). Invasive alien species are capable of disrupting food webs and causing extirpations of local species (Hutchison & Armstrong, 1993). Human activity has in recent years resulted in a rise in both frequency and scale of biological invasions (Carpio et al., 2019; Seebens et al., 2020). Compounding the threat of invasive alien species has been the onset of climate change, which (a) has allowed invasive alien species to colonise new areas and (b) allowed established invasive alien species to expand their distribution further (Sharma et al., 2007; Walther et al., 2002).

Invasive aliens can spread quickly and escalate from a local to a regional problem subsequent to establishment. Actions targeted at dispersal vectors or eradication measures may then rapidly become impossible or insufficient. As such, early detection through intensive monitoring and immediate eradication are often the only effective management options available (Mack et al., 2000). Identifying areas at high risk of invasion is therefore invaluable for management (Hulme, 2015) as subsequent targeting and management of dispersal vectors can be used to limit the spread of invasive aliens. This is particularly useful in ecosystems where species are required to move through environmental corridors (Hulme, 2009).

Introductions and subsequent establishments of invasive alien species often occur in a distinctly non-random manner. This allows identification of invasion pathways and drivers (García-Díaz et al., 2018; Strayer, 2010). The invasion process can be divided into three stages; introduction, establishment and spread (Blackburn et al., 2011), providing invasion biology with a theoretical framework (Uden et al., 2015). From an applied perspective, the establishment marks the most important stage, as practical consequences for the relevant ecosystem beforehand are rare. In the current study, we therefore focus on factors affecting the likelihood of establishment. This involves the joint probability of establishment and introduction (either through anthropogenic translocation or through natural dispersal from a connected translocated population; Uden et al., 2015). Another motivation for the focus on establishment is the general lack of available data needed to identify the different stages of the invasion process (Allen et al., 2013). This is often due to initial observations usually being made after the establishment of a population, particularly in environments which are not directly observable, such as aquatic ecosystems.

Freshwater ecosystems are particularly useful for studying invasion pathways of invasive alien species, as their dendritic nature means that well-defined pathways between two points often exist (Fagan, 2002). Hence, when a species has been introduced at one location, avenues for further establishments up or downstream can be identified, preventing both invasion and its associated management

costs (Gallardo & Aldridge, 2018; Keller et al., 2008; establishments which occur as a result of spreading from an introduction at a connected location will henceforth be referred to as secondary dispersal).

Invasive alien freshwater fish are often translocated directly to new rivers or lakes as a result of being desired recreational fishing objects, preferred food or as bait for existing species (Carpio et al., 2019; García-Díaz et al., 2018; Strayer, 2010). It is therefore necessary to consider anthropogenic covariates such as human population density or the presence of nearby roads when modelling the establishment and subsequent spread of invasive alien species, rather than just natural environmental predictors (García-Díaz et al., 2018; Leathwick et al., 2016; Miró & Ventura, 2015; Rodríguez-Rey et al., 2019).

In this study, we aim to forecast potential risk of establishment of five invasive alien freshwater species through freshwater lakes in southern Norway. This area has seen much illegal transportation of freshwater species in recent decades (Hesthagen & Sandlund, 2012) with a warming climate likely resulting in further changes in interactions between invasive aliens and native species (Hein et al., 2013; Ohlund et al., 2014). We show that the combination of both population and human connectivity with occurrence-only data can be used to determine areas at risk of establishment by invasive alien species, creating a useful basis for local and regional managers to stop the spread of invasive alien species.

2 | MATERIALS AND METHODS

2.1 | Study design

Our study system consists of freshwater lakes located between 57.99 and 65.14 degrees latitude and 4.66 and 14.30 degrees longitude, and with an altitudinal range of 0.1 to 1,835 m. The area is divided by a mountain range running from north to south. Drainage basins to the east flow either into the Baltic Sea or the North Sea on Norway's south coast and are generally larger and less steep than those to the west.

The aforementioned steep topography in much of western Norway has made natural dispersal into much of this region impossible for non-anadromous species. This has left many catchments in this region with only one or two native fish species, which makes these ecosystems particularly vulnerable to invasion (Sandlund & Hesthagen, 2011).

In the present study we first parameterise models describing the likelihood of establishment of five invasive fish species outside of their historical native range in Norway. Second, we forecast establishments as the basis for maps of future establishment

risk. The species were the northern pike (*Esox lucius*, L.), European perch (*Perca fluviatilis*, L.), common roach (*Rutilus rutilus*, L.), European whitefish (*Coregonus lavaretus*, Valenciennes 1848) and the common rudd (*Scardinius erythrophthalmus*, L.). While all species have native distributions which intersect with Norway, all are alien to large parts of the country (Huitfeldt-Kaas, 1918). The intensification of human translocation over recent decades has led to the species studied here expanding their distribution range far outside their native distributions (Hesthagen & Sandlund, 2015).

All species are capable of disrupting native food webs upon establishment in naïve systems, and pike, roach and rudd have been placed on the Norwegian Alien Species List as regionally alien species of high or extremely high risk (Sandvik et al., 2020), and both perch and whitefish have had negative impacts on species in other countries in the same climatic zone (Eloranta et al., 2016; Hayden et al., 2013; Sandlund et al., 2013). Further information on species taxonomy and naming authorities can be found in Table S1.1 in Appendix S1.

2.2 | Species occurrence data

Occurrence data were downloaded from the Global Biodiversity Information Facility (GBIF, 2019). The initial download included data on occurrences of the focal species from 495 different datasets. A comprehensive description of the biodiversity informatics pipeline is given in Perrin, 2021 (<https://doi.org/10.5281/zenodo.4556024>). Data were filtered to exclude data outside Norway.

Lakes and occurrence records from the two northernmost Norwegian counties (Troms & Finnmark, Nordland) were also excluded. The low human population density and the vast area of northern Norway is likely to render registering of occurrences in this region much less complete than information on occurrences in southern Norway. The filtered dataset included data from 13 different sources, detailed in Appendix S2.

Species modelling entails large data requirements, and sources like GBIF, which contains over 1 billion georeferenced species records, make data accessible and free world-wide. As such, its effective integration into ecological studies could substantially assist our ability to model species distributions.

To determine the nature of an occurrence, we used historical native range of each species from maps initially published in Huitfeldt-Kaas (1918), which have subsequently been digitised and georeferenced (Daverdin et al., 2019). Any occurrence which fell outside of this range was considered to be the result of either secondary dispersal from a connected population or direct translocation by a human. Occurrences within the native range for all species were discounted. Native distribution ranges and subsequent establishments are shown in Figure 1a–e.

In order to bring the dataset to a less computationally demanding level, particularly with regard to inferred absence data (see below), we disregarded lakes with a surface area under two hectares.

Further information on species selection and occurrence data can be found in Appendix S1.

2.3 | Absence data design

There are valid concerns regarding observation bias when using occurrence-only data. This is particularly true when using compiled data sources which do not include information on sampling design, such as opportunistically collected citizen science data (Moudrý & Devillers, 2020; Troia & McManamay, 2016). For example, citizen scientists are likely to report observations from lakes with higher human accessibility with a higher effort than less accessible lakes. To account for potential spatial bias, we drew a restricted number of pseudo-absences from background data points representing similar spatial bias as the occurrence data using the sampling design described by Barbet-Massin et al. (2012). Further information can be found in Appendix S3. We compared the species models which used this sampling design to a second set of models, which simply used all lakes outside of the species' native ranges where occurrences were not registered as pseudo-absences.

2.4 | Environmental data

The environmental data consisted of seven covariates describing environmental properties, including lake connectivity or human impact properties with the potential to affect establishment likelihood. This included the environmental covariates lake area and average air temperature of the warmest annual quarter (henceforth referred to simply as temperature). Temperature data were derived from the EuroLST dataset for the centrepoint of each lake (Metz et al., 2014). Human Footprint Index (henceforth referred to as HFI) was used to measure human impact (Venter et al., 2016).

Connectivity covariates related to secondary dispersal included the presence or absence of extant populations upstream or downstream of the focal lake at year of establishment ($\text{population}_{\text{up}}$ and $\text{population}_{\text{down}}$).

The inverse distances to all lakes with an extant population of the focal species within 100 kilometres were aggregated using the FNN library (Beygelzimer et al., 2019). This formed the covariate proximity of extant populations, $\text{population}_{\text{prox}}$. This added covariate was necessary to account for likelihood of human-assisted translocation, as catchment or drainage based connectivity covariates are unlikely to impose the same hindrances on human movement as they do for fish. An additional covariate measuring the shortest distance between any point on the lake boundary and the nearest road was also used to measure human connectivity, forming the covariate $\text{road}_{\text{prox}}$. Comprehensive summaries of environmental covariates can be found in Appendix S1.

Interaction terms were added between HFI and $\text{population}_{\text{prox}}$, $\text{road}_{\text{prox}}$ and $\text{population}_{\text{up}}/\text{population}_{\text{down}}$ and between $\text{population}_{\text{up}}/\text{population}_{\text{down}}$ and $\text{population}_{\text{prox}}$, to measure whether the effect of

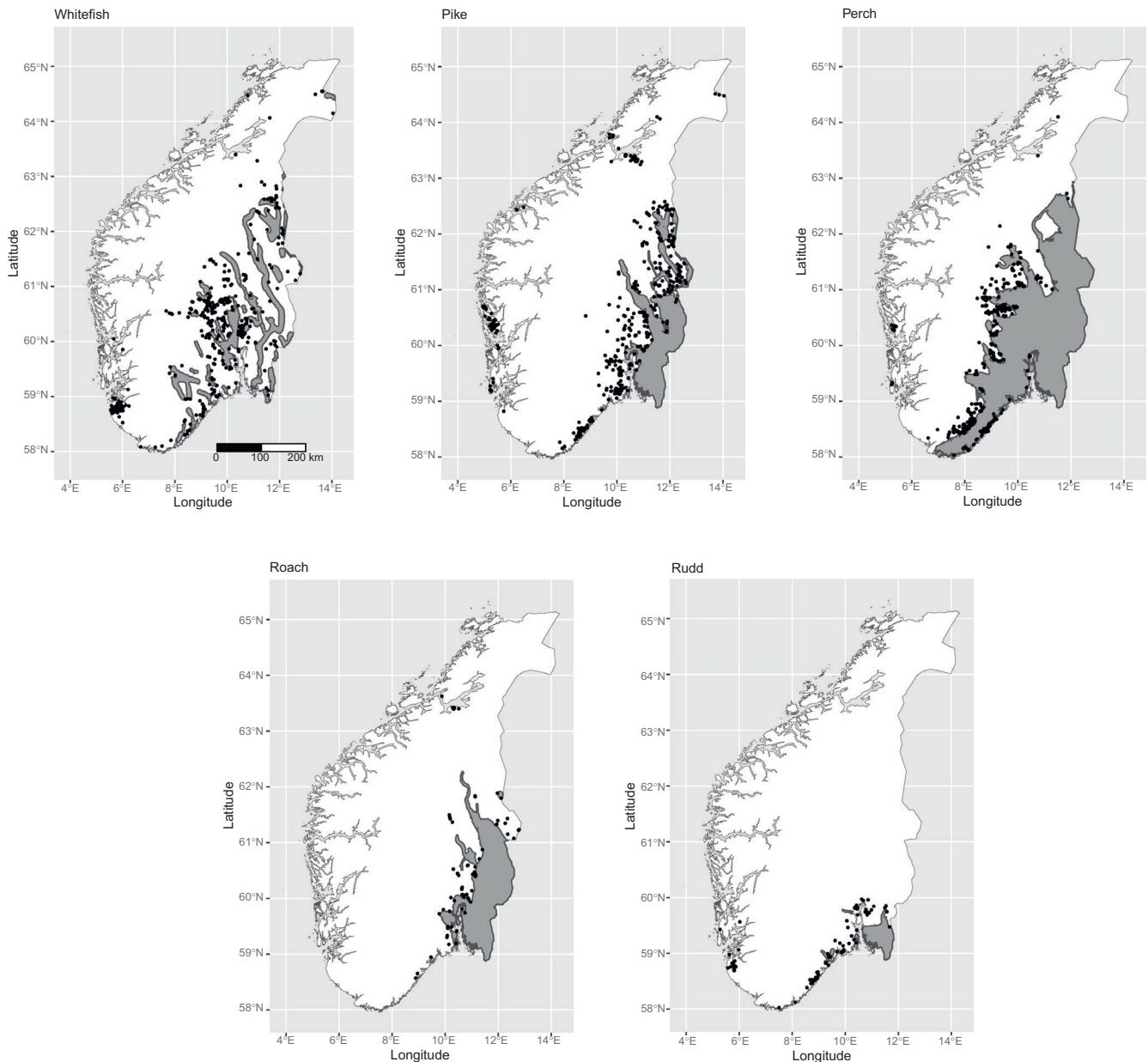


FIGURE 1 Establishments outside of native distribution range since 1918 for five species of Norwegian freshwater fish. Native distribution range was determined by digitisation of Huitfeldt-Kaas (1918) and subsequent establishments were taken from GBIF

human-related connectivity covariates changed in areas with higher HFI, or where species were already present within the watershed.

Original data can be downloaded and models can be replicated using R scripts found in Perrin, 2021 (<http://doi.org/10.5281/zenodo.4556024>). All statistical modelling used R (R Core Team, 2015) and RStudio (RStudio Team, 2020).

2.5 | Statistical modelling

We modelled establishment of each species as:

$$E \sim \text{Bernoulli}(p),$$

with

$$p_{ij} = \varphi^{-1}(\eta_{ij}),$$

where p_{ij} denotes the probability of a species being present at site i , and φ denotes our probit link function. We then denote η_i as:

$$\eta_i = \alpha + \sum_{k=1}^{n_c} x_{ik} \beta_k,$$

where α denotes the intercept, x_{ik} denotes the value of the environmental covariates/interaction terms k at each site i and β_k denotes the regression coefficient of environmental covariate k . We fitted the model with a Bayesian approach, utilising the GRETA package (Golding,

TABLE 1 Goodness-of-fit for models analysing the effect of seven environmental covariates on likelihood of establishment of five different invasive alien freshwater fish species in lakes outside of their historical native range. Mean values given for each measure of goodness-of-fit with accompanying standard deviations. Values given for models which used all background lakes as pseudo-absences, and for models which accounted for spatial bias by using a targeted background design

Species	AUC		Proportion deviance explained		Sensitivity	
	All background samples	Target background design	All background samples	Target background design	All background samples	Target background design
Whitefish	96.2 (± 0.3)	95.3 (± 0.6)	53.7 (± 2.8)	53.6 (± 3.5)	86.6 (± 2.3)	83.4 (± 1.8)
Pike	98.5 (± 0.1)	96.9 (± 0.2)	65.8 (± 1.1)	61.8 (± 1.2)	95.0 (± 1.1)	93.2 (± 3.1)
Roach	99.7 (± 0.2)	99.1 (± 0.5)	75.6 (± 8.0)	71.6 (± 9.7)	94.7 (± 5.0)	94.7 (± 5.0)
Rudd	99.8 (± 0.1)	99.7 (± 0.2)	79.8 (± 6.3)	79.6 (± 6.6)	98.8 (± 2.4)	97.7 (± 2.3)
Perch	99.3 (± 0.2)	98.6 (± 0.3)	74.4 (± 1.8)	70.7 (± 2.0)	97.8 (± 1.1)	95.6 (± 2.3)

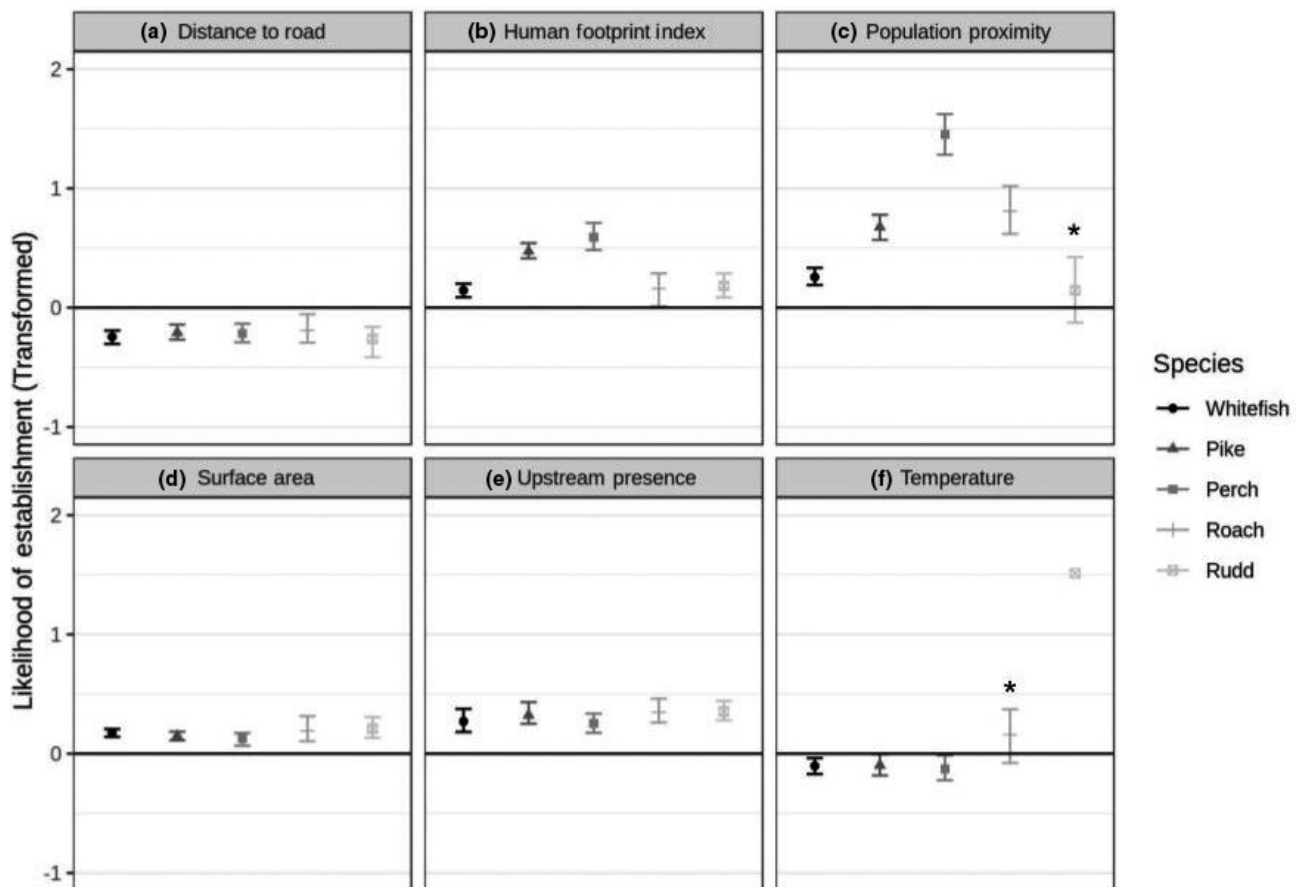


FIGURE 2 95% credible intervals of effects of different environmental and anthropogenic covariates on likelihood of establishment of five different invasive alien freshwater fish species in lakes outside their native range in southern Norway. Covariates include (a) shortest distance from lake perimeter to road, (b) Human Footprint Index, (c) proximity of nearby extant populations at year of observation, (d) surface area of lake, (e) the presence of an extant upstream population at year of observation, (f) average temperature of warmest annual quarter. Variables (a), (c) and (d) were heavily right skewed and thus log transformed to aid with model convergence. All variables were then scaled to a mean of zero and a standard distribution of one. Asterisk indicates that the effect's credible interval did not intersect with zero and was therefore not considered significant. Error bar for the effect of temperature on rudd was excluded as it was significantly larger than other credible intervals (lower bound = 0.76, upper bound = 2.12). Likelihood of establishment shown here logit transformed

2019). Each covariate was scaled to a mean of zero and a standard deviation of one.

Markov chain Monte Carlo (MCMC) sampling was done using 2,000 samples on four chains, giving a total of 8,000 samples, with

a burn-in of 1,000 samples on each chain. We used a Hamiltonian Monte Carlo sampler, sampling the number of leapfrog steps at each iteration uniformly between 40 and 60 (these numbers were manually tuned to achieve efficient sampling). Covariate effects were

considered to be significant if their 95% credible intervals did not intercept with zero. We used Rhat values to assess convergence, with values below 1.1 being considered acceptable. All covariate estimates converged acceptably for every model.

2.6 | Model evaluation

To validate our model and compare our sampling techniques, we used a fivefold cross-validation approach, splitting the dataset into five training and validation sets. To measure goodness-of-fit we calculated the proportion of explained deviance using an adjusted Dsquared function from the R package MODEvA (Márcia Barbosa et al., 2013). We also calculated sensitivity using:

$$\frac{a}{a + c},$$

where a is the number of true presences correctly predicted, and c is the number of false absences predicted by our model. The threshold for determining predicted presence or absence was set using individual species prevalence, as recommended by Liu et al. (2005). As an additional indicator of the model's ability to separate true presences from pseudo-absences, we calculated area under operating curve (AUC).

2.7 | Scenario modelling

We utilised the covariate estimates obtained from the regression model to predict future establishment risk for each species outside of their native range in Norway. To do this we ran 1,000 simulations. Our scenario comprised five 10-year intervals, over each of which the likelihood of establishment was determined based on covariate estimates taken randomly from one of the 8,000 MCMC draws. The interval length was chosen arbitrarily in order to allow us to observe the effect of secondary dispersal over a relevant time period. The rbinom function was then used to decide whether or not a lake was colonised based on that likelihood. After every step, our three connectivity covariates were recalculated for each lake to reflect new establishments of the focal species. The final establishment risk for a specific lake was then calculated based on the number of establishment events recorded during the 1,000 simulations.

3 | RESULTS

We found varying numbers of establishments across species, with lower rates of establishment for rudd and roach (Table S1.1, Appendix S1). Interaction terms between HFI and $road_{prox}$, between HFI and $population_{up}/population_{down}$ and between $population_{down}$ and $population_{prox}$ were non-significant and did not contribute to model fit, and as such were discarded from further analysis.

Our fivefold cross validation showed that AUC, deviance explained and sensitivity values were high across all species, although

less so for whitefish (Table 1). Furthermore, it revealed that model fit did not improve when using restricted pseudo-absences drawn from a target background sampling design. As such, all analysis henceforth used our model with full pseudo-absence sampling.

3.1 | Statistical modelling

The effects of most environmental covariates showed similar trends across species (Figure 2a–e). $population_{down}$ had a strong positive effect on likelihood of establishment of all species, while $population_{up}$ and lake area had a weaker positive (yet still significant) effect across all species. $road_{prox}$ had a significant negative effect across all species. $population_{prox}$ and HFI had positive effects of varying strength across all species, with stronger effects across both covariates on perch and pike, and $population_{prox}$ also having a strong effect on roach.

Temperature was the only covariate which clearly varied in effect direction between species (Figure 2f). It had a significant negative effect on whitefish, pike and perch, and a very strong significant positive effect on rudd, although the credible interval on this was extremely large. It also had a non-significant positive effect on roach.

The interaction term between $population_{prox}$ and HFI for whitefish, pike and perch showed a significant negative effect (Figure 3). There was a significant negative effect of the interaction between $population_{up}$ and $population_{down}$ (Figure 4). The interaction term between $population_{up}$ and $population_{prox}$ across all species showed a significant interaction term, yet it was extremely weak and as such was not included in our Figures.

Variance in covariate effects was generally higher for rudd and roach, particularly for temperature and $population_{prox}$. Full credible intervals across all variables and species can be found in Appendix S4.

3.2 | Scenario modelling

We forecasted much larger extents of areas of high establishment risk (orange–red) for pike, perch and roach (Figures 5b,c,e). Pike showed a larger extent of areas of medium establishment risk (yellow), although within these areas there were much smaller and less concentrated extents of areas of high establishment risk than for perch and roach. All three showed a spread mainly limited to eastern Norway, with limited capacity to spread in the west of Norway outside of the areas where establishments had already taken place.

Although whitefish showed a similar extent of medium risk areas to pike, their concentration was far less dense, and high risk areas were present infrequently throughout this extent (Figure 5a). Rudd showed very low risk of further establishment (Figure 5d).

An example detailing the use of the establishment risk in regional forecasting tools has been provided in Appendix S5 (Figure S5.1).

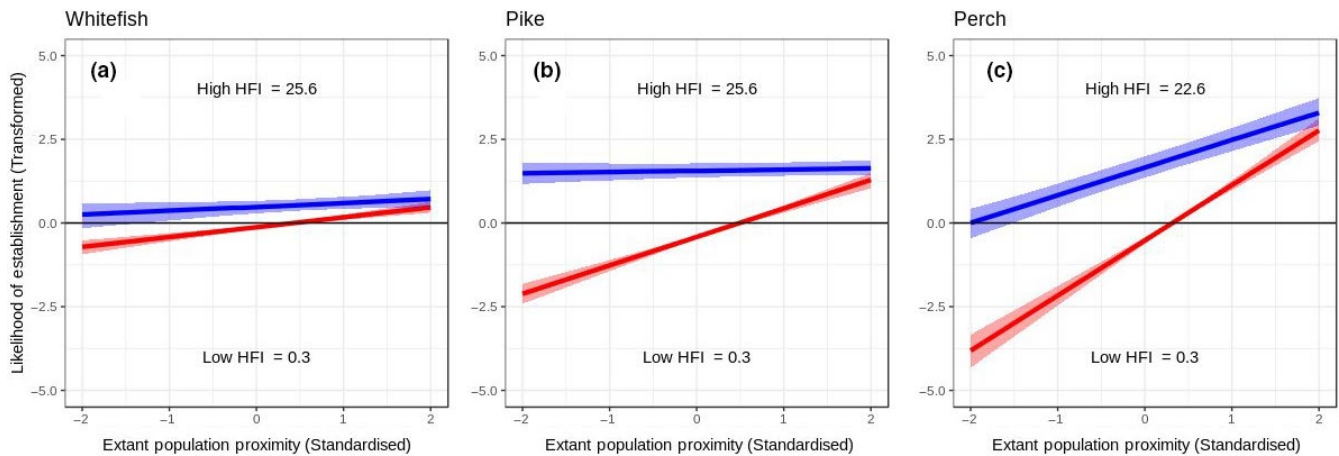


FIGURE 3 Interaction between effects of Human Footprint Index (HFI) and proximity of extant populations on the likelihood of establishment of three different invasive alien freshwater fish species in freshwater lakes in southern Norway. Species are (a) whitefish, (b) pike and (c) perch. Proximity values are standardised to a mean of zero and standard deviation of one, with likelihood values logit transformed. Blue trends indicate effect of proximity of extant populations on likelihood of introduction at high HFI scores, while red indicates the same trend at low HFI scores. HFI is an index ranging from 0 to 50 which qualifies human impact on local environment. Low and high HFI scores used are 2.5% and 97.5% quantiles of scores for lakes used in models

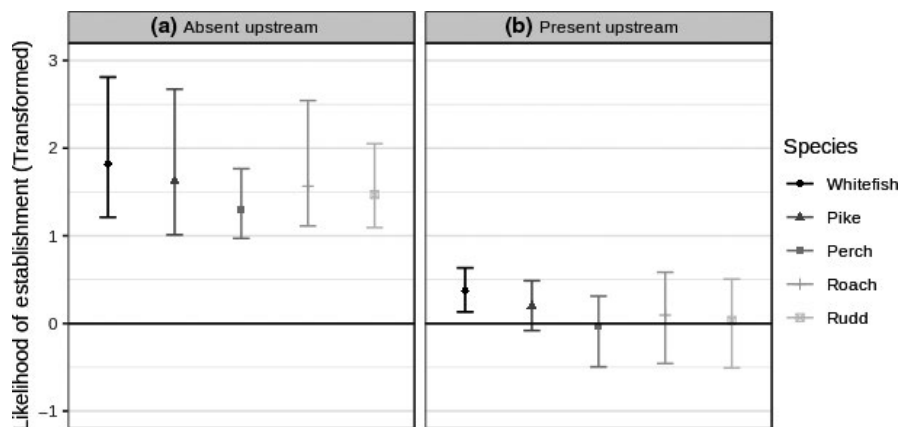


FIGURE 4 Effect of downstream presence of an extant population on the likelihood of establishment of five different species of regionally invasive alien fish in freshwater lakes in southern Norway. Effects are shown for when an extant population is (a) absent and (b) present upstream. Effects are the product of species distribution models which model likelihood of establishment of species in lakes outside of the species native range based on seven different covariates, including the two on display here. Likelihood of establishment is shown here logit transformed

4 | DISCUSSION

Understanding where and why biological invasions are likely to occur is a necessity if we are to make informed management decisions. In this study we have used a combination of anthropogenic and environmental covariates to model likelihood of establishment in lakes outside of the native range of five species of invasive alien freshwater fish in Norway. Our results show that the majority of these covariates, along with several interaction terms, affect the modelled likelihood of establishment in a consistent way across the five species. Increases in lake surface area, Human Footprint Index, proximity of extant populations and the presence of extant populations both upstream and downstream had positive effects on the likelihood of establishment for all species. Similarly, all species

decreased in likelihood of establishment with an increased distance to the nearest road. Temperature was the only covariate which varied in its effect direction across species.

We used this model to forecast establishments of the five invasive alien species over a 50-year timespan, showing that the future spatial distribution of establishment risk did vary somewhat between species. However, across all species, areas of high establishment risk seemed confined to regions close to populations established prior to the beginning of the scenario modelling, although the scenario did produce some areas of high establishment risk located considerably further from these populations.

Many of these results are in line with expectations. Larger, more accessible lakes are likely to have higher angler occurrence, and subsequent higher likelihoods of establishment of an invasive alien

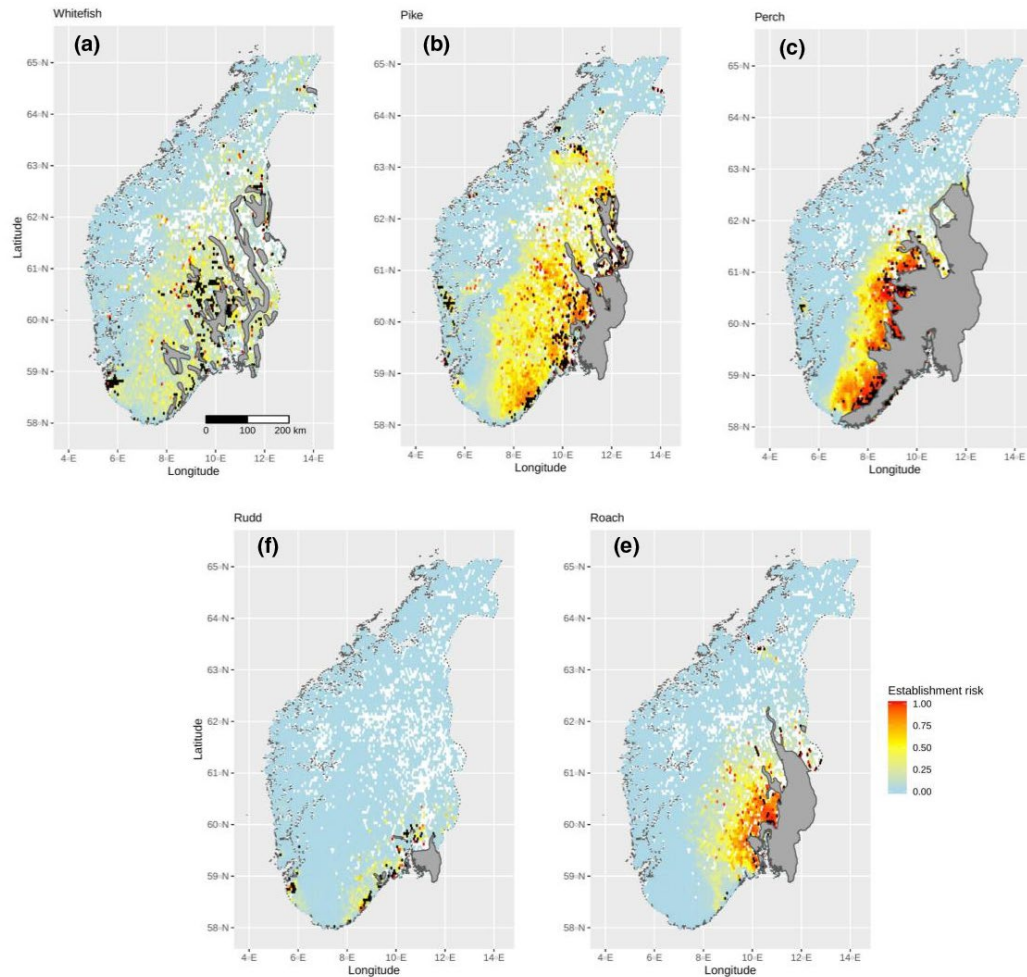


FIGURE 5 Visualisations of forecasted establishment risk for five freshwater fish species, including (a) whitefish, (b) pike, (c) perch, (d) rudd and (e) roach. Maps show a 50-year forecast of establishment risk. Grey polygons indicate each species' native range as inferred by Huitfeldt-Kaas (1918). Black hexagons indicate areas outside of the native range where a species has established since 1918, all other hexagons dictate risk of establishment on a scale from 0 to 1

species (Hunt et al., 2019; Post et al., 2008). Additionally, higher proximity of nearby extant populations make both human-assisted and natural secondary dispersal more likely. Both factors are likely to increase propagule pressures on lakes. This is consistent with previous studies showing human activity to be responsible for the majority of invasive alien fish distribution (Leprieur et al., 2008). The presence of an extant population of a species both upstream and downstream enables natural dispersal through waterways. Dispersal from an upstream population in particular is likely to be straightforward, particularly when taking into account Norway's steep topography, making establishments downstream of this population likely given sufficient time.

The effects of the interaction terms for whitefish, pike and perch indicate that in areas further from human activity, the proximity of extant populations has a stronger positive effect on likelihood of establishment. These three species have been introduced in attempts to establish populations, so this could be a product of anglers being less inclined to establish populations into areas less accessible to humans (Hunt et al., 2019). It could also be a product of a higher

population density increasing the potential occurrence of humans willing to translocate fish from regions further away. This would render the proximity to extant populations less important in densely human populated areas.

The variation in the spatial distribution of establishment risk across species seemed to more or less correspond to the variation in the effect of extant population proximity on likelihood of establishment (Figure 5). This effect was strongest for pike, perch and roach. This variation may result from differences in dispersal ability between species (Perrin et al., 2020). However, if an increase in extant population proximity only affected establishment via secondary dispersal, we would have expected the effect of extant population proximity to be more strongly influenced by whether or not there was an extant population either upstream or downstream. It is therefore likely that at least some of the effect of proximity to extant populations is caused by human activity. Variation in rates of translocation could be brought about due to differences in cultural attitudes and motivations for translocation which occur across these five species. For instance, it has been hypothesised that pike and

perch are generally illegally translocated by fishers to establish new populations, while rudd and roach are unintentionally introduced when used as bait (Hesthagen & Sandlund, 2012, 2015). Additionally, the low effect of population proximity on whitefish is potentially a product of extensive government translocation to fishless lakes in the early 1900s (Sandlund et al., 2013).

Previous studies from northern regions show that most species benefit from increased temperatures (Byström et al., 2007; Hayden et al., 2017). The heterogeneous response to temperature on likelihood of establishment across species observed in the current study may therefore appear counter-intuitive. However, the spatial, and hence climatic, extent of the current study was limited. All focal species, with exception of rudd, are likely to be well within their thermal tolerance throughout our study area. Furthermore, the current observed change in distribution of many freshwater fishes is likely to be driven more by human translocation than climatic shifts (Cazelles et al., 2019; Leprieur et al., 2008). This is not to suggest that managers should not be concerned about future climate change. Aside from increased temperatures potentially allowing for successful establishments in new lakes as well as increasing the dispersal capacity of more warm-tolerant species (Rahel & Olden, 2008), an increase in temperature could exacerbate the effects of the species studied here in the lakes they are presently established in (Ohlund et al., 2014; Rolls et al., 2017). This could result in increased homogenisation and reduction of species richness in freshwater communities across Norway.

The data used in the current study were compiled from 13 different openly published datasets, and metadata was in most instances limited. Thus, it was not feasible to estimate date of introduction or size of the initial propagule, data which can aid invasive species modelling considerably (Uden et al., 2015). Including populations that are potentially not yet established in our data may increase the extent of forecasted areas at high risk of establishment, however this potentially lessens the likelihood of false absences (Cordier et al., 2020).

An additional problem is the potential lack of data concerning sampling effort which is often present in compiled data sources with limited metadata. This lack of data can make it difficult to discern whether or not datasets are subject to sampling bias, a trend which has been noted before, including in studies of freshwater fish (Troia & McManamay, 2016). However, accounting for spatial bias here did not increase model fit, and the exclusion of inaccessible areas may even reduce model fit if accessibility is an importable variable (Chapman et al., 2019). Additionally, several of the studies included here were designed to ensure representative coverage of lakes across a range of human accessibility covariates, such as the 1995 Nordic Freshwater Fish Survey (Tammi et al., 2003).

Our results show that it is possible to identify lakes which are particularly vulnerable to the establishment of invasive alien species from available occurrence-only data, such as those originating from citizen science initiatives. This is an important step forward, as eradication treatment is a resource-intensive process, and the ability to conduct early monitoring using citizen science or eDNA techniques

in vulnerable lakes and identify where establishments are most likely to lead to a species spreading to further locations can help prioritisation of resources significantly. Given sufficient temperature data, it could also allow managers to determine where species are spreading as a result of direct human translocation, and where they are expanding their distribution as a result of increased temperatures.

We also note that we were unable to employ the current approach to several Norwegian species due to data deficiency. As such, expanding data material is an important first step, particularly for models that rely on occurrence-only data (Simmonds et al., 2020). While species here differed in their predicted spread, the covariates which informed our predictions of the subsequent spread affected the study species in the same way for the most part. This provides a useful tool both for managers within areas that already contain populations of invasive alien freshwater fish, and to those within areas vulnerable to invasion.

ACKNOWLEDGEMENTS

We thank Nick Golding for his help in the use of the GRETA package and for advancing our understanding of connectivity modelling, and Marc Daverdin for his help mapping the native ranges of species included in this paper. We also thank Antti Eloranta and Odd-Terje Sandlund for their contributions to our understanding of Arctic and sub-Arctic fish biology. Initial planning was funded through the INVAFISH project (Norwegian Research Council 243910). Sam Perrin was supported by a PhD grant from the ERA-Net BiodiverERsA project ODYSSEUS (Norwegian Research Council 266574).

AUTHORS' CONTRIBUTIONS

S.W.P., K.M.B., I.P.H. and A.G.F. conceived the ideas; S.W.P. and K.M.B. designed methodology; S.W.P. analysed the data and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

Data available via the Global Biodiversity Information Facility at <https://doi.org/10.15468/dl.ws1uri> (GBIF.org, 2019). All scripts used for producing likelihood of establishment models and future risk scenarios are available via Zenodo at <http://doi.org/10.5281/zenodo.4556024> (Perrin, 2021).

ORCID

Sam Wenaas Perrin  <https://orcid.org/0000-0002-1266-1573>
 Kim Magnus Bærum  <https://orcid.org/0000-0003-1197-9608>
 Ingeborg Palm Helland  <https://orcid.org/0000-0001-7380-3152>
 Anders Gravbrøt Finstad  <https://orcid.org/0000-0003-4529-6266>

REFERENCES

- Allen, C. R., Nemeč, K. T., Wardwell, D. A., Hoffman, J. D., Brust, M., Decker, K. L., Fogell, D., Hogue, J., Lotz, A., Miller, T., Pummill, M., Ramirez-Yañez, L. E., & Uden, D. R. (2013). Predictors of regional establishment success and spread of introduced non-indigenous vertebrates. *Global Ecology and Biogeography*, 22(7), 889–899. <https://doi.org/10.1111/geb.12054>

- Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3(2), 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., & Li, S. (2019). *FNN: Fast nearest neighbour search algorithms and applications*. R package version 1.1.3. Retrieved from <https://CRAN.R-project.org/package=FNN>
- Blackburn, T. M., Pyšek, P., Bacher, S., Carlton, J. T., Duncan, R. P., Jarošík, V., Wilson, J. R. U., & Richardson, D. M. (2011). A proposed unified framework for biological invasions. *Trends in Ecology & Evolution*, 26(7), 333–339. <https://doi.org/10.1016/j.tree.2011.03.023>
- Byström, P., Karlsson, J., Nilsson, P., van Kooten, T., Ask, J., & Olofsson, F. (2007). Substitution of top predators: Effects of pike invasion in a subarctic lake. *Freshwater Biology*, 52(7), 1271–1280. <https://doi.org/10.1111/j.1365-2427.2007.01763.x>
- Carpio, A. J., De Miguel, R. J., Oteros, J., Hillström, L., & Tortosa, F. S. (2019). Angling as a source of non-native freshwater fish: A European review. *Biological Invasions*, 21, 3233–3248. <https://doi.org/10.1007/s10530-019-02042-5>
- Cazelles, K., Bartley, T., Guzzo, M. M., Brice, M.-H., MacDougall, A. S., Bennett, J. R., Esch, E. H., Kadoya, T., Kelly, J., Matsuzaki, S.-I., Nilsson, K. A., & McCann, K. S. (2019). Homogenization of freshwater lakes: Recent compositional shifts in fish communities are explained by gamefish movement and not climate change. *Global Change Biology*, 25(12), 4222–4233. <https://doi.org/10.1111/gcb.14829>
- Chapman, D., Pescott, O. L., Roy, H. E., & Tanner, R. (2019). Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. *Journal of Biogeography*, 46(5), 1029–1040. <https://doi.org/10.1111/jbi.13555>
- Cordier, J. M., Loyola, R., Rojas-Soto, O., & Nori, J. (2020). Modeling invasive species risk from established populations: Insights for management and conservation. *Perspectives in Ecology and Conservation*, 18(2), 132–138. <https://doi.org/10.1016/j.pecon.2020.06.001>
- Daverdin, M., Finstad, A. G. & Blumentrath, S. (2019). Freshwater fish native distribution map transcriptions from Huitfeldt-Kaas (1918). <https://doi.org/10.21400/1mw73950>
- Eloranta, A. P., Helland, I. P., Sandlund, O. T., Hesthagen, T., Ugedal, O., & Finstad, A. G. (2016). Community structure influences species' abundance along environmental gradients. *Journal of Animal Ecology*, 85(1), 273–282. <https://doi.org/10.1111/1365-2656.12461>
- Fagan, W. F. (2002). Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology*, 83(12), 3243–3249. [https://doi.org/10.1890/0012-9658\(2002\)083\[3243:CFAERI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[3243:CFAERI]2.0.CO;2)
- Gallardo, B., & Aldridge, D. C. (2018). Inter-basin water transfers and the expansion of aquatic invasive species. *Water Research*, 143, 282–291. <https://doi.org/10.1016/j.watres.2018.06.056>
- García-Díaz, P., Kerecsy, A., Unmack, P. J., Lintermans, M., Beatty, S. J., Butler, G. L., Freeman, R., Hammer, M. P., Hardie, S., Kennard, M. J., Morgan, D. L., Pusey, B. J., Raadik, T. A., Thiem, J. D., Whiterod, N. S., Cassey, P., & Duncan, R. P. (2018). Transport pathways shape the biogeography of alien freshwater fishes in Australia. *Diversity and Distributions*, 24(10), 1405–1415. <https://doi.org/10.1111/ddi.12777>
- GBIF.org. (2019). *GBIF occurrence download*. Retrieved from <https://doi.org/10.15468/dl.ws1uri>
- Golding, N. (2019). greta: Simple and scalable statistical modelling in R. *Journal of Open Source Software*, 4(40), 1601. <https://doi.org/10.21105/joss.01601>
- Hayden, B., Holopainen, T., Amundsen, P.-A., Eloranta, A. P., Knudsen, R., Praebel, K., & Kahilainen, K. K. (2013). Interactions between invading benthivorous fish and native whitefish in subarctic lakes. *Freshwater Biology*, 58(6), 1234–1250. <https://doi.org/10.1111/fwb.12123>
- Hayden, B., Myllykangas, J.-P., Rolls, R. J., & Kahilainen, K. K. (2017). Climate and productivity shape fish and invertebrate community structure in subarctic lakes. *Freshwater Biology*, 62(6), 990–1003. <https://doi.org/10.1111/fwb.12919>
- Hein, C. L., Ohlund, G., & Englund, G. (2013). Fish introductions reveal the temperature dependence of species interactions. *Proceedings of the Royal Society B: Biological Sciences*, 281(1775), 20132641. <https://doi.org/10.1098/rspb.2013.2641>
- Hesthagen, T. H., & Sandlund, O. T. (2012). *Gjedde, sørv og suter: Status, vektorer og tiltak mot uønsket spredning*. 669. Norsk institutt for naturforskning. Available at <https://brage.nina.no/nina-xmlui/handle/11250/2447615>
- Hesthagen, T., & Sandlund, O. T. (2015). *Spredning av ferskvannsfisk i Norge*. 1205. Norsk institutt for naturforskning. Retrieved from <http://hdl.handle.net/11250/2383974>
- Huitfeldt-Kaas, H. (1918). *Ferskvandfiskenes utbredelse og indvandring i Norge, med et tillæg om krebsen*. Centraltrykkeriet.
- Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1), 10–18. <https://doi.org/10.1111/j.1365-2664.2008.01600.x>
- Hulme, P. E. (2015). Invasion pathways at a crossroad: Policy and research challenges for managing alien species introductions. *Journal of Applied Ecology*, 52(6), 1418–1424. <https://doi.org/10.1111/1365-2664.12470>
- Hunt, L. M., Camp, E. D., van Poorten, B., & Arlinghaus, R. (2019). Catch and non-catch-related determinants of where anglers fish: A review of three decades of site choice research in recreational fisheries. *Reviews in Fisheries Science & Aquaculture*, 27(3), 261–286. <https://doi.org/10.1080/23308249.2019.1583166>
- Hutchison, M. J., & Armstrong, P. H. (1993). The invasion of a South-Western Australian river system by *Perca fluviatilis*: History and probable causes. *Global Ecology and Biogeography Letters*, 3(3), 77–89. <https://doi.org/10.2307/2997549>
- Keller, R. P., Frang, K., & Lodge, D. M. (2008). Preventing the spread of invasive species: Economic benefits of intervention guided by ecological predictions. *Conservation Biology*, 22(1), 80–88. <https://doi.org/10.1111/j.1523-1739.2007.00811.x>
- Leathwick, J. R., Collier, K. J., Hicks, B. J., Ling, N., Stichbury, G., & Winton, M. (2016). Predictions of establishment risk highlight biosecurity priorities for invasive fish in New Zealand lakes. *Freshwater Biology*, 61(9), 1522–1535. <https://doi.org/10.1111/fwb.12792>
- Leprieux, F., Beauchard, O., Blanchet, S., Oberdorff, T., & Brosse, S. (2008). Fish invasions in the world's river systems: When natural processes are blurred by human activities. *PLOS Biology*, 6(2), e322.
- Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28(3), 385–393. <https://doi.org/10.1111/j.0906-7590.2005.03957.x>
- Mack, R. N., Simberloff, D., Mark Lonsdale, W., Evans, H., Clout, M., & Bazzaz, F. A. (2000). Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications*, 10(3), 689–710. [https://doi.org/10.1890/1051-0761\(2000\)010\[0689:BICEG C\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2000)010[0689:BICEG C]2.0.CO;2)
- Márcia Barbosa, A., Real, R., Muñoz, A.-R., & Brown, J. A. (2013). New measures for assessing model equilibrium and prediction mismatch in species distribution models. *Diversity and Distributions*, 19(10), 1333–1338. <https://doi.org/10.1111/ddi.12100>
- Metz, M., Rocchini, D., & Neteler, M. (2014). Surface temperatures at the continental scale: Tracking changes with remote sensing at unprecedented detail. *Remote Sensing*, 6(5), 3822–3840. <https://doi.org/10.3390/rs6053822>
- Miró, A., & Ventura, M. (2015). Evidence of exotic trout mediated minnow invasion in Pyrenean high mountain lakes. *Biological Invasions*, 17, 791–803. <https://doi.org/10.1007/s10530-014-0769-z>

- Moudrý, V., & Devillers, R. (2020). Quality and usability challenges of global marine biodiversity databases: An example for marine mammal data. *Ecological Informatics*, 56, 101051.
- Ohlund, G., Hedström, P., Norman, S., Hein, C. L., & Englund, G. (2014). Temperature dependence of predation depends on the relative performance of predators and prey. *Proceedings of the Royal Society B: Biological Sciences*, 282(1799), 20142254. <https://doi.org/10.1098/rspb.2014.2254>
- Perrin, S. W. (2021). samaperrin/freshwaterForecasting: ThymallusLeMeek. *Zenodo*, <https://doi.org/10.5281/zenodo.4556024>
- Perrin, S. W., Englund, G., Blumentrath, S., Brian O'Hara, R., Amundsen, P.-A., & Gravbrøt Finstad, A. (2020). Integrating dispersal along freshwater ecosystems into species distribution models. *Diversity and Distributions*, 26(11), 1598–1611. <https://doi.org/10.1111/ddi.13112>
- Post, J. R., Persson, L., Parkinson, E. A., & van Kooten, T. (2008). Angler numerical response across landscapes and the collapse of freshwater fisheries. *Ecological Applications*, 18(4), 1038–1049. <https://doi.org/10.1890/07-0465.1>
- R Core Team. (2015). *R: A language and environment for statistical computing*. R Core Team.
- Rahel, F. J., & Olden, J. J. (2008). Assessing the effects of climate change on aquatic invasive species. *Conservation Biology*, 22(3), 521–533. <https://doi.org/10.1111/j.1523-1739.2008.00950.x>
- Rodríguez-Rey, M., Consuegra, S., Börger, L., & García de Leaniz, C. (2019). Improving species distribution modelling of freshwater invasive species for management applications. *PLoS ONE*, 14(6), e0217896.
- Rolls, R. J., Hayden, B., & Kahilainen, K. K. (2017). Conceptualising the interactive effects of climate change and biological invasions on subarctic freshwater fish. *Ecology and Evolution*, 7(12), 4109–4128. <https://doi.org/10.1002/ece3.2982>
- RStudio Team. (2020). *RStudio: Integrated development environment for R*. RStudio, Inc. Retrieved from <http://www.rstudio.com/>
- Sandlund, O. T., & Hesthagen, T. (2011). Fish diversity in Norwegian lakes: Conserving species poor systems. In M. Jankun, G. Furghala-Selezniov, M. Wozniak, & A. M. Wisniewska (Eds.), *Water biodiversity assessment and protection* (pp. 7–20). University of Warmia and Mazury.
- Sandlund, O. T., Hesthagen, T., & Brabrand, Å. (2013). Coregonid introductions in Norway: Well-intended and successful, but destructive. *Advances in Limnology*, 64, 341–358. <https://doi.org/10.1127/1612-166X/2013/0064-0013>
- Sandvik, H., Hilmo, O., & Henriksen, S. (2020). Alien species in Norway: Results from quantitative ecological impact assessments. *Ecological Solutions*, 1(1), e12006. <https://doi.org/10.1002/2688-8319.12006>
- Seebens, H., Bacher, S., Blackburn, T. M., Capinha, C., Dawson, W., Dullinger, S., Genovesi, P., Hulme, P. E., Kleunen, M., Kühn, I., Jeschke, J. M., Lenzner, B., Liebhold, A. M., Pattison, Z., Pergl, J., Pyšek, P., Winter, M., & Essl, F. (2020). Projecting the continental accumulation of alien species through to 2050. *Global Change Biology*, 27(5), 970–982. <https://doi.org/10.1111/gcb.15333>
- Sharma, S., Jackson, D. A., Minns, C. K., & Shuter, B. J. (2007). Will northern fish populations be in hot water because of climate change? *Global Change Biology*, 13(10), 2052–2064. <https://doi.org/10.1111/j.1365-2486.2007.01426.x>
- Simmonds, E. G., Jarvis, S. G., Henrys, P. A., Isaac, N. J. B., & O'Hara, R. B. (2020). Is more data always better? A simulation study of benefits and limitations of integrated distribution models. *Ecography*, 43(10), 1413–1422. <https://doi.org/10.1111/ecog.05146>
- Strayer, D. L. (2010). Alien species in fresh waters: Ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biology*, 55(Suppl. 1), 152–174. <https://doi.org/10.1111/j.1365-2427.2009.02380.x>
- Tammi, J., Appelberg, M., Beier, U., Hesthagen, T., Lappalainen, A., & Rask, M. (2003). Fish status survey of Nordic lakes: Effects of acidification, eutrophication and stocking activity on present fish species composition. *Ambio*, 32(2), 98–105. <https://doi.org/10.1579/0044-7447-32.2.98>
- Troia, M. J., & McManamay, R. A. (2016). Filling in the GAPS: Evaluating completeness and coverage of open-access biodiversity databases in the United States. *Ecology and Evolution*, 6(14), 4654–4669.
- Uden, D. R., Allen, C. R., Angeler, D. G., Corral, L., & Fricke, K. A. (2015). Adaptive invasive species distribution models: A framework for modeling incipient invasions. *Biological Invasions*, 17(10), 2831–2850. <https://doi.org/10.1007/s10530-015-0914-3>
- Venter, O., Sanderson, E. W., Magrath, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., Watson, J. E. M., & Watson, J. E. M. (2016). Global terrestrial human footprint maps for 1993 and 2009. *Scientific Data*, 3, 160067.
- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, 416, 389–395. <https://doi.org/10.1038/416389a>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Perrin, S. W., Bærum, K. M., Helland, I. P., & Finstad, A. G. (2021). Forecasting the future establishment of invasive alien freshwater fish species. *Journal of Applied Ecology*, 58, 2404–2414. <https://doi.org/10.1111/1365-2664.13993>