Integrating dispersal along freshwater ecosystems into species distribution models

Sam Wenaas Perrin¹ | Göran Englund² | Stefan Blumentrath³ | Robert Brian O’Hara⁴ | Per-Arne Amundsen⁵ | Anders Gravbrøt Finstad¹

¹Department of Natural History, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway
²Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden
³Norwegian Institute for Nature Research, Trondheim, Norway
⁴Department of Mathematical Sciences, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, Trondheim, Norway
⁵Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics, UiT The Arctic University of Norway, Tromsø, Norway

Correspondence
Sam Wenaas Perrin, Department of Natural History, Centre for Biodiversity Dynamics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway. Email: sam.perrin@ntnu.no

Funding information
Norges Forskningsråd, Grant/Award Number: 243910 and 266574

Abstract

Aim: Our ability to model species distributions and abundances is a valuable ecological tool in predicting future distributions of species. Effectively incorporating connectivity into these predictions is crucial; however, many connectivity measures utilize metrics which may not have a direct relation to the dispersal capacity of the species they are attempting to model. The identification of more relevant metrics is therefore a vital step forward in species distribution modelling.

Location: 85 freshwater lakes across a latitudinal gradient in Sweden, and an additional 282 freshwater lakes in one drainage basin in northern Norway.

Methods: To investigate the effect of different connectivity measures, we first record recolonization of fish into lakes previously treated with the piscicide rotenone. Two invasive fish species, the northern pike (Esox lucius) and the European perch (Perca fluviatilis), were used as focal study species. We model the distributions of these species in a drainage basin with snapshot data of present-day distributions to see how well the effects of the different connectivity measures correspond to the effects seen in our recolonization study. Connectivity is quantified using slope and distance along streams connecting lacustrine populations.

Results: The effects of connectivity variables were similar in both the recolonization study and the species distribution modelling. Incorporation of connectivity improved species distribution models significantly. There was little evidence for the inclusion of distance between populations, while there was strong evidence for the inclusion of different slope parameters for both species.

Main conclusions: Our study demonstrates the need to ensure the relevance of connectivity measures when accounting for dispersal limitation in distribution models. The correspondence of estimated connectivity measures from recolonization studies to those estimated from species distribution models demonstrates a link between species dispersal capacity and the connectivity measures employed, and is likely to improve our ability to predict species future distributions.
1 | INTRODUCTION

In the wake of global trends in species distribution range shifts and population decline, providing reliable estimates of shifts in species distribution and abundance has become one of the foremost goals of ecology (Briscoe et al., 2019; Rahel & Olden, 2008; Spooner, Pearson, & Freeman, 2018). Species distribution models help provide key management solutions in cases where such changes threaten ecosystem stability (Araújo et al., 2019). These models have improved significantly over time and are now able to incorporate multiple species simultaneously, while accounting for phylogeny and trait variation (Ovaskainen et al., 2017).

Accounting for dispersal is integral when modelling species distributions. Even if a particular region's abiotic characteristics may make it suitable for a species to colonize, physical barriers may prevent this species from entering the region. Dispersal is often accounted for in models through the use of population connectivity. This can significantly boost the predictive power of the models (Attum, Lee, Roe, & Kingsbury, 2008; Bonte, Baert, Lens, & Maelfait, 2004; Olden, Jackson, & Peres-Neto, 2001). However most distribution models use measures of connectivity based on Euclidean distance, which may in some cases be a poor representation of species dispersal limitations (Calabrese & Fagan, 2004). For example, distance over land between freshwater habitats is unlikely to relate to any tangible effect on the dispersal ability of fish. In this context, length and passability of the pathway between habitat patches may only be vaguely related to the Euclidean distance between lakes (Fagan, 2002).

In order to incorporate distribution limitations into a species distribution model, an understanding of the relationship between species dispersal capacity and habitat connectivity is necessary (Caplat et al., 2016; Ellis, Václavík, & Meentemeyer, 2010; Vasudev, Fletcher Jr., Goswami, & Krishnadas, 2015). Choice of parameters for connectivity in species distribution models can alter the outcomes of forecasting considerably (Fletcher, Reichert, & Holmes, 2018; Hermoso, Linke, Prenda, & Possingham, 2011; Václavík, Kupfer, & Meentemeyer, 2012). An often occurring impediment to the incorporation of distribution limitations into species distribution modelling is the lack of biological understanding of dispersal which underpins connectivity measures. As such, a vital step forward would be to demonstrate direct links between connectivity measures and species dispersal capacity when including such measures in species distribution models.

Here, we investigate the effect of introducing connectivity measures to species distribution modelling and directly test for how these measures relate to species dispersal capacity. We take a two-stage approach. Our first stage involves a controlled, large-scale experimental setting. We take data from a series of lakes over a latitudinal gradient in Sweden, which have previously been treated with the piscicide rotenone. We investigate whether or not fish species have been able to recolonize these lakes. We compare recolonization to several connectivity measures derived from terrain maps, including the slope and length of streams between populations, as slope has previously been shown to be a limiting factor for fish dispersal (Forget, Baglinière, Marchand, Richard, & Neuvoux, 2018; Hein, Öhlund, & Englund, 2011; Spens, Englund, & Lundqvist, 2007). Our second stage involves the use of a more pristine study system, a series of lakes in northern Norway which have not been treated with rotenone. We use joint species distribution models (JSDMs) to assess whether or not the same measures affected fish species distribution in a similar fashion as to our first system, despite a plausible history of human introductions. By investigating the effects of several connectivity measures on two species over the two study systems, we will test not only the effect of using connectivity measures in modelling distributions, but also how the effects of these same measures correspond to their direct impact on species dispersal ability. This will indicate how effectively and accurately connectivity measures can be incorporated into and interpreted in larger-scale distribution modelling.

Our focal species are the northern pike (Esox lucius, hereafter referred to simply as pike) and the European perch (Perca fluviatilis, hereafter referred to simply as perch). Pike and perch are native to both study regions. However, with increased temperatures in these freshwater networks and stocking by fishermen, there is potential for both species to move into ecosystems previously inaccessible and untenable for them. Both are capable of invading ecosystems with negative impacts on native top predators and cascading effects on the entire local ecosystem (Byström et al., 2007; Hein, Öhlund, & Englund, 2014; Hutchison & Armstrong, 1993). Freshwater ecosystems provide an ideal example for comparing the effects of connectivity parameters, as the dendritic nature of river networks means that well-defined pathways between populations exist (Fagan, 2002). We expect our selected connectivity measures to show similar effects across study systems, with different connectivity measures potentially varying in their effects across species.

2 | METHODS

2.1 | Recolonization analysis

2.1.1 | Study system

The study system consisted of 85 lakes (hereafter termed treated lakes) located between 56.9° and 67.9° N and 12.3 to 22.6° E in 30 drainage basins throughout Sweden (Figure 1a). Both focal species are common throughout all drainage basins. Between 1955 and 1980 the lakes were treated with the piscicide rotenone to
remove invasive species. Rotenone-treated lakes were generally small and high up in drainage basins, to avoid the possibility of (a) rotenone treatment failing and (b) recolonization of focal species from upstream populations. Data on rotenone treatments were extracted from the Swedish rotenone archive, which is kept at the Swedish Agency for Marine and Water Management, Gothenburg. The archive contains information about the date of treatment, the original species composition of the lake, and whether or not the treatment was successful.

Data on the subsequent recolonization success of pike and perch were obtained from the databases PIKE (Englund & Finstad, 2019) and NORS (Shah & Kinnerbäck, 2019), which are available online. Presences or absences were recorded until as late as 2015, with no presence recorded after 2015 being taken as an absence. The databases include information from gill net data, governmental records, mailed surveys, as well as interviews with local fishermen, fishery rights owners and other private citizens. Absence records only included lakes where either species were confirmed by locals to be absent, as gillnet data often misses pike. All lakes were occupied by our focal species directly before treatment. The source of colonization (hereon called the source lake) was determined by taking the nearest downstream lake that surveys showed had the focal species present when the rotenone treatment occurred. In some cases this involved taking lakes that were upstream of a confluence situated downstream from the treated lake. Some lakes also had different colonization sources for pike and perch, and one lake was discounted for perch as its source lake was uncertain. Presence and nature of dams were ascertained from the Swedish Water archive, however the use of dams in our models was ultimately discarded, as of our 85 lakes, only seven had dams downstream, and all had fish ladders built into them. While we acknowledge that fish ladders may have different effects on a species ability to disperse upstream than a pathway unhindered by a dam, we here make the assumption that over a long period the fish ladders would have afforded both species upstream access (Noonan, Grant, & Jackson, 2012).

It is important to note that while recolonization of an upstream lake could theoretically have happened via reintroduction by local fishers or other similar means, there are concerted efforts throughout the affected regions to reduce anthropogenic reintroductions. Introductions by means other than dispersal upstream are likely to be rare, and reintroductions that were thought to be results of human-assisted dispersal were excluded from the initial datasets. Thus, going forward we will consider successful recolonization to be equivalent to successful upstream dispersal.
2.1.2 | Estimation of connectivity parameters

Slope and distance between lakes were chosen as our connectivity parameters, as slope has previously been shown to be a limiting factor for fish dispersal (Forget et al., 2018; Hein et al., 2011; Spens et al., 2007), and distance between populations is often used as a measure of connectivity in species distribution models (Ovaskainen et al., 2017). A stream network dataset is required in order to identify and characterize connections between lakes in terms of slope and distance. The stream network data in Sweden have only been available at a very coarse scale and for main watercourses. Therefore, a more detailed river network was derived from a transnational, composite 10m digital elevation model (DEM). The DEM in Sweden was aggregated from very-high-resolution laser scanning data. The stream extraction was conducted with the r.stream.extract module (Jasiewicz & Metz, 2011) in GRASS GIS 7.4 (GRASS Development Team, 2018). Although the stream network derived from a terrain model does not necessarily match with the stream locations, it achieves the closest possible correspondence between terrain and stream network and avoids spikes in slope measurements that can occur where, for example, manually digitized stream network data does not match with the DEM.

In the next step, the generated stream network was rasterized, and local slope was calculated for each 10 x 10m pixel as elevation difference of subsequent pixels along the stream (Figure 2), using the “gradient” measure of the GRASS GIS 7.4 module r.stream.slope (Jasiewicz & Metz, 2011). Local slope at pixel level was then further aggregated to stream section level. Stream sections are stretches connecting two of the following elements in the stream network: source, junction, lake (boundary) or outlet. For each section, statistics were computed from the stream slope raster dataset, using the v.rast.stats module in GRASS GIS 7.4, after stream sections within lakes were removed from the network dataset. This process resulted in a set of slope parameters for each section, including minimum, 5% percentile, 1st quartile, mean, median, 3rd quartile, 95% percentile, maximum as well as standard deviation and variance. Finally, all stream sections that connected source and treated lakes were identified using the python-igraph network analysis package (Csardi & Nepusz, 2006), and slope characteristics of all sections along the pathway connecting these lakes were further aggregated using either the mean weighted by length of the sections or the maximum of the section.

Pathways between source lake and focal lake often had lakes or confluences interspersed among them. In these cases, the pathways between each lake and confluence were considered sections of a larger stream (Figure 3). We aggregated the statistics calculated for each section that made up a pathway between two lakes, in order to obtain different stream level summaries of the slope, as different aspects of a stream’s hydrology may affect species in different manners. For example, we know that pike have demonstrated better acceleration ability than most salmonids, which may allow them to overcome short steep stretches, but have demonstrated poorer sustained swimming ability than the same salmonids (Blake, 2004). This may suggest that they would struggle to overcome longer stretches of steeper inclines. As such, the different parameters derived from extended univariate statistics included (a) mean slope of aggregated stream sections connecting the treated and source lake (henceforth referred to as slopeavg) and (b) third quartile of the slope of all stream sections connecting the treated lake and source lake (slope3Q). If these two parameters had a negative effect on recolonization, it could suggest that a species struggled over extended uphill stretches. Also calculated were two steeper gradient measurements, (c) averaged maximum slope of every stream section connecting the treated lake and source lake (slope_max) and (d) absolute maximum slope at any point between the treated lake and source lake (slopeabs_max). If these two parameters had a negative effect on recolonization, it could suggest that while gradual extended slopes did not prove a barrier for fish, shorter steeper inclines could have prevented successful colonization. Total stream distance between the treated lake and source lake (distance_lake) was also calculated. Visual representation of the different measures can be found in Figure 3.

![Figure 2](image-url) - (a) Schematic raster representation of a stream (blue cells) imposed on a digital elevation model (DEM) and (b) resulting slope measures in degree * 100 computed as elevation difference between neighbouring pixels along the stream. Slope of grid cell $x_{i}y_{i}$ and next cell in the stream $x_{j}y_{j}$ is computed using: round(degrees(atan((12 - 9)/sqrt(102 + 103)))*100)
We used a logistic regression with the model

\[ \text{Colonization} \sim \text{Slope} + \text{Distance} \]

and, to ensure consistency with our joint species distribution analysis, we fit the model with a Bayesian approach, utilizing the R2jags package (Su & Yajima, 2015). We used SSVS priors (George & McCulloch, 1993), meaning that if the probability of inclusion was low, the parameter’s effect on species presence or absence will be subsequently reduced.

Four different models were used. Each used the logistic regression highlighted above, with one of our four slope parameters being used as each model’s single independent slope covariate. Distance\text{lake} was then used as an additional independent covariate on each of these four models. All slope parameters were scaled to a mean of zero and a standard deviation of one. Distance\text{lake} was log-transformed to account for its distribution being heavily right-skewed. Significance of slope and distance parameters was determined using Bayes factors, with Kass and Raftery’s (1995) guidelines to interpret the strength of evidence for an effect. As Bayes factors increase, so does the evidence for inclusion of the corresponding parameter in a model (i.e. a Bayes factor > 3 would provide positive evidence, greater than 10 provide strong evidence and >150 would be very strong evidence for the inclusion of a parameter into the model in question). Bayes factors were calculated using

\[ \text{BF} = p / (1 - p) \]

where \( p \) is the posterior probability of the parameter’s inclusion in the model. Prior probability of inclusion was set to 0.5, so the prior odds are 1 and thus omitted from the calculation of the Bayes Factor. The interaction term between each slope variable and distance was also tested in a further four models; however, there was no evidence of inclusion for the interaction term in any of the models, and it was subsequently discarded.

Monte Carlo Markov Chain (MCMC) sampling was done with three chains, 10,000 iterations, a burn-in of 1,000 and no thinning. Rhat values were used to estimate convergence, with 1.1 considered as an acceptable upper limit (Gelman, Carlin, Stern, & Rubin, 2004), and all models converged adequately, with \( R \) values all below 1.01 (see Appendix S1 in Supporting information).

All statistical modelling for both study systems used R version 3.4.2 (R Core Team, 2017) and RStudio version 1.1.383 (RStudio Team, 2016).

### 2.2 Joint species distribution analysis

#### 2.2.1 Study system

Study area 2 (hereafter termed the Kautokeino drainage basin) is a network of 282 focal lakes, nested within a larger network of lakes in a 5,800-km\(^2\) large drainage basin in the Kautokeino municipality, Troms and Finnmark county, northern Norway (Figure 1b). The lakes are located from 300 to 700 m above sea level over a geographical gradient ranging in latitude from 68.6 to 69.5\(^\circ\) N and in longitude from 22.2 to 24.2\(^\circ\) E. Their surface areas range from 10 to 2,100 ha. The lake network contains metapopulations of the fish displayed in Table 1, including pike and perch. The drainage basin was chosen due to the extensive presence/absence data available for all fish species that inhabit it. The use of data on other fish species allowed us to construct a joint species distribution model (JSDM), which allows us
to account for the effects of environmental variables as well as other species when modelling our focal species’ distributions in a drainage basin that has been disturbed relatively little in comparison with those found in other regions.

Data on whether or not the seven fish species listed in Table 1 were present in individual lakes of the drainage basin were obtained from interviews with local Sami fishermen in 1981 and 1982. The Kautokeino municipality is a core area for the indigenous Sami people, who have strong traditions related to subsistence on local natural resources, including freshwater fish populations. Different Sami family groups have been using these areas for decades, resulting in strong traditional knowledge about local extant populations. Two employees of the Kautokeino municipality carried out the interviews. They interviewed local fishermen from 16 settlements in the municipality, with most interviews taking place in person, and a smaller number being carried out by phone. Along with additional gillnet surveys, they ascertained information for a total of 1,321 lakes, ponds and rivers. From this data, only lakes with a total area of more than 10 hectares that drained into the Kautokeino-Alta watercourse were used, which reduced the number of lakes to 282.

| Species common name | Species Latin name | Species family | Lakes with presence |
|---------------------|--------------------|----------------|---------------------|
| Arctic charr        | Salvelinus alpinus | Salmonidae     | 90                  |
| Brown trout         | Salmo trutta       | Salmonidae     | 11                  |
| Northern pike       | Esox lucius        | Esocidae       | 188                 |
| European perch      | Perca fluviatilis  | Percidae       | 161                 |
| Burbot              | Lota lota          | Lotidae        | 132                 |
| Grayling            | Thymallus thymallus| Salmonidae     | 16                  |
| Whitefish           | Coregonus lavaretus| Salmonidae     | 126                 |

Note: Naming authority for all species is Linnaeus, 1758, with the exception of Coregonus lavaretus, the authority for which is Valenciennes, 1848.

| Variable                  | Transformation | Unit       | Notes                                                                 |
|---------------------------|----------------|------------|----------------------------------------------------------------------|
| Elevation                 | None           | m          |                                                                    |
| Temperature               | None           | Degrees Celsius | Mean temperature during the warmest quarter                      |
| Lake area                 | Logarithmic    | m²         |                                                                    |
| Shoreline complexity      | Logarithmic    | unit less  | (lake perimeter/1,000)/(2*sqrt(lake area))                         |
| Closed vegetation         | None           | Percentage | Aggregates broad leaved forest, mixed forest and transitional woodland shrub |
| Population density        | Logarithmic    | People     | Within 5 km radius of lake                                          |

Note: All variables were scaled.
environmental variable for all focal lakes, making the use of this variable uninformative.

2.2.3 | Statistical modelling

The gradient parameters between our source lake and our 282 focal lakes were determined in the manner described for the recolonization analysis, and combined with the environmental variables listed in Table 2 to model all seven species’ distributions in our JSDM, specifically a Bayesian Ordinal Regression and Analysis model from the “boral” package (Hui, 2016). The BORAL model fitted correlated response generalized linear models with two latent variables to account for any residual correlation between our explanatory variables. This was modelled against presence/absence data for all seven species. By using a JSDM, we were able to simultaneously predict species distributions for both of our focal species while accounting for both abiotic parameters, and biotic factors produced by species co-occurrence. We used SSVS priors as described in our recolonization analysis for all environmental variables listed in Table 2 except for lake area and temperature. As the posterior probability of inclusion of lake area

**Figure 4** Presence/absences of pike (Esox lucius) compared to gradient measurements for the streams connecting lakes and their nearest source populations in lakes previously treated by the piscicide rotenone throughout Sweden (white) and in a selection of lakes throughout the Kautokeino drainage basin in northern Norway (grey). Parameters include (a) mean gradient of aggregated streams, (b) gradient third quartile of aggregated streams, (c) averaged maximum gradient of all streams, (d) maximum gradient at any point along all streams, (e) total distance between two points. Maximum gradient at any point along all streams was not used in the Kautokeino drainage basin and is therefore not included. All parameters have been scaled to a mean of zero and a standard distribution of one to account for the large variation in their distributions.
and temperature was one for all models, they were assigned normal priors with mean zero and variance 10. There was no notable correlation between our gradient and distance parameters and the additional environmental variables.

MCMC sampling was used to estimate posterior distribution, with one chain, 2,000 iterations, a burn-in of 1,000 and no thinning. Four separate models were analysed, three which included distance, one of the three slope parameters and the environmental variables included in Table 2, and one which only used the environmental variables. Significance of parameters was determined as for our re-colonization analysis. Deviance, $D$, for each model was calculated separately for both species using the equation

$$D = -2 \sum (y_i \log (\pi_i) + (1 - y_i) \log (1 - \pi_i))$$

where $y_i$ is the presence or absence of a species at site $i$, and $\pi_i$ is the probability of presence of a species at site $i$ as dictated by BORAL’s inbuilt prediction function. As such, deviance is a measure of how well our model is able to explain the variance within our data, with lower values indicating better fit. To ensure that use of a JSDM was warranted, we created a single species distribution model for comparison and tested its deviance as well. While our deviance value for pike improved when using a JSDM, it worsened for perch. This is a result of the use of SSVS priors in our JSDM, which increased deviance slightly by reducing the contribution of parameters that had a moderate but not significant impact on model fit. We also calculated deviance for a null model, in which $p_i$ was simply the number of lakes with species presence divided by the total number of lakes.

### 3 | RESULTS

#### 3.1 | Recolonization analysis

Of the treated lakes in the study of re-colonization, pike were able to re-colonize 49 of the 85 lakes they were initially found in, while perch were able to re-colonize 42 of 84.

Preliminary observation of raw data suggested that increasing slope between treated and source lake appeared to limit pike dispersal, although we were not able to estimate a distinct threshold through which pike could not disperse (Figure 4). There were very few re-colonizations for pike beyond certain points for both of our more gradual slope parameters (slope$^{\text{mean}}$ and slope$^{\text{Q}}$), and the average slope for re-colonized lakes was noticeably lower than for lakes which remained uncolonized. This trend was less pronounced for slope$^{\text{abs, max}}$ and slope$^{\text{abs, max}}$, although the average slope was still lower for re-colonized lakes (Figure 4a–d). Our models confirmed this with all Bayes factors, suggesting the evidence for an effect of slope on re-colonization was “positive” (for slope$^{\text{max}}$ and slope$^{\text{abs, max}}$) or “very strong” (for the other variables: Table 3). The estimates were all negative, meaning a steeper slope reduced pike re-colonization (Figure 5). All Bayes factors for the effect of distance on pike dispersal were lower than one. Subsequent calculation of the evidence for no effect (division of 1 by the Bayes factor) showed values between 10 and 15, giving strong evidence for no effect of distance on re-colonization.

Similar observations for perch indicated that distance was a more limiting factor than for pike, although again, there did not appear to be a distinct dispersal threshold for perch in any of the slope parameters or distance (Figure 6). Slope measurements for lakes did not seem to vary to a large degree between lakes that were colonized and lakes that remained uncolonized, with the possible exception of slope$^{\text{abs, max}}$. The estimates of all slope parameters slope$^{\text{max}}$ and slope$^{\text{abs, max}}$ (with the possible exception of slope$^{\text{Q}}$) were negative, as were the estimates of distance in each model, suggesting that an increase in our more gradual slope parameters and in distance reduced perch colonization (Figure 7). Our models, however, suggested that, with the exception of slope$^{\text{abs, max}}$, none of our statistics were worthy of inclusion into the model, nor were the accompanying distance statistic (all Bayes factors were below 1); for slope$^{\text{abs, max}}$, Bayes factor provided strong evidence for the inclusion of slope into the model (Figure 7, Table 3).

---

**Table 3** Bayes factors of slope and distance parameters in two different models explaining re-colonization ability of the species *Esox lucius* (pike) and *Perca fluviatilis* (perch)

| Covariates     | Bayes factor (Recolonization analysis) | Bayes factor (JSDM) |
|----------------|----------------------------------------|---------------------|
|                | Slope | Distance | Slope | Distance |
| Pike           |       |          |       |          |
| slope$^{\text{mean}}$ | >150  | 0.070    | >150  | 0.351    |
| slope$^{\text{Q}}$     | >150  | 0.082    | >150  | 0.101    |
| slope$^{\text{max}}$   | >150  | 0.081    | 2.745 | 0.792    |
| slope$^{\text{abs, max}}$ | 7.850 | 0.096 | NA     | NA       |
| Perch          |       |          |       |          |
| slope$^{\text{mean}}$ | 0.255 | 0.099    | >150  | 0.131    |
| slope$^{\text{Q}}$     | 0.274 | 0.143    | >150  | 0.033    |
| slope$^{\text{max}}$   | 0.812 | 0.139    | 2.040 | 0.825    |
| slope$^{\text{abs, max}}$ | 25.316 | 0.085 | NA     | NA       |
3.2 | Joint species distribution analysis

Pike were found in 188 of our 282 focal lakes in the Kautokeino drainage basin, while perch were found in 161.

Preliminary observation of presence against connectivity parameters suggested that slope had a negative effect on dispersal for both species (Figures 4 and 6). Distance did not appear to show any negative effects on colonization for either species.

Slope had a negative effect on pike and perch recolonization for both slope mean and slope max, with Bayes factors indicating strong evidence for inclusion of both parameters in the model, and estimates that were negative (Figures 5 and 7; Table 3). Slope max appeared to have a negative effect on both pike and perch dispersal, though not to the same extent, and Bayes factors showed no evidence for inclusion in connectivity models. Distance appeared to have a positive effect on perch dispersal when models included slope mean or slope max; however, Bayes factors showed no evidence for the inclusion of distance in connectivity models (Figures 5 and 7; Table 3).

Our models showed marked improvements in deviance from the null model. Adding environmental covariates to our null model improves model fit by 65.2 log-likelihood units (pike) and 67.8 (perch). Adding the connectivity parameters improved our models further, with improvements of between 19.2 and 31.6 (pike), and between 10.1 and 14.3 (perch, Table 4). While deviance for perch barely varied between parameters, deviance for pike showed better results when slope mean was used as opposed to slope max and slope max. Spatial visualizations of model improvement with inclusion of connectivity parameters can be found in Supplementary Information for both pike (Figure S2.1) and perch (Figure S2.2), with increases and decreases in likelihood of presence showing spatial similarities across both species.

The estimates of the slope effects from the JSDM were broadly consistent with those from the recolonization model, for pike in particular, with very strong evidence for a negative effect of slope across both models. The lack of an effect of distance was mirrored across all species as well. There was slight variation in the effect of slope on perch dispersal across models; however, both give more evidence for inclusion of slope parameters than for distance.

4 | DISCUSSION

In this study, we validate the estimated effects of connectivity parameters in a species distribution model by comparing them with the effects in a model, which observed dispersal directly through monitoring recolonization of rotenone-treated lakes. By monitoring the effects of connectivity on dispersal through recolonization...
ability, we show that stream slope has a negative effect on presence for pike, with longer stretches of steep slope providing more of a barrier than short sharp inclines. We show similar effects in our more comprehensive species distribution models. Conversely, perch dispersal appears to be more constrained by shorter, steeper slopes in our controlled study system, and by more gradual slopes in our species distribution models. Similar results for pike across both study systems, combined with consistent model improvement upon introduction of connectivity parameters to our JSDMs, confirm that connectivity measures, which can be closely linked to dispersal ability, can improve the accuracy of species distribution modelling. However, the divergence in results for perch between the two systems provides a strong argument for further mechanistic approaches to dispersal alongside broad species distribution models, in an effort to understand the biology behind the effects of connectivity.

There is a lack of evidence for the inclusion of distance in our models. Effects of distance may have been negated by historical human introductions, as human activities have been shown to alter species distributions across large regions (Garcia-Diaz et al., 2018; Peoples & Midway, 2018). However, the fact that slope metrics affect
distance must affect dispersal on short time-scales, but it seems plausible that the time-scales covered by our data (approximately 50 years for the recolonization data and probably much longer for the distribution data) are long enough to ensure dispersal to the most distant lakes.

Similar effects of slope on pike in both our recolonization analysis and JSDM provide support for the relevance of slope as a connectivity parameter in the distribution modelling of pike. This corroborates existing ecological studies, which show that slope limits pike dispersal (Hein et al., 2011; Spens et al., 2007). The negative effect of more prolonged slopes on pike recolonization is corroborated by mechanistic studies, which showed that while they do possess strong acceleration ability, this corresponds to a decrease in their prolonged swimming capabilities (Harper & Blake, 1991; Webb, 1988).

Recolonization analysis suggests that while maximum slope was an important factor, other slope parameters have a negligible effect on perch distribution. There are several potential explanations for this lack of correspondence between different metrics. Perch has a shorter body length than pike (Davies et al., 2004), which has been suggested to result in lower maximum swimming speeds and higher endurance (Beach, 1984). Perch have also been shown to be adept at both overcoming barriers and maintaining high swimming speeds for longer stints (Starrs, Starrs, Lintermas, & Fulton, 2017; Tudorache, Viaene, Vereecken, & De Boeck, 2008), so longer stretches of more gradual slopes may not present a barrier. However, our JSDM suggests that higher values of our more gradual parameters are more significant limiters of dispersal. This may again be a result of the cumulative effect of several short, steeper stretches of river pathway making it more difficult for perch to traverse certain streams. Previous work has suggested that the cumulative effect of dispersal barriers can limit perch dispersal (Shaw, Lange, Shucksmith, & Lerner, 2016). Future experiments should take into account a wider range of connectivity parameters, including variation in slope.

The correspondence of the effects of connectivity across both types of data on pike dispersal, alongside the lack thereof on perch dispersal, demonstrates the need to understand dispersal mechanisms when including dispersal limitation in species distribution models. Pike swimming abilities have been somewhat better studied mechanistically than those of perch, and thus, we are able to make reasonable assumptions about the physiological factors, which may contribute to our results here for pike. Further study of percid swimming abilities may allow for expansion of our connectivity parameters to include those which may have more effect on the dispersal of this species.

The use of survey data here could potentially reduce the certainty of our results. For instance, undetected presences since
TABLE 4 Deviance for joint species distribution model (JSDM) estimating presence or absence of two different fish species (Esox lucius, pike and Perca fluviatilis, perch) as a function of connectivity parameters and other abiotic variables in a series of lakes in northern Norway

| Covariates                        | Deviance | Pike | Perch |
|-----------------------------------|----------|------|-------|
| null model                        | 0        | 0    |       |
| single species distribution model | −61.5    | −80.7|       |
| environmental covariates          | −65.2    | −67.8|       |
| environmental covariates + slope mean + distance | −84.4 | −78.3 | |
| environmental covariates + slope square + distance | −91.8 | −82.1 | |
| environmental covariates + slope maximum + distance | −96.8 | −77.9 | |

Note: Models which incorporate connectivity involve three different aggregations of the slope parameters of the connections between a focal lake and a lake downstream with known extant population of the species. Slope mean is the average slope between two lakes, slope square is the third quartile, and slope maximum is the average of the maximum slopes of the rivers, which make up the connection. The JSDM measures species response at a community level to environmental variables, including lake area, shoreline complexity, elevation, temperature, percentage of closed vegetation in the surrounding area and local human population density. Deviance given as deviation from null model, −313.3 for pike and −287.7 for perch.

roteneone treatment in our recolonization analysis could alter our estimates significantly. However, the ground-truthing carried out in our JSDM dataset supports the use of such data, as do other studies which support the use of local knowledge in identifying ecological trends (Hesthagen & Sandlund, 2004; Turvey et al., 2013).

The nature of defined aquatic pathways between freshwater sites provides a uniquely appropriate ecosystem for studying the incorporation of connectivity into different models. Our results support those found in other studies in ecosystems with less well-defined pathways (Correa-Ayram, Mendoza, Etter, & Pérez-Salicrup, 2019; Václavík et al., 2012), and as such supports the call for more investigation into the relevance of connectivity parameters to species dispersal.

Our results suggest that the introduction of biologically relevant connectivity parameters has the ability to improve distribution modelling. This has important consequences for conservation, with end users now being able to prioritize efforts on areas that are vulnerable to invasion on account of high dispersal ability of invaders. The extra forecasting power provided by the addition of relevant connectivity parameters into these models can also help identify areas where dispersal barriers can be removed—increasing connectivity and thus the ability of local populations to persist—without facilitating the introduction of invaders. Knowing where natural barriers to immigration occur also precludes the need to introduce more artificial barriers and enables management to focus limited resources where they are most needed.

ACKNOWLEDGEMENTS
We would like to thank Francis Hui for his help in advancing our initial understanding of his BORAL model. We would also like to thank Antti Eloranta and Trygve Hesthagen for their contributions to our understanding of Arctic and sub-Arctic fish biology. Finalisation of the the Kautokeino dataset was funded through the INVAFISH project (Norwegian Research Council 243910). Sam Perrin was supported by a PhD grant from the ERA-Net BiodivERsA project ODYSSEUS (Norwegian Research Council 266574).

AUTHOR CONTRIBUTIONS
S.P. and A.F. conceived the ideas, S.B. constructed the connectivity matrix, G.E. and P.A. helped collate the datasets used, S.P. and R.B.O. analysed the data, S.P. led the writing.

DATA AVAILABILITY STATEMENT
Data used for statistical analysis are openly available at https://doi.org/10.5281/zenodo.1491966, and the original raw data for the recolonisation analysis and joint species distribution analysis, including individual methods of collection and sources are available upon reasonable request from Göran Englund of Umeå University and Per-Arne Amundsen of the Arctic University of Oslo.

ORCID
Sam Wenaas Perrin https://orcid.org/0000-0002-1266-1573
Göran Englund https://orcid.org/0000-0001-5634-8602
Robert Brian O’Hara https://orcid.org/0000-0001-9737-3724
Per-Arne Amundsen https://orcid.org/0000-0002-2203-8216
Anders Gravbret Finstad https://orcid.org/0000-0003-4529-6266

REFERENCES
Araújo, M. B., Anderson, R. P., Márca Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., ... Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. Science Advances, 5(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858
Attum, O., Lee, Y. M., Roe, J. H., ... Kingsbury, B. A. (2008). Wetland complexes and upland-wetland linkages: Landscape effects on the distribution of rare and common wetland reptiles. Journal of Zoology, 275, 245–251. https://doi.org/10.1111/j.1469-7998.2008.00435.x
Beach, M. H. (1984). Fish pass design - criteria for the design and approval of fish passes and other structures to facilitate the passage of migratory fish in rivers. Minister of Agriculture, Fisheries and Food, 44pp.
Blake, R. W. (2004). Fish functional design and swimming performance. Journal of Fish Biology, 65, 1193–1222. https://doi.org/10.1111/j.0022-1112.2004.00568.x
Bonte, D., Baert, L., Lens, L., & Maelfait, J.-P. (2004). Effects on aerial dispersal, habitat specialisation, and landscape structure on spider distribution across fragmented grey dunes. Ecosphere, 27, 343–349.
Briscoe, N. J., Elith, J., Salguero-Gómez, R., Lahoz-Monfort, J. J., Camac, J. S., Giljohann, K. M., ... Guillerma-Arroita, G. (2019). Forecasting species range dynamics with process-explicit models: Matching methods to applications. Ecology Letters, 22, 1940–1956. https://doi.org/10.1111/ele.13348
Byström, P., Karlsson, J., Nilsson, P., Van Kooten, T., Ask, J., & Olofsson, F. (2007). Substitution of top predators: Effects of pike invasion in...
European freshwater fish species. *Ecology of Freshwater Fish*, 17, 284–291. https://doi.org/10.1111/j.1600-0633.2007.00280.x

Turvey, S. T., Risley, C. L., Moore, J. E., Barrett, L. A., Yujiang, H., Xiujiang, Z., ... Ding, W. (2013). Can local ecological knowledge be used to assess status and extinction drivers in a threatened freshwater cetacean? *Biological Conservation*, 157, 352–360. https://doi.org/10.1016/j.biocon.2012.07.016

Václavík, T., Kupfer, J. A., & Meentemeyer, R. K. (2012). Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). *Journal of Biogeography*, 39, 42–55. https://doi.org/10.1111/j.1365-2699.2011.02589.x

Vasudev, D., Fletcher, R. J. Jr, Goswami, V. G., & Krishnadas, M. (2015). From dispersal constraints to landscape connectivity: Lessons from species distribution modeling. *Ecography*, 38, 967–978. https://doi.org/10.1111/ecog.01306

Webb, P. W. (1988). “Steady” swimming kinematics of tiger musky, an esociform accelerator, and rainbow trout, a generalist cruiser. *Journal of Experimental Biology*, 138, 51–69.

BIOSKETCH

**Sam Perrin** is a freshwater ecologist currently writing his PhD thesis at the Norwegian university of Science and Technology. He is interested in the effects of connectivity on the spread of alien freshwater species, particularly in light of climate change. The general focus of the research team is to help conservationists and wildlife managers make decisions about freshwater connectivity in the face of climate change, with a particular focus on the effects of invasive species on freshwater fish communities.

SUPPORTING INFORMATION

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

**How to cite this article:** Perrin SW, Englund G, Blumentrath S, Brian O’Hara R, Amundsen P-A, Gravbrøt Finstad A. Integrating dispersal along freshwater ecosystems into species distribution models. *Divers Distrib*. 2020;26:1598–1611. https://doi.org/10.1111/ddi.13112