



RESEARCH PAPER

Global Ecology
and BiogeographyA Journal of
Macroecology

WILEY

Invasion of freshwater ecosystems is promoted by network connectivity to hotspots of human activity

Daniel S. Chapman^{1,2}  | Iain D. M. Gunn¹ | Henrietta E. K. Pringle³ |
Gavin M. Siriwardena³ | Philip Taylor¹ | Stephen J. Thackeray⁴ | Nigel J. Willby² |
Laurence Carvalho¹

¹Centre for Ecology & Hydrology, Penicuik, United Kingdom

²Biological and Environmental Sciences, University of Stirling, Stirling, United Kingdom

³British Trust for Ornithology, Thetford, United Kingdom

⁴Centre for Ecology & Hydrology, Lancaster Environment Centre, Lancaster, United Kingdom

Correspondence

Daniel S. Chapman, Biological and Environmental Sciences, University of Stirling, Stirling, FK9 4LA, UK.
Email: daniel.chapman@stir.ac.uk

Funding information

Natural Environment Research Council, Grant/Award Number: NE/N006437/1

Editor: Melodie McGeoch

Abstract

Aim: Hotspots of human activity are focal points for ecosystem disturbance and non-native introduction, from which invading populations disperse and spread. As such, connectivity to locations used by humans may influence the likelihood of invasion. Moreover, connectivity in freshwater ecosystems may follow the hydrological network. Here we tested whether multiple forms of connectivity to human recreational activities promotes biological invasion of freshwater ecosystems.

Location: England, UK.

Time period: 1990–2018.

Major taxa studied: One hundred and twenty-six non-native freshwater birds, crustaceans, fish, molluscs and plants.

Methods: Machine learning was used to predict spatial gradients in human recreation and two high risk activities for invasion (fishing and water sports). Connectivity indices were developed for each activity, in which human influence decayed from activity hotspots according to Euclidean distance (spatial connectivity) or hydrological network distance (downstream, upstream and along-channel connectivity). Generalized linear mixed models identified the connectivity type most associated to invasive species richness of each group, while controlling for other anthropogenic and environmental drivers.

Results: Connectivity to humans generally had stronger positive effects on invasion than all other drivers except recording effort. Recreation had stronger influence than urban land cover, and for most groups high risk activities had stronger effects than general recreation. Downstream human connectivity was most important for invasion by most of the groups, potentially reflecting predominantly hydrological dispersal. An exception was birds, for which spatial connectivity was most important, possibly because of overland dispersal capacity.

Main conclusions: These findings support the hypothesis that freshwater invasion is partly determined by an interaction between human activity and species dispersal in the hydrological network. By comparing alternative connectivity types for different

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. Global Ecology and Biogeography published by John Wiley & Sons Ltd

human activities, our approach could enable robust inference of specific pathways and spread mechanisms associated with particular taxa. This would provide evidence to support better prioritization of surveillance and management for invasive non-native species.

KEYWORDS

anthropogenic, biological invasion, connectivity, dispersal, fishing, human influence, recreation, river catchment, species richness, water sports

1 | INTRODUCTION

Human transportation of species beyond their native ranges is an important driver of global ecological change, leading to biological invasions that impact biodiversity and ecosystem function (Vilà et al., 2011). As such, substantial effort has been dedicated to understanding the risk factors explaining invasion and using these relationships to prioritize surveillance and management (McGeoch et al., 2016). Previous studies have linked variation in invasion at biogeographic scales to interactions between environmental, biotic and anthropogenic drivers (Catford, Jansson, & Nilsson, 2009; Chapman et al., 2016; Essl et al., 2011; Gallardo, Zieritz, & Aldridge, 2015; Pyšek et al., 2010; Theoharides & Dukes, 2007). Among these, human activities clearly play a major role. Firstly, humans are responsible for the propagule pressure by which species are introduced into new non-native regions and dispersed regionally from established non-native populations (Bullock et al., 2018; Chapman, Purse, Roy, & Bullock, 2017; Lockwood, Cassey, & Blackburn, 2005). Secondly, humans cause the ecosystem disturbance and resource inputs that are thought to lessen biotic resistance of recipient communities to establishment of non-native species (Davis, Grime, & Thompson, 2001). As such, quantifying the link between human activity and invasion is an important challenge for invasion biology.

Previous studies linking invasion rates to human activities have generally used proxies of human activity in a location to explain its level of invasion. For example, human population density was correlated to non-native fish occurrence in England (Copp, Vilizzi, & Gozlan, 2010), a human influence index combining information on local land cover, population density and transport infrastructure explained non-native freshwater species occurrence in north-west Europe (Gallardo et al., 2015) and non-native plant cover in river catchments was correlated to catchment-scale human indicators such as numbers of roads and buildings (Catford, Vesk, White, & Wintle, 2011). While local proxies for human activity have clearly proven to be useful predictors of invasion, here we argue for a more nuanced assessment of the relationship between humans and invasion rates. In particular, we identify three ways to improve such efforts. Firstly, specific activities with a high risk of non-native species transport or ecosystem disturbance should be better predictors of invasion than general human activity patterns. Secondly, aggregate levels of human activity in the surrounding landscape may be a better predictor of invasion than local human activity. This is because

invading populations and the disturbance impact of human activities may disperse from hotspots of human activity. Thirdly, dispersal processes may mediate the way in which human influence on invasion rates percolates from locations of high activity. However, to our knowledge no previous studies have investigated whether consideration of the mechanistic interactions between these three effects provides a better explanation of observed invasion than simple proxies for overall human activity.

Concepts of spatial connectivity, developed in metapopulation ecology, provide a useful framework for inferring signals of spatially mediated processes in biodiversity data (Moilanen & Nieminen, 2012). Adopting a connectivity framework involves the calculation of quantitative site-based indices for the proximity of a location to multiple sources of connectivity distributed across a landscape. In its original metapopulation formulation, the sources of connectivity were populations of the focal species, and connectivity to these sources declined with increasing distance, according to the species' dispersal potential (Ovaskainen & Hanski, 2004). However, these connectivity indices can easily be generalized to accommodate alternative measures of the source of connectivity, such as the level of human activity, and alternative measures for the distance-decay in connectivity, such as those informed by hydrological or other movement networks (Altermatt, 2013; Heino et al., 2014). The strength of the correlation between variously defined connectivity indices and the biological observation of interest can then be compared in order to infer the underlying process that is most consistent with the observations. For example, use of alternative connectivity indices has previously identified barriers to dispersal and colonization within river networks (Chapman, Oxford, & Dytham, 2009) and identified international trade networks driving national-scale patterns of invasion (Chapman et al., 2017).

In this study, we tested whether connectivity to different types of human activity explains the invasion of freshwater ecosystems. Freshwater ecosystems are disproportionately impacted by biological invasion, having high propagule pressure from intentional and accidental introduction pathways, major habitat alteration and large impacts of invasion (Ricciardi & Maclsaac, 2011). Indeed, there are many examples of ecosystem transformation caused by predatory or competitively dominant non-native species (Strayer, Caraco, Cole, Findlay, & Pace, 1999). Here, we analysed the relationships between human activity and invasion rates by non-native birds, crustaceans, fish, molluscs and plants across England.

To develop national-scale spatial gradients in connectivity to humans, we first used machine learning to generalize from known locations where recreation occurs and predict spatially comprehensive gradients in the relative likelihoods of three types of relevant human activity (all recreation, fishing and water sports). Next, we developed a suite of indices for connectivity to these human activities that varied in terms of the measure of human activity used, the manner in which human influence percolated through the surrounding landscape and the rate at which human influence decayed with increasing distance. Using these human connectivity indices, we tested which ones best explained the species richness of each group of freshwater non-native species. The results were interpreted in terms of the predominant introduction pathway and dispersal mode of each group of species.

2 | MATERIALS AND METHODS

The analysis was designed to test the link between human activities and the spread of non-native species in freshwater environments in England. To achieve this we characterized national gradients of recreational activity in freshwater environments, derived multiple indices for connectivity to human activity hotspots and tested which human connectivity indices best explained the richness patterns of non-native freshwater species (see Figure 1).

2.1 | Machine learning to predict human activity

To predict spatial gradients of human activity in freshwater environments, machine learning was used to model the locations of known

recreational visits to freshwater ecosystems in England. Data on the locations of human recreational visits were obtained from the Monitor of Engagement with Natural Environment (MENE) survey (Natural England, 2017). MENE is an ongoing random stratified survey of the population of England, in which people are asked about occasions when they spent time outdoors. From the 2009–2016 MENE database, information on 11,567 visits to rivers, lakes or canals were extracted, all of which contained information on the distance travelled and most (9,952) had georeferenced destinations. MENE also includes information on visit activity. Therefore, as well as modelling all recreation, we also modelled subsets of the data representing fishing (608 visits) and water sports (228 visits), known to be two high risk activities for introducing and spreading non-native species (Anderson, White, Stebbing, Stentiford, & Dunn, 2014; Peoples & Midway, 2018).

MENE visit locations were modelled using the *h2o* R package for scalable open source machine learning (Landry, 2018) with R version 3.5.3 (R Core Team, 2019). Specifically, supervised binary classifiers were used to predict the relative probability of an English 1 km × 1 km grid cell (Ordnance Survey British National Grid) being visited in the MENE database. Since the MENE visits are an unbiased sample of all visits in England, the model-fitted values provide a relative measure of the likelihood of human activity. To perform the classification, a suite of potentially relevant features with which to model visit locations was assembled on the same spatial grid (Table 1). These features broadly represented accessibility, infrastructure and freshwater conditions. Continuous features were logged if strongly right skewed and all were centred on zero and scaled to unit variance prior to the classification.

Using *h2o*, five machine learning algorithms were trained to classify visited and unvisited grid cells in the MENE database. The

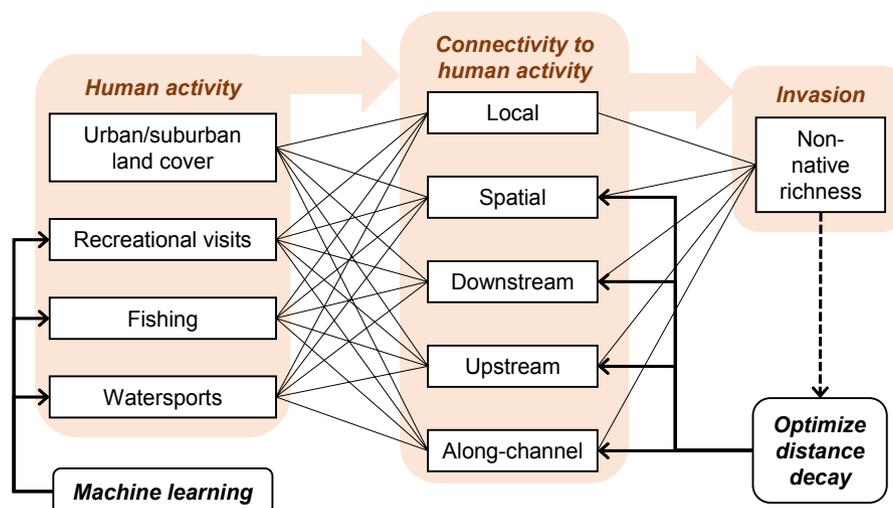


FIGURE 1 Schematic overview of the analysis to test whether human activities explain the diversity of non-native freshwater species in England. First, machine learning was used to model gradients in general recreation, fishing and water sports. Urban land cover was also included as a direct indicator of human presence. Next, indices of connectivity to human activity hotspots were derived. These indices scaled the distance decay in connectivity by the spatial (Euclidean) distance or by downstream, upstream or along-channel hydrological flow distance. Finally, generalized linear mixed models were used to optimize the distance decay in each connectivity index and compare the associations between the optimized connectivity indices and the richness of non-native freshwater taxa

TABLE 1 Variables used for machine learning of spatial gradients in human recreation, fishing and water sports in freshwater ecosystems in England. All features were assembled on a 1 km × 1 km grid

Feature	Source and details
Human population density (per km ²)	Population density disaggregated with Corine land cover 2,000 (Gallego, 2010)
Human population in potential range of the grid cell	Focal smoothing of the human population grid using log-normal distributions fitted to the distribution of travel distances for each activity type in the Monitor of Engagement with Natural Environment (MENE) data
Urban and suburban cover in the grid cell (km ²)	UK land cover map (LCM2007) (Morton et al., 2011)
Distance to nearest car park (km)	Calculated from 128,430 car parks in OpenStreetMap (https://www.openstreetmap.org)
Distance to nearest boat access point (km)	Calculated from the locations of sailing, yachting and canoeing clubs listed in five online directories (Royal Yachting Association, Yachts and Yachting magazine, British Canoeing, findasailingclub.com and Noble Marine) and from marinas and slipways in OpenStreetMap
Distance to nearest navigation infrastructure (km)	Calculated from the locations of lock gates in OpenStreetMap and locks in Scottish Environment Protection Agency and Environment Agency databases of obstacles to fish passage
Distance to nearest navigation barrier (km)	Calculated from the locations of weirs and dams in OpenStreetMap and weirs, dams and sluices in Scottish Environment Protection Agency and Environment Agency databases of obstacles to fish passage
Distance to nearest inland fishery (km)	Calculated from the locations of inland fisheries listed in three online directories (www.wheretofish.co.uk , www.fisharound.net and www.getfishing.org.uk)
Lake area (km ²)	From UK Lakes portal (https://eip.ceh.ac.uk/apps/lakes)
Lake perimeter (km)	From UK Lakes portal (https://eip.ceh.ac.uk/apps/lakes)
River length (km)	From Centre for Ecology & Hydrology (CEH) digital river network of Great Britain
Protected area	Grid cell under any level of protection in the World Database on Protected Areas (UNEP-WCMC & IUCN, 2018)
Elevation (m)	Mean elevation from the shuttle radar topography mission digital elevation model (Jarvis et al., 2008)
Easting (m)	To represent unexplained spatial gradients
Northing (m)	To represent unexplained spatial gradients

algorithms used were: generalized linear model (GLM); generalized boosting model (GBM) with tuning of the number of trees, maximum tree depth and learning rate; distributed random forest (DRF) with tuning of the number of trees, maximum tree depth and number of variables used per split; deep learning (DL) neural network with tuning of the number and size of hidden layers and the input dropout ratio; and super learning (SL) through creation of a stacked ensemble that combines the former algorithms using a GLM constrained to fit positive coefficients (Landry, 2018).

In all cases, balance sampling was used to equalize the overall influence of the small number of MENE visits with that of the large number of grid cells without recorded visits (Landry, 2018). As such, the models predicted relative rather than absolute likelihoods of human activity. To avoid overfitting, we excluded predictor features performing no better than chance by first fitting the algorithms with an additional predictor containing normally distributed random numbers. Then, the algorithms were re-fitted using only those predictor features performing better than the random numbers, according to h2o's variable importance measure.

The best performing of the five algorithms was selected by ranking predictive discrimination performance. This was assessed through

fivefold latitudinal block cross-validation using predictive area under the receiver operating characteristic curve (AUC) as the performance measure. In other words, England was divided into five equal area regions by latitude and predictions were made and evaluated for each region using models trained on the other four regions. From the best performing trained algorithms, gridded relative likelihoods of recreational visitation, fishing and water sports were then produced for the whole of Great Britain, providing maps of relative human activity gradients.

2.2 | Human connectivity indices

We derived a range of indices for the connectivity of grid cells to human activity in the surrounding landscape. These were all derived from the same general connectivity model based on the summation of proximity-weighted contributions from all possible sources of human influence (Chapman et al., 2009, 2017; Moilanen & Nieminen, 2012). In this model, the connectivity to human activity S_i of grid cell i is:

$$S_i = \sum_{j \in N_i} h_j e^{-\frac{D_{ij} \ln 0.5}{D_{50}}} / a_{ij} \tag{1}$$

N_i represents the grid cells in the connectivity neighbourhood of the focal cell i , while j indexes over all cells belonging to this neighbourhood. For this application, the focal cell i was included within N_i so that the connectivity measure captured both the local and neighbouring human influence. The term h_j is the human activity level in grid cell j , while D_{ij} represents the distance between the two grid cells. An exponential decay in connectivity with increasing distance is scaled by the parameter D_{50} , the distance at which connectivity falls to 50% of the value at zero distance (i.e., at the focal cell). Finally, a_{ij} normalizes connectivity for the dimensionality of the system in which spread occurs, for example $a_{ij} = 2\pi D_{ij}^2$ for spread in two dimensions or $a_{ij} = 1$ for spread constrained within a one-dimensional system.

As indicated in Figure 1, a suite of connectivity measures can be produced by combining different variables for human activity (h), distance (D) and connectivity neighbourhood (N) and by varying the distance weight parameter (D_{50}). In this study, four different options for h were used, namely the modelled relative likelihood of recreation, fishing and water sports, as well as the proportion of urban and suburban land (henceforth “urban cover”), derived from the 2007 UK land cover map (Morton et al., 2011). Urban cover was included to test whether modelled human activity had greater predictive value for invasion than a simpler, directly observed proxy.

Five combinations of distances (D) and connectivity neighbourhoods (N) were considered, resulting in five alternative metrics for connectivity to human activity h that represent a range of possibilities for dispersal and spread in hydrological networks (Altermatt, 2013; see Supporting Information Appendix S3 for example maps). Firstly, as an index of local connectivity to humans we simply used the values of h in the focal grid cell. This is a special case of Equation 1 in which $N_i = \{i\}$, $D_{ii} = 0$ and so $S_i = h_i$. Secondly, to develop a simple spatial connectivity index, D was set to the Euclidean distance between grid cells and N included all grid cells in Great Britain and connectivity was normalized in two dimensions. This may be an appropriate connectivity measure for species dispersing over land. Thirdly, we included one-dimensional connectivity measures informed by the topology of the hydrological network that may be more appropriate for species predominantly dispersing within or along river channels. To model downstream connectivity, N_i included all cells upstream of the focal cell in the same hydrological catchment and D was set to the downstream hydrological flow path distance (Domisch, Amatulli, & Jetz, 2015). To model upstream connectivity, N_i included all cells downstream of the focal cell in the same hydrological catchment and D was set to the upstream flow path distance. Finally, an along-channel connectivity index was produced by setting N_i to the entire hydrological catchment and using either the direct up- or downstream flow paths for D . Note that this simple measure does not capture connectivity caused by downstream movement to a confluence and then upstream movement in another tributary channel.

In all cases, hydrological catchments and flow paths were calculated from the HydroSHEDS flow grid (Lehner, Verdin, & Jarvis, 2008) using the R packages `gdistance` v. 1.2.-2 (van Etten, 2017) and `raster` v. 2.6-7 (Hijmans, 2019). To tune the distance weighting parameter (D_{50}), the spatial and hydrologically informed connectivity

indices were calculated using a range of D_{50} values from 1–25 km (specifically we used $D_{50} = 1, 2, 3, \dots, 10, 15, 20$ and 25 km). R code for calculating the connectivity indices is available in Supporting Information Appendix S1.

2.3 | Effect of human connectivity on invasion

Generalized linear mixed models (GLMMs) with Poisson errors were used to test the effect of human connectivity on the non-native richness of freshwater species in five taxonomic groups—birds, crustaceans, fish, molluscs and plants. Other taxonomic groups were considered for analysis but yielded too few non-native species or occurrence records. The goal of the analysis was to identify the human connectivity index most strongly correlated to invasion by each taxonomic group.

For each group, a list of non-native species classified as occurring in UK freshwater habitats was compiled from multiple sources (Gunn et al., 2018; Hill, 2005; Katsanevakis et al., 2015; McInerney et al., 2018; Roy et al., 2014; Zieritz, Armas, & Aldridge, 2014). This list was filtered to remove predominantly terrestrial or marine species and some potentially native species (e.g., those with unclear status or that are native to parts of the UK). For all groups, other than birds, recent georeferenced occurrence records were obtained from the UK's National Biodiversity Network (NBN) Atlas (<http://www.nbnatlas.org>), with minimum coordinate precision of 1 km and year no earlier than 1990. For birds, equivalent occurrence data were obtained from the British Trust for Ornithology's BirdTrack database (<https://app.bto.org/birdtrack2>), which we considered to be of higher quality than NBN for this group. From the occurrence records, we calculated the observed non-native richness of each taxonomic group in each 1 km \times 1 km grid cell in England. As a proxy for spatial gradients in recording effort, we also obtained gridded densities of NBN or BirdTrack records for each taxonomic group (all native and non-native species).

To test the effect of human connectivity on non-native richness, GLMMs were fitted using the `lme4` v. 1.1-21 R package (Bates, Mächler, Bolker, & Walker, 2015). In addition to the fixed effect of human connectivity, the models included a random effect of hydrological catchment identity and the following fixed effects, mainly derived as in Table 1. Record density of the focal taxonomic group was included to capture effects of recording effort. River length, lake presence and flow accumulation calculated using the HydroSHEDS flow grid (Lehner et al., 2008) were included as measures of freshwater habitat amount. Proportion cover of agricultural land and improved grassland and presence of protected areas were included as measures of human disturbance. Elevation and annual mean temperature for 1980–2010 derived from HadUK-Grid data (Hollis, McCarthy, Kendon, Legg, & Simpson, 2019) were included as primary environmental controls on freshwater habitats. We expected these fixed effects may have a strong influence on observed invasion in addition to any effects of human connectivity, so included them in the analysis as “nuisance variables”. Prior to the analysis, all fixed effect covariates were Box-Cox transformed

to normality and then centred and scaled to zero mean and unit variance to reduce heteroscedasticity and aid GLMM convergence.

The GLMMs were fitted on the subset of 1 km × 1 km grid cells satisfying the following criteria for analysis of each group. Firstly, we restricted the models to England, from where the MENE data came. Next, we selected grid cells on the hydrological network (river length or lake area greater than zero). Then, we selected cells with at least one record from the entire taxonomic group to exclude entirely unsuitable or un-surveyed grid cells. Next we excluded all hydrological catchments not known to have been invaded by any member of the focal taxonomic group to prevent problems with random effect convergence. Finally, we excluded very small catchments (<10 km²) and catchments with <5 remaining valid grid cells, in order to satisfy minimum requirements for replication of random effect levels (Bolker et al., 2009). The species remaining in the analysis and numbers of valid grid cells for each are given in Supporting Information Appendix S4.

With the remaining data, a total of 1,060 GLMMs were fitted for every combination of taxonomic group, human activity measure (proportion urban land cover or modelled recreation, fishing and water sports) and connectivity type (local, spatial, downstream, upstream and along-channel) with distance decay parameters varying between 1 and 25 km. For each connectivity specification, the optimal distance decay parameter was selected based on minimal

GLMM Akaike information criterion (AIC; Burnham, Anderson, & Huyvaert, 2011). Then, the optimized connectivity models for each taxonomic group were compared based on their AIC to identify the connectivity type best explaining invasion.

3 | RESULTS

3.1 | Machine learning to predict human activity

The best performing machine learning algorithm for predicting recreational visitation in the MENE database was GBM, with 1,000 trees, a learning rate of 0.01 and a maximum tree depth of three splits. For modelling fishing, the best algorithm was a DRF with 100 trees, five randomly selected variables per tree split and a maximum tree depth of five. DRF was also the best algorithm for modelling water sports, with 500 trees, one variable per split and a maximum tree depth of 10. Block cross-validated AUC values of the best models were 0.853 for recreational visitation, 0.837 for fishing and 0.907 for water sports, indicating a high accuracy for predicting human usage in new regions of England.

The importance of predictor features in the machine learning models varied between activity types (Figure 2). For predicting

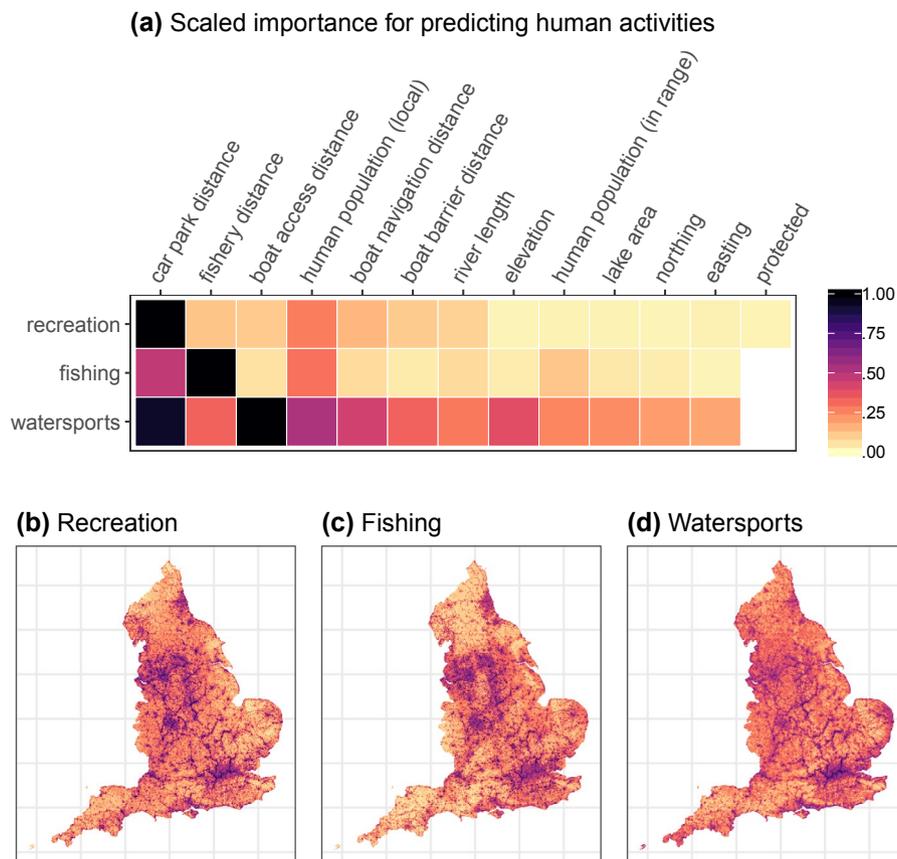


FIGURE 2 (a) Variable importance in the machine learning models for general recreation, fishing and water sports activity in freshwater environments in England. Importance values are scaled relative to the most important feature. See Table 1 for a full explanation of the features. (b–d) Model predicted gradients in relative activity produced from the machine learning models, shaded linearly on the logit scale. See Supporting Information Appendix S2 for larger maps

recreational visitation, distance from car parks and local population density were most important, and strong effects of inland fisheries, boating infrastructure and river length were also detected. For predicting fishing, distance to inland fisheries was the strongest feature, while the model for water sports was most influenced by distance to boating infrastructure, car parks, population density and elevation.

3.2 | Effect of human connectivity on invasion

For all the taxonomic groups analysed, the GLMMs detected highly significant positive associations between non-native species richness and human activity, after accounting for recording effort and other environmental and anthropogenic drivers of invasion (Table 2, see also Supporting Information Appendix S5). For all groups other than fish, non-local human connectivity measures clearly explained invasion better than local human activity, although in general the best fitting connectivity indices had rapid optimal distance decay (small value of the D_{50} parameter; Table 2 and Figure 3). The different non-native taxonomic groups varied in the measure of human connectivity that best explained their invasion (Table 2 and Figure 3). Water sports was the human activity most closely associated with invasion by birds, fishing was most strongly correlated to non-native richness of fish and molluscs, while non-native crustacean and plant richness was associated most strongly with all recreation (Table 2). In general, the best performing connectivity indices were computed using distance measures informed by downstream flow within the hydrological network, rather than spatial (Euclidean), upstream or along-channel distances (Figure 3). The one exception to this was for birds, for which spatial connectivity gave the best fitting model (Table 2).

In the best fitting GLMMs, connectivity to humans had a stronger effect on invasion than all other environmental or anthropogenic predictors considered, with the exception of recording effort and also lake presence for non-native birds (Figure 4a and Supporting Information Appendix S5). It was clearly evident that heavily invaded

grid cells tended to have higher connectivity values across all taxon groups (Figure 4b and Supporting Information Appendix S6), despite influences of additional predictors that were controlled for in the GLMMs, but not in Figure 4b. Regarding these additional predictors, recording effort always had a strong positive effect on observed non-native richness, river length and disturbed land use types (arable and improved) generally had positive effects on invasion, while elevation and protected areas generally had negative effects. Effects of lake presence and flow accumulation were less consistent across groups, while temperature surprisingly had no significant effect on any group (Figure 4a and Supporting Information Appendix S5).

4 | DISCUSSION

Our findings support the hypothesis that measures of human activity in the surrounding landscape, quantified using connectivity indices informed by hydrological network topology, provide a better explanation of the invasion of freshwater ecosystems than local human activity. This was the case for all taxonomic groups analysed other than fish, for which non-native richness was better explained by local fishing activity than any connectivity metric. This may be explained by comparatively low recorded invasion rates among this group (see Table 2 and Supporting Information Appendix S4), anglers favouring areas with suitable habitat for non-native fish or, potentially, fish containing disproportionately more casual species whose presence relies on repeated stocking by anglers.

Previous studies testing correlations between invasion of freshwater habitats and human activities have generally used proxies for local human activity (Catford et al., 2011; Copp et al., 2010; Gallardo et al., 2015; Johnson, Olden, & Vander Zanden, 2008). Therefore, our findings suggest that previous analyses could have under-estimated the role of humans in promoting invasion, relative to other environmental risk factors. However, it is worth noting that this study used a relatively high spatial resolution (1 km × 1 km) and that the optimal

TABLE 2 Summary of the analysis to identify human connectivity measures with the strongest correlations to species richness of five groups of non-native freshwater taxa. First, the number of species and grid cells in the analyses and their observed invasion rates are given. Then the effect of human connectivity in the best-fitting generalized linear mixed models (GLMMs) is reported (see Figure 3). Akaike weights (w_{AIC} ; Burnham et al., 2011) compare each connectivity specification with optimized distance decay (D_{50}) and all models with reasonable support ($w_{AIC} > .05$) are reported. The effect of the human connectivity measure is given as its z value (fixed effect coefficient divided by standard error), indicating the significance and direction of the effect. All z values are highly significant positive effects ($p < .001$)

Taxon group	Non-native species	1 km × 1 km grid cells	% invaded	Best GLMMs for non-native species richness				
				Human activity	Connectivity type	D_{50} (km)	w_{AIC}	Human effect (z)
Birds	54	55,724	15.5	Water sports	Spatial	2	0.630	10.73
				Water sports	Downstream	25	0.352	10.62
Crustaceans	17	14,233	42.3	Recreation	Downstream	3	1	13.28
Fish	13	14,832	8.9	Fishing	Local	0	0.752	8.17
				Fishing	Spatial	1	0.247	8.00
Molluscs	10	20,850	52.6	Fishing	Downstream	2	1	10.70
Plants	32	40,408	21.5	Recreation	Downstream	7	1	37.33

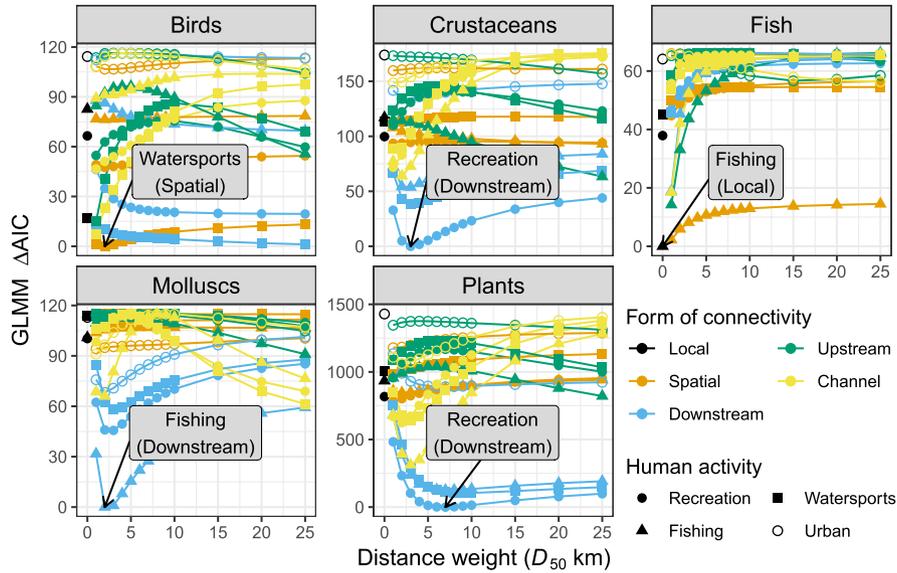


FIGURE 3 Optimization of indices for connectivity to human activity hotspots that predict non-native species richness of five freshwater taxon groups. The connectivity indices combine different measures of human activity, the form of connectivity to that activity and the distance weight parameter (D_{50}). Model fits are compared by the difference in Akaike information criteria (ΔAIC) for generalized linear mixed models (GLMMs) fitted to non-native species richness using each connectivity index. $\Delta AIC = 0$ indicates the connectivity index giving the best explanation of invasion (Burnham et al., 2011). Grey text labels point to the optimal connectivity index for each group

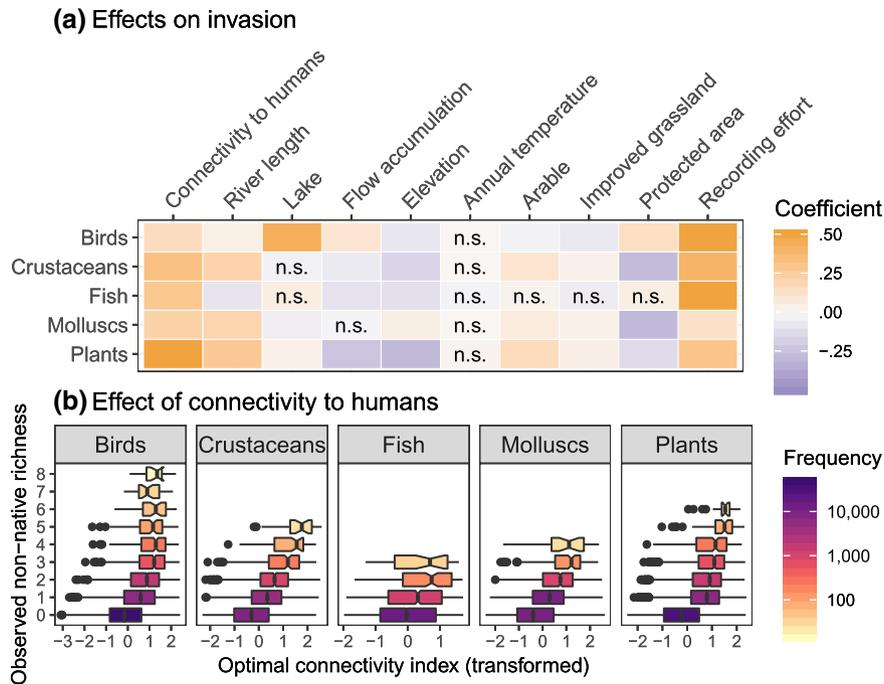


FIGURE 4 (a) Standardized effects of anthropogenic and environmental drivers of non-native freshwater species richness in 1 km × 1 km grid cells in England, from the optimal generalized linear mixed models (GLMMs) (Figure 3). Shading shows the GLMM coefficients for scaled predictors (for visualization, values > 0.5 are shaded as 0.5). Connectivity to humans and recording effort generally have the strongest effect sizes. All effects have $p < .05$ unless labelled as not significant (n.s.). (b) Boxplots showing the raw relationships between non-native richness and the connectivity measures, after transformation to normality. Boxes are shaded according to the frequency of 1 km × 1 km grid cells used in the analysis, and boxes with frequencies < 10 are not drawn. Boxes show the median (notches approximate its 95% confidence interval) and extend from the 25th and 75th percentiles. Whiskers extend no further than 1.5 interquartile ranges from the box edges in both directions, and points show outliers

connectivity indices generally selected rapid distance decay in the human connectivity. As such, direct measures of human activity may have performed better at predicting invasion patterns at the coarser spatial resolutions typically used in previous studies. With coarser resolutions, data were aggregated across a larger region and hence neighbourhood effects at higher resolutions become local.

The best fitting connectivity indices generally used distance decay informed by the hydrological network, and especially represented downstream connectivity rather than simple Euclidean distances. This finding is consistent with other studies showing that network topology has better explanatory power than Euclidean distance for native biodiversity patterns in aquatic systems (Altermatt, 2013; Chapman et al., 2009; Heino et al., 2014). One explanation for our result is that non-native species are predominantly introduced at hotspots of human activity and then disperse and spread unassisted into the surrounding landscape following the hydrological network. Supporting this interpretation, downstream connectivity most strongly promoted invasion by crustaceans, molluscs and plants, which are predominantly passively dispersed by downstream flow, albeit with some potential for upstream dispersal (Hänfling, Edwards, & Gherardi, 2011). By contrast non-native birds were best explained by spatial connectivity, consistent with their ability to disperse via overland flights. A second explanation for our results is that the connectivity measures also reflect the percolation of human disturbance that facilitates the establishment of non-native species (Davis et al., 2001). This may explain why bird invasion was most strongly associated with connectivity to locations used for water sports, as recreational boats are not considered a pathway for introducing or spreading non-native birds. Instead, large lakes and rivers used for boating, canoeing and other sports may be disproportionately invaded by non-native birds because they have features not captured in the models that are attractive to non-native water fowl (e.g., shallow lake depth, reservoir usage) or are more disturbed through hydrological modification, urbanization, noise pollution or supplementary feeding. These two explanations for our results are not mutually exclusive and it seems plausible that both have helped to drive the patterns observed.

Our analysis revealed fishing and water sports to be two human activities strongly linked to invasion, vindicating a strong focus on public warnings about invasion risks at facilities used by these activities. Both activities have previously been identified as carrying high risks for accidentally introducing and spreading non-native aquatic species because of frequent inter-catchment and international movements by anglers and boaters, and an ability of many non-native taxa to survive for periods of time in transported equipment (Anderson et al., 2014; Peoples & Midway, 2018). In addition, fishing plays a direct role in invasion through the deliberate stocking of non-native fish and live baiting (Ricciardi & MacIsaac, 2011). However, we also found that all recreation was more strongly linked to invasion than either fishing or water sports for crustaceans and plants. Although they include a minority of fishing and water sports observations, the recreation data were numerically dominated by general recreational activities that have less direct interaction with the water (e.g., walking). This suggests that these groups may be introduced and spread

through pathways linked to general accessibility, as well as the two high risk activities of fishing and water sports. For example, releases from personal indoor aquaria are a major pathway for introducing non-native freshwater species (Hänfling et al., 2011; Zieritz et al., 2017) and it seems likely that these are most likely to occur in locations accessible for general recreational usage.

A novel feature of this study was that we used machine learning to model survey data on individual human visits made for different purposes. Conceptually, our approach to predict human activity gradients has a strong affinity to presence-only species distribution modelling (Pearce & Boyce, 2006). However, because the visit data were derived from a large nationally representative survey rather than the biased field observations usually used for species modelling (Chapman, Pescott, Roy, & Tanner, 2019; Phillips et al., 2009), the data available to us for modelling human activity are of much higher quality (Guillera-Arroita et al., 2015). In addition we used spatially blocked cross-validation of the models to ensure extrapolation within England was robust (Chapman, 2010). Extrapolation far outside England was less reliable however. For example, the algorithms predicted water sports activity in north-west Scotland where many remote lakes occur but with little human activity.

Importantly, our modelling revealed that different human activities in freshwater habitats vary in distribution and underlying drivers. Unsurprisingly, general recreation was most strongly predicted by vehicle parking facilities and local human population density, fishing was most strongly related to inland fisheries and water sports were strongly associated with boat access infrastructure and vehicle parking. More interestingly, human population densities in the surrounding region, weighted by travel distances for each activity type, were generally not strong predictors of human activities. This suggests that widely used but relatively simple indices of human influence based on proximity to human populations (Uchida & Nelson, 2009; WCS & CIESIN, 2005) could be improved by modelling data on where different human activities actually occur and using predictors of higher risk human activities that are more relevant than proximity.

Comparing the correlations we found between non-native richness of multiple taxonomic groups and alternative forms of connectivity to different human activities suggests that our analysis detected signals of introduction pathways and spread mechanisms specific to the different groups. For example, fishing explained invasion of non-native fish, which are predominantly introduced by deliberate stocking for angling. By contrast, general recreation was more important for groups such as crustaceans and plants that are predominantly introduced and spread unintentionally. Furthermore, downstream connectivity was important for crustaceans, molluscs and plants, which have strongly downstream-biased dispersal, while spatial connectivity was important for birds with overland movement capability. This suggests that our approach could be applied for robust inference of the specific introduction or long-distance human dispersal pathways and spread mechanisms of individual taxa. Doing so could involve applying similar analyses to the occurrence or abundance of individual non-native species. Also, given the importance of multiple introduction pathways in freshwater invasions (Copp et al., 2010;

Zieritz et al., 2017), extending our current models to include connectivity to multiple human activities could be of interest (Chapman et al., 2017). We suggest that this could provide a better evidence base for risk mapping to support invasive non-native species surveillance and management, where high resolution predictions of invasion risk are valuable for prioritizing resources (McGeoch et al., 2016). Future efforts should therefore develop the connectivity-based approach of this study into a general framework for inferring invasion processes from spatial data and improving predictions of risk.

ACKNOWLEDGMENTS

This study was funded by the Natural Environment Research Council Highlight Topics grant NE/N006437/1: Hydroscape–connectivity x stressor interaction in freshwater habitats. We acknowledge the project consortium for useful discussions on the analyses. We thank the volunteer recorders who collected the NBN and British Trust for Ornithology (BTO) species data used in the analysis.

DATA ACCESSIBILITY

The data that support the findings of this study mostly came from openly accessible repositories including the National Biodiversity Network Atlas (www.nbnatlas.org) and the sources in Table 1. The BirdTrack bird distribution data were available on request from the British Trust for Ornithology (<https://www.bto.org/our-science/data/data-request-system>). Raster files for the human activity gradients and connectivity to them are available on Dryad at <https://doi.org/10.5061/dryad.18931zcs6>.

ORCID

Daniel S. Chapman  <https://orcid.org/0000-0003-1836-4112>

REFERENCES

- Altermatt, F. (2013). Diversity in riverine metacommunities: A network perspective. *Aquatic Ecology*, 47, 365–377. <https://doi.org/10.1007/s10452-013-9450-3>
- Anderson, L. G., White, P. C. L., Stebbing, P. D., Stentiford, G. D., & Dunn, A. M. (2014). Biosecurity and vector behaviour: Evaluating the potential threat posed by anglers and canoeists as pathways for the spread of invasive non-native species and pathogens. *PLoS ONE*, 9, e92788. <https://doi.org/10.1371/journal.pone.0092788>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution*, 24, 127–135. <https://doi.org/10.1016/j.tree.2008.10.008>
- Bullock, J. M., Bonte, D., Pufal, G., da Silva Carvalho, C., Chapman, D. S., García, C., ... Delgado, M. M. (2018). Human-mediated dispersal and the rewiring of spatial networks. *Trends in Ecology and Evolution*, 33, 958–970. <https://doi.org/10.1016/j.tree.2018.09.008>
- Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65, 23–35. <https://doi.org/10.1007/s00265-010-1029-6>
- Catford, J. A., Jansson, R., & Nilsson, C. (2009). Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. *Diversity and Distributions*, 15, 22–40. <https://doi.org/10.1111/j.1472-4642.2008.00521.x>
- Catford, J. A., Veski, P. A., White, M. D., & Wintle, B. A. (2011). Hotspots of plant invasion predicted by propagule pressure and ecosystem characteristics. *Diversity and Distributions*, 17, 1099–1110. <https://doi.org/10.1111/j.1472-4642.2011.00794.x>
- Chapman, D. S. (2010). Weak climatic associations among British plant distributions. *Global Ecology and Biogeography*, 19, 831–841. <https://doi.org/10.1111/j.1466-8238.2010.00561.x>
- Chapman, D. S., Makra, L., Albertini, R., Bonini, M., Páldy, A., Rodinkova, V., ... Bullock, J. M. (2016). Modelling the introduction and spread of non-native species: International trade and climate change drive ragweed invasion. *Global Change Biology*, 22, 3067–3079. <https://doi.org/10.1111/gcb.13220>
- Chapman, D. S., Oxford, G. S., & Dytham, C. (2009). Process from pattern in the distribution of an endangered leaf beetle. *Ecography*, 32, 259–268. <https://doi.org/10.1111/j.1600-0587.2008.05576.x>
- Chapman, D. S., 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, 1029–1040. <https://doi.org/10.1111/jbi.13555>
- Chapman, D. S., Purse, B. V., Roy, H. E., & Bullock, J. M. (2017). Global trade networks determine the distribution of invasive non-native species. *Global Ecology and Biogeography*, 26, 907–917. <https://doi.org/10.1111/geb.12599>
- Copp, G. H., Vilizzi, L., & Gozlan, R. E. (2010). The demography of introduction pathways, propagule pressure and occurrences of non-native freshwater fish in England. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 20, 595–601. <https://doi.org/10.1002/aqc.1129>
- Davis, M. A., Grime, J. P., & Thompson, K. (2001). Fluctuating resources in plant communities: A general theory of invasibility. *Journal of Ecology*, 88, 528–534. <https://doi.org/10.1046/j.1365-2745.2000.00473.x>
- Domisch, S., Amatulli, G., & Jetz, W. (2015). Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. *Scientific Data*, 2, 150073.
- Essl, F., Dullinger, S., Rabitsch, W., Hulme, P. E., Hulber, K., Jarosik, V., ... Pysek, P. (2011). Socioeconomic legacy yields an invasion debt. *Proceedings of the National Academy of Sciences USA*, 108, 203–207. <https://doi.org/10.1073/pnas.1011728108>
- Gallardo, B., Zieritz, A., & Aldridge, D. C. (2015). The importance of the human footprint in shaping the global distribution of terrestrial, freshwater and marine invaders. *PLoS ONE*, 10, e0125801. <https://doi.org/10.1371/journal.pone.0125801>
- Gallego, F. J. (2010). A population density grid of the European Union. *Population and Environment*, 31, 460–473. <https://doi.org/10.1007/s11111-010-0108-y>
- Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., ... Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24, 276–292. <https://doi.org/10.1111/geb.12268>
- Gunn, I. D. M., Carvalho, L., Davies, C. E., Edwards, F. K., Furse, M. T., Maitland, P. S., ... Winfield, I. J. (2018). UK Checklist of freshwater species. NERC Environmental Information Data Centre. <https://doi.org/10.5285/57653719-434b-4b11-9f0d-3bd76054d8bd>
- Hänfling, B., Edwards, F., & Gherardi, F. (2011). Invasive alien Crustacea: Dispersal, establishment, impact and control. *BioControl*, 56, 573–595. <https://doi.org/10.1007/s10526-011-9380-8>
- Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S., & Bini, L. M. (2014). Metacommunity organisation, spatial extent and dispersal in aquatic systems: Patterns, processes and prospects. *Freshwater Biology*, 60, 845–869. <https://doi.org/10.1111/fwb.12533>
- Hijmans, R. J. (2019). Raster: Geographic data analysis and modeling. R package version 2.3–33. Retrieved from <https://CRAN.R-project.org/package=raster>
- Hill, M. O. (2005). *Audit of non-native species in England*. Peterborough, UK: English Nature.

- Hollis, D., McCarthy, M., Kendon, M., Legg, T., & Simpson, I. (2019). HadUK-grid—A new UK dataset of gridded climate observations. *Geoscience Data Journal*, 6(2), 151–159. <https://doi.org/10.1002/gdj3.78>
- Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe: Version 4. CGIAR-CSI SRTM 90m Database. Retrieved from <http://srtm.csi.cgiar.org>
- Johnson, P. T. J., Olden, J. D., & Vander Zanden, M. J. (2008). Dam invaders: Impoundments facilitate biological invasions into freshwaters. *Frontiers in Ecology and the Environment*, 6, 357–363. <https://doi.org/10.1890/070156>
- Katsanevakis, S., Deriu, I., D'Amico, F., Nunes, A. L., Pelaez Sanchez, S., Crocetta, F., ... Curto, G. (2015). European alien species information network (EASIN): Supporting European policies and scientific research. *Management of Biological Invasions*, 6, 147–157. <https://doi.org/10.3391/mbi.2015.6.2.05>
- Landry, M. (2018). *Machine learning with R and H2O* (7th ed.). Mountain View, CA: H2O.ai Inc.
- Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions, AGU*, 89(10), 93–94.
- Lockwood, J. L., Cassey, P., & Blackburn, T. (2005). The role of propagule pressure in explaining species invasions. *Trends in Ecology and Evolution*, 20, 223–228. <https://doi.org/10.1016/j.tree.2005.02.004>
- McGeoch, M. A., Genovesi, P., Bellingham, P. J., Costello, M. J., McGrannachan, C., & Sheppard, A. (2016). Prioritizing species, pathways, and sites to achieve conservation targets for biological invasion. *Biological Invasions*, 18, 299–314. <https://doi.org/10.1007/s10530-015-1013-1>
- McInerney, C. J., Musgrove, A. J., Stoddart, A., Harrop, A. H. J., Dudley, S. P., The British Ornithologists' Union Records Committee (BOURC), ... French, P. (2018). The British list: A checklist of birds of Britain. *Ibis*, 160, 190–240.
- Moilanen, A., & Nieminen, M. (2012). Simple connectivity measures in spatial ecology. *Ecology*, 83, 1131–1145. [https://doi.org/10.1890/0012-9658\(2002\)083\[1131:SCMISE\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[1131:SCMISE]2.0.CO;2)
- Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., ... Simpson, I. (2011). Final report for LCM2007—the new UK land cover map. Countryside Survey Technical Report No 11/07. Wallingford, UK: Centre for Ecology and Hydrology (Natural Environment Research Council).
- Natural England. (2017). *Monitor of engagement with the natural environment: The national survey on people and the natural environment. Technical report to the 2009–16 surveys*. Peterborough, UK: Author.
- Ovaskainen, O., & Hanski, I. (2004). Metapopulation dynamics in highly fragmented landscapes. In I. Hanski & O. E. Gaggiotti (Eds.), *Ecology, genetics and evolution of metapopulations* (pp. 73–103). Burlington, MA: Elsevier Academic Press.
- Pearce, J. L., & Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43, 405–412. <https://doi.org/10.1111/j.1365-2664.2005.01112.x>
- Peoples, B. K., & Midway, S. R. (2018). Fishing pressure and species traits affect stream fish invasions both directly and indirectly. *Diversity and Distributions*, 24, 1158–1168. <https://doi.org/10.1111/ddi.12747>
- Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, 19, 181–197. <https://doi.org/10.1890/07-2153.1>
- Pyšek, P., Jarošík, V., Hulme, P. E., Kühn, I., Wild, J., Arianoutsou, M., ... Winter, M. (2010). Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences USA*, 107, 12157–12162. <https://doi.org/10.1073/pnas.1002314107>
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Ricciardi, A., & MacIsaac, H. J. (2011). Impacts of biological invasions on freshwater ecosystems. In D. M. Richardson (Ed.), *Fifty years of invasion ecology: The legacy of Charles Elton* (pp. 211–224). Oxford, UK: Blackwell Publishing.
- Roy, H. E., Preston, C. D., Harrower, C. A., Rorke, S. L., Noble, D., Sewell, J., ... Booy, O. (2014). GB Non-native species information portal: Documenting the arrival of non-native species in Britain. *Biological Invasions*, 16, 2495–2505. <https://doi.org/10.1007/s10530-014-0687-0>
- Strayer, D. L., Caraco, N. F., Cole, J. J., Findlay, S., & Pace, M. L. (1999). Transformation of freshwater ecosystems by bivalves: A case study of zebra mussels in the Hudson River. *BioScience*, 49, 19–27. <https://doi.org/10.2307/1313490>
- Theoharides, K. A., & Dukes, J. S. (2007). Plant invasion across space and time: Factors affecting nonindigenous species success during four stages of invasion. *New Phytologist*, 176, 256–273. <https://doi.org/10.1111/j.1469-8137.2007.02207.x>
- Uchida, H., & Nelson, A. (2009). *Agglomeration index: Towards a new measure of urban concentration*. Washington, DC: World Bank. Retrieved from <https://openknowledge.worldbank.org/handle/10986/9039>
- UNEP-WCMC & IUCN. (2018). *Protected planet: The world database on protected areas (WDPA)*. Cambridge, UK: UNEP-WCMC. Retrieved from <https://www.protectedplanet.net/>
- van Etten, J. (2017). R package gdistance: Distances and routes on geographical grids. *Journal of Statistical Software*, 76, 1–21.
- Vilà, M., Espinar, J. L., Hejda, M., Hulme, P. E., Jarošík, V., Maron, J. L., ... Pyšek, P. (2011). Ecological impacts of invasive alien plants: A meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters*, 14, 702–708. <https://doi.org/10.1111/j.1461-0248.2011.01628.x>
- Wildlife Conservation Society - WCS, and Center for International Earth Science Information Network - CIESIN - Columbia University. (2005). *Last of the Wild Project, version 2, 2005 (LWP- 2): Global human influence index (HII) dataset (geographic)*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4BP00QC>
- Zieritz, A., Armas, B., & Aldridge, D. C. (2014). Registry of non-native species in the Two Seas region countries (Great Britain, France, Belgium and the Netherlands). *Neobiota*, 23, 65–80. <https://doi.org/10.3897/neobiota.23.5665>
- Zieritz, A., Gallardo, B., Baker, S. J., Britton, J. R., van Valkenburg, J. L. C. H., Verreycken, H., & Aldridge, D. C. (2017). Changes in pathways and vectors of biological invasions in Northwest Europe. *Biological Invasions*, 19, 269–282. <https://doi.org/10.1007/s10530-016-1278-z>

BIOSKETCH

The research team is currently investigating how biodiversity of freshwater habitats is influenced by interactions between connectivity and anthropogenic stressors.

SUPPORTING INFORMATION

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

How to cite this article: Chapman DS, Gunn IDM, Pringle HEK, et al. Invasion of freshwater ecosystems is promoted by network connectivity to hotspots of human activity. *Global Ecol Biogeogr*. 2019;00:1–11. <https://doi.org/10.1111/geb.13051>