Paper:
Januchowski-Hartley, S., Jézéquel, C. & Tedesco, P. (2019). Modelling built infrastructure heights to evaluate common assumptions in aquatic conservation. *Journal of Environmental Management, 232*, 131-137.
http://dx.doi.org/10.1016/j.jenvman.2018.11.040

Title page

Modelling built infrastructure heights to evaluate common assumptions in aquatic conservation

Stephanie R Januchowski-Hartley[a,b], Céline Jézéquel[b] and Pablo A Tedesco[b]

[a] Department of Biosciences, Swansea University, Swansea, United Kingdom SA2 8PP

[b] Laboratoire Evolution et Diversité Biologique (EDB UMR 5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, 118 route de Narbonne, Bat 4R1, 31062 Toulouse Cedex 9, France

Abstract

Built infrastructure, such as dams and weirs, are some of the most impactful stressors affecting aquatic ecosystems. However, data on the distribution and characteristics of small built infrastructure that often restrict fish movement, impede flows, and retain sediments and materials, remain limited. Collection of this necessary information is challenged by the large number of built infrastructure with unknown dimensions (e.g., height), which means scientists and practitioners need to make assumptions about these characteristics in research and decision-making. Evaluating these common assumptions is essential for advancing conservation that is more effective. We use a statistical modelling approach to double the number of small (≤5 m high) built infrastructure with height values in France. Using two scenarios depicting common assumptions (all infrastructure without height data are impassable, or all are passable for all species) and one based on our modelled heights, we demonstrate how assumptions can influence our understanding of river fragmentation. Assuming all built infrastructure without height data are passable results in a 5-fold reduction

in estimated river fragmentation for fish species that cannot pass built infrastructure ≥1.0 m. The opposite is true for fish species that cannot pass ≥2.0 m, where assuming all built infrastructure without height data are impassable results in a 7-fold increase in fragmentation compared to the scenario with modelled heights to attribute built infrastructure passability. Our findings suggest that modelled height data leads to better understanding of river fragmentation, and that knowledge of different fish species' abilities to pass a variety of built infrastructure is essential to guide more effective management strategies. Our modelling approach, and results, are of particular relevance to regions where efforts to both remediate and remove built infrastructure is occurring, but where gaps in data on characteristics of built infrastructure remain, and limit effective decision making.

Highlights

- We double the number of small (≤5 m) built infrastructure with height values in France.
- Common assumptions affect our understanding of river fragmentation.
- Modelled height data leads to better understanding of river fragmentation.
- We provide essential information for protocols evaluating river ecological continuity.

1. Introduction

Scientists and practitioners require information on the characteristics of built infrastructure, such as dams and weirs, to better understand associated impacts, costs, and benefits, in relation to ecological processes, services, and human values (Poff & Hart, 2002; Januchowski-Hartley et al., 2013; Major et al., 2017). Characteristics of larger built infrastructure are increasingly well understood, because of improved identification via remotely sensed imagery (Mantel et al., 2017), and superior record keeping due to the importance of size and water holding capacity for monitoring energy production and water storage (e.g., Carvajal et al., 2017). Despite likely impacts from small built infrastructure which often restrict fish movement (i.e., being impassable), impede river flows, and retain sediments and materials, data on their distribution and characteristics remain limited (Januchowski-Hartley et al., 2013; Couto & Olden, 2018). Collection of this necessary information is challenged by the large number of built infrastructure with unknown dimensions (e.g., height), which means that assumptions are often necessary in research and decision making (e.g., assume binary passability or impassability of built infrastructure) when height data are unmeasured (Cote et al., 2009; Perkin & Gido, 2012; Radinger et al., 2017). This raises the question of how common assumptions about characteristics of built infrastructure affect estimates of habitat fragmentation, and the potential implications of this for fishes with different abilities to pass over infrastructure.

Here, we investigate existing data and data gaps for built infrastructure (Fig. 1a), and evaluate how these influence measures of river fragmentation when considering passability (the ability of a fish species to pass built infrastructure in an upstream direction) for native fishes in France. We do this by bringing together a database of built infrastructure, and associated environmental data to model and predict heights to fill data gaps for small built infrastructure (≤5 m in height; Fig. 1b). We then develop three alternative scenarios with the first two representing common assumptions used when height data are unmeasured: 1) all built infrastructure without height data are impassable, 2) all built infrastructure without height data are passable, and 3) all built infrastructure without height data are allocated median height prediction from our model. We evaluate differences between these three scenarios when quantifying two catchment-level metrics of river fragmentation (percentage of and distance between impassable built infrastructure) for fish species when built infrastructure with heights ≥1.0, 1.5, or 2.0 m (our three passability thresholds) are impassable. Our three passability thresholds are based on the ecological continuity protocol established by the French National Agency for Water and Aquatic Environments (Baudoin et al., 2014). France's ecological continuity protocol is aimed at evaluating built infrastructure passability for fish species, and knowledge of the heights of different built infrastructure are both a major consideration in evaluation and a critical data gap in implementing the protocol at a national scale. We discuss the implications of common assumptions made about built infrastructure, and our modelling technique, for determining the effects of built infrastructure on aquatic ecosystems, and our ability to address impacts more effectively.

2. Methods

2.1. Built infrastructure and environmental data

We analyzed publicly available data for 76,292 built infrastructure from the French National Agency for Water and Aquatic Environments (http://www.onema.fr/le-roe). We excluded any records listed as destroyed, planned, under construction, invalid, or duplicated in the database. After these exclusions we had a total of 19,302 records with height data (Fig. 1a). A further 882 built infrastructure had available height data, but were without values for environmental data, and so were not included in our modelling of height, but retained for our assessment of infrastructure passability. For subsequent modelling we created a training dataset based on built infrastructure ≤5 m in height. We did this because < 1% (461) of built infrastructure with height and environmental data were greater than 5 m. Given the common dependence by humans on larger built infrastructure, we assumed that height values for these structures were well documented, and not likely unmeasured in our database. We retained these larger built infrastructure to include in our estimations of passability and calculations of catchment-level fragmentation.

The starting point for our model training dataset was 17,959 built infrastructure with heights ≤5 m and environmental variable data attributed to stream reaches available from the French Theoretical Hydrographic Network (Pella et al., 2012). There were an additional 20,077 built infrastructure without height values, but with environmental variable values, and we used our models to predict their heights (Fig. 1a). Environmental data were not available for all stream reaches with built infrastructure in place, but we initially considered 11 variables available for all stream reaches and included the percentage of land cover that was urban or agriculture within a 1 km circular buffer around each structure for initial consideration in our modelling (Table 1). We included agriculture and urban cover to account for landscape factors that can influence the distribution of infrastructure. Smaller and more frequent infrastructure, such as weirs, tend to occur in agriculture-dominated landscapes, and higher and less frequent infrastructure tend to occur in steeper landscapes with less human modification.

2.2. Modelling and predicting built infrastructure heights

We used Boosted Regression Trees (BRT; Elith et al., 2008) to model and predict infrastructure heights using the *dismo* package 2.1 (Hijmans et al., 2016) in R Statistical Package 3.2.2 (http://www.R-project.org/). We briefly describe BRT models; technical details and applications of these models have been widely presented in environmental and ecological science literature (e.g., Elith et al., 2008; Bhatt et al., 2013; Soykan et al., 2014; Hain et al., 2017). BRTs are part of the classification and regression tree family; techniques used to advance single classification or regression trees by averaging the results for each binary split from numerous trees or forests. Boosted tree models retain the positive aspects of single trees seen in classification and regression tree models, but provide improved predictive performance, nonlinearities and interactions are easily assessed, and the models can provide an ordered list of the importance of the explanatory variables (Elith et al. 2008; De'ath 2007).

For our BRT models, height values were rounded to the nearest half-meter for modelling (e.g., 0-0.24m = 0 m; 0.25-0.74m = 0.5m; 0.75-1.24m = 1.0m, etc), because there were likely moderate levels of uncertainty around the estimated heights supplied in the original database, and preliminary modelling demonstrated improved model performance when using rounded height values. Training our models with all 17,959 built infrastructure was impractical because of the computation time required, and previous work by Elith et al. (2008) demonstrated trade-offs with sample size and computing time, where modelling with a sub-sample of 6,000 sites showed high predictive performance and moderate computation time. Therefore, we randomly selected three sub-samples consisting of 5,000 built infrastructure records, and used these as our training datasets for subsequent modelling. With the three training data sub-samples, we fitted three BRT models, assuming the response followed a Gaussian distribution. We tested combinations of tree complexity (tc) (10-15), learning rate (lr) (0.001, 0.005) and bag fraction (bf) (0.5, 0.75). The learning rate determines the contribution of each tree to the growing model. Tree complexity controls whether interactions are fitted in the model: a tree complexity of one fits an additive model, a tree complexity of two fits a model with up to two-way interactions, and so on. Introducing some randomness into a boosted model can improve accuracy and speed and reduce over-fitting (Elith et al. 2008), but this can also introduce variance in fitted values and predictions between runs. The bag fraction controls stochasticity in the model, specifying the proportion of data to be used at each step; a bf of 0.75 means that 75% of the data are randomly drawn from the full model training dataset without replacement (Elith et al. 2008). We determined that for all three of our BRT models the following parameters returned highest model performance: tc = 15; lr = 0.005; bf = 0.75. We predicted height values for the 20,077 built infrastructure without values, giving three height predictions for each. For each of the three BRT models, we used a tenfold cross-validation (CV; Elith et al. 2008), evaluating model CV correlation (where higher values indicate a better model) and standard error, to assess model predictive performance to withheld portions of data (Elith et al. 2008).

We initially considered 11 environmental variables in each of the three BRT models (Table 1), and the importance of each environmental variable in each of the three models was evaluated based on its contribution to model fit. Strahler stream order and percentage urban cover were dropped from final models, leaving nine environmental variables, because they contributed <2% to each model, and model performance was the same without their inclusion.

2.3. Built infrastructure passability and catchment-level fragmentation

Applying the assumptions of our three scenarios for built infrastructure without heights, we determined if each of the 39,379 built infrastructure with known or predicted heights were passable or impassable for fish species unable to pass ≥1.0, 1.5, or 2.0 m heights. Our three built infrastructure passability thresholds (1.0, 1.5, and 2.0 m) were based on the most conservative estimates of fish species swimming and jumping capacities (i.e., their ability to pass built infrastructure or not) determined by Baudoin et al. (2014) for fishes moving in an upstream direction in favourable hydrologic conditions. We chose infrastructure height as an indicator of a fish species ability to pass over built infrastructure or not because: 1) we had access to height information in our database, and 2) Baudoin et al. (2014) established that for vertical, sub-vertical or inclined dams and weirs (those built infrastructure considered in our analysis), an extreme height value is the first element that determines whether or not a structure is likely to be passable for a particular fish species. Baudoin et al. (2014) determined built infrastructure passability thresholds for fish species in France that are unable to pass ≥1.0, 1.5, or 2.0 m heights, and we present 30 of the native species for which these thresholds are applicable in Table 2. For example, built infrastructure at 1 m or more are impassable for fish species such as Three-spined Stickleback (*Gasterosteus gymnurus*), those at 1.5 m or more are impassable for species like Burbot (*Lota lota*), and those at 2 m or more are impassable for species like Twait Shad (*Alosa fallax*).

Using our built infrastructure data, and the French hydrographical network (https://www.data.gouv.fr/fr/datasets/bd-carthage-onm) to represent rivers, we then determined and compared river fragmentation across 26 major catchments based on two metrics: the percentage of impassable built infrastructure and average distance (km) between impassable built infrastructure. We evaluated differences in the resulting values for each fragmentation metric when applying our three scenarios and the built infrastructure passability thresholds (1.0, 1.5, and 2.0 m). We used analysis of covariance (ANCOVA) to investigate catchment-level differences for both of our river fragmentation metrics, comparing between scenarios for each of the passability thresholds, and with river length within each catchment as a co-variate. ANCOVA was conducted for both fragmentation metrics using the function *lm* from the *base* package, and Tukey's post-hoc tests using the *glht* function from the *multcomp* (Hothorn 2008) package in R Statistical Software (version 3.2.2) (http://www.R-project.org/). It was necessary to log transform average distance between impassable built infrastructure for each catchment to meet assumptions of normality and homogeneity.

3. Results

3.1. Modelling and predicting built infrastructure heights

Our three BRT models showed similar and reasonable discrimination and predictive performances for small built infrastructure in France (Table 3). The final predicted heights for built infrastructure ranged from 0 to 4 m across France (all modelled data available at: https://figshare.com/s/617347a78cc27f419023). Regardless of the model considered, we found that four of the nine environmental variables had at least 12% relative influence on infrastructure height (Fig. 2; Table 4). Higher infrastructure tended to occur on shorter stream reaches (19% relative influence on average between the three models), at lower (<500 m) and higher elevation (>1000 m) (14% on average), and on stream reaches with higher gradient (change in elevation per reach length; 13% on average) (Fig. 2; Table 4). Infrastructure height also rapidly increased with increasing average annual flow and tended to level off at flows above 100 m^3/s (12% on average) (Fig. 2; Table 4). Median height values across our three models were consistent, with half the predicted values having zero standard deviation, and the majority (18,105; 90%) of built infrastructure had median predicted height values of 1.0 (n = 10,749) or 1.5 meters (n = 7,356). Our full database of built infrastructure, including known heights, predicted height values for built infrastructure from all three models, modelled median height values for built infrastructure, and model deviation are available at: https://figshare.com/s/617347a78cc27f419023.

3.2. Built infrastructure passability and catchment-level fragmentation

We found significant differences in catchment-level fragmentation between our three scenarios, the pattern of which varied with passability threshold (Fig. 3a-c; Table S1). For a passability threshold of 1.0 m, on average 85% ± 2 (SE) of built infrastructure were impassable under scenarios 1 and 3 across catchments (see Table S1), and distance between impassable structures also did not differ (18.0 km ± 2.8 on average), whereas significantly fewer built infrastructure were impassable under scenario 2 (29% ± 3.0 on average; see Table S1) (ANCOVA: $F_{2,74} = 214.53$, $p < 0.001$), and the distance between impassable built infrastructure (106 km ± 41.3 on average) was significantly greater (ANCOVA: $F_{2,74} = 18.7$, $p < 0.001$) than under scenarios 1 and 3 (Fig. 3a). We found that for a passability threshold of 1.5 m all three scenarios differed significantly both in terms of percentage (scenario 1: 74% ± 3.0; scenario 2: 18% ± 2.0; scenario 3: 41% ± 3.0 on average; Table S1) (ANCOVA: $F_{2,74} = 123.25$, $p < 0.001$) and distance (scenario 1: 20.7 km ± 2.9; scenario 2: 143 km ± 42.0; scenario 3: 46.5 km ± 11.1 on average) between impassable infrastructure across catchments (ANCOVA: $F_{2,74} = 23.0$, $p < 0.001$) (Fig. 3b). For a passability threshold of 2.0 m, scenarios 2 and 3 showed no difference on average across catchments either in terms of percentage (scenario 1: 11% ± 1.0; scenario 2: 17% ± 2.0 on average; Table S1) or distance between (197.0 km ± 60.6, and 160.9 km ± 61.9 on average; Fig. 3c) impassable built infrastructure, but both the percentage (67% ± 4.0 on average) and distance between impassable built infrastructure (23.3 km ± 3.2 on average) were significantly different for scenario 1 (ANCOVA: $F_{2,74} = 148.2$ and $F_{2,74} = 37.9$, $p < 0.001$; see Table S1; Fig. 3c). We found no effect of river length (km) on catchment-level river fragmentation regardless of fragmentation metric or the passability threshold.

4. Discussion

Drawing on remotely collected data we modelled and predicted built infrastructure height with reasonable certainty, doubling the number with height values across France. We further demonstrated that common assumptions made about built infrastructure when data gaps exist can result in significantly different estimates of river fragmentation for fish species with varied abilities to pass built infrastructure.

When large numbers of built infrastructure have unknown dimensions, such as height, we can be forced to make assumptions; either that all built infrastructure are passable, or impassable (e.g., Radinger et al., 2017). Our results suggest that these assumptions can result in opposite outcomes for measures of river fragmentation for fish species with varied abilities to pass built infrastructure. For example, assuming that all built infrastructure without height data were passable resulted in a 5-fold reduction in river fragmentation for species such as the Three-spined Stickleback (passability threshold ≥1.0 m) compared to using our predicted height values to measure distance between impassable built infrastructure. We found the opposite was true for species like the Twait Shad (passability threshold ≥2.0 m), where assuming all built infrastructure without height data were impassable resulted in a 7-fold increase in river fragmentation compared to using our predicted height values to measure distance between impassable built infrastructure. Our findings suggest that inclusion of built infrastructure height, and modelling height where necessary, can help to refine estimates of river fragmentation for fish species with varied abilities to pass built infrastructure. With an increased interest in modelling fish species' dispersal abilities (Radinger et al., 2017), and continued efforts to prioritize removal projects using indicators of built infrastructure passability (Neeson et al., 2015), our approach can be used to improve understanding of built infrastructure impact and inform the identification of priorities for restoring river connectivity to benefit different species.

Our results demonstrate a first step toward more explicit accountancy of built infrastructure impact on aquatic biodiversity. For example, our approach builds on earlier work by Perkin & Gido (2012) who noted that infrastructure passability for different fish species could be a function of both structure height and local hydrological regimes but did not explicitly account for such factors and instead assumed partial passability for all infrastructure. Refinements to our modelling approach that explicitly consider species' biological characteristics, which can influence their ability to pass built infrastructure, would likely further improve estimates of river fragmentation for individual species, but such data are not broadly available. We were able to account for a coarse estimation of river hydrology in our catchment-level fragmentation calculations, because hydrologic variability was integrated in the passability thresholds established by Baudoin et al. (2014). Finer-scale data on river discharge at individual infrastructure is currently not available, but explicit consideration of this factor would be useful in future iterations of this work. We emphasize that our models specifically address a need for overcoming gaps in knowledge about built infrastructure height, and additional considerations such as discharge and fish species' biological characteristics will only help to refine our modelling and findings. Further, mismatches in existing spatial data products did not allow us to predict height values for all built infrastructure in France, and factors such as fish passage facilities that we were unable to account for in our assessment, could influence whether or not these are passable for different fish species. Uncertainty in infrastructure status and presence of fish passage facilities could be validated using a combination of finer-scale spatial data and field surveys.

Currently field surveys are being carried out across France, but the number of built infrastructure prevents assessments being completed in short time periods (e.g., 1 or 2 years). Coupling on ground work with acquisition of fine-scale spatial data could facilitate rapid, and cost-effective validation procedures, while our results could be used to systematically target potential problem areas.

Globally, built infrastructure removal and installation is occurring simultaneously (Hydropower Status Report 2017; Dam Removal Europe 2016) and methods similar to what we present here offer a starting point for improving our ability to quantify costs and benefits associated with these processes. Our results (i.e., known and predicted height values) could be integrated in conservation planning exercises, along with other ecological and socio-economic considerations, as a relative indicator of cost to remove built infrastructure. Built infrastructure height can also be used as an indicator of environmental benefit, such as downstream response to removal, where higher dams have been shown to have longer-lasting and more wide-spread downstream effects than shorter dams (Major et al., 2017). These examples demonstrate the wide-applicability of our approach and results to informing conservation decisions with broader considerations than fishes. Further, our approach could be used to inform future scenarios that consider how built infrastructure change over time with respect to removal, installation and other environmental and socio-political factors, such as changing climate and flows, and placement of fish passage facilities to reduce impact. We see particular relevance of our approach to other areas in Europe as well as North America where efforts to both remediate (in the form of including fish passage facilities) and remove built infrastructure is rapidly occurring (Foley et al., 2017; Dam Removal Europe 2016) but where gaps in data on characteristics of built infrastructure remain (e.g., Radinger et al., 2017; Januchowski-Hartley et al., 2013) and limit our ability to make effective decisions. We see further applicability of our modelling approach and results to other parts of the world as a global proliferation of smaller infrastructure continues with limited consideration or documentation of characteristics like height (Couto & Olden, 2018). Ultimately, as global change continues, approaches like ours will become increasingly important for guiding more proactive and effective strategies for built infrastructure management.

Author contributions

SRJ and PAT conceived the idea and designed the methods; SRJ and CJ collated the data; SRJ analysed the data; SRJ led the writing of the manuscript. All authors contributed critically to all manuscript drafts and gave final approval for publication.

Acknowledgements

This work was funded by the French Agency for Biodiversity and Research Foundation for Biodiversity Biodiversa project ODYSSEUS. Laboratoire Evolution et Diversité Biologique is part of ‘‘LABEX TULIP’’ and ‘‘LABEX CEBA’’ (ANR-10-LABX-41, ANR-10-LABX-25-01). SRJ also acknowledges funding from the Welsh European Funding Office and European Regional Development Fund under project number 80761-SU-140 (West). We thank F. Januchowski-Hartley and two anonymous reviewers for useful comments on earlier versions of this manuscript.

References

Baudoin JM, Burgun V, Chanseau M, *et al.* 2014. Assessing the passage of obstacles by fish. Concepts, design and application. Onema. 200 pages.

Bhatt S, Gething PW, Brady OJ, *et al.* 2013. The global distribution and burden of dengue. *Nature* **496**: 504-507.

Carvajal JD, Januchowski-Hartley SR, Maldonado-Ocampo JA, *et al.* 2017. Damming fragments species' ranges and heightens extinction risk. *Conserv Lett* **10**: 708-716.

Cote D, Kehler DG, Bourne C and Wiersma YF. 2009. A new measure of longitudinal connectivity for stream networks. *Landsc Ecol* **24**: 101-113.

Couto TBA and Olden JD. 2018. Global proliferation of small hydropower plants – science and policy. *Front Ecol Environ* **16**: 91-100.

Dam Removal Europe. 2016. What is Dam Removal Europe? http://damremoval.eu/ Viewed 1 May 2017.

De'ath G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* **88**: 243-251.

Elith J, Leathwick JR, and Hastie T. 2008. A working guide to boosted regression trees. *J Anim Ecol* **77**: 802–813.

ESRI. 2015. ArcGIS 10.3.1. Environmental Systems Research Institute, Redlands, CA.

Foley MM, Magilligan FJ, Torgersen CE. *et al*. 2017. Landscape context and the biophysical response of rivers to dam removal in the United States. *PLoS One* 12: e0180107.

Hain EF, Kennen JG, Caldwell PV. et al. 2017. Using regional scale flow-ecology modelling to identify catchments where fish assemblages are most vulnerable to changes in water availability. *Fresh Biol* **63**: 928-945.

Hijmans RJ, Phillips S, Leathwick J, and Elith J. 2016. Dismo: species distribution modeling. R package version 1.0–15. See https://CRAN.R-project.org/package=dismo.

Hothorn T, Bretz F, and Westfall P. 2008. Simultaneous inference in general parametric models. *Biom J* **50**: 346-363.

Hydropower Status Report. 2017. https://www.hydropower.org/2017-hydropower-status-report. Viewed 1 June 2017.

Januchowski-Hartley,SR, McIntyre PB, Diebel M. *et al*. 2013. Restoring aquatic ecosystem connectivity requires expanding inventories of both dams and road crossings. *Front Ecol Environ* **11**: 211-217.

Januchowski-Hartley, SR, Diebel, M, Doran PJ, and McIntyre PB. 2014. Predicting road culvert passability for migratory fishes. *Divers Distrib* **20**: 1414-1424.

Major JJ, East AE, O'Connor JE. *et al*. 2017. Geomorphic responses to dam removal in the United States - a two-decade perspective. In: Gravel-bed rivers: processes and disasters. Oxford: John Wiley & Sons Ltd.

Mantel SK, Rivers-Moore N, and Ramulifho P. 2017. Small dams need consideration in riverscape conservation assessments. *Aquatic Conserv: Mar Freshw Ecosyst* **27**: 748-754.

Neeson TM, Ferris MC, Diebel MW, *et al*. 2015. Enhancing ecosystem restoration efficiency through spatial and temporal coordination. *Proc Natl Acad Sci* **112**: 6236-6241.

Pella H, Lejot J, Lamouroux N, and Snelder TH. 2012. Le réseau hydrographique théorique (RHT) français et ses attributs environnementaux. *Geomorphologie* **3**: 317-336.

Perkin JS and Gido KB. 2012. Fragmentation alters stream fish community structure in dendritic ecological networks. *Ecol Appl* **22**: 2176-2187.

Poff LN and Hart DD. 2002. How dams vary and why it matters for the emerging science of dam removal. *BioScience* **52**: 659-667.

Radinger J, Essl F, Hölker F, *et al*. 2017. The future distribution of river fish: the complex interplay of climate and land use changes, species dispersal and movement barriers. *Global Change Biol* https://doi.org/10.1111/gcb.13760.

Soykan CU, Eguchi T, Kohin S, and Dewar H. 2014. Prediction of fishing effort distributions using boosted regression trees. *Ecol Appl* **24**: 71-83.

Figure captions

Figure 1. Data on characteristics of (a) small (≤5 m in height) built infrastructure are often limited, resulting in scientists and practitioners needing to make assumptions about related impact on species like fishes. In France, (b) slightly more than half of the documented built infrastructure are without height data.

Figure 2. Partial dependency plots for environmental variables contributing >12% in three models (a-c) for small (≤5 m high) built infrastructure heights. Rug plots inside the top of each plot show the distribution of observations across the range of that variable, in deciles.

Figure 3. Catchment-level average distances between impassable (in an upstream direction) built infrastructure under three scenarios depicting common assumptions (Scenario 1 = all built infrastructure without height data assumed impassable; Scenario 2 = all built infrastructure without height data assumed passable), and Scenario 3 using median modelled height data from three Boosted Regression Tree models, compared for three passability thresholds: (a) 1.0 m, (b) 1.5 m, and (c) 2.0 m. For each passability threshold, boxplots show the median and 50% quartiles, whiskers are 1.5 times interquartile range, for log transformed average distance between impassable built infrastructure (km) under each scenario. Outlying values are not shown. Images of (a) *Gasterosteus gymnurus*, (b) *Lota lota*, and (c) *Alosa fallax* above boxplots depict the types of fish species for which passability thresholds are applicable. The *Gasterosteus gymnurus* image was created by Milton Tan, was unchanged, and is used under creative commons license (https://creativecommons.org/licenses/by-nc-sa/3.0/). The *Gasterosteus gymnurus* and *Lota lota* images were sourced from PhyloPic (phylopic.org).

Figure 1

Click here to download high resolution image

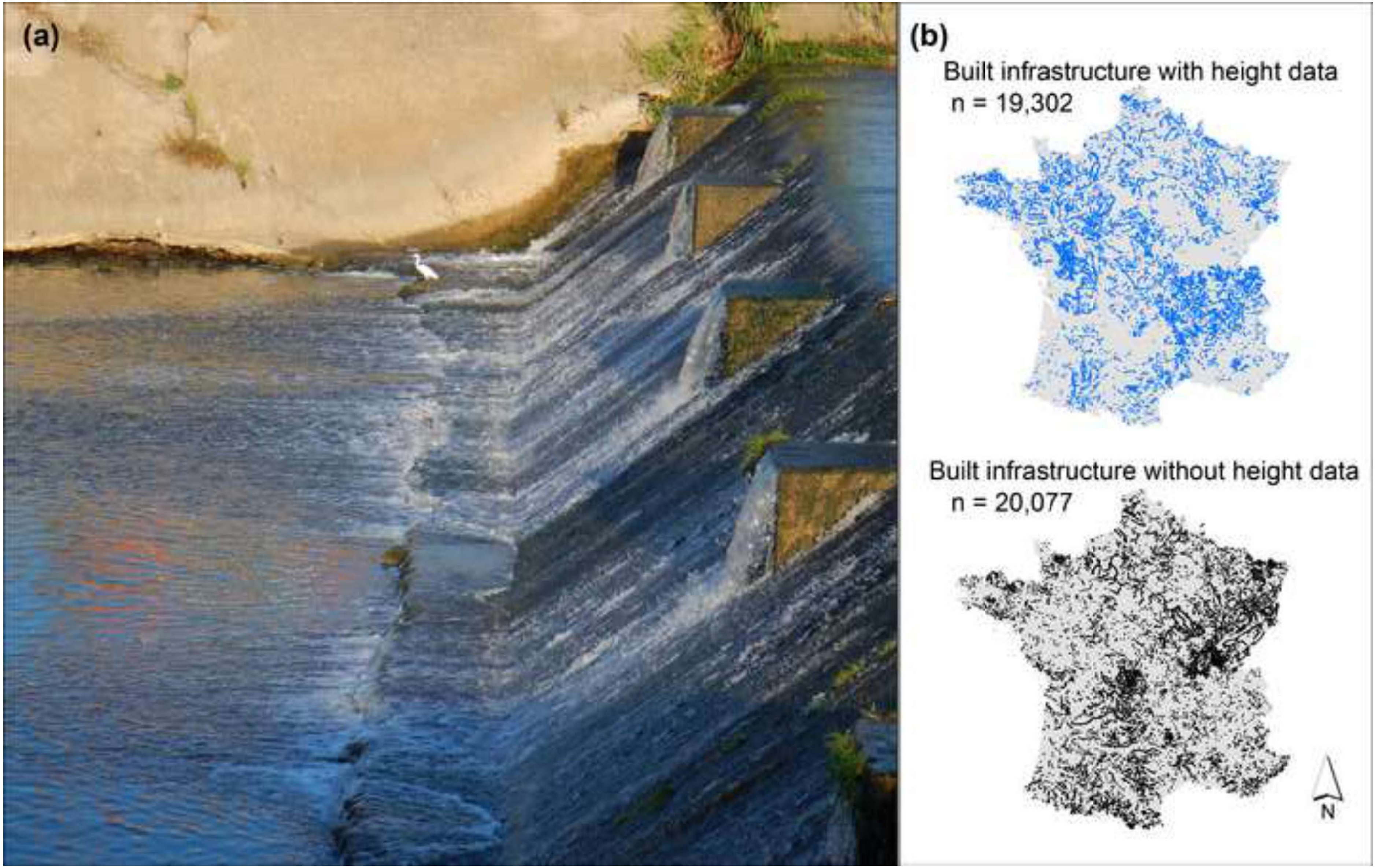

Figure 2

Click here to download high resolution image

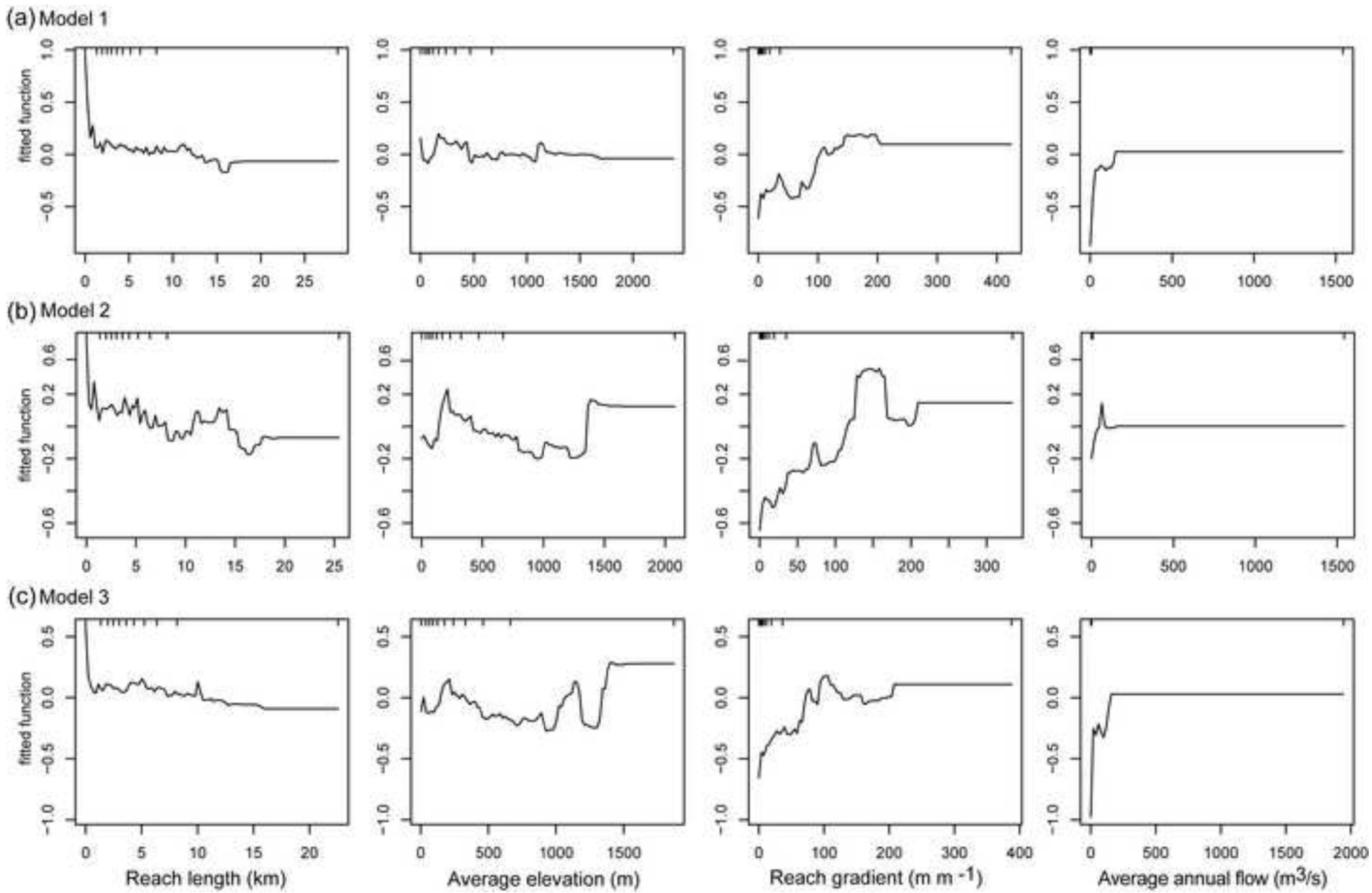

Figure 3
Click here to download high resolution image

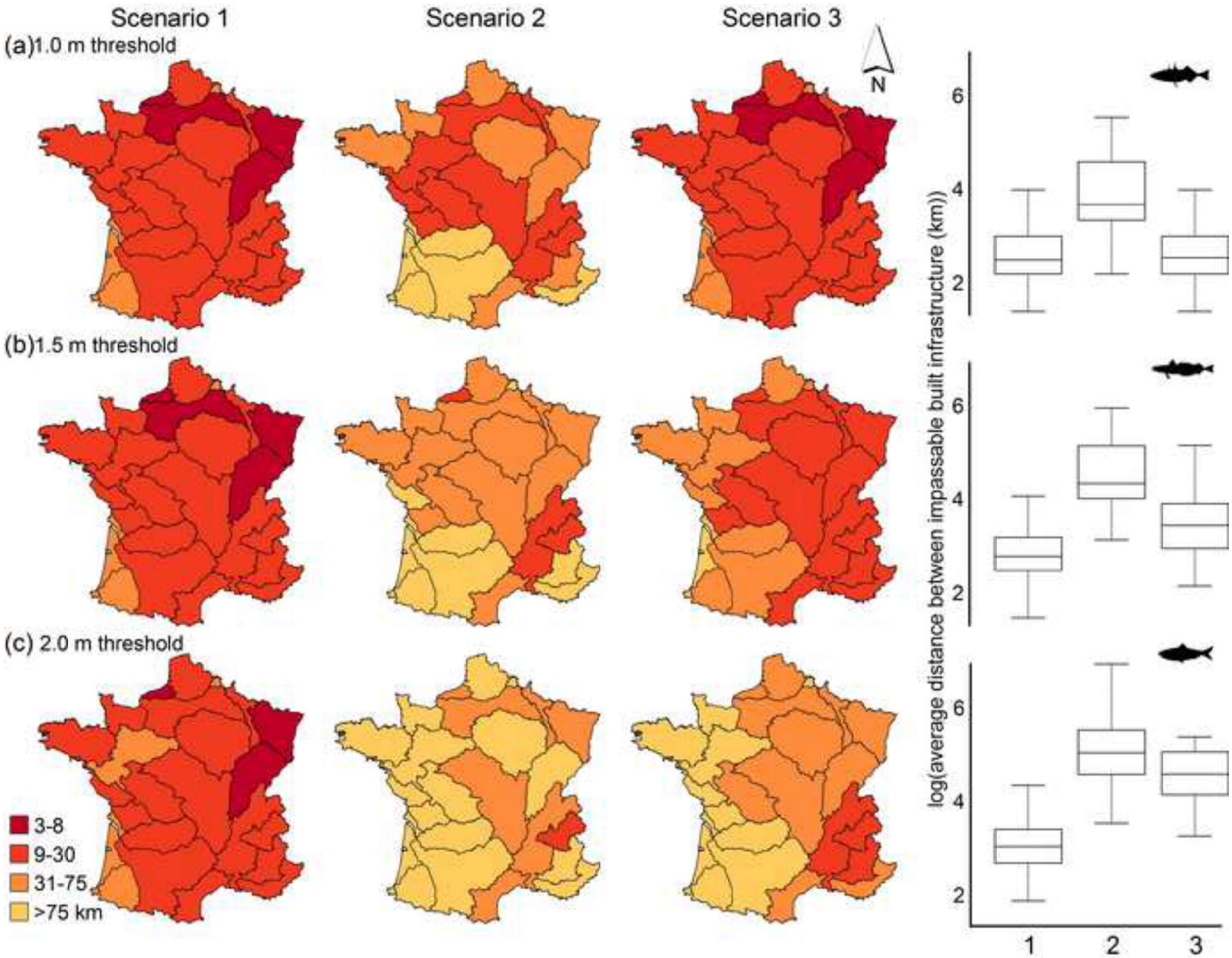

Table 1. Environmental variables used to characterize both the known and predicted built infrastructure heights in three Boosted Regression Tree models. Description of each environmental variable is provided.

Environmental Variable	Description
Stream reach gradient (m m^{-1})	Stream reach gradient where built infrastructure is located.
Average monthly minimum flow (m^3/s)	Average monthly minimum flow of a stream reach where built infrastructure is located.
Stream reach average elevation (m)	Average elevation of a stream reach where built infrastructure is located.
Average annual flow (m^3/s)	Average annual flow of a stream reach where built infrastructure is located.
Stream reach drainage area (km^2)	The amount of area locally draining to a stream reach where built infrastructure is located.
Distance to source (km)	Distance to the upstream source of the river network from a stream where built infrastructure is located.
Stream reach upstream drainage area (km^2)	The amount of upstream area draining to a stream reach where built infrastructure is located.
Stream reach length (km)	The length of a stream reach where built infrastructure is located.
Strahler stream order (categorical)	The Strahler stream order of a stream reach where built infrastructure is located.
Percentage agriculture cover (%)	Percentage of agricultural land cover within a 1 km circular buffer around built infrastructure.
Percentage urban cover (%)	Percentage of urban cover within a 1 km circular buffer around built infrastructure.

Table 2. Fish species native to France that are unable to pass (in an upstream direction) built infrastructure ≥ 1.0, 1.5 or 2.0 m in height.

Species	**1.0 m threshold**	**1.5 m threshold**	**2.0 m threshold**
Anguilla anguilla	1		
Rhodeus amarus	1		
Gasterosteus gymnurus	1		
Pungitius laevis	1		
Cobitis taenia	1		
Barbatula barbatula	1		
Lampetra planeri	1		
Zingel asper	1		
Parachondrostoma toxostoma	1		
Scardinius erythrophthalmus	1		
Rutilus rutilus	1		
Carassius gibelio	1		
Carassius carassius	1		
Telestes souffia	1		
Barbus meridionalis	1		
Alburnoides bipunctatus	1		
Alburnus alburnus	1		
Tinca tinca		1	
Perca fluviatilis		1	
Lota lota		1	
Blicca bjoerkna		1	
Abramis brama		1	
Lampetra fluviatilis		1	
Squalius cephalus		1	
Barbus barbus		1	
Thymallus thymallus		1	
Aspius aspius			1
Esox lucius			1
Petromyzon marinus			1
Alosa fallax			1
Total	17	9	4

Table 3. Parameters and performance for three Boosted Regression Tree models of built infrastructure heights.

Model	Number of records	Environmental variables	Training data correlation (based on 5000 sites)	Cross validation correlation	Cross validation standard error	Number of trees
Model 1	5000	9	0.79	0.40	0.01	3750
Model 2	5000	9	0.75	0.35	0.01	2950
Model 3	5000	9	0.73	0.37	0.02	2550

Table 4. Environmental variable contributions to three Boosted Regression Tree models of built infrastructure heights.

Environmental variable	**Model 1**	**Model 2**	**Model 3**	**Model average**
Stream reach length (km)	18%	19%	19%	19%
Stream reach average elevation (m)	14%	15%	14%	14%
Stream reach gradient (m m^{-1})	12%	12%	14%	13%
Average annual flow (m^3/s)	12%	12%	12%	12%
Percentage agriculture cover (%)	10%	10%	10%	10%
Stream reach drainage area (km^2)	10%	9%	9%	9%
Average monthly minimum flow (m^3/s)	9%	9%	9%	9%
Stream reach upstream drainage area (km^2)	8%	7%	7%	7%
Distance to source (km)	7%	7%	6%	7%

Table S1. Total river length, number of built infrastructure (BI count), percentage of impassable built infrastructure, and average distance between impassable built infrastructure for 26 river catchments in France. The percentage of, and average distance between, impassable built infrastructure is presented for three different passability thresholds (≥1.0, 1.5, or 2.0 m, representing heights at which built infrastructure are impassable for different fish species) when considering three scenarios of built infrastructure heights: two based on common assumptions (all infrastructure without height data are impassable (scen 1) or passable (scen 2)), and one based on modeled heights (scen 3).

ID	River length (km)	BI count	Impass_1m_scen1 (%)	Impass_1_5m_scen1 (%)	Impass_2m_scen1 (%)	Impass_1m_scen2 (%)	Impass_1_5m_scen2 (%)	Impass_2m_scen2 (%)	Impass_1m_scen3 (%)	Impass_1_5m_scen3 (%)	Impass_2m_scen3 (%)
1	20546.0	751	89%	82%	78%	20%	13%	9%	89%	54%	28%
2	1271.0	46	85%	63%	48%	54%	33%	17%	85%	48%	20%
3	26772.0	1825	79%	66%	57%	37%	25%	15%	79%	35%	16%
4	835.0	338	72%	57%	51%	25%	10%	4%	72%	14%	4%
5	1608.0	34	100%	100%	100%	9%	9%	9%	100%	24%	9%
6	8052.0	125	93%	91%	90%	6%	5%	4%	93%	26%	4%
7	717.0	23	100%	96%	91%	17%	13%	9%	100%	26%	9%
8	7286.0	641	98%	95%	92%	11%	8%	6%	98%	58%	15%
9	14933.0	1315	91%	83%	77%	30%	22%	17%	91%	62%	27%
10	13487.0	880	86%	75%	67%	32%	21%	13%	86%	37%	14%
11	7217.0	741	71%	46%	40%	36%	13%	8%	70%	23%	10%
12	5933.0	820	68%	46%	34%	36%	15%	5%	68%	27%	5%
13	20175.0	1278	92%	84%	77%	20%	12%	6%	90%	42%	12%
14	9354.0	731	92%	86%	83%	18%	12%	10%	92%	61%	42%
15	10719.0	1083	81%	70%	65%	28%	18%	13%	80%	26%	15%