



**HAL**  
open science

## Modelling built infrastructure heights to evaluate common assumptions in aquatic conservation

Stephanie R. Januchowski-Hartley, Céline Jézéquel, Pablo A. Tedesco

### ► To cite this version:

Stephanie R. Januchowski-Hartley, Céline Jézéquel, Pablo A. Tedesco. Modelling built infrastructure heights to evaluate common assumptions in aquatic conservation. *Journal of Environmental Management*, 2019, 232, pp.131–137. 10.1016/j.jenvman.2018.11.040 . hal-04716888

**HAL Id: hal-04716888**

**<https://cnrs.hal.science/hal-04716888v1>**

Submitted on 1 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Swansea University  
Prifysgol Abertawe



## Cronfa - Swansea University Open Access Repository

---

This is an author produced version of a paper published in:  
*Journal of Environmental Management*

Cronfa URL for this paper:  
<http://cronfa.swan.ac.uk/Record/cronfa48243>

---

### **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>

---

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

## **Modelling built infrastructure heights to evaluate common assumptions in aquatic conservation**

Stephanie R Januchowski-Hartley<sup>a,b</sup>, Céline Jézéquel<sup>b</sup> and Pablo A Tedesco<sup>b</sup>

<sup>a</sup> Department of Biosciences, Swansea University, Swansea, United Kingdom SA2 8PP

<sup>b</sup> 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 ( $\leq 5$  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  $\geq 1.0$  m. The opposite is true for fish species that cannot pass  $\geq 2.0$  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 ( $\leq 5$  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 1. Introduction

2  
3  
4  
5 3 Scientists and practitioners require information on the characteristics of built infrastructure,  
6  
7 4 such as dams and weirs, to better understand associated impacts, costs, and benefits, in  
8  
9  
10 5 relation to ecological processes, services, and human values (Poff & Hart, 2002;  
11  
12 6 Januchowski-Hartley et al., 2013; Major et al., 2017). Characteristics of larger built  
13  
14 7 infrastructure are increasingly well understood, because of improved identification via  
15  
16  
17 8 remotely sensed imagery (Mantel et al., 2017), and superior record keeping due to the  
18  
19  
20 9 importance of size and water holding capacity for monitoring energy production and water  
21  
22 10 storage (e.g., Carvajal et al., 2017). Despite likely impacts from small built infrastructure  
23  
24 11 which often restrict fish movement (i.e., being impassable), impede river flows, and retain  
25  
26  
27 12 sediments and materials, data on their distribution and characteristics remain limited  
28  
29  
30 13 (Januchowski-Hartley et al., 2013; Couto & Olden, 2018). Collection of this necessary  
31  
32 14 information is challenged by the large number of built infrastructure with unknown  
33  
34 15 dimensions (e.g., height), which means that assumptions are often necessary in research and  
35  
36  
37 16 decision making (e.g., assume binary passability or impassability of built infrastructure) when  
38  
39 17 height data are unmeasured (Cote et al., 2009; Perkin & Gido, 2012; Radinger et al., 2017).  
40  
41 18 This raises the question of how common assumptions about characteristics of built  
42  
43  
44 19 infrastructure affect estimates of habitat fragmentation, and the potential implications of this  
45  
46 20 for fishes with different abilities to pass over infrastructure.

47  
48  
49 21 Here, we investigate existing data and data gaps for built infrastructure (Fig. 1a), and  
50  
51 22 evaluate how these influence measures of river fragmentation when considering passability  
52  
53 23 (the ability of a fish species to pass built infrastructure in an upstream direction) for native  
54  
55  
56 24 fishes in France. We do this by bringing together a database of built infrastructure, and  
57  
58  
59 25 associated environmental data to model and predict heights to fill data gaps for small built  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

infrastructure ( $\leq 5$  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  $\geq 1.0$ , 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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51 further 882 built infrastructure had available height data, but were without values for  
52 environmental data, and so were not included in our modelling of height, but retained for our  
53 assessment of infrastructure passability. For subsequent modelling we created a training  
54 dataset based on built infrastructure  $\leq 5$  m in height. We did this because  $< 1\%$  (461) of built  
55 infrastructure with height and environmental data were greater than 5 m. Given the common  
56 dependence by humans on larger built infrastructure, we assumed that height values for these  
57 structures were well documented, and not likely unmeasured in our database. We retained  
58 these larger built infrastructure to include in our estimations of passability and calculations of  
59 catchment-level fragmentation.

60         The starting point for our model training dataset was 17,959 built infrastructure with  
61 heights  $\leq 5$  m and environmental variable data attributed to stream reaches available from the  
62 French Theoretical Hydrographic Network (Pella et al., 2012). There were an additional  
63 20,077 built infrastructure without height values, but with environmental variable values, and  
64 we used our models to predict their heights (Fig. 1a). Environmental data were not available  
65 for all stream reaches with built infrastructure in place, but we initially considered 11  
66 variables available for all stream reaches and included the percentage of land cover that was  
67 urban or agriculture within a 1 km circular buffer around each structure for initial  
68 consideration in our modelling (Table 1). We included agriculture and urban cover to account  
69 for landscape factors that can influence the distribution of infrastructure. Smaller and more  
70 frequent infrastructure, such as weirs, tend to occur in agriculture-dominated landscapes, and  
71 higher and less frequent infrastructure tend to occur in steeper landscapes with less human  
72 modification.

73

## 74 2.2. Modelling and predicting built infrastructure heights

75



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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

76 For our BRT models, height values were rounded to the nearest half-meter for modelling  
77 (e.g., 0-0.24m = 0 m; 0.25-0.74m = 0.5m; 0.75-1.24m = 1.0m, etc), because there were likely  
78 moderate levels of uncertainty around the estimated heights supplied in the original database,  
79 and preliminary modelling demonstrated improved model performance when using rounded  
80 height values. Training our models with all 17,959 built infrastructure was impractical  
81 because of the computation time required, and previous work by Elith et al. (2008)  
82 demonstrated trade-offs with sample size and computing time, where modelling with a sub-  
83 sample of 6,000 sites showed high predictive performance and moderate computation time.  
84 Therefore, we randomly selected three sub-samples consisting of 5,000 built infrastructure  
85 records, and used these as our training datasets for subsequent modelling. With the three  
86 training data sub-samples, we fitted three BRT models, assuming the response followed a  
87 Gaussian distribution. We tested combinations of tree complexity (tc) (10-15), learning rate  
88 (lr) (0.001, 0.005) and bag fraction (bf) (0.5, 0.75). The learning rate determines the  
89 contribution of each tree to the growing model. Tree complexity controls whether interactions

101 are fitted in the model: a tree complexity of one fits an additive model, a tree complexity of  
102 two fits a model with up to two-way interactions, and so on. Introducing some randomness  
103 into a boosted model can improve accuracy and speed and reduce over-fitting (Elith et al.  
104 2008), but this can also introduce variance in fitted values and predictions between runs. The  
105 bag fraction controls stochasticity in the model, specifying the proportion of data to be used  
106 at each step; a bf of 0.75 means that 75% of the data are randomly drawn from the full model  
107 training dataset without replacement (Elith et al. 2008). We determined that for all three of  
108 our BRT models the following parameters returned highest model performance: tc = 15; lr =  
109 0.005; bf = 0.75. We predicted height values for the 20,077 built infrastructure without  
110 values, giving three height predictions for each. For each of the three BRT models, we used a  
111 tenfold cross-validation (CV; Elith et al. 2008), evaluating model CV correlation (where  
112 higher values indicate a better model) and standard error, to assess model predictive  
113 performance to withheld portions of data (Elith et al. 2008).

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

120

### 121 2.3. Built infrastructure passability and catchment-level fragmentation

122

123 Applying the assumptions of our three scenarios for built infrastructure without heights, we  
124 determined if each of the 39,379 built infrastructure with known or predicted heights were  
125 passable or impassable for fish species unable to pass  $\geq 1.0$ , 1.5, or 2.0 m heights. Our three

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
126 built infrastructure passability thresholds (1.0, 1.5, and 2.0 m) were based on the most  
127 conservative estimates of fish species swimming and jumping capacities (i.e., their ability to  
128 pass built infrastructure or not) determined by Baudoin et al. (2014) for fishes moving in an  
129 upstream direction in favourable hydrologic conditions. We chose infrastructure height as an  
130 indicator of a fish species ability to pass over built infrastructure or not because: 1) we had  
131 access to height information in our database, and 2) Baudoin et al. (2014) established that for  
132 vertical, sub-vertical or inclined dams and weirs (those built infrastructure considered in our  
133 analysis), an extreme height value is the first element that determines whether or not a  
134 structure is likely to be passable for a particular fish species. Baudoin et al. (2014)  
135 determined built infrastructure passability thresholds for fish species in France that are unable  
136 to pass  $\geq 1.0$ , 1.5, or 2.0 m heights, and we present 30 of the native species for which these  
137 thresholds are applicable in Table 2. For example, built infrastructure at 1 m or more are  
138 impassable for fish species such as Three-spined Stickleback (*Gasterosteus gymmurus*), those  
139 at 1.5 m or more are impassable for species like Burbot (*Lota lota*), and those at 2 m or more  
140 are impassable for species like Twait Shad (*Alosa fallax*).

36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
141 Using our built infrastructure data, and the French hydrographical network  
142 (<https://www.data.gouv.fr/fr/datasets/bd-carthage-onm>) to represent rivers, we then  
143 determined and compared river fragmentation across 26 major catchments based on two  
144 metrics: the percentage of impassable built infrastructure and average distance (km) between  
145 impassable built infrastructure. We evaluated differences in the resulting values for each  
146 fragmentation metric when applying our three scenarios and the built infrastructure  
147 passability thresholds (1.0, 1.5, and 2.0 m). We used analysis of covariance (ANCOVA) to  
148 investigate catchment-level differences for both of our river fragmentation metrics,  
149 comparing between scenarios for each of the passability thresholds, and with river length  
150 within each catchment as a co-variate. ANCOVA was conducted for both fragmentation

151 metrics using the function *lm* from the *base* package, and Tukey's post-hoc tests using the  
152 *glht* function from the *multcomp* (Hothorn 2008) package in R Statistical Software (version  
153 3.2.2) (<http://www.R-project.org/>). It was necessary to log transform average distance  
154 between impassable built infrastructure for each catchment to meet assumptions of normality  
155 and homogeneity.

156

### 157 3. Results

158

#### 159 3.1. Modelling and predicting built infrastructure heights

160

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

176 modelled median height values for built infrastructure, and model deviation are available at:  
177 <https://figshare.com/s/617347a78cc27f419023>.

178

### 179 3.2. Built infrastructure passability and catchment-level fragmentation

180

181 We found significant differences in catchment-level fragmentation between our three  
182 scenarios, the pattern of which varied with passability threshold (Fig. 3a-c; Table S1). For a  
183 passability threshold of 1.0 m, on average  $85\% \pm 2$  (SE) of built infrastructure were  
184 impassable under scenarios 1 and 3 across catchments (see Table S1), and distance between  
185 impassable structures also did not differ ( $18.0 \text{ km} \pm 2.8$  on average), whereas significantly  
186 fewer built infrastructure were impassable under scenario 2 ( $29\% \pm 3.0$  on average; see Table  
187 S1) (ANCOVA:  $F_{2,74} = 214.53$ ,  $p < 0.001$ ), and the distance between impassable built  
188 infrastructure ( $106 \text{ km} \pm 41.3$  on average) was significantly greater (ANCOVA:  $F_{2,74} = 18.7$ ,  
189  $p < 0.001$ ) than under scenarios 1 and 3 (Fig. 3a). We found that for a passability threshold of  
190 1.5 m all three scenarios differed significantly both in terms of percentage (scenario 1:  $74\% \pm$   
191  $3.0$ ; scenario 2:  $18\% \pm 2.0$ ; scenario 3:  $41\% \pm 3.0$  on average; Table S1) (ANCOVA:  $F_{2,74} =$   
192  $123.25$ ,  $p < 0.001$ ) and distance (scenario 1:  $20.7 \text{ km} \pm 2.9$ ; scenario 2:  $143 \text{ km} \pm 42.0$ ;  
193 scenario 3:  $46.5 \text{ km} \pm 11.1$  on average) between impassable infrastructure across catchments  
194 (ANCOVA:  $F_{2,74} = 23.0$ ,  $p < 0.001$ ) (Fig. 3b). For a passability threshold of 2.0 m, scenarios  
195 2 and 3 showed no difference on average across catchments either in terms of percentage  
196 (scenario 1:  $11\% \pm 1.0$ ; scenario 2:  $17\% \pm 2.0$  on average; Table S1) or distance between  
197 ( $197.0 \text{ km} \pm 60.6$ , and  $160.9 \text{ km} \pm 61.9$  on average; Fig. 3c) impassable built infrastructure,  
198 but both the percentage ( $67\% \pm 4.0$  on average) and distance between impassable built  
199 infrastructure ( $23.3 \text{ km} \pm 3.2$  on average) were significantly different for scenario 1  
200 (ANCOVA:  $F_{2,74} = 148.2$  and  $F_{2,74} = 37.9$ ,  $p < 0.001$ ; see Table S1; Fig. 3c). We found no

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

201 effect of river length (km) on catchment-level river fragmentation regardless of fragmentation  
202 metric or the passability threshold.

203

#### 204 4. Discussion

205

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

211         When large numbers of built infrastructure have unknown dimensions, such as height,  
212 we can be forced to make assumptions; either that all built infrastructure are passable, or  
213 impassable (e.g., Radinger et al., 2017). Our results suggest that these assumptions can result  
214 in opposite outcomes for measures of river fragmentation for fish species with varied abilities  
215 to pass built infrastructure. For example, assuming that all built infrastructure without height  
216 data were passable resulted in a 5-fold reduction in river fragmentation for species such as the  
217 Three-spined Stickleback (passability threshold  $\geq 1.0$  m) compared to using our predicted  
218 height values to measure distance between impassable built infrastructure. We found the  
219 opposite was true for species like the Twait Shad (passability threshold  $\geq 2.0$  m), where  
220 assuming all built infrastructure without height data were impassable resulted in a 7-fold  
221 increase in river fragmentation compared to using our predicted height values to measure  
222 distance between impassable built infrastructure. Our findings suggest that inclusion of built  
223 infrastructure height, and modelling height where necessary, can help to refine estimates of  
224 river fragmentation for fish species with varied abilities to pass built infrastructure. With an  
225 increased interest in modelling fish species' dispersal abilities (Radinger et al., 2017), and

226 continued efforts to prioritize removal projects using indicators of built infrastructure  
1 passability (Neeson et al., 2015), our approach can be used to improve understanding of built  
2  
3  
4  
5 228 infrastructure impact and inform the identification of priorities for restoring river connectivity  
6  
7 229 to benefit different species.

9  
10 230 Our results demonstrate a first step toward more explicit accountancy of built  
11  
12 231 infrastructure impact on aquatic biodiversity. For example, our approach builds on earlier  
13  
14 232 work by Perkin & Gido (2012) who noted that infrastructure passability for different fish  
15  
16 233 species could be a function of both structure height and local hydrological regimes but did  
17  
18  
19 234 not explicitly account for such factors and instead assumed partial passability for all  
20  
21  
22 235 infrastructure. Refinements to our modelling approach that explicitly consider species'  
23  
24 236 biological characteristics, which can influence their ability to pass built infrastructure, would  
25  
26 237 likely further improve estimates of river fragmentation for individual species, but such data  
27  
28  
29 238 are not broadly available. We were able to account for a coarse estimation of river hydrology  
30  
31  
32 239 in our catchment-level fragmentation calculations, because hydrologic variability was  
33  
34 240 integrated in the passability thresholds established by Baudoin et al. (2014). Finer-scale data  
35  
36 241 on river discharge at individual infrastructure is currently not available, but explicit  
37  
38  
39 242 consideration of this factor would be useful in future iterations of this work. We emphasize  
40  
41 243 that our models specifically address a need for overcoming gaps in knowledge about built  
42  
43 244 infrastructure height, and additional considerations such as discharge and fish species'  
44  
45  
46 245 biological characteristics will only help to refine our modelling and findings. Further,  
47  
48  
49 246 mismatches in existing spatial data products did not allow us to predict height values for all  
50  
51 247 built infrastructure in France, and factors such as fish passage facilities that we were unable  
52  
53 248 to account for in our assessment, could influence whether or not these are passable for  
54  
55  
56 249 different fish species. Uncertainty in infrastructure status and presence of fish passage  
57  
58 250 facilities could be validated using a combination of finer-scale spatial data and field surveys.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

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

256 Globally, built infrastructure removal and installation is occurring simultaneously  
257 (Hydropower Status Report 2017; Dam Removal Europe 2016) and methods similar to what  
258 we present here offer a starting point for improving our ability to quantify costs and benefits  
259 associated with these processes. Our results (i.e., known and predicted height values) could  
260 be integrated in conservation planning exercises, along with other ecological and socio-  
261 economic considerations, as a relative indicator of cost to remove built infrastructure. Built  
262 infrastructure height can also be used as an indicator of environmental benefit, such as  
263 downstream response to removal, where higher dams have been shown to have longer-lasting  
264 and more wide-spread downstream effects than shorter dams (Major et al., 2017). These  
265 examples demonstrate the wide-applicability of our approach and results to informing  
266 conservation decisions with broader considerations than fishes. Further, our approach could  
267 be used to inform future scenarios that consider how built infrastructure change over time  
268 with respect to removal, installation and other environmental and socio-political factors, such  
269 as changing climate and flows, and placement of fish passage facilities to reduce impact. We  
270 see particular relevance of our approach to other areas in Europe as well as North America  
271 where efforts to both remediate (in the form of including fish passage facilities) and remove  
272 built infrastructure is rapidly occurring (Foley et al., 2017; Dam Removal Europe 2016) but  
273 where gaps in data on characteristics of built infrastructure remain (e.g., Radinger et al.,  
274 2017; Januchowski-Hartley et al., 2013) and limit our ability to make effective decisions. We  
275 see further applicability of our modelling approach and results to other parts of the world as a



276 global proliferation of smaller infrastructure continues with limited consideration or  
277 documentation of characteristics like height (Couto & Olden, 2018). Ultimately, as global  
278 change continues, approaches like ours will become increasingly important for guiding more  
279 proactive and effective strategies for built infrastructure management.

280

### 281 **Author contributions**

282

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

286

### 287 **Acknowledgements**

288

289 This work was funded by the French Agency for Biodiversity and Research Foundation for  
290 Biodiversity Biodiversa project ODYSSEUS. Laboratoire Evolution et Diversité Biologique  
291 is part of “LABEX TULIP” and “LABEX CEBA” (ANR-10-LABX-41, ANR-10-LABX-  
292 25-01). SRJ also acknowledges funding from the Welsh European Funding Office and  
293 European Regional Development Fund under project number 80761-SU-140 (West). We  
294 thank F. Januchowski-Hartley and two anonymous reviewers for useful comments on earlier  
295 versions of this manuscript.

296

### 297 **References**

298

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

- 301 Bhatt S, Gething PW, Brady OJ, *et al.* 2013. The global distribution and burden of dengue.  
302 *Nature* **496**: 504-507.
- 303 Carvajal JD, Januchowski-Hartley SR, Maldonado-Ocampo JA, *et al.* 2017. Damming  
304 fragments species' ranges and heightens extinction risk. *Conserv Lett* **10**: 708-716.
- 305 Cote D, Kehler DG, Bourne C and Wiersma YF. 2009. A new measure of longitudinal  
306 connectivity for stream networks. *Landsc Ecol* **24**: 101-113.
- 307 Couto TBA and Olden JD. 2018. Global proliferation of small hydropower plants – science  
308 and policy. *Front Ecol Environ* **16**: 91-100.
- 309 Dam Removal Europe. 2016. What is Dam Removal Europe? <http://damremoval.eu/> Viewed  
310 1 May 2017.
- 311 De'ath G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* **88**: 243-251.
- 312 Elith J, Leathwick JR, and Hastie T. 2008. A working guide to boosted regression trees.  
313 *J Anim Ecol* **77**: 802–813.
- 314 ESRI. 2015. ArcGIS 10.3.1. Environmental Systems Research Institute, Redlands, CA.
- 315 Foley MM, Magilligan FJ, Torgersen CE. *et al.* 2017. Landscape context and the biophysical  
316 response of rivers to dam removal in the United States. *PLoS One* **12**: e0180107.
- 317 Hain EF, Kennen JG, Caldwell PV. *et al.* 2017. Using regional scale flow-ecology modelling  
318 to identify catchments where fish assemblages are most vulnerable to changes in  
319 water availability. *Fresh Biol* **63**: 928-945.
- 320 Hijmans RJ, Phillips S, Leathwick J, and Elith J. 2016. Dismo: species distribution modeling.  
321 R package version 1.0–15. See <https://CRAN.R-project.org/package=dismo>.
- 322 Hothorn T, Bretz F, and Westfall P. 2008. Simultaneous inference in general parametric  
323 models. *Biom J* **50**: 346-363.
- 324 Hydropower Status Report. 2017. [https://www.hydropower.org/2017-hydropower-status-](https://www.hydropower.org/2017-hydropower-status-report)  
325 [report](https://www.hydropower.org/2017-hydropower-status-report). Viewed 1 June 2017.

- 326 Januchowski-Hartley,SR, McIntyre PB, Diebel M. *et al.* 2013. Restoring aquatic ecosystem  
1 connectivity requires expanding inventories of both dams and road crossings. *Front*  
2  
3 327  
4  
5 328 *Ecol Environ* **11**: 211-217.
- 6  
7 329 Januchowski-Hartley, SR, Diebel, M, Doran PJ, and McIntyre PB. 2014. Predicting road  
8  
9 330 culvert passability for migratory fishes. *Divers Distrib* **20**: 1414-1424.
- 10  
11 331 Major JJ, East AE, O'Connor JE. *et al.* 2017. Geomorphic responses to dam removal in the  
12  
13 332 United States - a two-decade perspective. In: Gravel-bed rivers: processes and  
14  
15 333 disasters. Oxford: John Wiley & Sons Ltd.
- 16  
17 334 Mantel SK, Rivers-Moore N, and Ramulifho P. 2017. Small dams need consideration in  
18  
19 335 riverscape conservation assessments. *Aquatic Conserv: Mar Freshw Ecosyst* **27**: 748-  
20  
21 336 754.
- 22  
23 337 Neeson TM, Ferris MC, Diebel MW, *et al.* 2015. Enhancing ecosystem restoration efficiency  
24  
25 338 through spatial and temporal coordination. *Proc Natl Acad Sci* **112**: 6236-6241.
- 26  
27 339 Pella H, Lejot J, Lamouroux N, and Snelder TH. 2012. Le réseau hydrographique théorique  
28  
29 340 (RHT) français et ses attributs environnementaux. *Geomorphologie* **3**: 317-336.
- 30  
31 341 Perkin JS and Gido KB. 2012. Fragmentation alters stream fish community structure in  
32  
33 342 dendritic ecological networks. *Ecol Appl* **22**: 2176-2187.
- 34  
35 343 Poff LN and Hart DD. 2002. How dams vary and why it matters for the emerging science of  
36  
37 344 dam removal. *BioScience* **52**: 659-667.
- 38  
39 345 Radinger J, Essl F, Hölker F, *et al.* 2017. The future distribution of river fish: the complex  
40  
41 346 interplay of climate and land use changes, species dispersal and movement barriers.  
42  
43 347 *Global Change Biol* <https://doi.org/10.1111/gcb.13760>.
- 44  
45 348 Soykan CU, Eguchi T, Kohin S, and Dewar H. 2014. Prediction of fishing effort distributions  
46  
47 349 using boosted regression trees. *Ecol Appl* **24**: 71-83.
- 48  
49 350

351 **Figure captions**

352

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

357

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

361

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

376

1

2

3

377

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

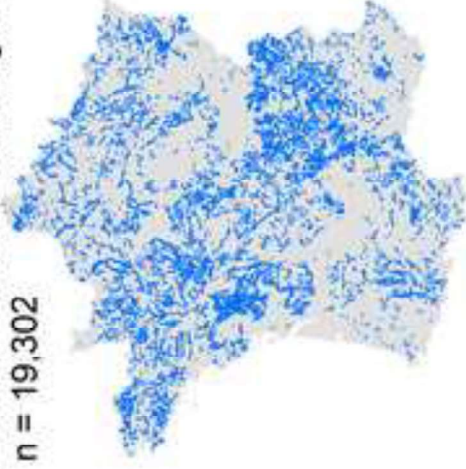
65

Figure 1  
[Click here to download high resolution image](#)



(b)

Built infrastructure with height data  
 $n = 19,302$



Built infrastructure without height data  
 $n = 20,077$

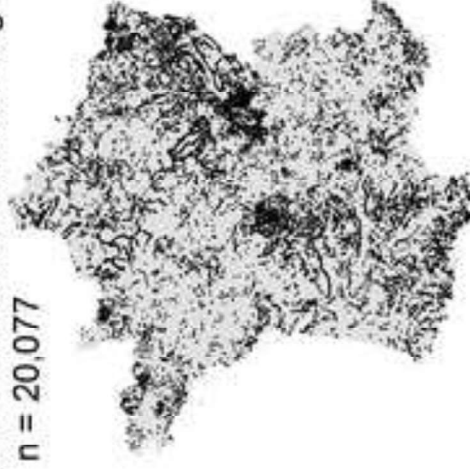


Figure 2  
[Click here to download high resolution image](#)

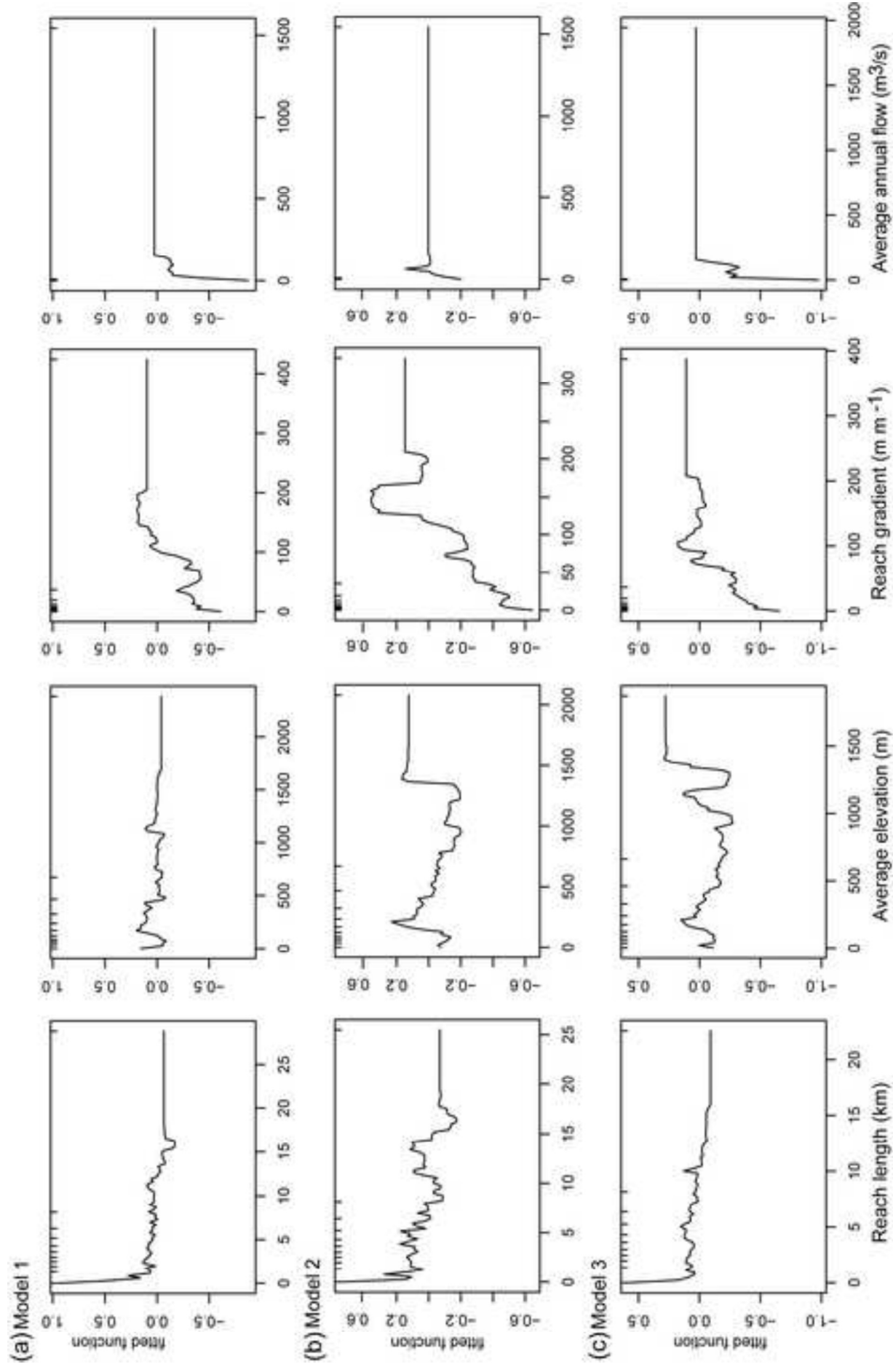
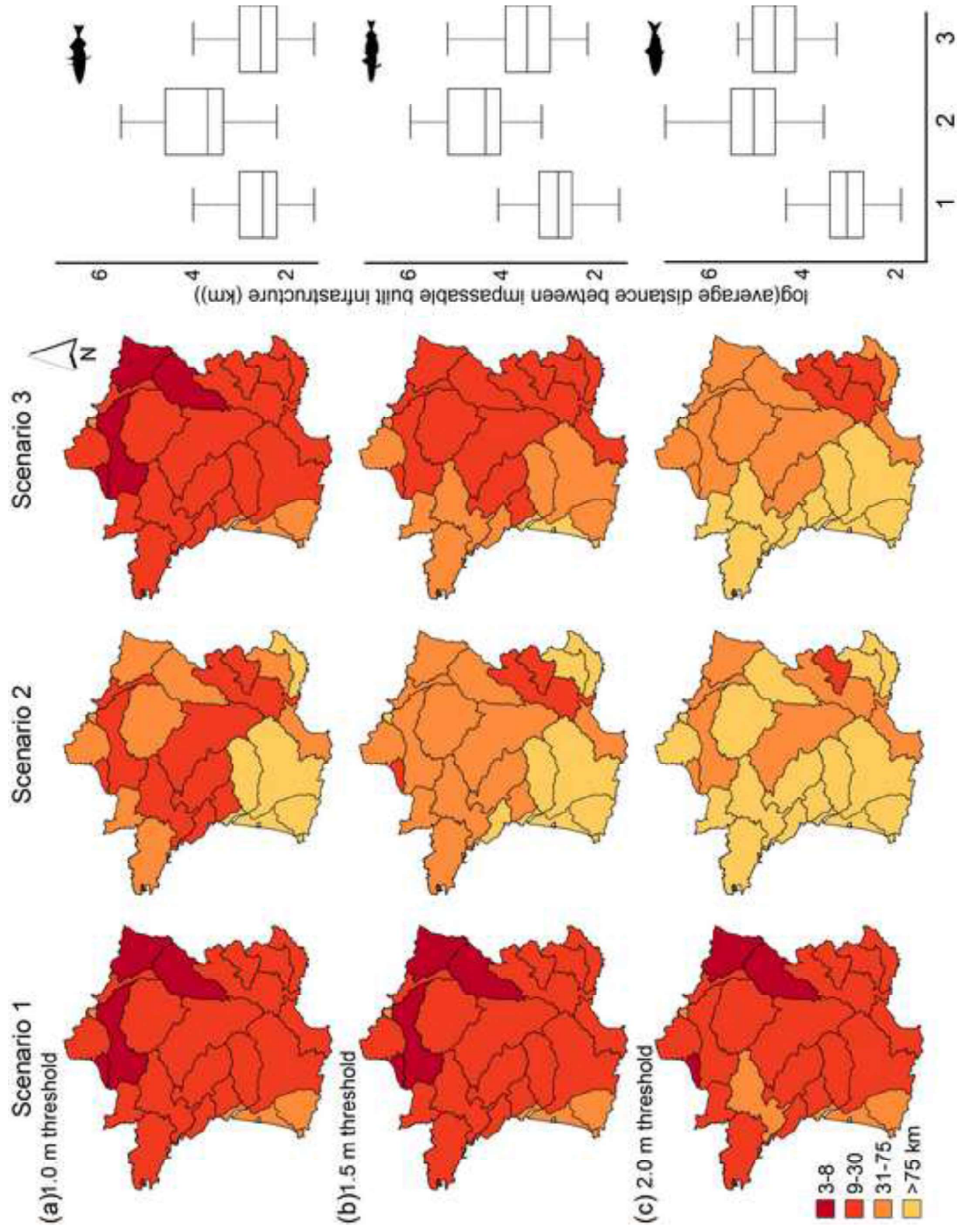


Figure 3  
[Click here to download high resolution image](#)





**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.

<b>Environmental Variable</b>	<b>Description</b>
Stream reach gradient ( $\text{m m}^{-1}$ )	Stream reach gradient where built infrastructure is located.
Average monthly minimum flow ( $\text{m}^3/\text{s}$ )	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 ( $\text{m}^3/\text{s}$ )	Average annual flow of a stream reach where built infrastructure is located.
Stream reach drainage area ( $\text{km}^2$ )	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 ( $\text{km}^2$ )	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**[Click here to download Table: Table\\_2.docx](#)

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

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

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

<b>Model</b>	<b>Number of records</b>	<b>Environmental variables</b>	<b>Training data correlation (based on 5000 sites)</b>	<b>Cross validation correlation</b>	<b>Cross validation standard error</b>	<b>Number of trees</b>
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.

<b>Environmental variable</b>	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model average</b>
Stream reach length (km)	18%	19%	19%	19%
Stream reach average elevation (m)	14%	15%	14%	14%
Stream reach gradient (m m <sup>-1</sup> )	12%	12%	14%	13%
Average annual flow (m <sup>3</sup> /s)	12%	12%	12%	12%
Percentage agriculture cover (%)	10%	10%	10%	10%
Stream reach drainage area (km <sup>2</sup> )	10%	9%	9%	9%
Average monthly minimum flow (m <sup>3</sup> /s )	9%	9%	9%	9%
Stream reach upstream drainage area (km <sup>2</sup> )	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 ( $\geq 1.0$ , 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_5m_scen1 (%)	Impass_1m_scen2 (%)	Impass_5m_scen2 (%)	Impass_1m_scen3 (%)	Impass_5m_scen3 (%)	Impass_1m_scen3 (%)	Impass_5m_scen3 (%)	Impass_1m_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%