ENVIRONMENTAL RESEARCH
LETTERS

PURPOSE-LED PUBLISHING™

Part of the
ENVIRONMENTAL RESEARCH
SERIES

LETTER • OPEN ACCESS

Modelling remote barrier detection to achieve free-flowing river targets

To cite this article: Millicent V Parks *et al* 2024 *Environ. Res. Lett.* **19** 084055

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
17 October 2023

REVISED
18 June 2024

ACCEPTED FOR PUBLICATION
17 July 2024

PUBLISHED
1 August 2024

LETTER

Modelling remote barrier detection to achieve free-flowing river targets

Millicent V Parks[1], **Carlos Garcia de Leaniz**[1,2,*], **Peter E Jones**[1,3] **and Josh Jones**[1,4]

[1] Department of BioSciences, Centre for Sustainable Aquatic Research (CSAR), Swansea University, Swansea, United Kingdom
[2] Marine Research Centre (CIM-UVIGO), Universidade de Vigo, Vigo, Spain
[3] Natural Resources Wales, Swansea, United Kingdom
[4] The Rivers Trust, Callington, United Kingdom
[*] Author to whom any correspondence should be addressed.

E-mail: c.garciadeleaniz@swansea.ac.uk

Keywords: river restoration, barrier prioritization, remote sensing, river fragmentation, barrier detection, modelling

Supplementary material for this article is available online

Abstract

Fragmentation caused by artificial barriers is one of the main stressors of rivers worldwide. However, many barrier inventories only record large barriers, which underestimates barrier numbers, and hence fragmentation. Corrected barrier numbers can be obtained via river walkovers, but these are costly and time consuming. We assessed the performance of remote sensing as an alternative to river walkovers for barrier discovery by comparing the number and location of barriers detected in the field with those detected using Google Earth imagery. Only 56% of known barriers could be detected remotely, but machine learning models predicted the likelihood of remote detection with 62%–65% accuracy. Barriers located downstream were twice as likely to be detected remotely than those in the headwaters, the probability of detection diminishing by 3%–4% for every decrease in Strahler stream order and for every 10 km increase in distance from the river mouth. Barriers located in forested reaches were 35% less likely to be detected than those in open reaches. Observer skills also affected the ability to locate barriers remotely and detection rate varied by 11% between experienced and less experienced observers, suggesting that training might improve barrier detection. Our findings have implications for estimates of river fragmentation because they show that the most under-represented structures in barrier inventories, i.e. small barriers located in forested headwaters, are unlikely to be detected remotely. Although remote sensing cannot fully replace 'boots on the ground' field surveys for filling barrier data gaps, it can reduce the field work necessary to improve barrier inventories and help inform optimal strategies for barrier removal under data-poor scenarios.

1. Introduction

To meet the targets of the EU Green Deal Agenda, at least 25 000 km of rivers will need to be reconnected and be made free-flowing by 2030 [1], which will require detailed knowledge of the extent of fragmentation and the location of instream barriers that can be removed or mitigated. However, restoration of river connectivity can be hampered by epistemic uncertainty caused by data deficiencies and imperfect knowledge of barrier numbers [2], as well as by ontological uncertainty resulting from the inherent variability and unpredictability of river systems [3]. Many official barrier databases are outdated and incomplete [4–6], but the degree of barrier under-reporting varies widely from place to place and also depends on barrier size. For example, 68% of barriers in Europe's rivers are less than 2 m in height and are thought to be underestimated by ∼75% in official barrier inventories, compared to large dams whose number is underestimated by ∼12% [4]. Because the impact of barriers on river fragmentation is determined mostly by the number and location of barriers, not by their height [2, 7], accurate estimates of barrier numbers (both large and small) are essential for maximizing the benefits of restoring connectivity.

Synoptic barrier inventories have tended to focus on large and medium size dams that create reservoirs detectable from satellite images [8, 9] but this can greatly underestimate the true extent of fragmentation [4, 10–12] because most barriers are not ponding structures. For example, in Europe there are ∼0.7 M unique longitudinal barrier records, but the true number is estimated to exceed 1.2 million, and could be as high as 3.7 million barriers [4], as small structures are difficult to locate and are grossly under-represented. Low-head barriers tend to have lower per capita impacts on rivers than large dams, but they are typically much more numerous [5, 13] and their cumulative impacts on river fragmentation are typically greater [14–16], an example of 'death by a thousand cuts' [2].

Efforts to remove barriers and restore river connectivity can be ineffectual if barrier numbers are underestimated [2]. For example, although more than 150 dams and weirs have been removed in Spain over the last decade this has not improved connectivity in any significant way [17]. Fragmentation estimates based on incomplete barrier data could result in suboptimal decisions on barrier prioritization, as they will typically underestimate the true extent of fragmentation and exaggerate the predicted gains accrued from barrier removal. If left uncorrected, these efforts are unlikely to achieve free-flowing river targets. Incomplete barrier data can still be useful to plan river restoration programs, but only if the subjacent biases are known and can be accounted for explicitly [2, 16, 18, 19].

River walkovers provide the most reliable way of locating missing barriers and assessing the true extent of fragmentation [4–6, 10, 11], but these are labour intensive and can be too expensive to undertake at large spatial scales. Alternative ways of locating missing barriers are needed. Remote sensing has proved useful in remote areas and can greatly increase the completeness of barrier inventories at a fraction of the cost of river walkovers [20–22], but its accuracy in barrier discovery has rarely been tested. Estimating the reliability of remote sensing for barrier discovery is particularly important for small barriers as these are the most numerous, and are generally located in the headwaters [23, 24] where remote detection could be more challenging.

Our objective was to determine the feasibility and limitations of using remote sensing for identifying instream barriers, as an alternative to costly 'boots on the ground' field surveys. To this end, we assessed to what extent barriers identified via river walkovers could be detected using Google Earth, modelled barrier detection, and estimated the associated errors. Our ultimate aim was to reduce epistemological and operational uncertainties regarding river fragmentation in order to facilitate progress on free-flowing rivers targets.

2. Methods

We defined longitudinal barriers as 'any built structure that interrupts or modifies the flow of water, the transport of sediments, or the movement of organisms and can cause longitudinal discontinuity'; these can be classified into six main types based on key features and the extent of habitat modification [2, 4, 25]: dam, weir, sluice, ramp, ford, and culvert. There were no sluices in our data set, and we excluded five dams as these were too few for analysis.

2.1. Ground truthing remote barrier detection with field data from river walkovers

We used Google Earth (version 9.136.0.2) for remote barrier detection because this service is free, easy to use, and provides convenient access to stable, high-spatial resolution imagery [26], representing the most popular tool for remote detection of river infrastructures [4, 11, 26–29]. Google Earth imagery is built from multiple datasets and its resolution typically varies from ∼10–30 m px^{-1} for satellite imagery [30] to ∼0.15–0.50 m px^{-1} for aerial photography [31].

We used two types of datasets ('exhaustive' and 'sample') to model the probability of remote detection. An 'exhaustive' dataset was obtained by conducting a full walkover of an entire catchment in Wales (River Afan, 171 km of total river length; figure 1) and recording and photographing all artificial barriers present ($n = 279$) using the protocol described in [4, 5]. Three observers working independently (and without information on the number or location of the barriers detected in the field) then used Google Earth to systematically scan through the catchment and record the coordinates of all visible barriers. We chose a 500 m eye (camera) altitude to detect barriers, as this generally resulted in the best image quality and reduced variation between observers, but barriers were zoomed at a closer eye altitude (typically 100–300 m) to help determine barrier type (figure S1). To ascertain barrier detection, a 10 m buffer radius was used to snap barrier coordinates to the river network [29], and a matrix of pairwise haversine (great circle) distances [32] was calculated between the coordinates of barriers located in Google Earth and those detected in the field. The Google Earth barrier located closest (within a 100 m radius) of a barrier detected in the field was considered to be most likely match.

In addition to the exhaustive river walkover, a second, 'sample' dataset was used consisting of 208 barriers that had been recorded during partial (i.e. incomplete) river walkovers of the principal salmon rivers of Wales during 2018–2021 [33]. Barrier coordinates were then uploaded into Google Earth and an attempt was made to locate the barriers remotely, using the same procedure as above.

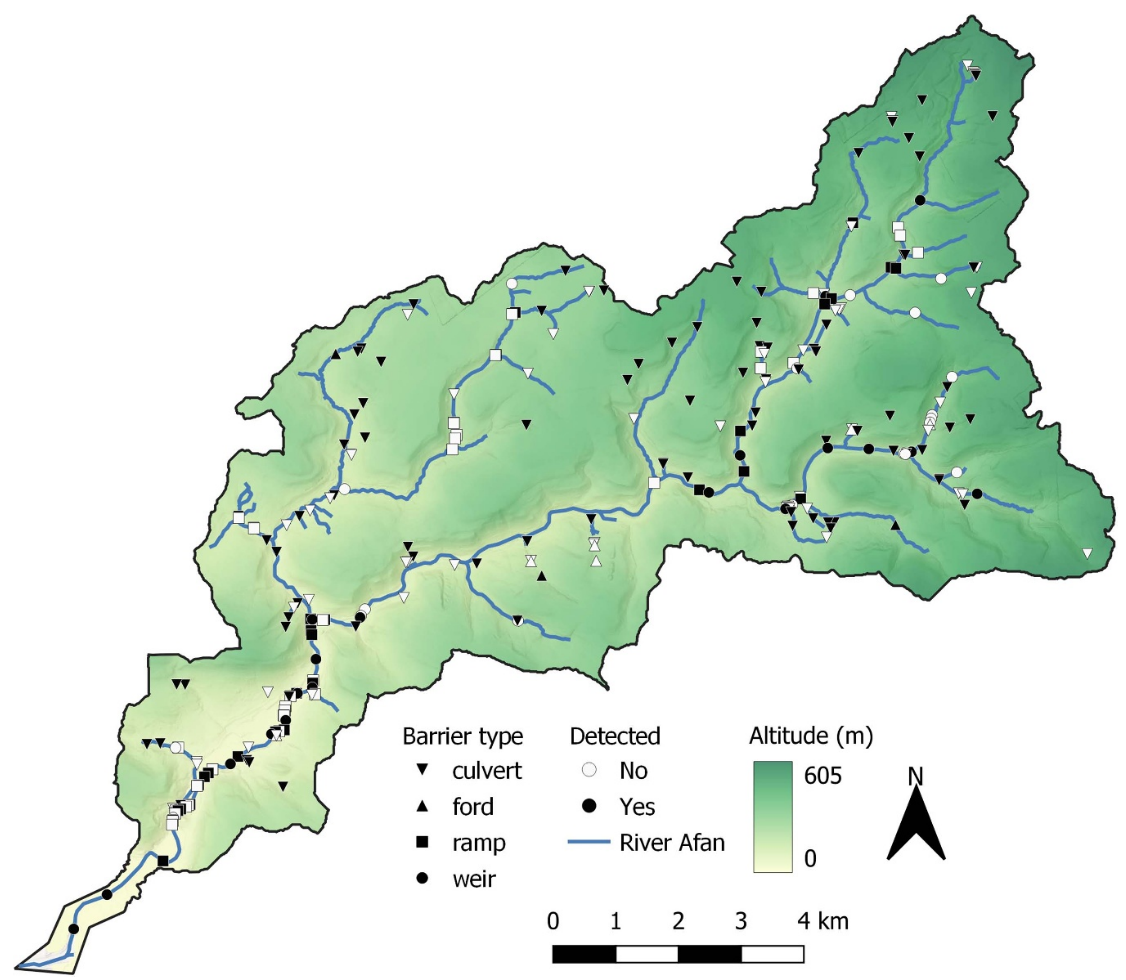

Figure 1. Location of the four main barrier types (culvert, ford, ramp, weir. $N = 260$) found in the complete walkover of the River Afan catchment showing whether they were detected (closed fill) or undetected (open fill) in Google Earth.

Therefore, each of the 487 barriers detected in the field during the river walkovers was cross-referenced with Google Earth imagery. However, while the exhaustive field dataset contained all the barriers present in a catchment, the sample data set contained only a sample of barriers present in some river reaches.

2.2. Barrier detection performance metrics

To estimate the performance of remote barrier detection we calculated four classification metrics and associated 95% confidence intervals (95 CIs) [34, 35]. We calculated 'detection rate' as the proportion of each barrier type that was detected remotely using Google Earth (regardless of whether they were classified into the correct barrier type or not). 'Sensitivity' measures the likelihood that a barrier would be detected remotely when a barrier existed in the field (true positive) and was calculated as the ratio of true positives to the sum of true positives (tp) and false negatives (fn). 'Specificity' measures the likelihood that a barrier would be detected remotely when there was no barrier (false positive) and was calculated as the ratio of true negatives (tn) to the sum of true negatives (tn) and false positives (fp). 'Accuracy' refers to the likelihood that a barrier would be classified correctly by remote sensing. Perfect barrier detection would yield 100% accuracy, detecting all true barriers (100% sensitivity), without any false positives (100% specificity).

2.3. Repeatability of barrier detection

To assess the repeatability of barrier detection using Google Earth (i.e. to what extent barrier detection depended on the skills of the observer) we calculated detection agreement between three raters using the exhaustive data set and the S-statistic in the '*raters*' R-package [36], along with bootstrapped 95 CIs. The S-statistic is a modification of the Fleiss–Kappa agreement which is suitable for nominal data like ours (i.e. barriers were 'detected' or 'undetected') and is unaffected by paradoxes that sometimes introduce biases in the Kappa index [37].

2.4. Modelling barrier detection via machine learning

We modelled remote barrier detection via binary logistic regression (yes/no) and Maximum

Likelihood, using 4 main barrier types [2, 4], survey type (exhaustive/sample), forest cover (yes/no), stream order (Strahler), barrier height (1–6), altitude (m), and distance to mouth (m) as predictors. Strahler, distance to mouth, and altitude of each barrier were calculated from the EU Digital Elevation Model (DEM) E30N30 and were extracted using 'fill sinks' [38] and the 'Strahler order' tool in the SAGA toolkit, QGIS version 3.10 [39]. Barrier height was grouped into six height classes: <0.5 m, 0.5 m–1 m, 1 m–2 m, 2 m–5 m, 5 m–10 m and >10 m, as used in the pan-European AMBER atlas of stream barriers [4]. There is no standard definition of how riparian buffers should be defined from remote imagery [40], so we adopted a simple operational definition whereby barrier locations with 5 or more trees within a 10 m radius of either bank were classified as 'forested', or as 'open' otherwise.

Continuous variables were scaled and centred before analysis. We used the '*glmulti*' R package [41] with an exhaustive search to evaluate all main effect models, using changes in AICc and the *anova* command to compare model fits. Standardized parameter estimates were obtained by fitting the most plausible model with an explicit intercept ($\sim$1) on a standardized version of the dataset. The Wald z-distribution approximation was used to compute 95% confidence intervals and significance values. The accuracy of model predictions was assessed by the area under curve (AUC) method using 10 000 bootstrap samples. Model assumptions were evaluated with the '*performance*' R package [42].

We complemented the analysis by logistic regression with two other machine learning methods: random forest and adaptive boosted trees using the '*randomForest*', '*adabag*' and '*rfPermute*' *R* packages. Logistic regression is useful because it provides interpretable coefficients that quantify the effects of each predictor on the binary response variable, but can have low predictive power when there are outliers, correlated predictors, and/or non-linear relationships [43], as was the case in our study. Random forest and adaptive boosted trees, on the other hand, do not provide interpretable coefficients, but can handle non-linearity, and are less affected by outliers and correlated predictors, typically resulting in better predictive performance [44]. Combining logistic regression with other machine learning algorithms can maximize predictive power and provide additional insights into remote sensing applications [45].

3. Results

3.1. Consistency in barrier detection from remote sensing

Consistency in barrier detection on the exhaustive data set between three observers using Goggle Earth was $S = 0.66$ (95 CI = 0.59–0.72, $P < 0.001$), which can be regarded as 'substantial' [46].

3.2. Metrics of barrier detection performance

Observers were able to detect on average 50% of the barriers found in the complete river walkover of the River Afan using images, but this varied between 47% and 55% depending on the individual observer, and approximately one quarter (26%) of barriers were not detected by any of the three observers (figure S2). Barrier detection and sensitivity were higher for barriers at river-road crossings (culverts and fords) than for other barriers (ramps and weirs), but at a cost of lower specificity and accuracy (figure 2). For example, sensitivity was 87% for river-road crossings compared to 69% for other barriers, while specificity was much higher for other barriers (92%) than for barriers at river-road crossings (18%).

Using both data sets (i.e. the exhaustive data set and the sample data set from partial river walkovers) barrier detection was 56%.

3.3. Predictors of barrier detection

We modelled the probability of detection of 468 barriers with complete data ($n = 260$ from the full survey and $n = 208$ from the partial surveys). We ran all 128 possible main-effects models with the seven predictors (i.e. 2^7), as there was no compelling reason to include some predictors but not others. The most plausible model contained five predictors (table S1) that were statistically significant according to an analysis of deviance (table 1). Three predictors (barrier type, stream order and forest cover) reached over 95% average model importance and two predictors (survey type and distance to mouth) reached over 75% average model importance (figure 3(a)). Although the model predictive power was low (R^2 Tjur $= 0.09$, $P < 0.001$), it was able to correctly predict whether a barrier could be detected remotely with 62% accuracy (95 CI = 57%–67%). The R^2 Tjur is a coefficient of discrimination for binary logistic regression [47] and represents the difference between the average fitted probability of the two logistic responses (in this case detected vs undetected barriers). Using two other ML approaches (figures 3(b) and (c)) did not improve model accuracy significantly (random forest accuracy = 57%; CI = 48%–65%; adaptive boosting accuracy = 65%, CI = 60%–69%), although the relative importance of the main predictors varied (figure 3).

Using the logistic model, the probability of barrier detection decreased significantly with increasing distance from the river mouth (figure 4(a), likelihood ratio test, LRT = 4.59, df = 1, $P = 0.032$) and forest cover (figure 4(e), LRT = 7.76, df = 1, $P = 0.005$), but increased with stream order (figure 4(b), LRT = 22.84, df = 5, $P < 0.001$) and varied also with barrier type (figure 4(c)), LRT = 20.56, df = 3, $P < 0.001$) and survey type (figure 4(d), LRT = 3.87, df = 1, $P = 0.049$).

Inspection of odds ratios (figure 5) indicated that barriers were twice more likely to be detected near the river mouth (P-detection = 0.70) than in the

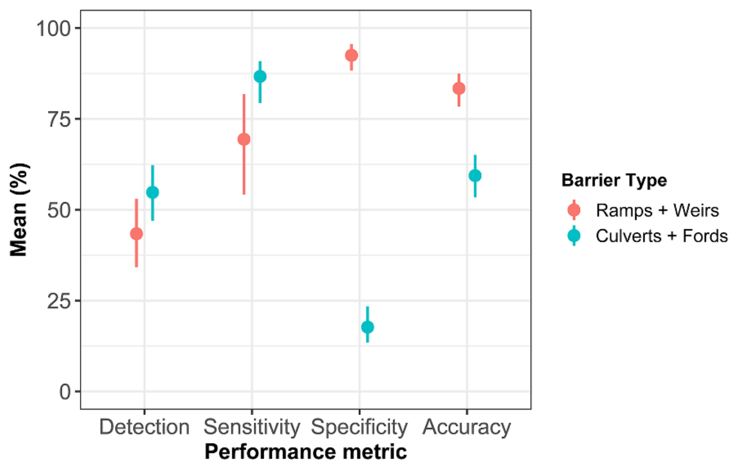

Figure 2. Performance metrics (%) of remote detection of barriers at river-road crossings (culverts and fords) and other barriers (ramps and weirs) found in an exhaustive river walkover. The results show the means (±95 CI) of three raters working independently.

Table 1. Results of single-term deletions using the drop1 command (stats R package) comparing the full model with all seven predictors with a reduced model where each predictor was removed in turn. Five significant predictors are highlighted in bold.

Predictor	df	Deviance	AIC	LRT	p-value
Full model (none dropped)		599.11	627.11		
Stream Order (Strahler)	5	621.95	639.95	22.84	**< 0.001**
Barrier Type	3	619.66	641.66	20.56	**< 0.001**
Forest Cover	1	606.86	632.86	7.76	**0.005**
Survey Type	1	602.97	628.97	3.87	**0.049**
Scale (Dist. to mouth)	1	603.70	629.70	4.59	**0.032**
Scale (Altitude)	1	599.17	625.17	0.07	0.796
Scale (Height class)	1	599.29	625.29	0.18	0.670

LRT = likelihood ratio test.

headwaters (P-detection = 0.33, odds-ratio = 0.78, $P = 0.032$), the probability of detection decreasing by ∼3% for every 10 km increase in distance moved upstream (figure 4(a)). This is also evident when results are analysed by stream order (figure 4(b)). As one moves upstream from sixth-order reaches (P-detection = 0.90) to first order streams (P-detection = 0.64) the probability of detection diminishes by about 4% for every single decrease in stream order (odds-ratio = 5.34, $P < 0.01$). Barrier type was an important predictor of barrier detection (figure 4(c)). Barriers at river-road crossings such as culverts (P-detection = 0.63) and fords (P-detection = 0.48) were significantly more likely to be detected remotely than ramps (P-detection = 0.33) or weirs (P-detection = 0.34, odds ratio = 0.28, $P < 0.001$, figures 4(c) and S2). Barriers detected in the exhaustive river walkover were less likely to be detected in Google Earth (P-detection = 0.64) than those found in less stringent, partial river surveys (P-detection = 0.77; odds-ratio = 1.90, $P = 0.049$; figure 4(d)). Forest cover had also a strong negative effect on barrier detection (figure 4(e)), and barriers located in forested reaches (P-detection = 0.45) were 35% less likely to be detected than those located in open reaches (P-detection = 0.64, odds ratio = 0.46, $P = 0.005$).

4. Discussion

Restoration of stream connectivity can be hindered by the existence of unrecorded barriers, as these can reduce the ability to realistically simulate the gains and costs of barrier removal and therefore its effectiveness [2]. Two approaches that could help overcome barrier data deficiencies are river walkovers and remote detection.

River walkovers are arguably the gold standard for ground-truthing barrier inventories [4, 5], but these are costly and time consuming [5, 10, 11], and impractical at large spatial scales [48]. Remote barrier detection offers an alternative to river walkovers for bridging data gaps in barrier inventories [9, 49, 50], but its reliability in barrier discovery has seldom been tested systematically [10]. Our results indicate that 44% of barriers could not be detected remotely, these being mostly small structures <2 m in height. Other studies have also found that only a small proportion of barriers could be detected remotely. For example, in Ireland only 30% of barriers could be

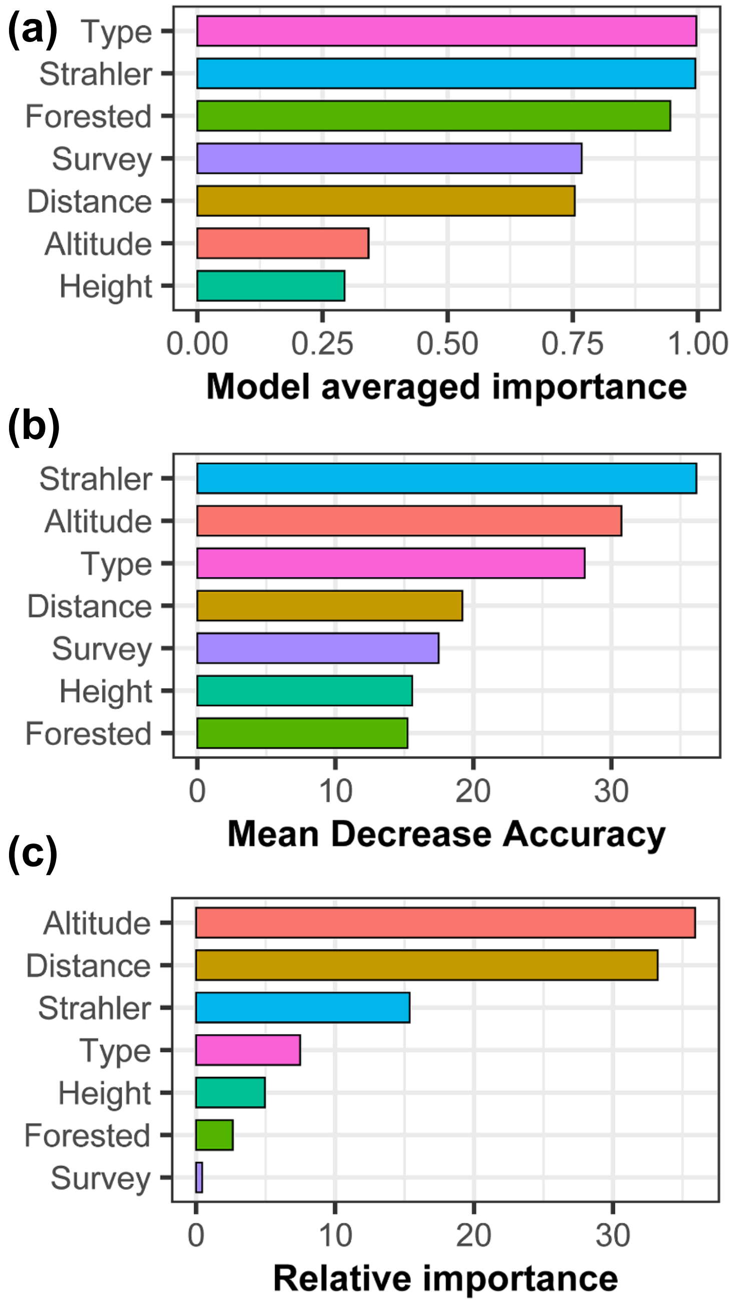

Figure 3. Predictor importance of three machine learning models of remote barrier detection: (a) logistic regression (AUC = 0.62, accuracy = 0.62 ± 0.05); (b) random forest (AUC = 0.70, accuracy = 0.57 ± 0.08), and (c) adaptive boosting (AUC = 0.85, accuracy = 0.65 ± 0.04).

located using remote sensing [11]. In other studies, the probability of barrier detection in different sub-catchments ranged from 0.67 to 1.00 [29], although this consisted of partial barrier surveys (i.e. not exhaustive ones) which, as our study shows, tend to overestimate the probability of detection, presumably because they miss the less conspicuous barriers.

We were able to predict whether a barrier could be detected remotely with 62%–65% accuracy, which is lower than the 73%–89% accuracy reported by others for predicting barrier sites remotely using machine learning [22, 51], perhaps because our barriers were smaller than in other studies. Our modelling indicates that the probability of barrier detection decreases with increasing distance from the river mouth and forest cover, and increases with stream order, varying also with barrier type and survey type, as well as with observers' skills. Barriers located in the main stem and the lower part of the river network were more than twice as likely to be detected remotely than barriers in the headwaters, suggesting that most barriers located in mountain streams might go undetected by remote sensing. The probability of detection in our study diminished by ~4% for every decrease in stream order. A lower probability of barrier detec-

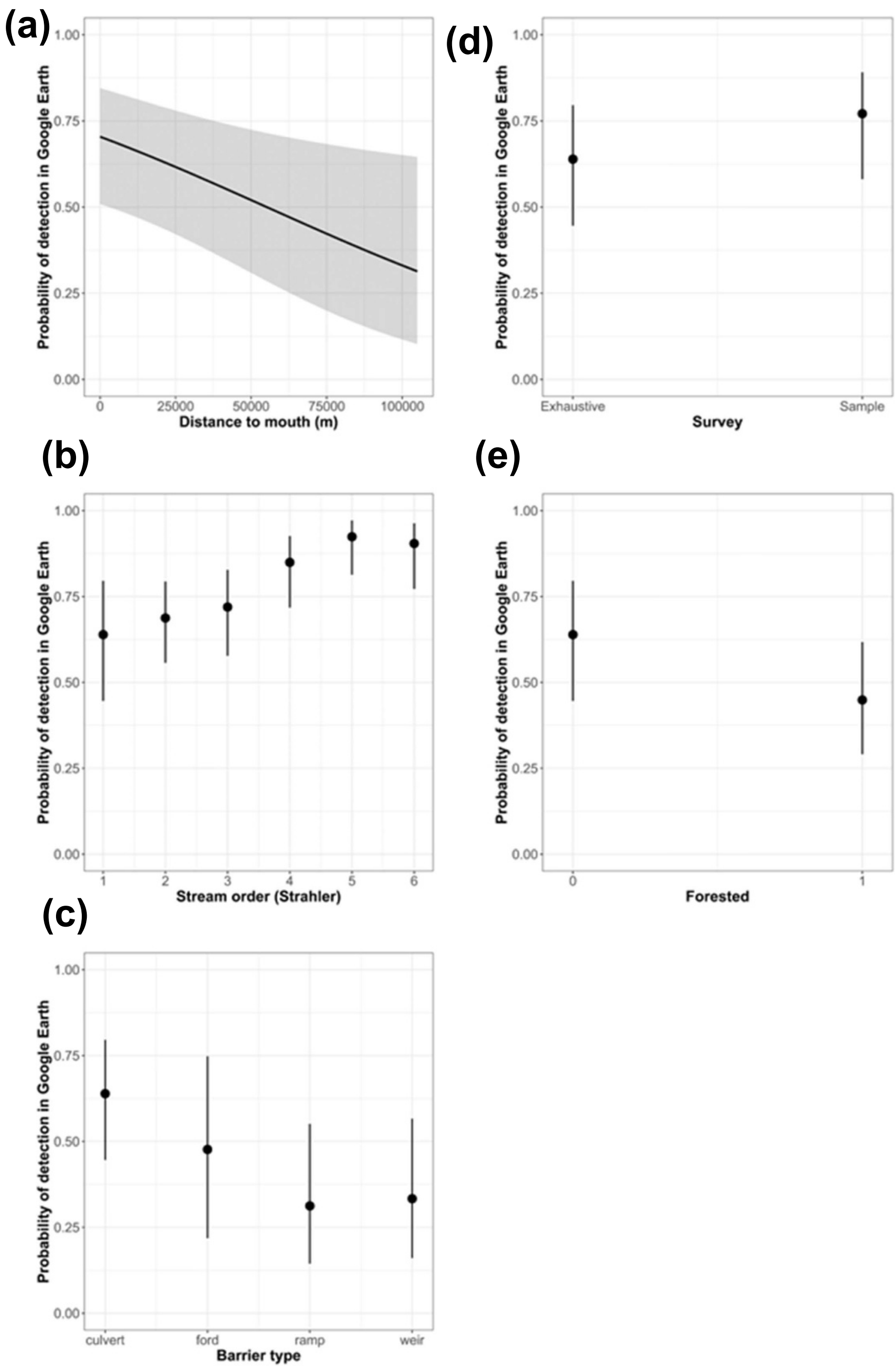

Figure 4. Parameter estimates (±95 CI) of five significant predictors of remote barrier detection using logistic regression.

tion in smaller streams has been reported previously [10, 11, 29], although the marginal effects of potential confounding variables (e.g. altitude, forest cover, and barrier type) had not been addressed.

We found that barriers located in forested stream reaches were 35% less likely to be detected than those located in open reaches. Tree cover has been found to reduce the probability of barrier detection previously [29], as does cloud cover [9]. As expected, barriers located in a complete river walkover were less likely to be detected remotely than those found in less stringent, partial river surveys, where presumably only the most conspicuous barriers would have been reported. This finding has implications for correcting barrier under-reporting errors found in national barrier inventories [4–6], as the value of using remote sensing for bridging barrier data gaps depends on the extent of under-reporting. Unlike most other studies, we were able to capitalize on a complete barrier inventory of a whole catchment to obtain unbiased estimates of barrier detection. Without a complete river survey, determining the probability of barrier detection without bias is difficult, as partial river surveys will inevitably miss some barriers [52]. In our

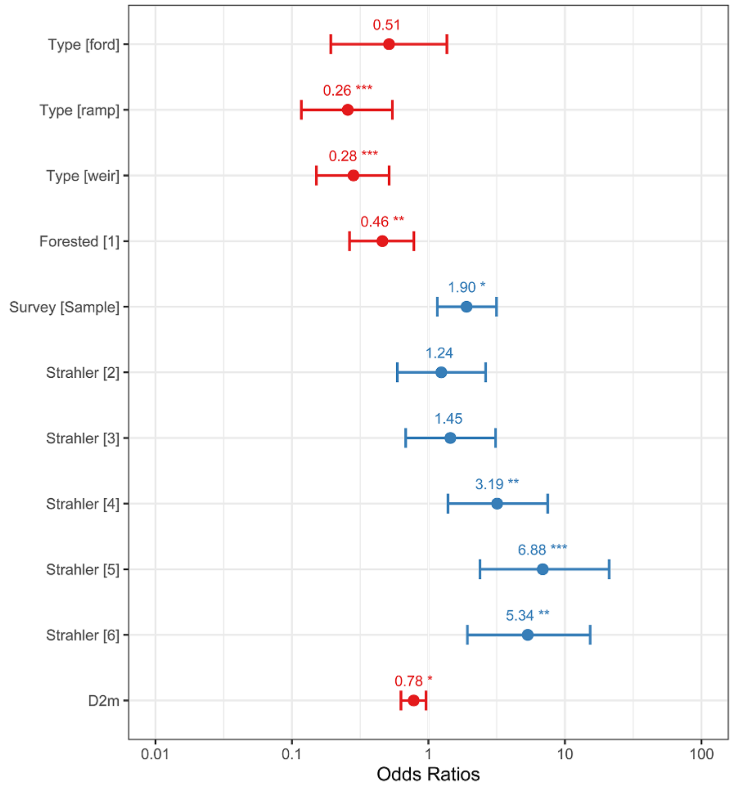

Figure 5. Odds ratios of five significant predictors of remote barrier detection using logistic regression compared to the reference values for barrier type (culvert), survey type (exhaustive), forest cover (0 = open), and stream order (first order). Predictors that increase the probability of barrier detection compared to the reference values are shown in blue, and those that decrease it are shown in red.

study, the probability of barrier detection was overestimated by ∼18% when ground-truthing was based on partial river surveys compared to a complete river walkover.

Barrier type was an important predictor of barrier detection. Barriers are not distributed at random in streams [2], and their locations can be predicted from knowledge of topography, anthropogenic pressures (including the road network) and land use [4]. In general, all types of artificial barriers are under-reported, but large dams are the easiest structures to be detected remotely [8, 9] and the best documented barrier type [5, 53–55]. Even so, almost 30% of dams are probably undetected [9], although their location can be predicted from knowledge of stream slope and upstream impoundments [51]. Collectively, impoundments have altered more than 10% of free-flowing habitat in Europe's rivers [56], although it is the small barriers that have probably caused the biggest cumulative impact on fragmentation [4, 12, 14]. There were few dams in our data set and dams were excluded from analysis, but other studies have shown that dams have the lowest under-reporting error [4, 5].

Barriers at river-road crossings, such as culverts and fords, were significantly more likely to be detected remotely than ramps or weirs in our study, although this was at a cost of lower accuracy and specificity. This finding might seem counterintuitive given their small size, upstream location [21] and under-representation in barrier inventories [4–6, 10]. We attribute this to the facility by which river-road crossings could be readily detected using remote sensing and GIS [20–22, 29]. Fords and culverts might not always be detected from aerial images, but bridges are much easier to see. Therefore, when a bridge is not detected at a river-road crossing, one can reasonably assume that a culvert or a ford is present. Data for England and Wales [22] indicates that ∼ 45% of all river-road crossings consist of culverts and fords (rather than bridges) and therefore constitute potential barriers to free flow, as well as an impediment for the movement of both migratory and resident fish [57]. Culverts rank among the most numerous and under-reported barriers across Europe [4] and North America [22], but their location can be predicted with some accuracy from knowledge of upstream drainage area [22] and their presence can be verified more easily than most other barriers because they are accessible by track or road [10].

Observer skills affected the ability to locate barriers remotely and detection accuracy varied by 11% between experienced and less experienced observers, suggesting that prior training might improve barrier detection. Two additional potential sources of error are the age of Google Earth images [29, 58]—which are typically 1–3 years old and could miss some new barriers—and differences in resolution and level of detail between locations [59], which could affect the probability of detection [29]. The resolution of Google Earth imagery (1 ∼ 0–30 m px^{-1} for satellite imagery [30] and ∼0.15–0.50 m px^{-1} for aerial photography) is similar to that provided by other services such as Planet Scope (3 m globally) [60] or ESA's 10 m Sentinel-1 synthetic aperture radar (SAR) [54]. Although Google Earth does not provide information on each image, the images we used to detect barriers were almost certainly derived from high resolution aerial photography, not from satellite images (figure S2). Satellite images with similar sub-metre resolution are available, but these are only offered as a pay-per-order service by some providers, are not available everywhere, and might be too expensive to be used across watersheds. In addition, improving image resolution alone might not be enough to detect barriers in heavily forested streams, which made up 75% of all the barriers in our study and reduced the probability of detection by ∼35%. The use of Light Detection and Ranging (LiDAR) could help detect barriers through the canopy, but only if barriers have a distinctive surface roughness (to differentiate them from natural rocks) or are at least 2 m in height [50], which would miss more than 68% of instream barriers in Europe [4]. The main advantage of Google Earth is that it is a free service providing high resolution images globally, and this makes it the tool of choice for many river applications [4, 11, 26–29]. Higher resolution could improve remote barrier detection in the future, but more significant gains can probably be obtained through the development of bespoke image analysis routines to identify and classify barriers from existing images, as developed for remote land use classification applications [34].

5. Conclusions

Numerous studies have reported a substantial underestimation of artificial river barriers, highlighting the need for better barrier inventories [5, 6, 9, 52]. Remote detection of barriers can help overcome data deficiencies, but our findings indicate that many barriers are not currently detectable from aerial or satellite images. Remote sensing, therefore, cannot fully replace 'boots on the ground' river walkovers for filling barrier data gaps and identifying barriers to free flow [20, 61], but can make their discovery more efficient. Thus, because the power to detect barriers remotely varies in predictable ways—being high for downstream barriers spanning across large rivers, and low for upstream barriers in forested streams—such knowledge can be used to improve strategies for bridging barrier data gaps. For example, knowledge of how different barriers are distributed in the stream, and their associated probabilities of detection, could be used to target costly river walkovers where they are most needed, and help derive more realistic measures of fragmentation that account for incomplete information [12, 51]. In this sense, the location of river-road crossings—readily obtained from existing digital maps or aerial imagery- can be used to account for the presence of fords and culverts with a higher degree of confidence [22, 48, 52]. Globally, these are the barriers most likely to be under-reported in existing barrier inventories [4], so addressing this knowledge gap is paramount for more accurate estimates of stream fragmentation and the prioritization of barrier removal [2]. More generally, our study provides a better understanding of the power and shortcomings of using remote sensing for filling barrier data gaps in estimates of river fragmentation, and can help to inform optimal strategies for barrier removal under data-poor scenarios [2].

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.6084/m9.figshare.24323506 [62].

Acknowledgments

This study was part-funded by the E C Horizon 2020 Research and Innovation Programme, AMBER (Adaptive Management of Barriers in European Rivers) Project, Grant Agreement Number 689682, led by C G L M V P was funded by a KESS-II scholarship sponsored by Natural Resources Wales. We are grateful to Jacqueline Rosette for help interpreting the limitations of remote sensing data.

Ethics Statement

The study was approved by Swansea University College of Science Ethics Committee with number 1 2023 8075 6909.

CRediT authorship contribution statement

Parks: Data collection, original draft preparation. Garcia de Leaniz: Conceptualization, Writing, Visualization, Data analysis, Supervision, Funding acquisition. Jones: Data collection, Supervision. Jones: Data collection, Supervision.

Conflict of interest

The authors declare no competing interests.

ORCID iD

Carlos Garcia de Leaniz https://orcid.org/0000-0003-1650-2729

References

[1] European Council 2021 *Fit for 55. The EU's Plan for a Green Transition* (European Council) (available at: www.consilium.europa.eu/en/policies/green-deal/eu-plan-for-a-green-transition)

[2] Garcia de Leaniz C and O'Hanley J R 2022 Operational methods for prioritizing the removal of river barriers: synthesis and guidance *Sci. Total Environ.* **848** 157471

[3] Hillman M and Brierley G 2008 *River Futures: An Integrative Scientific Approach to River Repair* ed G J Brierley and A F Kirstie ch 14, pp 257–72

[4] Belletti B *et al* 2020 More than one million barriers fragment Europe's rivers *Nature* **588** 436–41

[5] Jones J *et al* 2019 A comprehensive assessment of stream fragmentation in Great Britain *Sci. Total Environ.* **673** 756–62

[6] Sun J, Galib S M and Lucas M C 2020 Are national barrier inventories fit for stream connectivity restoration needs? A test of two catchments *Water Environ. J.* **34** 791–803

[7] Cote D, Kehler D G, Bourne C and Wiersma Y F 2009 A new measure of longitudinal connectivity for stream networks *Landsc. Ecol.* **24** 101–13

[8] Mulligan M *et al* 2021 Global Dam Watch: curated data and tools for management and decision making *Environ. Res.: Infrastruct. Sustain.* **1** 033003

[9] Mulligan M, van Soesbergen A and Saenz L 2020 GOODD, a global dataset of more than 38,000 georeferenced dams *Sci. Data* **7** 31

[10] Kelly-Quinn M *et al* 2022 *Assessment of the Extent and Impact of Barriers on Freshwater Hydromorphology and Connectivity in Ireland (Reconnect)* vol 73 (Environmental Protection Agency)

[11] Atkinson S, Bruen M, O' Sullivan J J, Turner J N, Ball B, Carlsson J, Bullock C, Casserly C M and Kelly-Quinn M 2020 An inspection-based assessment of obstacles to salmon, trout, eel and lamprey migration and river channel connectivity in Ireland *Sci. Total Environ.* **719** 137215

[12] Carolli M, Garcia de Leaniz C, Jones J, Belletti B, Huďek H, Pusch M, Pandakov P, Börger L and van de Bund W 2023 Impacts of existing and planned hydropower dams on river fragmentation in the Balkan Region *Sci. Total Environ.* **871** 161940

[13] Garcia de Leaniz C 2008 Weir removal in salmonid streams: implications, challenges and practicalities *Hydrobiologia* **609** 83–96

[14] Consuegra S *et al* 2021 Impacts of large and small barriers on fish assemblage composition assessed using environmental DNA metabarcoding *Sci. Total Environ.* **790** 148054

[15] Fencl J S, Mather M E, Costigan K H and Daniels M D 2015 How big of an effect do small dams have? Using geomorphological footprints to quantify spatial impact of low-head dams and identify patterns of across-dam variation *PLoS One* **10** e0141210

[16] Baumgartner L J, Marsden T, Duffy D, Horta A and Ning N 2021 Optimizing efforts to restore aquatic ecosystem connectivity requires thinking beyond large dams *Environ. Res. Lett.* **17** 014008

[17] Rodeles A A, Galicia D and Miranda R 2017 Recommendations for monitoring freshwater fishes in river restoration plans: a wasted opportunity for assessing impact *Aquat. Conserv.: Mar. Freshw. Ecosyst.* **27** 880–5

[18] King S, O'Hanley J R and Fraser I 2021 How to choose? A bioeconomic model for optimizing river barrier mitigation actions *Ecol. Econ.* **181** 106892

[19] King M, van Zyll de Jong M, Piercey D, Nunn A D and Cowx I G 2021 An integrated decision driven design framework to support the ecological restoration of rivers *J. Environ. Plan. Manage.* **65** 1–24

[20] Kroon F J and Phillips S 2016 Identification of human-made physical barriers to fish passage in the Wet Tropics region, Australia *Mar. Freshw. Res.* **67** 677–81

[21] Januchowski-Hartley S R *et al* 2022 Supporting proactive planning for climate change adaptation and conservation using an attributed road-river structure dataset *J. Environ. Manage.* **321** 115959

[22] Januchowski-Hartley S R, White J C, Pawar S K, Januchowski-Hartley F A and England J 2021 Addressing road-river infrastructure gaps using a model-based approach *Environ. Res.: Infrastruct. Sustain.* **1** 015003

[23] Neeson T M, Moody A T, O'Hanley J R, Diebel M, Doran P J, Ferris M C, Colling T and McIntyre P B 2018 Aging infrastructure creates opportunities for cost-efficient restoration of aquatic ecosystem connectivity *Ecol. Appl.* **28** 1494–502

[24] Diebel M W, Fedora M, Cogswell S and O'Hanley J R 2015 Effects of road crossings on habitat connectivity for stream-resident fish *River Res. Appl.* **31** 1251–61

[25] Jones J *et al* 2020 Quantifying river fragmentation from local to continental scales: data management and modelling methods *Authorea* (https://doi.org/10.22541/au.159612917.72148332/v2)

[26] Zhao Q, Yu L, Li X, Peng D, Zhang Y and Gong P 2021 Progress and trends in the application of Google Earth and Google Earth Engine *Remote Sens.* **13** 3778

[27] Kostianoy A G and Soloviev D M 2020 *The Handbook of Environmental Chemistry* vol 93, ed V Pešić, M Paunović and A Kostianoy (Springer)

[28] Yang X *et al* 2022 Mapping flow-obstructing structures on global rivers *Water Resour. Res.* **58** e2021WR030386

[29] Atkinson S, Bruen M, Turner J N, Ball B, Bullock C, O'Sullivan J J, Casserly C, King J J, Cullagh A and Kelly-Quinn M 2018 The value of a desk study for building a national river obstacle inventory *River Res. Appl.* **34** 1085–94

[30] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R 2017 Google Earth Engine: planetary-scale geospatial analysis for everyone *Remote Sens. Environ.* **202** 18–27

[31] Stack Exchange Inc 2024 *Geographic Information Systems. Spatial Resolution of Google Earth Imagery* (Stack Exchange Inc.) (available at: https://gis.stackexchange.com/questions/11395/spatial-resolution-of-google-earth-imagery/256850)

[32] Maria E, Budiman E and Taruk M *J. Phys.: Conf. Ser.* **1450** 012080

[33] Afonydd Cymru 2023 *Inland Fisheries Habitat Restoration Project 2020–21* (Afonydd Cymru) (available at: https://afonyddcymru.org/inland-fisheries-habitat-restoration-project-2020-21/)

[34] Rwanga S S and Ndambuki J M 2017 Accuracy assessment of land use/land cover classification using remote sensing and GIS *Int. J. Geosci.* **8** 611

[35] Zhu W, Zeng N and Wang N 2010 Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations *NESUG Proc.: Health Care and Life Sciences* (*Baltimore, Maryland*) vol 19 p 67

[36] Giardiello D, Quatto P, Ripamonti E and Vigliani S 2022 'raters': a modification of Fleiss' Kappa in case of nominal and ordinal variables *R package version 2.0.2*

[37] Cicchetti D V and Feinstein A R 1990 High agreement but low kappa: II. Resolving the paradoxes *J. Clin. Epidemiol.* **43** 551–8

[38] Wang L and Liu H 2006 An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling *Int. J. Geogr. Inf. Sci.* **20** 193–213

[39] QGIS Geographic Information System 2021 *Open Source Geospatial Foundation Project v. version 3.10 A*

[40] Clerici N, Weissteiner C J, Paracchini M L, Boschetti L, Baraldi A and Strobl P 2013 Pan-European distribution modelling of stream riparian zones based on multi-source earth observation data *Ecol. Indic.* **24** 211–23

[41] Calcagno V 2020 'glmulti': model selection and multimodel inference made easy *R package v. version 1.0.8*

[42] Lüdecke D, Ben-Shachar M S, Patil I, Waggoner P and Makowski D 2021 'performance': an R package for

assessment, comparison and testing of statistical models *J. Open Source Softw.* **6** 3139
[43] Hosmer D J, Lemeshow S and Sturdivant R X 2013 *Applied Logistic Regression* vol 528 (Wiley)
[44] Bradley B and Brandon G 2020 *Hands-On Machine Learning with R* 1st edn (CRC Press)
[45] Mohajane M, Costache R, Karimi F, Bao Pham Q, Essahlaoui A, Nguyen H, Laneve G and Oudija F 2021 Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area *Ecol. Indic.* **129** 107869
[46] Landis J R and Koch G G 1977 The measurement of observer agreement for categorical data *Biometrics* **33** 159–74
[47] Tjur T 2009 Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination *Am. Stat.* **63** 366–72
[48] Januchowski-Hartley S R, Diebel M, Doran P J and McIntyre P B 2014 Predicting road culvert passability for migratory fishes *Divers. Distrib.* **20** 1414–24
[49] Donchyts G, Winsemius H, Baart F, Dahm R, Schellekens J, Gorelick N, Iceland C and Schmeier S 2022 High-resolution surface water dynamics in Earth's small and medium-sized reservoirs *Sci. Rep.* **12** 13776
[50] Swan B and Griffin R 2020 A LiDAR–optical data fusion approach for identifying and measuring small stream impoundments and dams *Trans. GIS* **24** 174–88
[51] Buchanan B P *et al* 2022 A machine learning approach to identify barriers in stream networks demonstrates high prevalence of unmapped riverine dams *J. Environ. Manage.* **302** 113952
[52] Januchowski-Hartley S R, McIntyre P B, Diebel M, Doran P J, Infante D M, Joseph C and Allan J D 2013 Restoring aquatic ecosystem connectivity requires expanding inventories of both dams and road crossings *Front. Ecol. Environ.* **11** 211–7
[53] Thieme M L, Khrystenko D, Qin S, Golden Kroner R E, Lehner B, Pack S, Tockner K, Zarfl C, Shahbol N and Mascia M B 2020 Dams and protected areas: quantifying the spatial and temporal extent of global dam construction within protected areas *Conserv. Lett.* **13** e12719
[54] Lehner B *et al* 2011 High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management *Front. Ecol. Environ.* **9** 494–502
[55] Grill G *et al* 2019 Mapping the world's free-flowing rivers *Nature* **569** 215–21
[56] Parasiewicz P *et al* 2023 Over 200,000 kilometres of free-flowing river habitat in Europe is altered due to impoundments *Nat. Commun.* **14** 6289
[57] De Fries L, Camana M, Guimarães M and Becker F G 2023 Road crossings hinder the movement of a small non-migratory stream fish *Environ. Biol. Fishes* **106** 1–17
[58] Clarkin K *et al* 2005 *National Inventory and Assessment Procedure* vol 76 (USDA-Forest Service)
[59] Google Earth Studio 2022 *Best Practices* (available at: https://earth.google.com/studio/docs/best-practices/)
[60] Valman S J, Boyd D S, Carbonneau P E, Johnson M F and Dugdale S J 2024 An AI approach to operationalise global daily PlanetScope satellite imagery for river water masking *Remote Sens. Environ.* **301** 113932
[61] Gargan P G, Roche W K, Keane S, King J J, Cullagh A, Mills P and O' Keeffe J 2011 Comparison of field-and GIS-based assessments of barriers to Atlantic salmon migration: a case study in the Nore catchment, Republic of Ireland *J. Appl. Ichthyol.* **27** 66–72
[62] Parks M, de Leaniz G, Carlos J, Jones P and Jones J 2023 Modelling barrier detection to achieve free-flowing river targets figshare *Dataset* (https://doi.org/10.6084/m9.figshare.24323506.v1)