# Predicting substrate size at a watershed scale to inform conservation planning for a declining salmonid species

Kyleisha J. Foote[1], Shawn J. Leroux[1], Ava J. Hart[1], Nick C. Murphy[1], Craig F. Purchase[1]

[1]Department of Biology, Memorial University of Newfoundland, 45 Arctic Ave, St. John's, NL A1C 5S7, Canada.

Corresponding author: Kyleisha J. Foote – kyleisha.foote@gmail.com

## Abstract

Good quality spawning habitat is critical for fish embryo development, survival, and overall population productivity. Appropriate riverbed substrate size is particularly important for riverine-spawning salmonids but the availability of suitable substrate may vary across a watershed. Predicting substrate size at watershed extents may therefore be critical to inform management and conservation of salmonid populations. Using a 30 m Digital Elevation Model (DEM), we modelled riverbed substrate size using the Shields equation at stream segments across a large watershed (~11,200 $km^2$) in Newfoundland, Canada. We use this model to predict the distributions of substrate suitable for Atlantic salmon (*Salmo salar*) spawning. We confront model predictions with field data on substrate size at 149 sites across the watershed. Our model predicted that 58% of segment length contained suitable (16-35 mm) or marginal (35-64 mm) substrate for salmon spawning. At these classes our model precision was between 24 and 39%. Overall, our model predicted the correct field-measured substrate class for 42% of segments. This was bolstered by higher precision (77%) at larger substrate sizes (>64 mm) possibly due to this class being over-represented in field data. Model predictions of substrate size increased alongside field measurements but generally under-estimated substrate size. Our model provides an important step toward watershed-scale assessments of potential salmon spawning habitat to help guide more efficient restoration and conservation planning. We surmise that higher resolution DEM data would allow us to fine-tune model predictions to areas more relevant to salmon spawning.

**Keywords:** Atlantic salmon, spawning habitat, GIS, gravel, Salmo salar, Exploits River, salmonid, model

**Acknowledgments**

We thank Pierce McNie, Connor Hanley, Emmerson Wilson, Chizurum Azu-Emuchay, Mikael Ranta, and Ky Schlosberg for assistance in the field. We thank Hannah Adams for initial GIS work and assistance with site selection, Mikael Ranta for assistance with site selection and Pierce McNie for help with coding and visualization. Niels van Miltenburg was instrumental in providing logistical support and resources. In addition, we received valuable logistical support from Terry Paul, Kevin Sheppard and Darren Ryan (ERMA). Funding for this research was provided by the Department of Fisheries and Oceans Canada - Aquatic Ecosystems Restoration Fund to the Environment and Resources Management Association (ERMA).

**Author contributions using CREdiT taxonomy**

Conceptualization: KF, CP, SL; Data collection and curation: KF, AH, NM; Formal analysis: KF, SL; Funding acquisition: CP, SL; Investigation: KF, SL, CP; Methodology: KF, SL, CP; Project administration: KF, SL, CP; Resources: SL, CP, KF; Software: KF, AH; Supervision: SL, CP; Validation: KF, SL; Visualization: KF; Writing – original draft: KF; Writing – review and editing: All authors.

## Introduction

Salmonids are widely recognized for their ecological, economic and cultural importance (Noble et al., 2016; Thorstad et al., 2021) and are one of the most researched fish taxa. However, many salmonid populations are declining over broad geographic areas, largely due to habitat loss and/or degradation (Dauwalter et al., 2020; Dudgeon et al., 2006; Jelks et al., 2008; Laramie et al., 2015). Salmonid biomass in rivers around the world has declined by 38% from pre-1980 levels to post-2000 (Foote et al., 2025). Diadromous species are in particular peril (Deinet et al., 2024; Joy et al., 2018) due to complex life histories that require the use of both freshwater and marine habitats (Limburg & Waldman, 2009; McDowall, 1999). Specifically, different life stages require different types of habitat (Armstrong et al., 2003; Bardonnet & Baglinière, 2000; Jonsson & Jonsson, 2011); limitations or bottlenecks at one stage may result in poor reproductive success, even if there is an abundance of good quality habitat for other stages (Heggenes, 1990). Thus, understanding where these limitations exist in populations is critical. The use of multiple habitat types by diadromous fishes complicates their management, as population declines are likely driven by factors in both freshwater and marine ecosystems that are not fully understood (Soulsby et al., 2024; Wilson et al., 2022). That said, anadromous salmonid population productivity is typically constrained during the freshwater, not marine, phase of the life cycle (Behnke, 2002). What is more, freshwater habitats are generally more amenable to management actions, such as restoration and enhancement, than marine habitats (Purchase, 2026).

The availability and quality of freshwater spawning habitat is critical for the development and survival of salmonid embryos (Bardonnet & Baglinière, 2000; Jonsson & Jonsson, 2011; Wilkins & Snyder, 2011), eventually influencing productivity. Important aspects of salmonid riverine spawning habitat include depth, flow, river-bed substrate (i.e., gravel), and cover (Armstrong et al., 2003; Crisp, 2000; Louhi et al., 2008). Different species and populations have different ideal requirements and conditions (Kondolf & Wolman, 1993). In particular, substrate size and mobility are essential as fish cannot effectively construct nests (i.e., redds) if they cannot move the river-bed substrate (Wilkins & Snyder, 2011). Salmonids generally spawn in gravels with a median diameter of around 1-10 % of their body length (Kondolf & Wolman, 1993). Larger fish can move larger substrates and thus can use a wider range of potential habitats (Buffington et al., 2004; Riebe et al., 2014).

Stream bed substrate size is controlled by channel hydraulics and sediment supply (Buffington et al., 2004; Wilkins & Snyder, 2011). Channel hydraulics control the force of water moving against the

channel bed (i.e., shear stress) and the largest substrate size a river can carry (Buffington & Montgomery, 1999). Larger substrate particles require higher shear stresses to initiate transport (Powell, 1998). Bankfull flows, occurring every 1-2 years for most rivers, often have the largest influence on the transport and size of bed substrate in alluvial rivers (Andrews, 1984; Wolman & Miller, 1960). The size and volume of substrate available in rivers is controlled by sediment supply (Buffington et al., 2004). Generally substrate is hydraulically sorted downstream with substrate decreasing in size (Powell, 1998), but this is disrupted by lakes dams and other barriers acting as sediment sinks (Arp et al., 2007; Fox et al., 2016; Hauer et al., 2018), as well as tributaries bringing different substrate loads (Dawson, 1988; Knighton, 1980). River confinement and land use disturbances often reduce the quantity or quality of spawning habitats (Barlaup et al., 2008; Hauer et al., 2020; Pulg et al., 2013), but some regions may also be naturally sparse in spawning sized gravel (Purchase, 2026). Identifying sites with suitable spawning gravel is an important research objective to support conservation of threatened or vulnerable salmonid populations. Similarly, identifying areas with the capacity to support suitable spawning gravel that have since been degraded is essential for guiding restoration or enhancement efforts.

Depending on the spatial extent and desired precision, different methods exist for measuring substrate size. At a small extent (i.e., 50-200 m), field measurements of bed substrate size (e.g. Wolman, 1954) are a common technique to determine substrate distributions and median substrate size (Kondolf & Li, 1992). For example, researchers have measured substrate sizes for stream habitat assessments (Quinn & Hickey, 1994; Sutherland et al., 2010). Field measurements, however, can be time-consuming, limited in spatial scope, and may include bias from different operators (Daniels & McCusker, 2010; Pearson et al., 2017). Remote sensing, on the other hand, has given us the ability to assess landscapes from afar. While remote analysis does not entirely substitute for intimate knowledge and understanding that can be gained from on the ground surveying and methods, remote sensing can widen the areas we assess that would not be possible due to time and physical constraints (Kerr & Ostrovsky, 2003; Osborne et al., 2001; Reif & Theel, 2017).

Geographic Information Systems (GIS) techniques are used to analyze data from remote sensors and GIS are widely used to predict habitat over large areas (Osborne et al., 2001; Valavanis et al., 2008; Weiers et al., 2004). In relation to spawning habitat, various GIS models have been developed to predict bed substrate size using geomorphic relationships and remote sensed data, particularly Digital Elevation Models (DEMs) (Buffington et al., 2004; Gorman et al., 2011; Haddadchi et al., 2018; Hedger et al., 2006; Snelder et al., 2011; Snyder et al., 2013; Wilkins & Snyder, 2011). High resolution (e.g. 3-10 cm) aerial

imagery can offer detailed analyses of river bed terrain used to estimate substrate size (e.g. Carbonneau et al., 2004; Hedger et al., 2006; Javernick et al., 2014; Pearson et al., 2017), but are often applied at small spatial extents (e.g. a single gravel bar up to reaches at the kilometer scale). Likewise, fine scale DEMs created from Light Detection and Ranging (LiDAR) are available at a relatively high resolution (1-2 m) for many areas, including some parts of Canada (Natural Resources Canada, 2025a). While useful for detailed, small-scale analyses, high resolution data can be impractical for large watershed analyses due to the processing power required. Given that salmonids may migrate long distances often in remote areas, comprehensive field surveys of spawning habitat are often impractical. Consequently, there is a need for remote sensed models of spawning gravel at watershed extents where these fish are distributed.

Process based models to predict substrate size using DEMs and hydrologic data have produced accurate results over the extent of small watersheds (< 700 $km^2$) (Buffington et al., 2004; Gorman et al., 2011; Snyder et al., 2013; Wilkins & Snyder, 2011) and have great potential to be applied in larger watersheds. However, these models are seldom used at larger watershed extents. Thus, we investigate whether a process-based model can accurately predict riverbed substrate at a large spatial extent. We implement one of these models (Wilkins & Snyder, 2011) to predict river bed substrate size suitable for spawning Atlantic salmon (*Salmo salar*) in a large watershed (11,250 $km^2$, Exploits River, Newfoundland, Canada). We use GIS to predict the median size (i.e., also known as $D_{50}$) of bed surface substrate for river segments between 0.03 and 6.5 km in length using low resolution elevation (30 m) and hydraulic data. We then test the accuracy of our GIS model with data collected from field surveys of substrate size at 50 m stream sections.

## Methods

We begin by describing the watershed (Figure 1) and Atlantic salmon spawning substrate habitat. We then describe data sources, GIS data processing, and data extraction and analysis to predict substrate size, followed by our field data collection methods and validation for the GIS model predictions. Figure 2 follows the steps used and more detailed descriptions are in the supplementary material (Section S1.2).

### *Study area*

The Exploits River flows from Beothuk Lake in central Newfoundland, Canada, and drains into the Bay of Exploits near Bishop's Falls (Figure 1). Upstream, Lloyd's River and Victoria River, along with many smaller tributaries, flow into Beothuk Lake. At 246 km long, draining an area of ~11,270 $km^2$, the Exploits River is the longest river with the largest drainage basin on the island of Newfoundland (see Table S1.1 in supplementary material for characteristics of the watershed). The Exploits Valley was formed by glaciation and is dominated by siliciclastic and volcanic marine bedrocks and surficial geology of glacial till and hummocky terrain (Newfoundland and Labrador Geological Survey, 2024). The watershed contains ~960 lakes or reservoirs over 10 ha (Lehner et al., 2022), covering around 9.8% of the total drainage area.

Freshwater fish species present in the Exploits watershed include Atlantic salmon, both anadromous and non-anadromous (known as ouananiche in Newfoundland) forms, brook trout (*Salvelinus fontinalis*), three-spined stickleback (*Gasterosteus aculeatus*), American eel (*Anguilla rostrata*), and Arctic charr (*Salvelinus alpinus*) (Mullms et al., 2007) although it is mostly dominated by Atlantic salmon and inter-specific competition is low. The current population of anadromous Atlantic salmon in the watershed is higher than historical level (COSEWIC, 2010; Fisheries and Oceans Canada, 2025), as waterfalls restricted access to around 90% of the watershed until enhancement activities and fish passages increased fish stocks and improved access around the 1960s to mid-1990s (O'Connell & Bourgeois, 1987; Scruton et al., 2008). The Exploits River currently has the highest known return of Atlantic salmon in Newfoundland (~25,000 fish, Fisheries and Oceans Canada, 2025), however, productivity is low relative to predicted potential based on habitat capacity (Davis & Riche, 1983; Scruton et al., 2003).

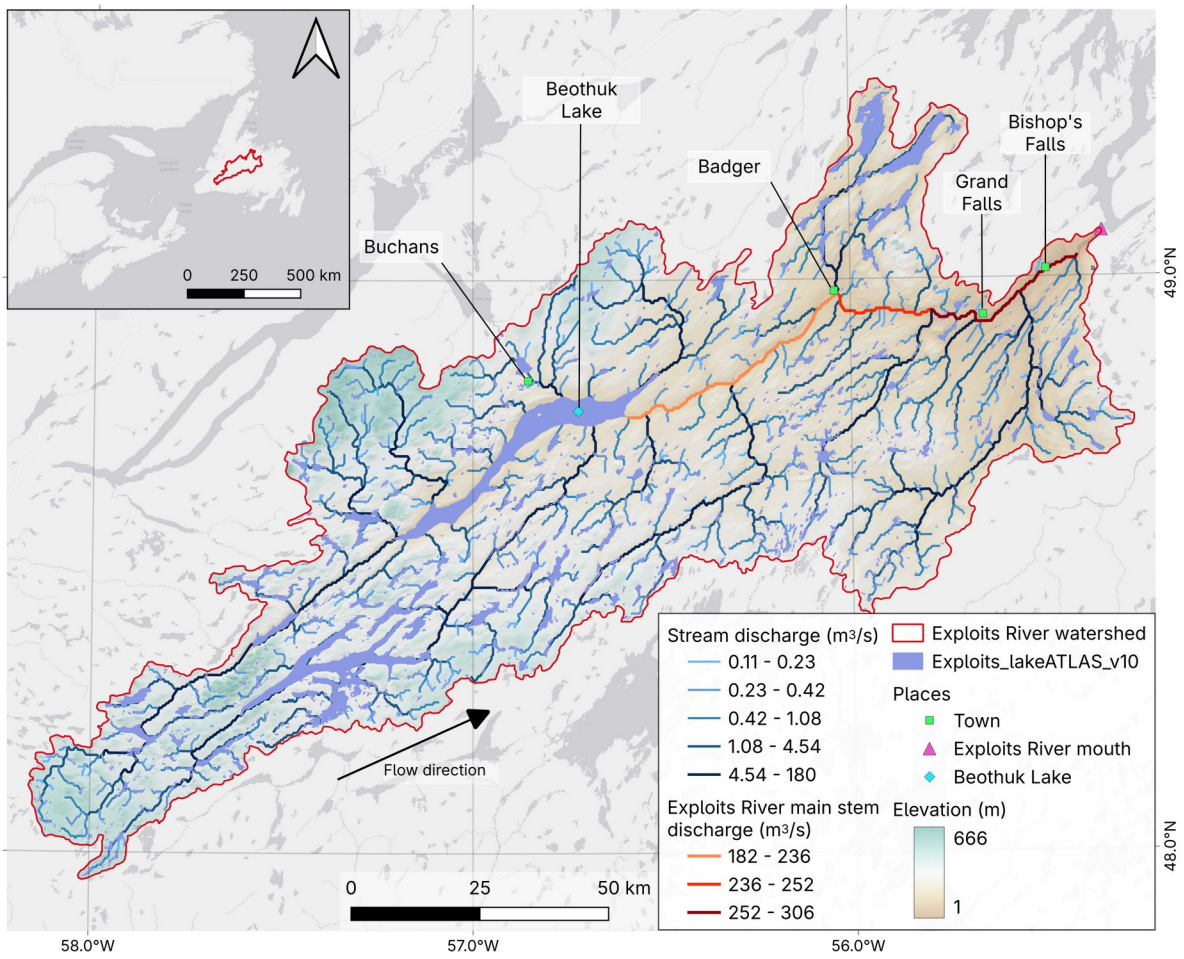



Figure 1. Exploits River watershed in Central Newfoundland (shown by red boundary in inset), with major towns and Beothuk Lake labelled. The main stem of the Exploits River is coloured orange/red by discharge while other streams are in blue. The black arrow indicates general flow direction towards the river mouth. Data sources: Basemap - Esri, HERE, Garmin, © OpenStreetMap contributors, and the GIS User Community; Elevation – DEM (Natural Resources Canada, 2025b); river lines with discharge – RiverATLAS (Linke et al., 2019); lakes – LakeATLAS (Lehner et al., 2022).

### *Atlantic salmon spawning habitat*

Atlantic salmon (hereafter salmon) show a wide range of life history strategies, including freshwater residents and sea-run anadromous forms, varying greatly in time spent at sea (for a overview of Atlantic salmon life history see Fleming, 1996; Gibson & Haedrich, 2006; Jonsson & Jonsson, 2007). Typically, females deposit their eggs in gravel substrate hollows, called redds, in flowing freshwater in Autumn (O'Connell et al., 2006). Suitable spawning substrate generally has an intermediate axis length (the second-longest dimension on an individual substrate particle) between 16 and 64 mm (Armstrong et al., 2003; Louhi et al., 2008), but habitat use largely depends on availability

(Heggenes, 1990) and female fish size (Kondolf & Wolman, 1993; Overstreet et al., 2016; Riebe et al., 2014). As is the case for most Newfoundland salmon populations, around 90 per cent or more of salmon returns to the Exploits River are fish under 63 cm, classed as small salmon (Department of Fisheries and Oceans, 2022; Fisheries and Oceans Canada, 2025), thus suitable spawning gravel may be on the smaller side of the typical range (Purchase, 2026). In this paper, we consider gravel sizes between 16 and 35 mm as suitable and sizes ranging from 35 to 64 mm as marginal.

### *GIS data processing*

To predict substrate size for the Exploits River watershed we used a combination of open access data sets including HydroSHEDS data (Lehner et al., 2022; Linke et al., 2019), a 30 m resolution digital elevation model (DEM) (Natural Resources Canada, 2025b), Google satellite imagery, and Geobase hydrography spatial data (Government of Canada et al., 2022) (Table 1). Because of its low resolution, HydroSHEDS data is better used for larger scale analysis, i.e., across watersheds and regions; however, in the absence of fine-scale hydrological data throughout the watershed, we used discharge and channel width (derived from dividing river area by length) from the HydroSHEDS RiverATLAS layer in our model (Table 1). Instead of using stream segments determined by RiverATLAS, we created a river network that contains a larger number of segments than RiverATLAS. With more segments over the same area, the accuracy of our substrate model was likely to be improved. We outline the process below and in Figure 2, while detailed methods are described in section S1.2 of the supplementary material.

Using QGIS tools and standard protocols, we created a river channel network for the Exploits watershed (termed DEM $Channels_{ER}$) using the 30 m DEM (Figure 2A and B). Due to the coarseness of the DEM, the created river network often does not follow the actual river channel, but is a higher resolution improvement from RiverATLAS segments. Thus, we manually corrected channels where obvious major deviations occurred using Google satellite imagery and Geobase hydrography as references for the true river channels. The DEM $Channels_{ER}$ output comprises ‘linestrings’ (segments or features) that represent a length of stream between two confluences, i.e. segments ended or started at every stream confluence as determined by the DEM. Therefore, segment lengths were variable throughout the watershed (Table 2).

Table 1. GIS Data layers used to create river lines and/or predict substrate size.

| Data | Description | Use of data | Source |
|---|---|---|---|
| Digital Elevation Model (DEM) | Raster of 30 m resolution elevation data. | Creating river network, calculating slope. | Natural Resources Canada (2025b) |
| HydroSHEDS data (see ATLAS layers below) | Global database of digital data layers consisting of hydrographic data. The ATLAS products of HydroSHEDS contain additional attribute data. | We refer to HydroSHEDS to encompass the whole database but refer to ATLAS products when they are used (see below). | See ATLAS sources |
| Riverlines - RiverATLAS | Vectorized line network of rivers that have a watershed area of at least 10 $km^2$ or an average river discharge of at least 0.1 $m^3/s$. Contains a large number of hydro-environmental characteristics. | Extracting hydro-environmental variables such as discharge and width. | Linke et al. (2019) |
| Lakes - LakeATLAS | Shoreline polygons of lakes with a surface area of at least 10 ha. | Used to exclude river lines inside lake polygons in our river network. | Lehner et al. (2022) |
| Basins - BasinATLAS | Watershed (basin) polygons at different scales, containing hydro-environmental characteristics. | General polygons of the watershed and sub-basins. | Lehner & Grill (2013) |
| Google Satellite | Photographic images of the Earth collected by imaging satellites. | Reference for modifying river lines. | Google Satellite added by QMS in QGIS |
| Geobase Hydrography | Geospatial digital data such as lakes, reservoirs, watercourses (rivers and streams), canals, islands, drainage linear network, geographical names, constructions and obstacles related to surface waters, etc. Created from existing data at the 1:50 000 scale or better. | Flow directions and watershed extent for modifying channel lines and watershed and sub-basin boundaries. | Government of Canada et al. (2022) |

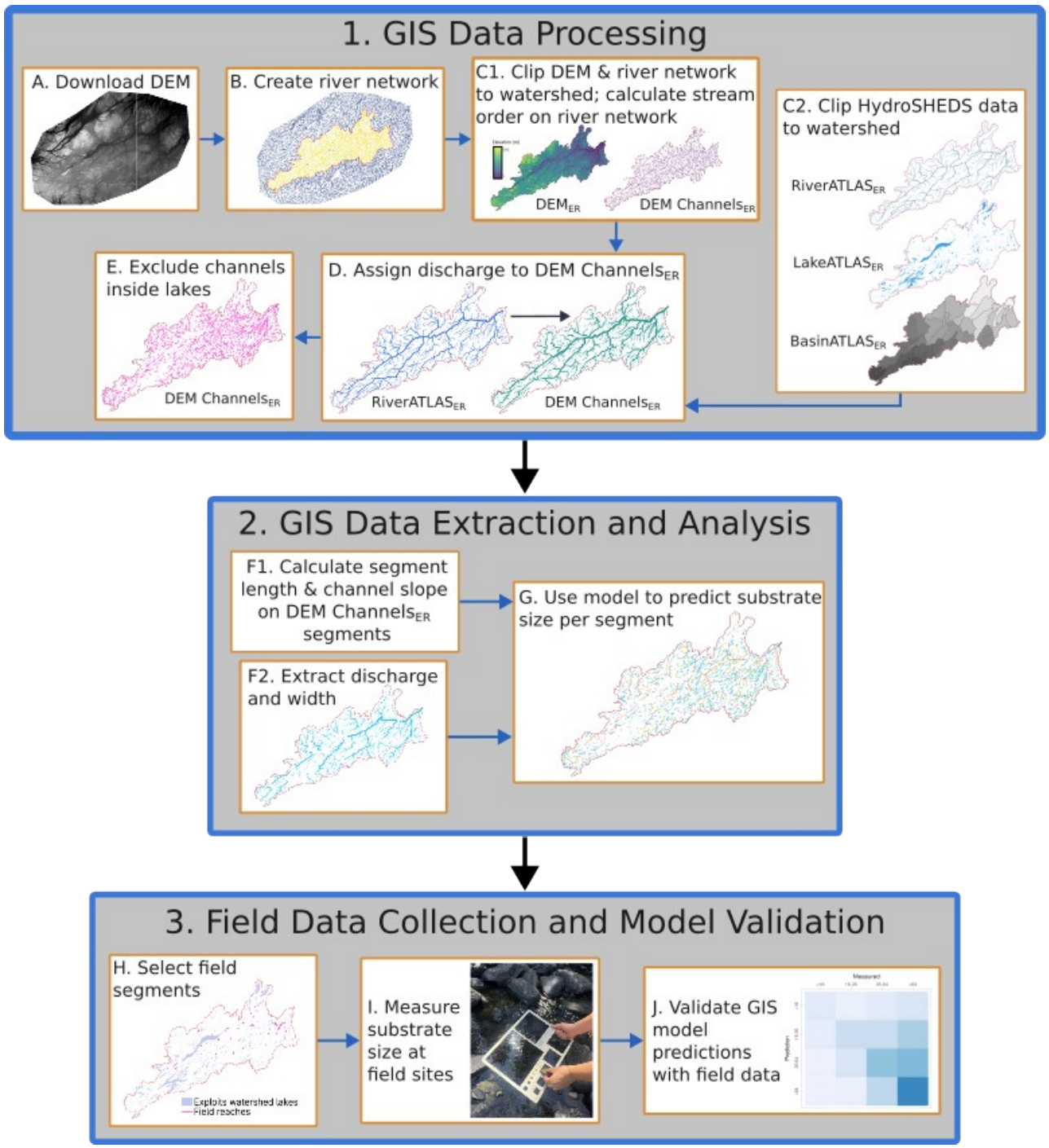



Figure 2. Overview of analytical methods to predict and validate the substrate model. The red polygon in all maps is the boundary of the Exploits River watershed, determined using Google satellite imagery and Geobase hydrography. The $_{ER}$ denotation on layers indicates the Exploits River watershed. Panel 1 outlines our GIS data processing steps used to prepare the data for model predictions (see GIS data processing section of the main text and supplementary material S1.2 for further details). A: We downloaded a DEM to cover a wider range than the Exploits River watershed. B: We created a river lines network using the DEM. Yellow lines depict streams that run to the Exploits River mouth while

blue lines are outside the watershed. C1: We clipped the DEM and river network to the watershed, termed $DEM_{ER}$ and DEM $Channels_{ER}$, respectively. Elevation (m) above sea level is shown on the $DEM_{ER}$ with darker blues representing lower elevations and yellows higher elevations. We then calculated stream order on the DEM $Channels_{ER}$ river network. C2: We clipped the river, lakes, and basin ATLAS products of HydroSHEDS to the watershed boundary. D: We assigned discharge from $RiverATLAS_{ER}$ to DEM $Channels_{ER}$, using calculated stream order in the DEM $Channels_{ER}$ layer to match with discharge. Thicker lines represent higher discharge and stream order for $RiverATLAS_{ER}$ and DEM $Channels_{ER}$, respectively. E: We excluded river lines inside lake polygons from $LakeATLAS_{ER}$ in DEM $Channels_{ER}$. Panel 2. We then extracted data from GIS layers to run the substrate model. F1: We calculated the length of channel segments on the DEM $Channels_{ER}$ layer, followed by channel slope using elevation from $DEM_{ER}$. F2: We extracted discharge and width assigned from $RiverATLAS_{ER}$ (see D). Segments are displayed by discharge with thicker lines representing higher discharges. G: We predicted substrate size and group outputs into four size classes, shown by different colours. Panel 3 outlines field data collection and model validation. H: We selected field segments that were accessible and wadeable, shown by pink lines. I: We used a gravelometer, shown in the image, to measure substrate size. J: We validated GIS model predictions using field data with a confusion matrix, shown on the right, and explained in the GIS model validation section of the methods.

To obtain an estimate of river size, we calculated stream order on DEM $Channels_{ER}$ (Figure 2C1). This allowed us to match streams based on size with RiverATLAS data for the Exploits River ($RiverATLAS_{ER}$) by comparing stream order with discharge. By matching similar stream sizes, we could assign discharge and other attributes from $RiverATLAS_{ER}$ to DEM $Channels_{ER}$ (Figure 2D). We clipped the DEM (Figure 2C1) and three ATLAS layers (Figure 2C2) to the Exploits River watershed to be used in various stages of GIS analysis (Table 1).

The model we used to predict substrate size is intended for use in stream channels rather than lakes. Therefore, to obtain more accurate substrate size estimates, we excluded segments that crossed through lakes (Figure 2E). Characteristics of the final layer used for substrate predictions are summarized in Table 2.

Table 2. Summary of features in the DEM $Channels_{ER}$ river network (5010 features or stream segments). Discharge and width were extracted from $RiverATLAS_{ER}$ and slope and segment length were calculated on the DEM $Channels_{ER}$ layer.

| Variable | Mean | Median | Min / Max | Standard deviation |
|---|---|---|---|---|
| Segment length (m) | 845 | 657 | 10.95 / 6200 | 722 |
| Width (m) | 11.58 | 6.64 | 2.45 / 105.44 | 16.628 |
| Slope | 0.0089 | 0.0054 | 2e-06 / 0.202 | 0.0122 |
| Mean annual discharge ($m^3/s$) | 10.62 | 0.386 | 0.106 / 305.5 | 42.93 |
| Maximum annual discharge ($m^3/s$) | 24.60 | 0.971 | 0.237 / 668.4 | 97.36 |

### *GIS data extraction and analysis*

To predict substrate size, we estimated channel hydraulics using bed shear stress and Shields parameter following Wilkins and Snyder (2011). Bed shear stress ($\tau_b$) is the force that flowing water applies to the channel bed to transport sediment, which we estimated by:

$$\tau_b = \rho g n^{(0.6)} \left[\frac{Q}{w}\right]^{(0.6)} S^{(0.7)}$$

where $\rho$ is water density (1000 $kg/m^3$), $g$ is acceleration due to gravity (9.80665 $m/s^2$), $n$ is the Manning's channel roughness coefficient, $Q$ is discharge ($m^3/s$), $w$ is channel width (m), and $S$ is channel slope. We used an $n$ value of 0.05, which is within estimates for gravel, cobble and boulder rivers (Aldridge & Garrett, 1973; Barnes, 1967). We derived $Q$ (average annual maximum) and $w$ from $RiverATLAS_{ER}$ data (Table 1 and Figure 2F2), and $S$ from the $DEM_{ER}$ (Figure 2F1; see supplementary material S1.2 for full details).

We then used $\tau_b$ to predict median grain size ($D_{50}$), following Wilkins and Snyder (2011):

$$D_{50} = \frac{\tau_b}{(\rho_s - \rho) g \tau^*}$$

where $\rho_s$ is sediment density (using the density of quartz, $\rho_s$ = 2650 $kg/m^3$), and $\tau^*$ is Shields parameter, an indicator for the onset of sediment transport. Estimates of $\tau^*$ vary with basin size and channel characteristics but generally range between 0.03 and 0.07 (Buffington & Montgomery, 1997; Mueller et al., 2005). As we did not have derived values of $\tau^*$ for our watershed, we used a $\tau^*$ value of 0.04, a median estimate of gravel-bed rivers by Mueller et al. (2005) and after Wilkins and Snyder (2011). We used R (R Core Team, 2025) to calculate substrate size across the Exploits River watershed based on

parameters described above, termed the model predicted $D_{50}$ (Figure 2G). Each river segment of the DEM $Channels_{ER}$ layer had one model predicted $D_{50}$, resulting in 5010 predictions. To investigate model predictions, we report the proportion of segment length predicted to be suitable for spawning in different sub-basins across the watershed.

We grouped substrate into four size classes that generally align with suitability for Atlantic salmon spawning: less than 16mm (too small), 16-35mm (suitable, includes values >16mm to <35mm), 35-64mm (marginal, includes values >35mm to <64mm), and greater than 64mm (too large) (Armstrong et al., 2003; Kondolf & Wolman, 1993). Substrate sizes between 35 and 64mm might be used for spawning if better habitat is limited, or by larger female fish (Purchase, 2026).

### *Field data collection*

To compare with model predictions, we conducted field measurements of substrate size at locations across the Exploits River watershed between June 12 and October 20, 2025. We selected 51 accessible and wadeable segments to cover a wide spatial range over the watershed (Figure 2H and Figure S2.1). Field segments ranged in length between 0.35 and 5.4km. For each segment, we digitized a centerline along the length of the river based on Google satellite imagery and topography maps. Typically, three sample sites were randomly selected within each segment, for a total of 149 sites (see supplementary material S2.1 for details of random site selection).

At each site, we performed a Wolman (1954) pebble count, randomly selecting a minimum of 100 clasts while walking in a zig-zag pattern over a 50 m length of river (Figure S2.2A and S2.2B). We used a gravelometer to measure the intermediate axis of each randomly selected individual substrate particle up to 300 mm (Figure 2I and S2.2C) and estimated size for larger substrates following the size categories in Table S2.1. We calculated the median ($D_{50}$) substrate size for each site using these data, termed the field-measured $D_{50}$. Note that we recorded field measurements of substrate in discrete bins (Table S2.1), while the model predicted substrate size along a continuous scale. Although we have one model-predicted $D_{50}$ for each DEM $Channels_{ER}$ segment, we have up to three field-measured $D_{50}$ values to capture the variability within each segment.

### *Field data analysis*

For each measured site, we plot the distribution of substrate size by the cumulative percentage of substrate finer than each gravelometer bin size. Sites were split into four facets based on the segment model-predicted size class that each site was in. For example, sites in the <16mm facet had a segment model-predicted $D_{50}$ less than 16 mm, but the site field-measured $D_{50}$ may not necessarily fall in this group. Note that model-predicted size classes which are based on fish size (Fisheries and Oceans Canada, 2025; Kondolf & Wolman, 1993; Purchase, 2026) loosely fit with gravelometer bins (see Table S2.2). Importantly, model bins align perfectly with gravelometer bins for the joint suitable and marginal spawning size classes (i.e., the gravelometer has bins of 16 mm and 64 mm). We plot the mean substrate distribution of sites within each facet group and therefore were able to see the combined field measured $D_{50}$ in each group.

### *GIS model validation*

We tested how well the GIS model predicts field-measured substrate classes with a confusion matrix (Figure 2I) using the caret (Kuhn, 2008) and cvms (Olsen & Zachariae, 2025) packages in R (R Core Team, 2025). Specifically, a confusion matrix allowed us to quantify the number of true positives/negatives (i.e., model-predicted substrate size = field-measured substrate size) and false positives/negatives (i.e., model-predicted substrate size ≠ field-measured substrate size). This enabled us to see where the model may be correctly and incorrectly predicting substrate classes. The accuracy of the GIS model predictions is the ratio of the number of correct predictions to the total number of predictions, measured from 0 to 1 (low to high). High accuracy does not necessarily indicate good model performance, as unbalanced classes (i.e., differing numbers of samples among classes) may allow the model to correctly predict the dominant class while performing relatively poorly in other classes, thereby inflating overall accuracy. Thus, we also measure the true positive rate (i.e., sensitivity or recall) and true negative rate (i.e., specificity), which represent correct predictions. A much higher value for one of these rates may indicate imbalances in the data. Precision measures how many positive predictions were actually correct; it is a measure of the accuracy relative to the prediction of a specific class. We report the F1 value, which is the weighted mean of precision and sensitivity, with 1 being the best score and 0 the worst. Finally, we report Cohen’s Kappa which identifies how well the predictions are compared to random guessing and ranges from -1 to 1, with 0 indicating no better than random

chance, and less than 0 worse than chance. A high Kappa score (closer to 1) indicates a strong agreement between model predictions and actual values. Table S2.3 gives detailed descriptions and bounds of confusion matrix metrics.

We tested for significant differences in the GIS model-predicted $D_{50}$ grouped by the measured substrate size classes using a Kruskall-Wallis, and Dunn's Test with a Bonferroni correction to adjust for multiple comparisons, using the rstatix package (Kassambara, 2025) in R (R Core Team, 2025). We used a non-parametric test as the data is not normally distributed. We used a significance threshold of alpha = 0.05.

## Results

Our model predicted that more than half of the stream length in the Exploits River watershed had a median grain size ($D_{50}$) in the suitable (16-35 mm) or marginal (35-64 mm) spawning suitability categories (Table 3). Note that the model predicts one $D_{50}$ estimate for each DEM $Channels_{ER}$ segment. This distribution was uneven across the watershed (Figure 3A) with the highest proportion of predicted suitable substrate (16-35 mm) in the Great Rattling Brook basin (Figure 3B and 3C).

Table 3. Summary of substrate $D_{50}$ model predictions on the DEM $Channels_{ER}$ layer split into size classes for the Exploits River watershed.

| **Predicted $D_{50}$ substrate class size (mm)** | **Suitability for spawning** | **Number of stream segments in class** | **Length of stream segments in class (km)** | **Percentage of stream length in class** |
|---|---|---|---|---|
| <16 | Too small | 1869 | 1093 | 25.8 |
| 16-35 | Suitable | 1318 | 1327 | 31.4 |
| 35-64 | Marginal | 1082 | 1109 | 26.2 |
| >64 | Too large | 741 | 703 | 16.6 |
| *Total* | | *5010* | *4232* | *100* |

The model under-estimated substrate sizes when compared to field measurements in the three lowest model-predicted size classes (Figure 4). The average field-measured $D_{50}$ for each size category, denoted by the black dashed lines in Figure 4 is consistently higher than the model-predicted size class. The exception was the >64mm predicted category. This group had a much larger size range compared to the smaller three groups, so it is somewhat unsurprising we see more agreement in this group. Nonetheless, the measured $D_{50}$ increased progressively with the predicted classes, i.e. the measured $D_{50}$ in the 16-35 mm predicted class is higher than the measured $D_{50}$ in the <16 mm predicted class, and this pattern follows for the 35-64 mm and >64 mm classes (Figure 4). Of note, the combined $D_{50}$ of all measured substrates in the predicted 35-64 mm group fell closer to 64 mm than the subsequent bin of 90 mm. The $D_{50}$ is determined by which bin the $50^{th}$ percentile of counts fall; because substrate field measurements are discrete, the $50^{th}$ percentile will be situated in one of the gravelometer substrate size categories (Table S2.2). While the $50^{th}$ percentile in the 35-64 mm group was in the 90 mm bin, the $48^{th}$ percentile fell in the 64 mm bin; thus the combined measured $D_{50}$ only just exceeds the bounds of the prediction.

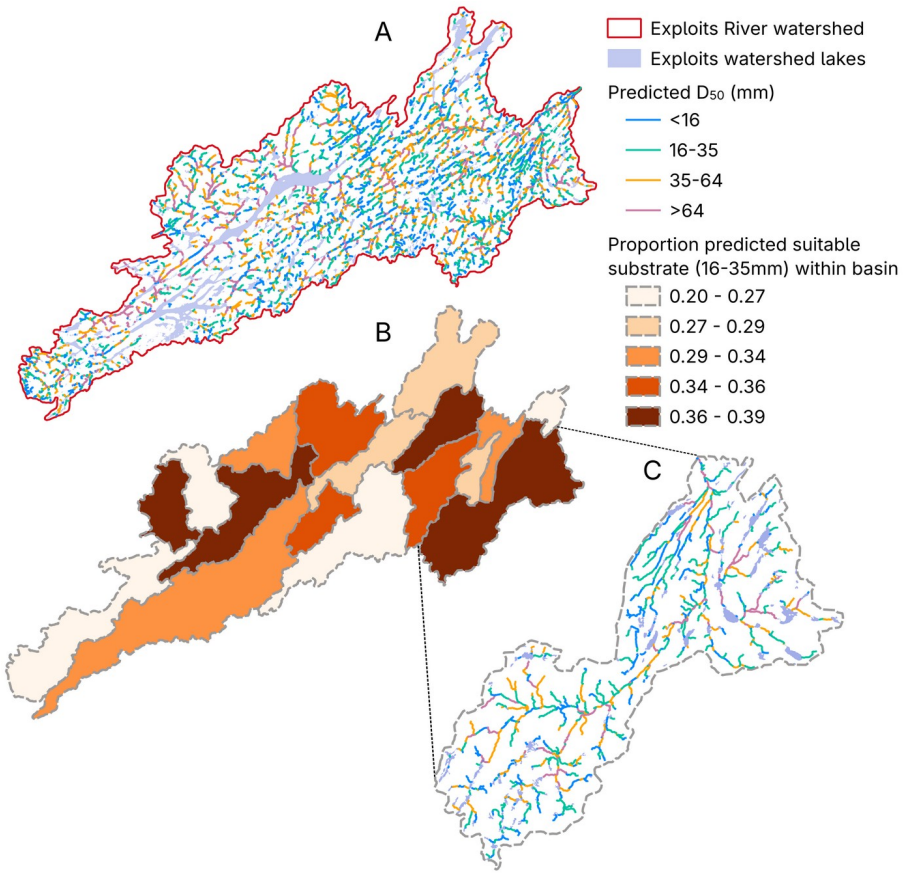



Figure 3. Model predictions of median grain size ($D_{50}$, mm) in the Exploits River watershed. Classes of median grain size (Predicted $D_{50}$) are calculated for each segment. A: Model predictions on the DEM $Channels_{ER}$ layer. B: The proportion of suitable spawning substrate (16-35 mm, green segments in A and C) by segment length within mapped sub-basins. C: Model predicted $D_{50}$ for segments in the Great Rattling Brook basin. GIS sources: watershed and basin polygons are modified from a BasinATLAS layer (Lehner & Grill, 2013); lakes from LakeATLAS (Lehner et al., 2022).

Sites that fall in the two larger predicted classes (35-64 mm and >64 mm) had lower confidence intervals in measured substrate distributions (Figure 4), indicating these sites are more similar in their distribution of bed substrate size within each class. In contrast, the two smaller groups had a wider distribution of substrate sizes across sites and more uncertainty. The greater uncertainty could be partly attributed to the smaller sample sizes in these groups. Importantly, sites with a large measured proportion of finer substrate (sites in the upper left hand corner of panel A and B, Figure 4), were also predicted to only fall into the two lowest substrate classes. Likewise, there were few sites in the <16 mm category that contained large substrates.

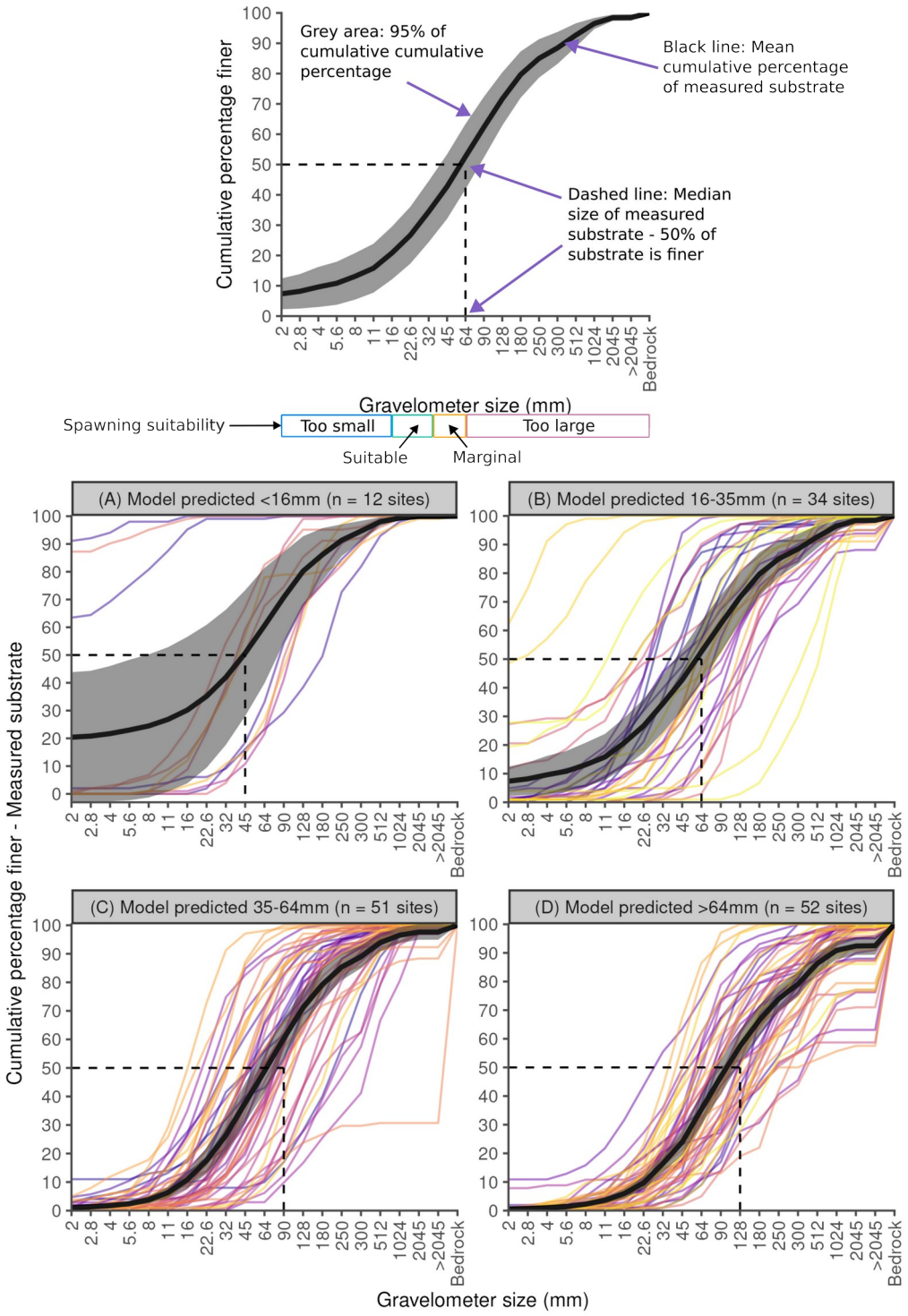


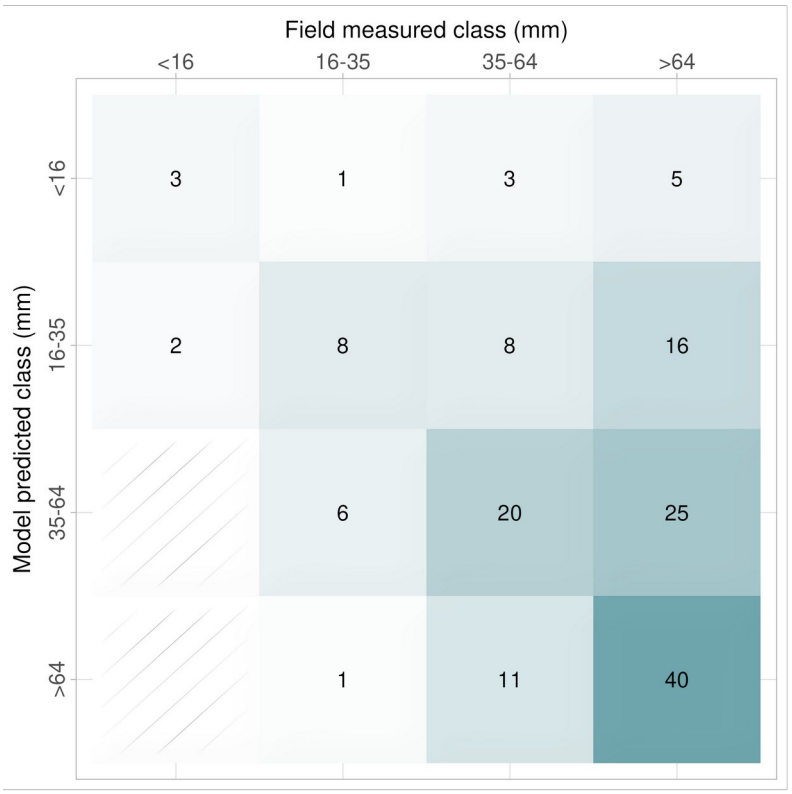

Figure 4. Cumulative percentage of substrate finer than each gravelometer size bin at 149 sites in the Exploits River watershed. Sites are grouped by the size class of the model-predicted median substrate size ($D_{50}$). Coloured lines in panels A-D are the distribution of field-measured substrate from Wolman pebble counts with each line representing one site. Twelve, 34, 51, and 52 sites were sampled in model-predicted substrate classes <16 mm, 16-35 mm, 35-64 mm, >64 mm, respectively. The top panel is to

guide interpretation of panels A-D. The solid black line and grey shaded area are the mean and 95% confidence interval (CI), respectively, of the cumulative percentage of field-measured substrate sizes within each model-predicted class. The black dashed lines represent the median field-measured $D_{50}$ of all sites within the class. For example, in Panel B, we had 34 field sites that were sampled at segments with a model predicted $D_{50}$ of 16-35 mm. The field-measured $D_{50}$ for these sites = 64 mm with a 95% CI = 45 mm, 90 mm. This shows that the model predicts lower $D_{50s}$ than observed in the field. The spawning suitability of substrate sizes are shown in the interpretation panel.

The overall accuracy for our model was 0.477, meaning almost half the time (48%) it predicted correctly, either true positive or true negative (Figure 5). True positives, where the model predicted the correct substrate size measured, are shown from the top left to bottom right diagonal of the confusion matrix (Figure 5). Higher numbers here indicate accurate predictions and thus a stronger model. Overall, specificity (0.82) was higher than sensitivity (0.51), meaning the model was better at correctly detecting the negative class compared to the positive class (Figure 5). For example, if the model predicted that the $D_{50}$ was not <16 mm and the measured substrate was also not <16 mm, it correctly detected the negative class. A true positive detection is where both the predicted and measured classes were the same.

Precision, meaning the accuracy of positive predictions, in the overall model was moderate (0.41; Figure 5). Thus, the model incorrectly predicted some classes. The F1 score may be a better indicator of the overall performance of the model, being a harmonic mean between precision and sensitivity. As our sensitivity and precision were moderate, the F1 score was also moderate (0.42; Figure 5). However, there were significant differences between classes (p-value = 0.0001, Figure 5; Figure S3.1). While sensitivity in the >64 mm classes was lower than all other classes (0.465), the F1 score was higher (0.579), as the model was better at predicting correct values (0.769) in this class (Table S3.1). In the other direction, the <16 mm class had a higher sensitivity rate (0.6), but because of its low precision (0.25), the F1 score was lower (0.353; Table S3.1). The Cohen's Kappa score (0.239, Figure 5) showed fair agreement between model predictions and field observations.

We found predicted $D_{50s}$ generally increased with each subsequent measured class (Figure S3.1). We found significant differences between the <16 mm class and both of the larger substrate classes (35-64 mm and >64 mm), as well as between the 16-35 mm and >64 mm size classes (Kruskal Wallis test statistic = 25, p-value < 0.001, Dunn's test results in Table S3.2).




| Metric | Value |
|---|---|
| Accuracy | 0.477 |
| Accuracy 95% CI | 0.401, 0.566 |
| Balanced Accuracy | 0.663 |
| Sensitivity | 0.51 |
| Specificity | 0.815 |
| Precision | 0.412 |
| F1 | 0.421 |
| Kappa | 0.239 |
| P-value | 0.0001 |

Figure 5. Matrix of field-measured and model-predicted median substrate size ($D_{50}$) classes at field sites in the Exploits River watershed. The colour intensity is based on the number of sites, shown in the middle of each tile. Metric outputs for the overall confusion matrix model are presented on the right side of the figure and explained in the text. CI is the confidence interval. See Table S2.3 for further descriptions and bounds of metrics.

## Discussion

Concern grows over the decline of many salmonid species across the range of the species, including Atlantic salmon (Clarke & Scruton, 2002; Dempson et al., 2024; Limburg & Waldman, 2009; Soulsby et al., 2024). Spawning is a key life stage; lack of good quality habitat at this stage could be a major limitation for population productivity (Smialek et al., 2021). Identifying areas of high quality spawning habitat, however, is a major challenge, particularly at a watershed extent. Given river bed substrate is one of the key aspects of riverine spawning habitat (Gibson, 1993; Louhi et al., 2008), we modelled substrate size in a large watershed (~11,000km$^2$) to better inform our understanding of spawning habitat distribution. Our key goal was to test the reliability of a substrate model applied at a large watershed extent. We found that the model was better at predicting large substrate and often under-estimated substrate size relative to field observations. We discuss aspects of our modelling framework that led to successful predictions and consider additional steps that may improve future model predictions.

Based on model validation data, we can consider our model to have fair agreement with field observations, on average (Kappa = 0.239 is in the range of fair agreement (0.21 to 0.4)) (Landis & Koch, 1977). Overall, our model successfully predicted the field-measured substrate class about 42% of the time (precision, Figure 5). The model precision varied substantially across substrate classes. For instance, the model was better at correctly identifying substrates in the >64 mm class (precision = 0.77) compared to the other three classes (precision = 0.25, 0.24, 0.41 for <16 mm, 16-35 mm, 35-64 mm, respectively; Table S3.1).

The island of Newfoundland has some of the Earth's most ancient geological formations (Twenhofel, 1939). As such, large weathered substrate and bedrock (>64 mm) were very common at our field sites. Resistant bedrock may limit the supply of finer substrate (Curtis et al., 2010). In practice, this skewed distribution of field-observed substrate size led to an imbalance in specificity and sensitivity (true negative and true positive rates, respectively). For instance, we sampled more sites with a $D_{50}$ greater than 64 mm compared to other size classes. Ideally we would have sampled sites from all size classes evenly, but this was difficult due to site accessibility, time constraints, and a lack of prior knowledge about substrate distribution in the watershed.

In contrast to our field observations, our model predicted that the median substrate size suitable (16-35 mm) or marginal (35-64 mm) for spawning covered 58 per cent of segments in the watershed.

Model segments varied in length between 11 m and 6.2 km; each segment had one model prediction which is the expected median substrate size for the segment. At the segment extent in the field, however, we observed large variation in substrate sizes. Thus, the model predicted $D_{50}$ is a conservative estimate that may not reflect within segment heterogeneity. Higher resolution GIS data (e.g., 1 m or 5 m) may improve predictions of spatial heterogeneity in substrate size by generating shorter stream segments and, therefore, a greater number of model predictions. We might better interpret this result as 58 per cent of segments are likely to have suitable or marginal substrate for spawning somewhere at the segment scale. What remains is to understand how aggregated suitable spawning substrate is in order to find sites where there is enough gravel coverage to facilitate spawning (Overstreet et al., 2016; Riebe et al., 2014). For example, larger anadromous female Atlantic salmon require around 1-5 $m^2$ of spawning gravel to construct a redd (Barlaup et al., 2008). By knowing how aggregated suitable substrate is, we could then estimate the capacity of river sections for spawning (for example see Overstreet et al., 2016; Riebe et al., 2014).

While the model predicted that median substrate size in the 16-35 mm range covered the greatest proportion of the watershed (Table 3), model precision was lowest for this class (Table S3.1). For almost half of the sites (16 of 34) that the model predicted a $D_{50}$ in the 16-35 mm class, field measurements of $D_{50}$ were greater than 64 mm (Figure 5). In fact, the model under-estimated substrate size in the three lowest size classes (Figure 4), indicating the model may perform poorly in these finer substrate classes. Importantly though, model predictions of substrate size increased as field measurements increased (Figure 4). All of the sites that had a low measured $D_{50}$ (i.e., fine substrate) were in the lowest two model-predicted classes (<16 mm and 16-35 mm), while the model predicted only one segment with a $D_{50}$ less than 35 mm at sites with a high proportion of large boulders and bedrock (Figure 5). This gives us confidence that areas where the model-predicted substrate sizes were greater than 64 mm are unlikely to provide suitable spawning habitat.

While there are few examples of substrate models applied to large watershed extents, we can compare our results to other models used to predict substrate size in smaller watersheds and river segments. Wilkins and Snyder (2011) used shear stress equations to successfully predict $D_{50}$ at 62-70 per cent of sites across two North American Atlantic coastal streams. Similar to our findings, model performance improved at sites with coarser substrates, with stronger agreement between predicted and observed values when the $D_{50}$ exceeded 64 mm. In contrast, their model generally over-predicted $D_{50}$, such that field-measured substrate sizes were finer than predicted values. In our study, field

measurements were typically coarser than model predictions. Building on their sites, Snyder et al. (2013) applied three different models across four rivers in Maine and reliably predicted $D_{50}$ at approximately 70 per cent of sites. Likewise, model performance improved in coarser-bed reaches ($D_{50}$ >16 mm), whereas model failures were more common in fine-grained (<16 mm) depositional reaches (Snyder et al., 2013). Similarly, Gorman et al. (2011) accurately predicted median substrate size across 20 stream reaches in northeastern Ohio, USA using GIS-derived variables to estimate stream power. The high success rate of these studies may partly reflect the relatively small spatial extents over which the models were applied: all of the study watersheds were less than 700 $km^2$. Furthermore, most studies used higher resolution DEMs (1-10 m) to calculate stream slope, a key predictor variable. Fine-resolution data are often only available at smaller spatial extents, but where they are available, they can better capture channel morphology and local hydraulic variability. For example, Hedger et al. (2006) used high-resolution imagery (3 cm) to accurately predict substrate size within relatively small study sites (90 m x 50 m) in the Sainte-Marguerite, Quebec. Applying this level of spatial resolution across larger extents, however, would likely be impractical due to data and processing demands. Nonetheless, incorporating higher-resolution data would allow us to fine tune predictions to an area more relevant to the resolution of field data collection and of salmon spawning, i.e. 1-100 m rather than 100-1000 m.

The model approach requires several parameters that are difficult to estimate, for example, the Shields parameter ($\tau^*$) and Manning's hydraulic roughness coefficient ($n$). We used values similar to other studies but future work could attempt to estimate these parameters with data. For instance, the availability of spawning gravel may be heavily influenced by hydraulic roughness ($n$) (Buffington et al., 2004). We used a standard value of $n$ across our watershed, but roughness varies at the reach or even site extent. More information is required at this extent to assign a reliable $n$ value (Aldridge & Garrett, 1973; Barnes, 1967). To improve model performance we could estimate $n$ at field sites and extrapolate across the watershed. In areas where hydraulic roughness may be limiting spawning gravel quantity, the addition of large wood could increase flow resistance and aid in textual fining of the bed substrate (Buffington et al., 2004).

Another consideration to improve model performance is to investigate factors that may be affecting sediment supply and transport. Lakes, barriers, land cover, and both local and catchment geology can affect sediment dynamics within a watershed. For example, Wilkins and Snyder (2011) found that lakes and the legacy effects of dams altered sediment flow in the Narraguagus watershed in Maine, contributing to discrepancies between modelled and field-measured $D_{50}$ values. Similarly, the

Exploits River watershed contains multiple dams and nearly 1000 lakes, many of which likely function as sediment sinks. Dams alter downstream flow regimes and disrupt sediment transport processes (Brandt, 2000; Curtis et al., 2010; Graf, 1999; Kondolf et al., 2014; Nilsson et al., 2005). Large reservoirs can trap nearly 100 per cent of incoming sediment, substantially reducing downstream sediment supply (Kondolf et al., 2014; Williams & Wolman, 1984). Sediment starvation downstream of dams often leads to channel incision, bed coarsening, and changes in channel slope (Brandt, 2000; Kondolf et al., 2014), resulting in reduced sediment mobility and larger median substrate sizes in regulated reaches (Cheng & Granata, 2007; Curtis et al., 2010). To a lesser degree, coarsening may occur downstream of lakes (Purchase, 2026), but to a lesser degree. Historical logging practices may also continue to influence riverine habitat conditions in the watershed. The Exploits River and its tributaries were widely used for log transport, often requiring the construction of dams to control flow and store logs (Scruton et al., 1998). Log driving typically involved the removal of natural channel obstructions, including boulders, large rocks, trees, or sunken logs, as well as more extensive channel modification (Sedell et al., 1991). These activities frequently altered stream morphology through channel scouring, channel widening, and gravel displacement (Sedell et al., 1991). Much of the associated infrastructure was abandoned and left to deteriorate in place (Scruton et al., 1998). The Exploits River may be still recovering from previous alterations in sedimentation and flow and may have not yet reached an equilibrium state. The predicted $D_{50}$ values from our model may better reflect conditions expected under a more natural or consistent sediment regime, without major disruptions to sediment transport.

Exploring and comparing alternative model approaches may also provide a useful direction for future research. Regression-based and summed normal distribution models that incorporate environmental variables have been used to estimate bed substrates with varying levels of success (Haddadchi et al., 2018; Snelder et al., 2011). For example, Snelder et al. (2011) applied random forest models to predict substrate size in rivers across France. Although they found high uncertainty at the reach scale, their model was useful for basin-scale assessments of substrate and therefore showed promise for watershed management and conservation planning. Consistent with our results and other studies (e.g. Snyder et al., 2013; Wilkins & Snyder, 2011), they found greater uncertainty in finer substrate classes (Snelder et al., 2011). With sufficient data, these modelling approaches could be applied at regional or even national extents; however, their applicability may be limited in areas where environmental data are sparse or unavailable (Haddadchi et al., 2018).

We set out to build models to predict salmon spawning substrate at large watershed extents, aligning with the extent at which restoration and conservation planning is increasingly being conducted or recommended (Biron et al., 2014; Hermoso et al., 2012; Leroux & Rayfield, 2014; Roni et al., 2002). Effective conservation and restoration planning requires managing populations at the spatial extent over which they occur (Linke et al., 2011). Given upstream processes strongly influence downstream habitat conditions, adopting a whole-watershed perspective is particularly important (Allan, 2004; Fuller et al., 2015; Wohl, 2019). This is especially relevant for anadromous salmon, which can migrate long distances upstream to spawn. Given the cumulative influence of watershed processes on downstream habitats, lower river reaches are often more influenced by anthropogenic disturbances such as agriculture, forestry, and development (Allan, 2004; Dudgeon et al., 2006; Maloney & Weller, 2011). Consequently, salmon may migrate farther upstream to access higher-quality spawning habitat, either as a preferential habitat selection strategy or because suitable habitats in lower reaches have been degraded (Thorstad et al., 2008). Understanding habitat distribution at the watershed extent allows us to target conservation and restoration efforts where they are likely to have the greatest impact. It can also help us to identify drivers of habitat loss, for example, by spatially correlating habitat and watershed disturbances over time.

Substrate size is an important part of spawning habitat, but it is only one component. We must also consider gravel quality and whether it is accessible to anadromous spawning salmon. Small amounts of fine sediment (<2 mm) within spawning gravels can significantly reduce egg survival (Soulsby et al., 2001), while small, disconnected patches of gravel may provide insufficient habitat for female salmon to construct redds and spawn successfully (Overstreet et al., 2016; Riebe et al., 2014). Additionally, stream flow, depth, temperature, cover, and connectivity are all important habitat aspects of spawning success (Armstrong et al., 2003; Louhi et al., 2008). Our substrate model can be used as an input for larger spawning habitat models as well as help inform habitat distribution for other life stages where substrate size is important, such as rearing (Gibson, 1993).

Our model is valuable because it operates at the watershed scale and uses readily available data. This makes it practical for broad-scale application by managers and conservationists. Furthermore, it is applied at the extent to which Atlantic salmon migrate. While we can narrow down potential spawning areas by excluding segments with predicted $D_{50}$ values greater than 64 mm, additional refinement is needed to improve predictions in smaller substrate size classes. Despite these limitations, our model provides an important step toward watershed-scale identification of potential spawning habitat. By

identifying areas that are more likely to provide suitable spawning substrate, it can help guide more efficient restoration and conservation planning.

# References


Aldridge, B. N., & Garrett, J. M. (1973). *Roughness coefficients for stream channels in Arizona* (p. 87) [Open-File Report]. U.S. Geological Survey.

Allan, J. D. (2004). Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics*, *35*(1), Article 1. https://doi.org/10.1146/annurev.ecolsys.35.120202.110122

Andrews, E. D. (1984). Bed-material entrainment and hydraulic geometry of gravel-bed rivers in Colorado. *GSA Bulletin*, *95*(3), 371–378. https://doi.org/10.1130/0016-7606(1984)95<371:BEAHGO>2.0.CO;2

Armstrong, J. D., Kemp, P. S., Kennedy, G. J. A., Ladle, M., & Milner, N. J. (2003). Habitat requirements of Atlantic salmon and brown trout in rivers and streams. *Fisheries Research*, *62*(2), Article 2. https://doi.org/10.1016/S0165-7836(02)00160-1

Arp, C. D., Schmidt, J. C., Baker, M. A., & Myers, A. K. (2007). Stream geomorphology in a mountain lake district: Hydraulic geometry, sediment sources and sinks, and downstream lake effects. *Earth Surface Processes and Landforms*, *32*(4), 525–543. https://doi.org/10.1002/esp.1421

Bardonnet, A., & Baglinière, J.-L. (2000). Freshwater habitat of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, *57*(2), 497–506. https://doi.org/10.1139/f99-226

Barlaup, B. T., Gabrielsen, S. E., Skoglund, H., & Wiers, T. (2008). Addition of spawning gravel—A means to restore spawning habitat of Atlantic salmon (*Salmo salar L.*), and Anadromous and resident brown trout (Salmo trutta L.) in regulated rivers. *River Research and Applications*, *24*(5), 543–550. https://doi.org/10.1002/rra.1127

Barnes, H. H. Jr. (1967). *Roughness characteristics of natural channels* (U.S. Geological Survey Water-Supply Paper 1849, p. 213). U.S. Geological Survey.

Behnke, R. J. (2002). *Trout and salmon of North America*. Fieldstone Publishing Inc.

Biron, P. M., Buffin-Bélanger, T., Larocque, M., Choné, G., Cloutier, C.-A., Ouellet, M.-A., Demers, S., Olsen, Y., Desjarlais, C., & Eyquem, J. (2014). Freedom Space for Rivers: A Sustainable Management Approach to Enhance River Resilience. *Environmental Management*, *54*(5), Article 5. https://doi.org/10.1007/s00267-014-0366-z

Brandt, S. A. (2000). Classification of geomorphological effects downstream of dams. *CATENA*, *40*(4), 375–401. https://doi.org/10.1016/S0341-8162(00)00093-X

Buffington, J. M., & Montgomery, D. R. (1997). A systematic analysis of eight decades of incipient motion studies, with special reference to gravel-bedded rivers. *Water Resources Research*, *33*(8), 1993–2029. https://doi.org/10.1029/96WR03190

Buffington, J. M., & Montgomery, D. R. (1999). Effects of hydraulic roughness on surface textures of gravel-bed rivers. *Water Resources Research*, *35*(11), 3507–3521. https://doi.org/10.1029/1999WR900138

Buffington, J. M., Montgomery, D. R., & Greenberg, H. M. (2004). Basin-scale availability of salmonid spawning gravel as influenced by channel type and hydraulic roughness in mountain catchments. *Canadian Journal of Fisheries and Aquatic Sciences*, *61*(11), 2085–2096. https://doi.org/10.1139/f04-141

Carbonneau, P. E., Lane, S. N., & Bergeron, N. E. (2004). Catchment-scale mapping of surface grain size in gravel bed rivers using airborne digital imagery. *Water Resources Research*, *40*(7). https://doi.org/10.1029/2003WR002759

Cheng, F., & Granata, T. (2007). Sediment transport and channel adjustments associated with dam removal: Field observations. *Water Resources Research*, *43*(3). https://doi.org/10.1029/2005WR004271

Clarke, K., & Scruton, D. (2002). Evaluating efforts to increase salmonid productive capacity through habitat enhancement in the low diversity/production systems of Newfoundland, Canada. *13th International Salmonid Habitat Enhancement Workshop*, 160–182.

COSEWIC. (2010). *COSEWIC assessment and status report on the Atlantic Salmon* Salmo salar *(Nunavik population, Labrador population, Northeast Newfoundland population, South Newfoundland population, Southwest Newfoundland population, Northwest Newfoundland population, Quebec Eastern North Shore population, Quebec Western North Shore population, Anticosti Island population, Inner St. Lawrence population, Lake Ontario population, Gaspé-Southern Gulf of St. Lawrence population, Eastern Cape Breton population, Nova Scotia Southern Upland population, Inner Bay of Fundy population, Outer Bay of Fundy population) in Canada.* (p. xlvii + 136). Committee on the Status of Endangered Wildlife in Canada.

Crisp, D. T. (2000). *Trout and Salmon. Ecology, Conservation and Rehabilitation*. Blackwell Science.

Curtis, K. E., Renshaw, C. E., Magilligan, F. J., & Dade, W. B. (2010). Temporal and spatial scales of geomorphic adjustments to reduced competency following flow regulation in bedload-dominated systems. *Geomorphology, 118*(1), 105–117. https://doi.org/10.1016/j.geomorph.2009.12.012

Daniels, M. D., & McCusker, M. H. (2010). Operator bias characterizing stream substrates using Wolman pebble counts with a standard measurement template. *Geomorphology, 115*(1), 194–198. https://doi.org/10.1016/j.geomorph.2009.09.038

Dauwalter, D. C., Duchi, A., Epifanio, J., Gandolfi, A., Gresswell, R., Juanes, F., Kershner, J., Lobón-Cerviá, J., McGinnity, P., Meraner, A., Mikheev, P., Morita, K., Muhlfeld, C. C., Pinter, K., Post, J. R., Unfer, G., Vøllestad, L. A., & Williams, J. E. (2020). A call for global action to conserve native trout in the 21st century and beyond. *Ecology of Freshwater Fish*, *29*, 429–432.

Davis, J. P., & Riche, D. (1983). *Exploits River salmon development program* (p. 47). Department of Fisheries and Oceans.

Dawson, M. (1988). Sediment size variation in a braided reach of the Sunwapta River, Alberta, Canada. *Earth Surface Processes and Landforms*, *13*(7), 599–618. https://doi.org/10.1002/esp.3290130705

Deinet, S., Flint, R., Puleston, H., Baratech, A., Royte, J., Thieme, M. L., Nagy, S., Hogan, Z. S., Januchowski-Hartley, S., & Wanningen, H. (2024). *The Living Planet Index (LPI) for migratory freshwater fish 2024 update*. World Fish Migration Foundation.

Dempson, J. B., Van Leeuwen ,Travis E., Bradbury ,Ian R., Lehnert ,Sarah J., Coté ,David, Cyr ,Frédéric, Pretty ,Christina, & and Kelly, N. I. (2024). A Review of Factors Potentially Contributing to the Long-Term Decline of Atlantic Salmon in the Conne River, Newfoundland, Canada. *Reviews in Fisheries Science & Aquaculture*, *32*(3), 479–504. https://doi.org/10.1080/23308249.2024.2341023

Department of Fisheries and Oceans. (2022). *Stock assessment of Newfoundland and Labrador Atlantic salmon in 2020.* (Science Advisory Report, 1919-5087; 2022/031, p. 35). Canadian Science Advisory Secretariat. https://publications.gc.ca/site/eng/9.913080/publication.html

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, *81*(2), Article 2. https://doi.org/10.1017/S1464793105006950

Fisheries and Oceans Canada. (2025). *2024 Stock status update of Atlantic salmon in Newfoundland and Labrador* (p. 32) [Canadian Science Advisory Secretariat Science Response 2025/007]. Fisheries and Oceans Canada.

Fleming, I. A. (1996). Reproductive strategies of Atlantic salmon: Ecology and evolution. *Reviews in Fish Biology and Fisheries*, *6*(4), 379–416. https://doi.org/10.1007/BF00164323

Foote, K. J., Grant, J. W. A., & Biron, P. M. (2025). Salmonid Biomass in Streams Around the World: A Quantitative Synthesis. *Fish and Fisheries*, *26*(3), 394–413. https://doi.org/10.1111/faf.12887

Fox, G. A., Sheshukov, A., Cruse, R., Kolar, R. L., Guertault, L., Gesch, K. R., & Dutnell, R. C. (2016). Reservoir Sedimentation and Upstream Sediment Sources: Perspectives and Future Research Needs on Streambank and Gully Erosion. *Environmental Management*, *57*(5), 945–955. https://doi.org/10.1007/s00267-016-0671-9

Fuller, M. R., Doyle, M. W., & Strayer, D. L. (2015). Causes and consequences of habitat fragmentation in river networks. *Annals of the New York Academy of Sciences*, *1355*(1), 31–51. https://doi.org/10.1111/nyas.12853

Gibson, R. J. (1993). The Atlantic salmon in fresh water: Spawning, rearing and production. *Reviews in Fish Biology and Fisheries*, *3*(1), 39–73. https://doi.org/10.1007/BF00043297

Gibson, R. J., & Haedrich, R. (2006). Life history tactics of Atlantic salmon in Newfoundland. *Freshwater Forum*, *26*, 38–45.

Gorman, A. M., Whiting, P. J., Neeson, T. M., & Koonce, J. F. (2011). Channel substrate prediction from GIS for habitat estimation in Lake Erie tributaries. *Journal of Great Lakes Research*, *37*(4), 725–731. https://doi.org/10.1016/j.jglr.2011.08.008

Government of Canada, Natural Resources Canada, Strategic Policy and Innovation Sector, Canada Centre for Mapping and Earth Observation, & Geobase. (2022). *National Hydro Network—NHN - GeoBase Series* [Dataset]. https://maps.geogratis.gc.ca/wms/hydro_network_en?version=1.3.0&legend_format=image%2Fpng&feature_info_type=text%2Fhtml&request=getcapabilities&service=wms

Graf, W. L. (1999). Dam nation: A geographic census of American dams and their large-scale hydrologic impacts. *Water Resources Research*, *35*(4), 1305–1311. https://doi.org/10.1029/1999WR900016

Haddadchi, A., Booker, D. J., & Measures, R. J. (2018). Predicting river bed substrate cover proportions across New Zealand. *CATENA*, *163*, 130–146. https://doi.org/10.1016/j.catena.2017.12.014

Hauer, C., Leitner, P., Unfer, G., Pulg, U., Habersack, H., & Graf, W. (2018). The Role of Sediment and Sediment Dynamics in the Aquatic Environment. In S. Schmutz & J. Sendzimir (Eds.), *Riverine Ecosystem Management: Science for Governing Towards a Sustainable Future* (pp. 151–169). Springer International Publishing. https://doi.org/10.1007/978-3-319-73250-3_8

Hauer, C., Pulg, U., Reisinger, F., & Flödl, P. (2020). Evolution of artificial spawning sites for Atlantic salmon (*Salmo salar*) and sea trout (*Salmo trutta*): Field studies and numerical modelling in Aurland, Norway. *Hydrobiologia*, *847*(4), 1139–1158. https://doi.org/10.1007/s10750-019-04173-1

Hedger, R. D., Dodson, J. J., Bourque, J.-F., Bergeron, N. E., & Carbonneau, P. E. (2006). Improving models of juvenile Atlantic salmon habitat use through high resolution remote sensing. *Ecological Modelling*, *197*(3), 505–511. https://doi.org/10.1016/j.ecolmodel.2006.03.028

Heggenes, J. (1990). Habitat utilization and preferences in juvenile Atlantic salmon (*Salmo salar*) in streams. *Regulated Rivers: Research & Management*, *5*, 341–354. https://doi.org/10.1002/rrr.3450050406

Hermoso, V., Kennard, M. J., & Linke, S. (2012). Integrating multidirectional connectivity requirements in systematic conservation planning for freshwater systems. *Diversity and Distributions*, *18*(5), 448–458. https://doi.org/10.1111/j.1472-4642.2011.00879.x

Javernick, L., Brasington, J., & Caruso, B. (2014). Modeling the topography of shallow braided rivers using Structure-from-Motion photogrammetry. *Geomorphology*, *213*, 166–182. https://doi.org/10.1016/j.geomorph.2014.01.006

Jelks, H. L., Walsh, S. J., Burkhead, N. M., Contreras-Balderas, S., Diaz-Pardo, E., Hendrickson, D. A., Lyons, J., Mandrak, N. E., McCormick, F., Nelson, J. S., Platania, S. P., Porter, B. A., Renaud, C. B., Schmitter-Soto, J. J., Taylor, E. B., & Warren, M. L. (2008). Conservation Status of Imperiled North American Freshwater and Diadromous Fishes. *Fisheries*, *33*(8), Article 8. https://doi.org/10.1577/1548-8446-33.8.372

Jonsson, B., & Jonsson, N. (2007). Life History of the Anadromous Trout *Salmo trutta*. In *Sea Trout: Biology, Conservation and Management* (pp. 196–223). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470996027.ch14

Jonsson, B., & Jonsson, N. (2011). *Ecology of Atlantic Salmon and Brown Trout: Habitat as a template for life histories*. Springer Netherlands. https://doi.org/10.1007/978-94-007-1189-1

Joy, M. K., Foote, K. J., McNie, P., & Piria, M. (2018). Decline in New Zealand's freshwater fish fauna: Effect of land use. *Marine and Freshwater Research*, *70*(1), Article 1. https://doi.org/10.1071/MF18028

Kassambara, A. (2025). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (R package version 0.7.3). https://CRAN.R-project.org/package=rstatix

Kerr, J. T., & Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing. *Trends in Ecology & Evolution*, *18*(6), 299–305. https://doi.org/10.1016/S0169-5347(03)00071-5

Knighton, A. D. (1980). Longitudinal changes in size and sorting of stream-bed material in four English rivers. *GSA Bulletin*, *91*(1), 55–62. https://doi.org/10.1130/0016-7606(1980)91<55:LCISAS>2.0.CO;2

Kondolf, G. M., & Li, S. (1992). The Pebble Count Technique for Quantifying Surface Bed Material Size in Instream Flow Studies. *Rivers*, *3*(2), 80–87.

Kondolf, G. M., Rubin, Z. K., & Minear, J. T. (2014). Dams on the Mekong: Cumulative sediment starvation. *Water Resources Research*, *50*(6), 5158–5169. https://doi.org/10.1002/2013WR014651

Kondolf, G. M., & Wolman, M. G. (1993). The sizes of salmonid spawning gravels. *Water Resources Research*, *29*(7), 2275–2285. https://doi.org/10.1029/93WR00402

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*, 1–26. https://doi.org/10.18637/jss.v028.i05

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*, 159–174.

Laramie, M. B., Pilliod, D. S., & Goldberg, C. S. (2015). Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biological Conservation*, *183*, 29–37. https://doi.org/10.1016/j.biocon.2014.11.025

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Lehner, B., Messager, M. L., Korver, M. C., & Linke, S. (2022). Global hydro-environmental lake characteristics at high spatial resolution. *Scientific Data*, *9*(1), 351. https://doi.org/10.1038/s41597-022-01425-z

Leroux, S. J., & Rayfield, B. (2014). Methods and tools for addressing natural disturbance dynamics in conservation planning for wilderness areas. *Diversity and Distributions*, *20*(3), 258–271. https://doi.org/10.1111/ddi.12155

Limburg, K. E., & Waldman, J. R. (2009). Dramatic Declines in North Atlantic Diadromous Fishes. *BioScience*, *59*(11), 955–965. https://doi.org/10.1525/bio.2009.59.11.7

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., & Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, *6*(1), 283. https://doi.org/10.1038/s41597-019-0300-6

Linke, S., Turak, E., & Nel, J. (2011). Freshwater conservation planning: The case for systematic approaches. *Freshwater Biology*, *56*(1), 6–20. https://doi.org/10.1111/j.1365-2427.2010.02456.x

Louhi, P., Mäki-Petäys, A., & Erkinaro, J. (2008). Spawning habitat of Atlantic salmon and brown trout: General criteria and intragravel factors. *River Research and Applications*, *24*(3), 330–339. https://doi.org/10.1002/rra.1072

Maloney, K. O., & Weller, D. E. (2011). Anthropogenic disturbance and streams: Land use and land-use change affect stream ecosystems via multiple pathways. *Freshwater Biology*, *56*(3), 611–626. https://doi.org/10.1111/j.1365-2427.2010.02522.x

McDowall, R. M. (1999). Different kinds of diadromy: Different kinds of conservation problems. *Journal of Marine Science*, *56*, 410–413. https://doi.org/10.1006/jmsc.1999.0450

Mueller, E. R., Pitlick, J., & Nelson, J. M. (2005). Variation in the reference Shields stress for bed load transport in gravel-bed streams and rivers. *Water Resources Research*, *41*(4). https://doi.org/10.1029/2004WR003692

Mullms, C. C., Bourgeois, C. E., & Porter, T. R. (2007). *Opening Up New Habitat: Atlantic Salmon (Salmo salar L.) Enhancement in Newfoundland*. 200–221. https://doi.org/10.1002/9780470995495.ch17

Natural Resources Canada. (2025a). *High Resolution Digital Elevation Model Mosaic (HRDEM Mosaic)—CanElevation Series* [Dataset]. https://open.canada.ca/data/en/dataset/0fe65119-e96e-4a57-8bfe-9d9245fba06b

Natural Resources Canada. (2025b). *Medium Resolution Digital Elevation Model (MRDEM)—CanElevation Series* [Dataset]. https://open.canada.ca/data/en/dataset/18752265-bda3-498c-a4ba-9dfe68cb98da

Newfoundland and Labrador Geological Survey. (2024). *Newfoundland and Labrador GeoScience Atlas OnLine* [Dataset]. https://geoatlas.gov.nl.ca/Default.htm

Nilsson, C., Reidy, C. A., Dynesius, M., & Revenga, C. (2005). Fragmentation and Flow Regulation of the World's Large River Systems. *Science*, *308*(5720), Article 5720. https://doi.org/10.1126/science.1107887

Noble, M., Duncan, P., Perry, D., Prosper, K., Rose, D., Schnierer, S., Tipa, G., Williams, E., Woods, R., & Pittock, J. (2016). Culturally significant fisheries: Keystones for management of freshwater social-ecological systems. *Ecology and Society*, *21*(2). https://www.jstor.org/stable/26270409

O'Connell, M. F., & Bourgeois, C. E. (1987). Atlantic Salmon Enhancement in the Exploits River, Newfoundland, 1957-1984. *North American Journal of Fisheries Management*, *7*(2), 207–214. https://doi.org/10.1577/1548-8659(1987)7<207:ASEITE>2.0.CO;2

O'Connell, M. F., Dempson, J. B., & Chaput, G. (2006). *Aspects of the life history, biology, and population dynamics of Atlantic salmon (Salmo salar L.) in Eastern Canada* (No. 2006/014; p. 51). Fisheries and Oceans Canada.

Olsen, L. R., & Zachariae, H. B. (2025). *cvms: Cross-Validation for Model Selection*. https://CRAN.R-project.org/package=cvms

Osborne, P. e., Alonso, J. c., & Bryant, R. g. (2001). Modelling landscape-scale habitat use using GIS and remote sensing: A case study with great bustards. *Journal of Applied Ecology*, *38*(2), 458–471. https://doi.org/10.1046/j.1365-2664.2001.00604.x

Overstreet, B. T., Riebe, C. S., Wooster, J. K., Sklar, L. S., & Bellugi, D. (2016). Tools for gauging the capacity of salmon spawning substrates. *Earth Surface Processes and Landforms*, *41*(1), 130–142. https://doi.org/10.1002/esp.3831

Pearson, E., Smith, M. W., Klaar, M. J., & Brown, L. E. (2017). Can high resolution 3D topographic surveys provide reliable grain size estimates in gravel bed rivers? *Geomorphology*, *293*, 143–155. https://doi.org/10.1016/j.geomorph.2017.05.015

Powell, D. M. (1998). Patterns and processes of sediment sorting in gravel-bed rivers. *Progress in Physical Geography: Earth and Environment*, *22*(1), 1–32. https://doi.org/10.1177/030913339802200101

Pulg, U., Barlaup, B. T., Sternecker, K., Trepl, L., & Unfer, G. (2013). Restoration of Spawning Habitats of Brown Trout (*Salmo trutta*) in a Regulated Chalk Stream. *River Research and Applications*, *29*(2), 172–182. https://doi.org/10.1002/rra.1594

Purchase, C. P. (2026). *A systematic review on the effectiveness of salmonid spawning habitat improvements, and recommendations to potentially increase productivity of depressed Newfoundland Atlantic salmon (*Salmo salar*) populations*. EcoEvoRxiv. https://doi.org/10.32942/X2V673

Quinn, J. M., & Hickey, C. W. (1994). Hydraulic parameters and benthic invertebrate distributions in two gravel-bed New Zealand rivers. *Freshwater Biology*, *32*(3), 489–500. https://doi.org/10.1111/j.1365-2427.1994.tb01142.x

R Core Team. (2025). *R: A language and environment for statistical computing (Version 4.5.2) [R Foundation for Statistical Computing*. https://www.R-project.org/

Reif, M. K., & Theel, H. J. (2017). Remote sensing for restoration ecology: Application for restoring degraded, damaged, transformed, or destroyed ecosystems. *Integrated Environmental Assessment and Management*, *13*(4), 614–630. https://doi.org/10.1002/ieam.1847

Riebe, C. S., Sklar, L. S., Overstreet, B. T., & Wooster, J. K. (2014). Optimal reproduction in salmon spawning substrates linked to grain size and fish length. *Water Resources Research*, *50*(2), 898–918. https://doi.org/10.1002/2013WR014231

Roni, P., Beechie, T., Bilby, R. E., Leonetti, F. E., Pollock, M. M., & Pess, G. R. (2002). A review of stream restoration techniques and a hierarchical strategy for prioritizing restoration in Pacific Northwest watersheds. *North American Journal of Fisheries Management*, *22*, 1–20. https://doi.org/10.1577/1548-8675(2002)022<0001:AROSRT>2.0.CO;2

Scruton, D. A., Anderson, T. C., & King, L. W. (1998). Pamehac Brook: A case study of the restoration of a Newfoundland, Canada, river impacted by flow diversion for pulpwood transportation. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *8*(1), 145–157. https://doi.org/10.1002/(SICI)1099-0755(199801/02)8:1<145::AID-AQC257>3.0.CO;2-7

Scruton, D. A., McKinley, R. S., Kouwen, N., Eddy, W., & Booth, R. K. (2003). Improvement and optimization of fish guidance efficiency (FGE) at a behavioural fish protection system for downstream migrating Atlantic salmon (<i> Salmo salar <i/>) smolts. *River Research and Applications*, *19*, 605–617. https://doi.org/10.1002/rra.735

Scruton, D. A., Pennell, C. J., Bourgeois, C. E., Goosney, R. F., King, L., Booth, R. K., Eddy, W., Porter, T. R., Ollerhead, L. M. N., & Clarke, K. D. (2008). Hydroelectricity and fish: A synopsis of comprehensive studies of upstream and downstream passage of anadromous wild Atlantic

salmon, Salmo salar, on the Exploits River, Canada. *Hydrobiologia*, *609*(1), 225–239. https://doi.org/10.1007/s10750-008-9410-4

Sedell, J. R., Leone, F. N., & Duval, W. S. (1991). Water Transportation and Storage of Logs. *American Fisheries Society Special Publication, Influences of Forest and Rangeland Management on Salmonid Fishes and Their Habitats*, *19*, 325–368.

Smialek, N., Pander, J., & Geist, J. (2021). Environmental threats and conservation implications for Atlantic salmon and brown trout during their critical freshwater phases of spawning, egg development and juvenile emergence. *Fisheries Management and Ecology*, *28*(5), 437–467. https://doi.org/10.1111/fme.12507

Snelder, T. H., Lamouroux, N., & Pella, H. (2011). Empirical modelling of large scale patterns in river bed surface grain size. *Geomorphology*, *127*(3), 189–197. https://doi.org/10.1016/j.geomorph.2010.12.015

Snyder, N. P., Nesheim, A. O., Wilkins, B. C., & Edmonds, D. A. (2013). Predicting grain size in gravel-bedded rivers using digital elevation models: Application to three Maine watersheds. *GSA Bulletin*, *125*(1–2), 148–163. https://doi.org/10.1130/B30694.1

Soulsby, C., Youngson, A. F., Moir, H. J., & Malcolm, I. A. (2001). Fine sediment influence on salmonid spawning habitat in a lowland agricultural stream: A preliminary assessment. *Science of The Total Environment*, *265*(1), 295–307. https://doi.org/10.1016/S0048-9697(00)00672-0

Soulsby, C., Youngson, A., & Webb, J. (2024). The ecohydrology of rewilding: A pressing need for evidence in the restoration of upland Atlantic salmon streams. *Hydrological Processes*, *38*(5), e15142. https://doi.org/10.1002/hyp.15142

Sutherland, A. B., Culp, J. M., & Benoy, G. A. (2010). Characterizing deposited sediment for stream habitat assessment. *Limnology and Oceanography: Methods*, *8*(1), 30–44. https://doi.org/10.4319/lom.2010.8.30

Thorstad, E. B., Bliss, D., Breau, C., Damon-Randall, K., Sundt-Hansen, L. E., Hatfield, E. M. C., Horsburgh, G., Hansen, H., Maoiléidigh, N. Ó., Sheehan, T., & Sutton, S. G. (2021). Atlantic

salmon in a rapidly changing environment—Facing the challenges of reduced marine survival and climate change. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *31*(9), 2654–2665. https://doi.org/10.1002/aqc.3624

Thorstad, E. B., Økland, F., Aarestrup, K., & Heggberget, T. G. (2008). Factors affecting the within-river spawning migration of Atlantic salmon, with emphasis on human impacts. *Reviews in Fish Biology and Fisheries*, *18*(4), 345–371. https://doi.org/10.1007/s11160-007-9076-4

Twenhofel, W. H. (1939). Newfoundland: Geology and Peoples. *Sigma Xi Quarterly*, *27*(2), 103–112, 121.

Valavanis, V. D., Pierce, G. J., Zuur, A. F., Palialexis, A., Saveliev, A., Katara, I., & Wang, J. (2008). Modelling of essential fish habitat based on remote sensing, spatial analysis and GIS. *Hydrobiologia*, *612*(1), 5–20. https://doi.org/10.1007/s10750-008-9493-y

Weiers, S., Bock, M., Wissen, M., & Rossner, G. (2004). Mapping and indicator approaches for the assessment of habitats at different scales using remote sensing and GIS methods. *Landscape and Urban Planning, Development of European Landscapes*, *67*(1), 43–65. https://doi.org/10.1016/S0169-2046(03)00028-8

Wilkins, B. C., & Snyder, N. P. (2011). Geomorphic comparison of two Atlantic coastal rivers: Toward an understanding of physical controls on Atlantic salmon habitat. *River Research and Applications*, *27*(2), 135–156. https://doi.org/10.1002/rra.1343

Williams, G. P., & Wolman, M. G. (1984). *Downstream Effects of Dams on Alluvial Rivers*. U.S. Government Printing Office.

Wilson, K. L., Bailey, C. J., Davies, T. D., & Moore, J. W. (2022). Marine and freshwater regime changes impact a community of migratory Pacific salmonids in decline. *Global Change Biology*, *28*(1), Article 1. https://doi.org/10.1111/gcb.15895

Wohl, E. (2019). Forgotten Legacies: Understanding and Mitigating Historical Human Alterations of River Corridors. *Water Resources Research*, *55*(7), Article 7. https://doi.org/10.1029/2018WR024433

Wolman, M. G. (1954). A method of sampling coarse river bed material. *American Geophysical Union Transactions*, *35*, 951–956.

Wolman, M. G., & Miller, J. P. (1960). Magnitude and Frequency of Forces in Geomorphic Processes. *The Journal of Geology*, *68*(1), 54–74.

# Supplementary material

## Predicting substrate size at a watershed scale to inform conservation planning for a declining salmonid species

Kyleisha J. Foote[1], Shawn J. Leroux[1], Ava J. Hart[1], Nick C. Murphy[1], Craig F. Purchase[1]

[1]Department of Biology, Memorial University of Newfoundland, 45 Arctic Ave, St. John's, NL A1C 5S7, Canada.

Corresponding author: K. J. Foote – kyleisha.foote@gmail.com

# Table of Contents

# S1. Study Area and GIS Data

## S1.1. Study area

Table S1.1. Characteristics of the Exploits River.

| Variable | Value |
|---|---|
| Watershed drainage area ($km^2$) | 11250 |
| Mean annual discharge[1] ($m^3/s$) | 305 |
| Maximum annual discharge[1] ($m^3/s$) | 668 |
| Mean annual precipitation[1] (mm) | 1215 |
| Mean channel gradient[2] | 0.0089 |

Data sources: [1]HydroSHEDS (Linke et al., 2019), note that the discharge is calculated for the Exploits River mouth; [2]GIS calculations (see section S1.2 Step 12), mean includes all river segments in the watershed.

## S1.2. GIS data processing and analysis

The following steps outline methods to create a river network using QGIS and obtain data for the GIS substrate model predictions.

1. We used the STAC API browser plugin in QGIS to connect to the NRCan CCMEO datacube and load the medium resolution Canada digital elevation model (MRDEM). We followed instructions in the product access guide (Natural Resources Canada, 2024) to download a 30 m DEM to cover the extent of the Exploits River watershed. We downloaded an area that covered a larger area than the watershed (Figure S1.1A). As the DEM was at a medium resolution (30 m), we wanted to ensure that river lines in the whole watershed would be included (for creating river lines in step 3). Rivers can be incorrectly excluded if the DEM boundaries are too close to the drainage divide, especially at coarse resolutions. The DEM will be clipped to the watershed later (red outlined polygon in Figure S1.1A).
2. Abrupt elevation changes between adjacent cells may occur in DEMs. To remedy this, we smoothed out elevation in the DEM by filling sinks (‘Raster Terrain Analysis – Fill Sinks (Wang & Liu)’ tool; output: filled DEM). A minimum slope of 0.001 degrees was used.
3. To obtain channel lines, we used the SAGA ‘Channel Network and Drainage Basins’ tool using the filled DEM with a threshold of 5. We saved the ‘DEM Channels’ output (blue and yellow lines in Figure S1.1B). The ‘DEM Channels’ output comprises ‘linestrings’ (segments or features) that represent a length of stream between two confluences, i.e. segments ended or started at every stream confluence as determined by the DEM. Therefore, segment lengths were

variable throughout the watershed. Using this output, we selected the most downstream river segment of the Exploits River (i.e. the mouth) and ran the 'flow trace' tool upstream. This tool selects all connecting segments that flow upstream or downstream (depending on selection). For the Exploits watershed, all segments were selected except for streams flowing towards a dam in the south of Victoria Lake which is diverting water to the south coast (yellow lines in Figure S1B). The resolution of the DEM meant that the vectorized Channels output was coarse, and as a result, streams that are not actually draining into the Exploits watershed were connected as if they were flowing into the watershed (see yellow lines outside the red boundary in Figure S1.1B). Additionally, many streams were digitized incorrectly in that segments representing streams often do not follow the actual streams, particularly for smaller streams, and segments did not always join at actual stream confluences (e.g. Figure S1.1D). Because of this, some manual correction was undertaken on the DEM Channels layer (see step 7).

4. Using Google Satellite imagery to view rivers and lakes, we modified the Exploits River watershed basin extent from BasinATLAS (level 5) to follow streams flowing into the watershed and remove areas that were not part of the drainage area (Figure S1.1C). We also did this at the sub-basin level using the BasinATLAS level 9 layer (Figure S1.2).
5. Layers from HydroSHEDS and the DEM were then clipped to the Exploits River watershed layer using 'Vector Geoprocessing – Clip Vector by Mask Layer' for the HydroSHEDS layers and 'Raster Extraction – Clip Raster by Mask Layer' for the DEM. Our outputs were $LakeATLAS_{ER}$, $RiverATLAS_{ER}$, and $DEM_{ER}$ (Figure S1.2) which will be used in later steps. We use the average annual maximum discharge, displayed in Figure S1.2, to predict substrate size.
6. All vector layers and the $DEM_{ER}$ were projected to a local coordinate reference system (CRS): EPSG:26921 - NAD83 / UTM zone 21N.
7. We clipped the DEM Channels output from step 3 to the modified Exploits watershed produced in step 4 (output: DEM $Channels_{ER}$ using 'Vector Geoprocessing – Clip Vector by Mask Layer'). This resulted in 13456 individual segments. The DEM $Channels_{ER}$ layer was then compared with water bodies in Google satellite imagery and major discrepancies (i.e. where channels flowed the wrong direction or were not part of the watershed) were modified (example in Figure S1.1D). These discrepancies largely occur because the elevation changes between stream lines are too small to be captured by the DEM. Channel networks created with finer resolution DEMs will result in reduced errors. We did few modifications on very small streams (i.e. first and second order), although stream lines tended to be more inaccurate at this size due to the coarse resolution. However, first and second order streams account for the majority of stream segment counts (73%), thus requiring much more processing time to modify. Furthermore, they are unlikely to support spawning salmon and are less important for our purpose. At this stage we kept diverted streams in the Victoria River watershed and ended with 13342 segments in the DEM $Channels_{ER}$ layer (Figure S1.1E).

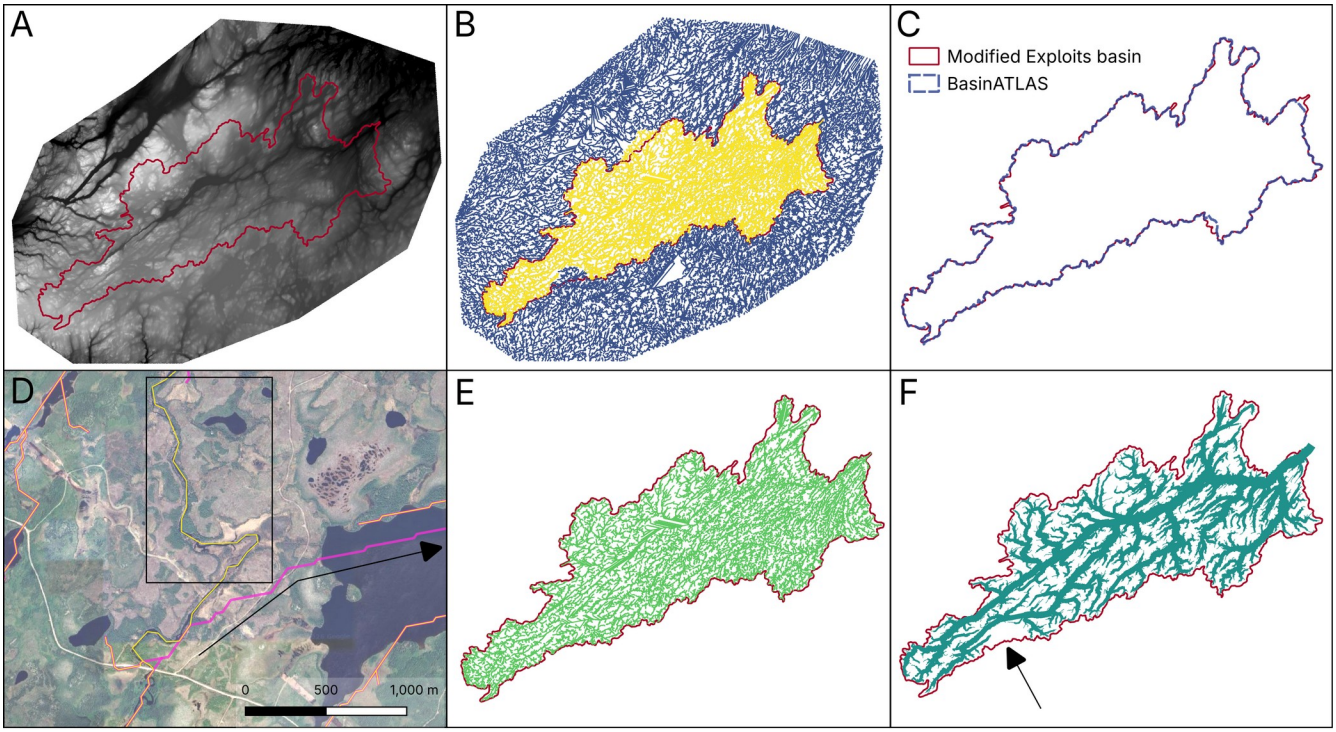



Figure S1.1. Examples of GIS processing for creating river channels. A: DEM of the watershed and surrounding area. B: Channels output derived from the DEM in A (blue and yellow lines). Yellow lines show all connected stream segments upstream from the Exploits River mouth, i.e. everything that flows to the mouth. C: Original BasinATLAS level 5 basin layer is shown by the blue dashed outline, while the red outline shows the modified Exploits River basin. D: Example of modifications made to the DEM ChannelsER output – the original DEM Channels output (pink line) has the stream incorrectly flowing to the right (following the arrow); we modified the segment to follow the channel shown in Google satellite imagery (mostly shown by the yellow line inside the black box). E: Modified DEM $Channels_{ER}$ output with 13342 segments (green lines). F: Modified DEM $Channels_{ER}$ layer displayed by Shreve order – larger segments are higher order streams. Streams flowing away from the watershed from Victoria Lake dam are not displayed (bottom right corner shown by the arrow) because stream order was not calculated as they do not flow to the Exploits River mouth.

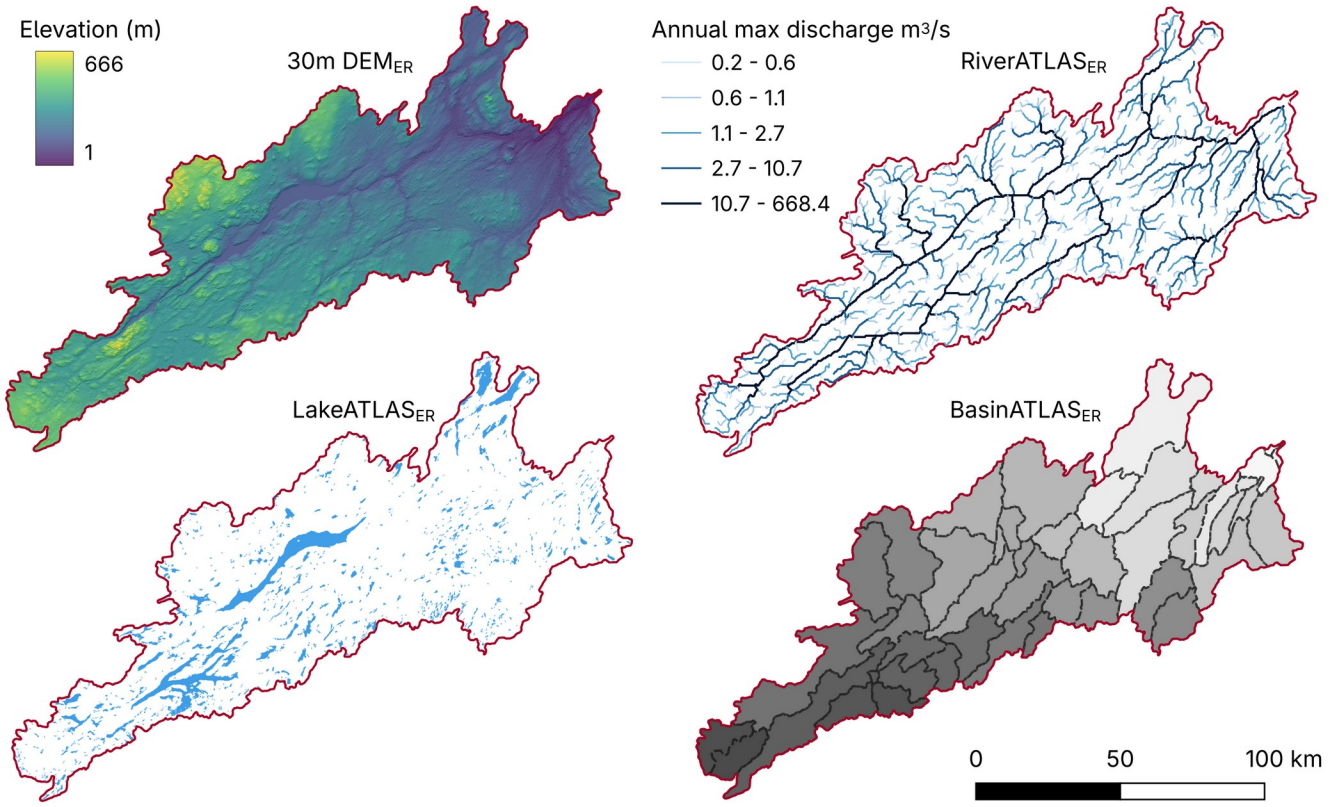



Figure S1.2. Clipped DEM and HydroSHEDS ATLAS layers used as data inputs and processing. Layer descriptions and sources are listed in Table 1 in the main manuscript. Lines in the RiverATLAS layer display the average annual maximum discharge, which we use to predict substrate size. We modified the BasinATLAS layer to represent drainage reflected by satellite imagery (see step 4).

8. We used the 'Lines Ranking' tool on DEM Channels$_{ER}$ to calculate stream orders. The Exploits River mouth was selected on the map as the Start Point Coordinate, thus, streams flowing downstream to this point will be included in the output. To check for major discrepancies in stream order we visually assessed whether stream order increased downstream by changing the symbology of the layer to display Shreve order by size (Figure S1.1F) and looking for obvious segments downstream that were larger than upstream. There were few (<10) instances where we modified stream order to ensure order was increasing downstream.
9. Because modifications were made to the length of segments in DEM Channels$_{ER}$, and thus coordinates for segments may have changed, we used the 'Vector Geometry – Fix Geometries' tool to update any geometries that had changed.
10. We calculated the length of segments using '@length' in a new column (Length) within the DEM Channels$_{ER}$ layer.

11. To get elevation data for each segment, we used the 'Vector Geometry – Drape (Set Z Value From Raster)' tool. The input layer was DEM $Channels_{ER}$ and the raster layer was $DEM_{ER}$ from step 5.
12. We use change in elevation to calculate channel slope for each segment in the DEM $Channels_{ER}$ layer. In the field calculator we created a new field (Slope) and used the formula: abs(z((start_point(@geometry)))-z((end_point(@geometry))))/Length. The start_point and end_point refers to the ends of the line segment; z is the elevation taken from the $DEM_{ER}$ in step 11; Length refers to the Length calculated in step 10; and abs converts the number to an absolute value to avoid negative slopes.
13. As the DEM $Channels_{ER}$ layer does not contain the data we use for substrate predictions (e.g. discharge), we join data from $RiverATLAS_{ER}$ onto the DEM $Channels_{ER}$ layer. Discharge and stream order are both representations of stream size - higher discharges generally correspond to higher stream orders. Thus, we assigned data from $RiverATLAS_{ER}$ to DEM $Channels_{ER}$ based on discharge (Figure S1.3) and order (Figure S1.3B) in the two layers, respectively. However, the layers are different resolutions, and segments do not overlap 100% (e.g. Figure S1.3C). Therefore, we need to proceed carefully to not assign discharge from $RiverATLAS_{ER}$ to the wrong DEM $Channels_{ER}$ segment. We do this in steps from largest to smallest discharge and then join the steps together at the end.
    a. Specifically, we categorize the data into graduated size classes for mean annual discharge ($RiverATLAS_{ER}$, Figure S1.3A) and Shreve order (DEM $Channels_{ER}$, Figure S1.3B) and match the segments in steps from the highest discharge/order to the smallest.
    b. Our highest category consists of segments with discharge of 180-306 $m^3/s$ and corresponding stream order of 2900-6596 (Figure S1.3D and S1.3E, respectively). The features of these categories in both layers were selected, and using the tool 'Vector General – Join Attributes by Nearest' we selected DEM $Channels_{ER}$ as the first input layer and $RiverATLAS_{ER}$ as input layer 2. For both layers the box 'Selected features only' was checked to ensure all features in the layer are not used. After running the tool, the resulting layer contains geometry and attributes from DEM $Channels_{ER}$ with attributes from $RiverATLAS_{ER}$ as added columns.
    c. We then selected features from the next category, order = 1100-2900 and discharge = 75-180 $m^3/s$ (Figure S1.3F and S1.3G, respectively), and followed the same steps.
    d. After these two categories, stream order and discharge did not follow segments consistently across the whole watershed. Therefore, to assign data from $RiverATLAS_{ER}$ that more accurately reflects the stream, we split segments from both layers into sub-basins or areas (basins shown in Figure S1.3J). This means that different basins have different order/discharge matches but are likely to reflect $RiverATLAS_{ER}$ data in their proximity.
    e. Matching order with discharge becomes more difficult with smaller stream orders as there are far fewer $RiverATLAS_{ER}$ segments. For instance, the DEM $Channels_{ER}$ layer contains 13456 segments, while the $RiverATLAS_{ER}$ layer contains 1083 segments. This means that,

when matched, many segments in the DEM Channels$_{ER}$ layer will contain the same data from Exploits RiverATLAS$_{ER}$. To reduce the number of segments with duplicate data, we excluded all first order (Shreve) streams in DEM Channels$_{ER}$. Furthermore, these first order streams are very small, often corresponding to ephemeral streams, and did not have corresponding segments in RiverATLAS$_{ER}$. The resulting layer contained 6529 segments.

f. An example of a sub-basin analysis using the Badger basin is shown in Figure S1.3H and S1.3I (discharge and stream order, respectively). Both layers in this basin were split into eight discharge/order classes and we proceeded by assigning each class separately (see Table S1.2). Categories were created iteratively by matching segments in similar proximity and/or size. This method was followed for each coloured basin in Figure S1.3J based on corresponding categories in Table S1.2.

g. Note on joining layers: The tool used to join layers, 'Vector General – Join Attributes by Nearest', joins layers based on the nearest proximity in distance. However, if the second joining layer has multiple features that are in equal distance from features in the first layer, several joins may occur. This results in multiple features with the same geometry but different attributes from the second joining layer. In our case, there may be duplicate segments that represent the same river stretch in DEM Channels$_{ER}$, but they will have different data attributes from RiverATLAS$_{ER}$. We remedy this problem in later steps (Step 18).

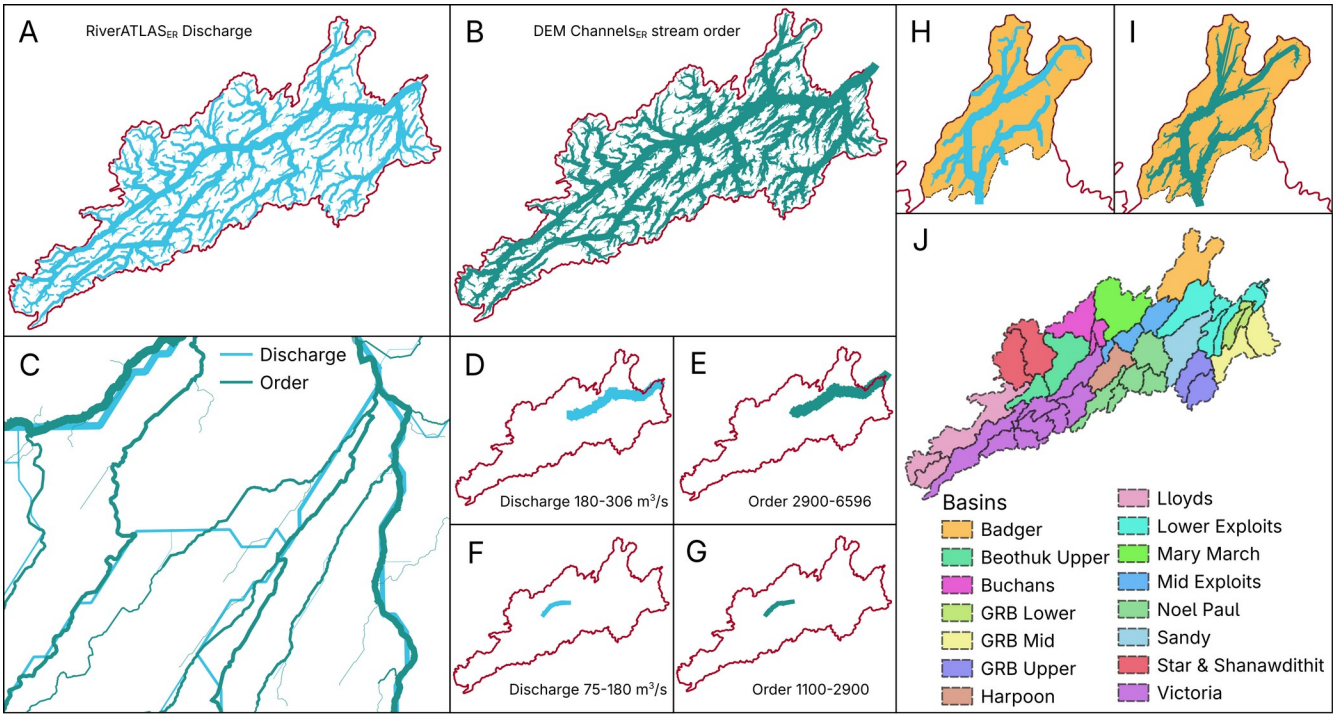



Figure S1.3. Examples of processing steps to join RiverATLAS$_{ER}$ data to DEM Channels$_{ER}$. A: RiverATLAS$_{ER}$ layer displayed by mean discharge with thicker lines representing higher discharges. B: Modified DEM Channels$_{ER}$ layer displayed by Shreve order (see Figure S1.1F for further explanation).

C: Example of how river lines from each layer are different, with fewer segments in the RiverATLAS$_{ER}$ layer (Discharge). D: Segments from DEM Channels$_{ER}$ with Shreve order between 2900 and 6596. E: Segments from RiverATLAS$_{ER}$ where discharge is between 180 and 306 $m^3/s$. Values in D were assigned to the nearest stream segments in E. F and G: Segments of RiverATLAS$_{ER}$ (F) and DEM Channels$_{ER}$ (G) that were matched following steps in D and E. H and I: Segments displayed by order and discharge, respectively, in the Badger basin (shown in J). Values in H were assigned to segments in I in multiple steps to avoid mis-assigning discharge to segments. J: Basins and areas displayed by colour that order/discharge joins were separated by.

14. The resulting layers (130 in total from matches in Table S1.2) from step 13 were merged using 'Vector General - Merge Vector Layers', creating a temporary layer. The layer was saved and under select fields to export, we unchecked boxes we did not require as the file exceeded the maximum number of attributes that can be saved as a shapefile. Note that the resulting merged layer contained 6683 features, indicating that there are duplicate features (there were 6529 features before discharge was attributed to the layer), which will be corrected in later steps. We visually checked that discharges were generally increasing downstream in the merged layer by displaying the layer by discharge with higher discharges represented by thicker lines (similar to Figure S1.3A). Inconsistencies can occur when joining layers if streams are selected in the wrong step for instance (step 13). Where we observed upstream discharges were higher than downstream (<10 instances), manual changes were made by increasing or decreasing discharge in the upstream or downstream segments, respectively. Generally, this involved switching values between upstream and downstream.
15. We predict substrate size in stream segments, not within lakes. In watersheds with few or no lakes, the following steps are not required. We excluded lake segments using the tool 'Vector Overlay – Difference' with the DEM Channels$_{ER}$ as the input layer and LakeATLAS$_{ER}$ polygon as the overlay layer. This tool cuts nodes from vectors inside polygons and deletes whole vectors if the entire line is inside a polygon, but does not delete the vectors if nodes exist outside of polygons. The resulting layer consisted of 5942 segments (Figure S1.4A).
16. The coarseness of the DEM Channels$_{ER}$ river segments means that there are many places where river lines do not cross through lake polygons, but rather run outside, i.e. they do not follow the rivers and lakes accurately (Figure S1.4B). Because of this, there were many fractured segments, i.e. nodes of vectors did not join or were in places with no rivers. To correct for this, we manually deleted vectors and/or nodes that represented lake segments (but that fell outside of polygons), and manually corrected stream segments that had major deviations from streams represented in Google Satellite imagery (Figure S1.4C), i.e., similar to step 7 in creating river networks. As this was conducted manually, errors may still occur. The final layer consists of 5142 segments.
17. Steps 9 to 12 above were then repeated for the DEM Channels$_{ER}$ layer to calculate new length and slope estimates for stream segments excluding lakes.

18. Where there were duplicate segments in the DEM Channels$_{ER}$ layer (explained in step 13g), we removed one of the entries, resulting in 5010 segments.
19. Width was calculated for each segment by dividing river area by segment length using attributes from RiverATLAS$_{ER}$ (river area and segment length).

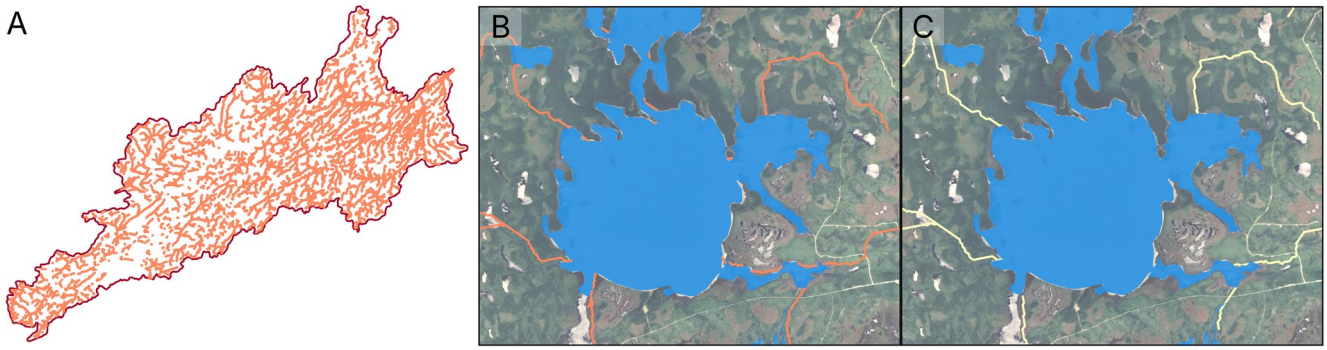



Figure S1.4. Exploits River channels with lakes exluded. A: Output of Exploits Channels with segments running through lake polygons excluded. B: Example of situation where river lines (orange lines) occur just outside of the boundaries of lake polygons (blue polygons) so vectors and/or nodes are not deleted and individual segments become disjointed. C: Fractured orange lines were deleted so that river channels (yellow lines) follow continuous paths.

Table S1.2. Matches of Shreve stream order and annual discharge used to assign RiverATLAS$_{ER}$ attributes to the DEM Channels$_{ER}$ layer.

| **Number** | **Shreve order - DEM Channels$_{ER}$** | **Annual discharge ($m^3/s$) - RiverATLAS$_{ER}$** | **Basin or area (see Figure S1.3J for locations in the watershed).** |
|---|---|---|---|
| 1 | 2900-6596 | 180-305.5 | Exploits main stem – Beothuk Dam to mouth |
| 2 | 1100-2900 | 75-180 | Mid Beothuk Lake to Beothuk Dam |
| 3 | 120-366 | 6-16 | Badger |
| 4 | 59-120 | 2.5-6 | Badger |
| 5 | 30-58 | 1.8-2.5 | Badger |
| 6 | 20-30 | 0.8-1.8 | Badger |
| 7 | 10-19 | 0.5-0.8 | Badger |
| 8 | 5-9 | 0.3-0.5 | Badger |
| 9 | 3-4 | 0.2-0.3 | Badger |
| 10 | 2 | 0-0.2 | Badger |
| 11 | 150-1003 | 55-74 | Beothuk Upper |
| 12 | 49-150 | 3-5 | Beothuk Upper |
| 13 | 30-48 | 2.1-3 | Beothuk Upper |
| 14 | 19-29 | 1.1-2.1 | Beothuk Upper |

| | | | |
|---|---|---|---|
| 15 | 11-18 | 0.6-1.1 | Beothuk Upper |
| 16 | 6-10 | 0.35-0.6 | Beothuk Upper |
| 17 | 4-5 | 0.26-0.35 | Beothuk Upper |
| 18 | 3 | 0.2-0.26 | Beothuk Upper |
| 19 | 2 | 0-0.2 | Beothuk Upper |
| 20 | 60-166 | 3.5-10 | Buchans |
| 21 | 20-59 | 2.1-3.5 | Buchans |
| 22 | 17-20 | 1.5-2.1 | Buchans |
| 23 | 12-16 | 0.8-1.5 | Buchans |
| 24 | 6-11 | 0.4-0.8 | Buchans |
| 25 | 4-5 | 0.25-0.4 | Buchans |
| 26 | 3 | 0.2-0.25 | Buchans |
| 27 | 2 | 0-0.2 | Buchans |
| 28 | 380-916 | 15.3-37.8 | GRB main stem |
| 29 | 125-380 | 5.5-15.3 | GRB main stem upper |
| 30 | 66-125 | 2.7-5.5 | GRB major tribs mouths |
| 31 | 20-66 | 0.8-2.7 | GRB Lower |
| 32 | 15-20 | 0.5-0.8 | GRB Lower |
| 33 | 7-14 | 0.25-0.5 | GRB Lower |
| 34 | 5-6 | 0.15-0.25 | GRB Lower |
| 35 | 2-4 | 0-0.15 | GRB Lower |
| 36 | 29-60 | 1.6-2.65 | GRB Mid |
| 37 | 13-28 | 0.6-1.6 | GRB Mid |
| 38 | 6-12 | 0.23-0.6 | GRB Mid |
| 39 | 3-5 | 0.2-0.23 | GRB Mid |
| 40 | 2 | 0-0.2 | GRB Mid |
| 41 | 28-66 | 1.29-2.7 | GRB Upper |
| 42 | 14-27 | 0.8-1.29 | GRB Upper |
| 43 | 5-13 | 0.32-0.8 | GRB Upper |
| 44 | 2-4 | 0-0.32 | GRB Upper |
| 45 | 139-244 | 5.3-9.2 | Harpoon |
| 46 | 90-138 | 3-5.3 | Harpoon |
| 47 | 51-90 | 1.6-3 | Harpoon |
| 48 | 32-50 | 1.3-1.6 | Harpoon |
| 49 | 16-31 | 0.8-1.3 | Harpoon |
| 50 | 11-15 | 0.4-0.8 | Harpoon |
| 51 | 7-10 | 0.25-0.4 | Harpoon |
| 52 | 4-6 | 0.2-0.25 | Harpoon |
| 53 | 2-3 | 0-0.2 | Harpoon |
| 54 | 260-603 | 13-38 | Lloyds |
| 55 | 82-260 | 3.7-13 | Lloyds |

| | | | |
|---|---|---|---|
| 56 | 60-81 | 2.9-3.7 | Lloyds |
| 57 | 33-59 | 1.9-2.9 | Lloyds |
| 58 | 17-32 | 1-1.9 | Lloyds |
| 59 | 11-17 | 0.61-1 | Lloyds |
| 60 | 8-10 | 0.45-0.61 | Lloyds |
| 61 | 6-7 | 0.4-0.45 | Lloyds |
| 62 | 4-5 | 0.3-0.4 | Lloyds |
| 63 | 2-3 | 0-0.3 | Lloyds |
| 64 | 45-122 | 2-4.915 | Lower Exploits |
| 65 | 25-45 | 1.1-2 | Lower Exploits |
| 66 | 18-24 | 0.7-1.1 | Lower Exploits |
| 67 | 14-17 | 0.49-0.7 | Lower Exploits |
| 68 | 9-13 | 0.36-0.4915 | Lower Exploits |
| 69 | 5-8 | 0.28-0.36 | Lower Exploits |
| 70 | 3-4 | 0.24-0.28 | Lower Exploits |
| 71 | 2 | 0-0.24 | Lower Exploits |
| 72 | 100-339 | 6.5-17.5 | Mary March |
| 73 | 60-100 | 2.7-6.5 | Mary March |
| 74 | 30-60 | 1.6-2.7 | Mary March |
| 75 | 25-30 | 1.2-1.6 | Mary March |
| 76 | 20-25 | 0.8-1.2 | Mary March |
| 77 | 11-20 | 0.5-0.8 | Mary March |
| 78 | 8-10 | 0.35-0.5 | Mary March |
| 79 | 4-7 | 0.25-0.35 | Mary March |
| 80 | 3 | 0.16-0.25 | Mary March |
| 81 | 2 | 0-0.16 | Mary March |
| 82 | 80-100 | 2-2.8 | Mid Exploits |
| 83 | 40-80 | 1.2-2 | Mid Exploits |
| 84 | 30-39 | 0.8-1.2 | Mid Exploits |
| 85 | 19-30 | 0.6-0.8 | Mid Exploits |
| 86 | 11-18 | 0.35-0.6 | Mid Exploits |
| 87 | 7-10 | 0.3-0.35 | Mid Exploits |
| 88 | 4-6 | 0.2-0.3 | Mid Exploits |
| 89 | 2-3 | 0-0.2 | Mid Exploits |
| 90 | 152-717 | 10-29 | Noel Paul |
| 91 | 71-151 | 3.5-10 | Noel Paul |
| 92 | 49-70 | 2-3.5 | Noel Paul |
| 93 | 36-48 | 1.5-2 | Noel Paul |
| 94 | 20-35 | 0.95-1.5 | Noel Paul |
| 95 | 13-19 | 0.7-0.95 | Noel Paul |
| 96 | 11-12 | 0.48-0.7 | Noel Paul |

| | | | |
|---|---|---|---|
| 97 | 6-10 | 0.43-0.48 | Noel Paul |
| 98 | 4-5 | 0.3-0.43 | Noel Paul |
| 99 | 2-3 | 0-0.3 | Noel Paul |
| 100 | 100-445 | 5-13.752 | Sandy |
| 101 | 65-100 | 2.3-5 | Sandy |
| 102 | 38-65 | 1.6-2.3 | Sandy |
| 103 | 21-37 | 0.8-1.6 | Sandy |
| 104 | 15-20 | 0.5-0.8 | Sandy |
| 105 | 10-14 | 0.34-0.5 | Sandy |
| 106 | 7-9 | 0.3-0.34 | Sandy |
| 107 | 4-6 | 0.25-0.3 | Sandy |
| 108 | 3 | 0.2-0.25 | Sandy |
| 109 | 2 | 0-0.2 | Sandy |
| 110 | 116-202 | 7.8-12.88 | Star & Shanawdithit |
| 111 | 51-110 | 3.1-7.8 | Star & Shanawdithit |
| 112 | 31-50 | 2.3-3.1 | Star & Shanawdithit |
| 113 | 25-30 | 1.9-2.3 | Star & Shanawdithit |
| 114 | 16-24 | 1.1-1.9 | Star & Shanawdithit |
| 115 | 10-15 | 0.7-1.1 | Star & Shanawdithit |
| 116 | 5-9 | 0.4-0.7 | Star & Shanawdithit |
| 117 | 3-4 | 0.25-0.4 | Star & Shanawdithit |
| 118 | 2 | 0-0.25 | Star & Shanawdithit |
| 119 | 403-1074 | 36-66 | Victoria |
| 120 | 100-402 | 7.5-36 | Victoria |
| 121 | 50-100 | 3.5-7.5 | Victoria |
| 122 | 36-50 | 2.39-3.5 | Victoria |
| 123 | 25-35 | 1.5-2.39 | Victoria |
| 124 | 21-24 | 1.2-1.5 | Victoria |
| 125 | 18-20 | 1.1-1.2 | Victoria |
| 126 | 13-17 | 0.54-1.1 | Victoria |
| 127 | 8-12 | 0.45-0.54 | Victoria |
| 128 | 5-7 | 0.4-0.45 | Victoria |
| 129 | 3-4 | 0.3-0.4 | Victoria |
| 130 | 2 | 0-0.3 | Victoria |

# S2. Field Data Collection and Model Validation

## S2.1. Random site selection

We sampled 51 DEM Channels$_{ER}$ river segments across the Exploits River watershed (Figure S2.1A). To select random sites for field sampling, we first divided each segment into three equal sections. In each section, two random points were created to guide substrate sampling sites using the QGIS tool ‘Vector Creation – Random Points Along Line’ (Figure S2.1B). Two points were created in case one was unsuitable or inaccessible for sampling. For sampling, we avoided ponds, water too deep to wade, waterfalls, directly under bridges or other crossings, and other human infrastructure such as dams and culverts. If both random points were in these areas, we sampled the closest suitable habitat to an access point (example in Figure S2.1C). All but nine of the 149 sites were within 1km of an access road or trail to allow timely access in this remote area. The remaining nine were within 1.8km of an access point.

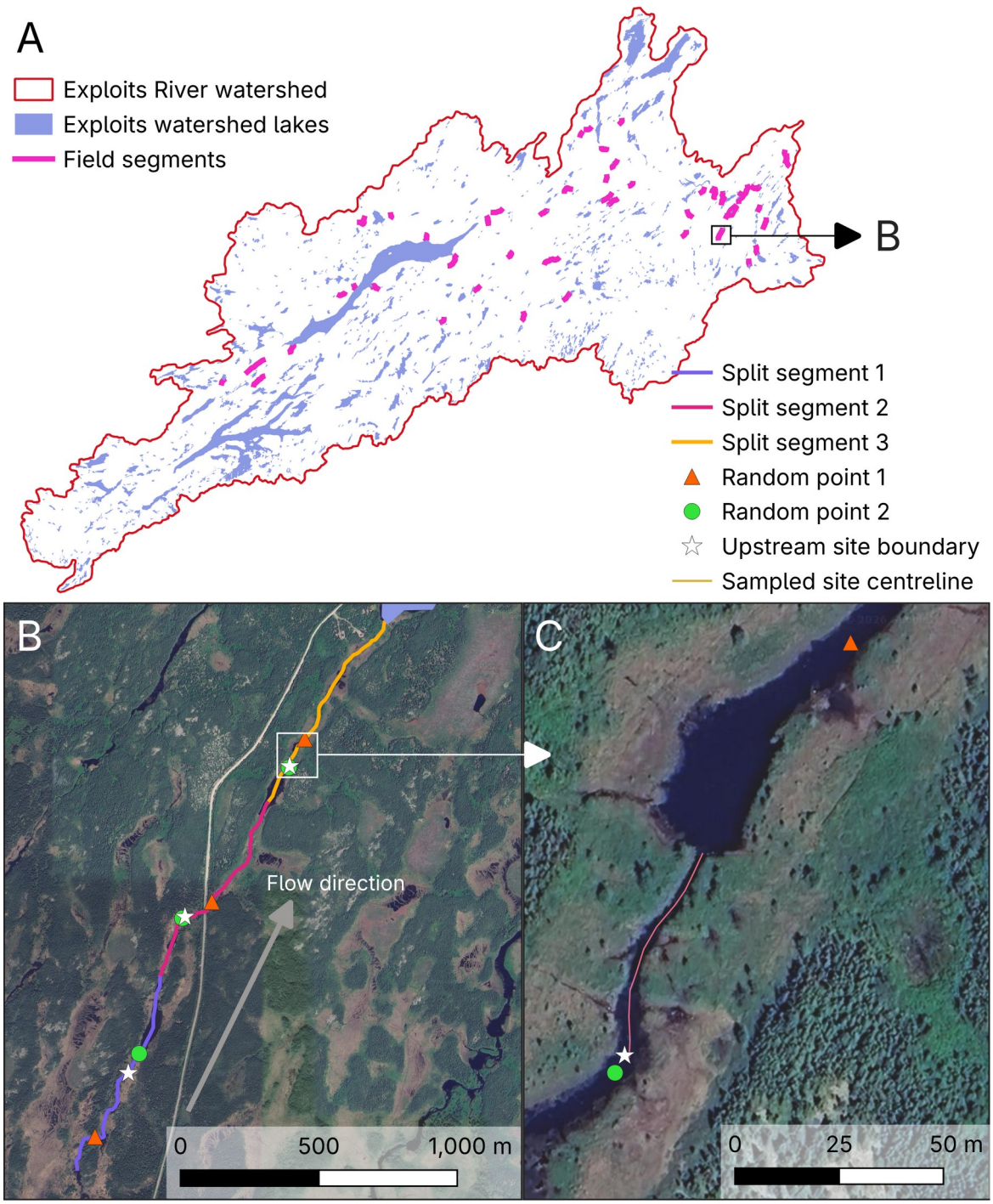



Figure S2.1. Overview of field segment locations and site selection. A: Segment locations in the Exploits River watershed. Black rectangle refers to extent in B. B: Example of a field segment split into three equal sections (Split segment 1-3). Random points to guide sampling are shown by orange triangles and green circles. White stars are the upstream points of actual sampling locations. White rectangle refers to extent in C. C: The sampled site centreline shows extent of sampled site where pebble counts were performed with the white star at the upstream point. The green circle and orange triangle are the two random points in this section. Basemap in B and C: Google Satellite imagery.

Figure S2.2. Measuring bed substrate at a field site. A: The 50m extent of a field site from upstream (white star) to downstream (white arrow). B: Undertaking a Wolman pebble count. C: Gravelometer showing different compartments for various sized substrates. Photos by C. Purchase (A and B) and A. Hart (C).

Table S2.1. Substrate size categories and descriptions of gravelometer size or size ranges of substrate.

| Size categories | Gravelometer size or size ranges (mm) | Value on figures |
|---|---|---|
| Silt/clay (smooth feel) | Very small | <2 |
| Sand (small grainy feel) | < 2 | <2 |
| Gravel (Pea to tennis ball diameter) | 2 - 2.8 | 2.8 |
| | 2.8 - 4 | 4 |
| | 4 - 5.6 | 5.6 |
| | 5.6 - 8 | 8 |
| | 8 - 11 | 11 |
| | 11 - 16 | 16 |
| | 16 - 22.6 | 22.6 |
| | 22.6 - 32 | 32 |
| | 32 - 45 | 45 |
| | 45 - 64 | 64 |
| Cobble (Tennis ball to basketball diameter) | 64 - 90 | 90 |
| | 90 - 128 | 128 |
| | 128 - 180 | 180 |
| | 180 - 250 | 250 |
| Boulder (basketball to car diameter) | 250 - 300 | 300 |
| | 300 - 512 | 512 |
| | 512 - 1024 | 1024 |
| | 1024 - 2048 | 2048 |
| | > 2048 | >2048 |
| Bedrock | Large unbroken rock surface | Bedrock |
| Woody debris | Sticks and wood | NA |

Note: Other field measurements of substrate usually do not include the size range of 250-300, and the range of 180-250 normally ends at 256, not 250. We switched to 250 and included the size range of 250-300 as we also used the long side of the gravelometer (ruler up to 300 mm) to measure substrate and these measures ensured quick measurements in the field.

Table S2.2. Relationship between gravelometer size bins and model prediction classes. Model predictions are continuous values that fall within defined prediction bins, while measured substrate sizes are discrete and correspond to specific bin values. The prediction bins are aligned with suitability for spawning.

| Gravelometer bin substrate size (mm) | Model prediction class (mm) | Suitability for spawning |
|---|---|---|
| <2 | <16 | No |
| 2.8 | | |
| 4 | | |
| 5.6 | | |
| 8 | | |
| 11 | | |
| 16 | | |
| 22.6 | 16 - 35 | Yes |
| 32 | | |
| 45 | 35 - 64 | Possible |
| 64 | | |
| 90 | > 64 | No |
| 128 | | |
| 180 | | |
| 256 | | |
| 300 | | |
| 512 | | |
| 1024 | | |
| 2048 | | |
| > 2048 | | |
| Bedrock | | |

Note: Except for <2, >2048, and Bedrock, gravelometer sizes are the size that substrate is finer than. For example in the 16 mm bin, substrate is between 11 mm and 16 mm.

## S2.2. Confusion matrix

Table S2.3. Description of multi class confusion matrix metrics and bounds used for validation analysis.

| Metric | Description | Bounds and interpretation |
|---|---|---|
| Accuracy | The ratio of the number of correct predictions (positive or negative) to the total number of predictions. | 0-1, 1 being all predictions are correct, 0 being no predictions are correct. |
| Sensitivity (or true positive rate or recall) | Proportion of positive correct predictions. | 0-1, 1 being all predictions = actual measurements. |
| Specificity (true negative rate) | Proportion of negative correct predictions. | 0-1, 1 being all predictions ≠ actual measurements. |
| Precision (or Positive Prediction Value) | Measures how many positive predictions are actually correct. | 0-1, 1 being all predictions for a class are actually that class. |
| Negative Prediction Value | Measures how many negative predictions are actually correct. | 0-1, 1 being all negative predictions for that class are correct. |
| F1 | Weighted mean of precision and sensitivity. | 0-1, 1 being there are no incorrect predictions. |
| Balanced accuracy | The average of the sensitivity for each class in a multi class confusion matrix. | 0-1, 1 being all predictions in all classes = actual predictions. |
| Kappa (Cohen's Kappa) | Indicates how well the predictions are compared to random guessing. Takes into account imbalanced datasets. More reliable that accuracy for assessing model performance. | -1-1; <0 = worse than random, 0 = no better than random, < 0 better than random with 1 being the best. |
| P-value | Significance value of McNemar's test to test for differences between classes. | <0.05 = significantly different. |
| Prevalence | The proportion of samples that belong to each measured (actual) class. Computed individually for each class. | 0-1, 1 being 100% of samples belonging to class, 0 being no samples in that class. |
| Detection rate | The proportion of the whole dataset that were correctly identified as each class. It combines the prevalence and sensitivity of each class. | 0-1, 1 being 100% of the dataset were correctly identified as class; unlikely. |
| Detection prevalence | The proportion of samples that were predicted as each class. | 0-1, 1 being 100% of the dataset were predicted as that class; unlikely. |

# S3. Results

## S3.1. Confusion matrix results

Table S3.1. Confusion matrix results for predictions in each measured substrate size class.

| Metric | Substrate class (mm) | | | |
|---|---|---|---|---|
| | **<16** | **16-35** | **35-64** | **>64** |
| Sensitivity (true positive rate) | 0.600 | 0.500 | 0.476 | 0.465 |
| Specificity (true negative rate) | 0.938 | 0.805 | 0.710 | 0.810 |
| Positive Prediction Value (precision) | 0.250 | 0.235 | 0.392 | 0.769 |
| Negative Prediction Value | 0.985 | 0.930 | 0.776 | 0.536 |
| F1 | 0.353 | 0.320 | 0.430 | 0.579 |
| Prevalence | 0.034 | 0.107 | 0.282 | 0.577 |
| Detection Rate | 0.020 | 0.054 | 0.134 | 0.269 |
| Detection Prevalence | 0.081 | 0.228 | 0.342 | 0.349 |
| Balanced Accuracy | 0.769 | 0.652 | 0.593 | 0.637 |

Note: Results derived from the ‘caret’ package in R. See Table S2.1 for descriptions of metrics.

## S3.2. Differences between substrate classes

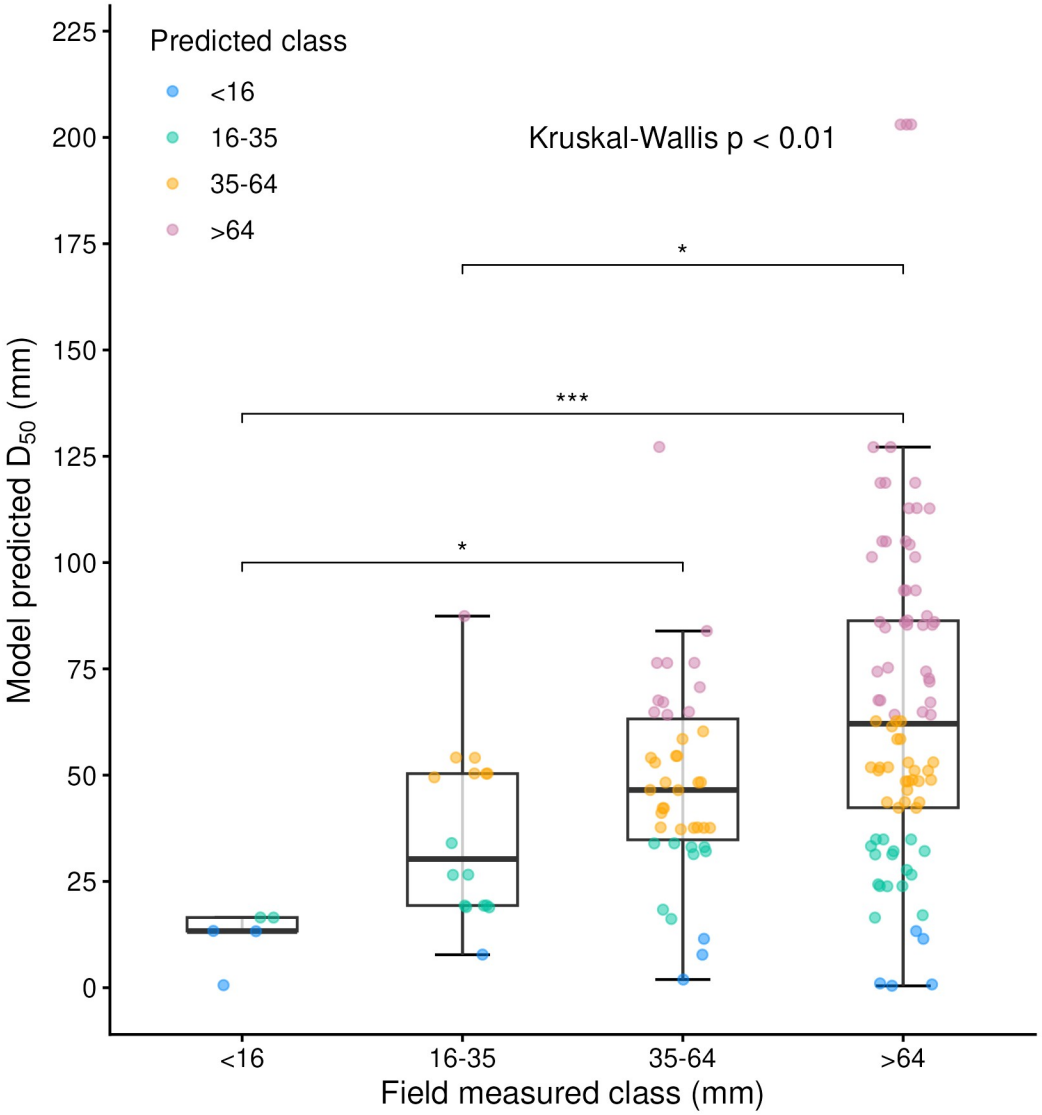



Figure S3.1. Boxplots of model-predicted substrate size ($D_{50}$) grouped by field-measured class. Points are coloured by model-predicted substrate class. Box plot explanation: Boxes show the median (centre line); and the interquartile range (IQR) of predicted D50 values. IQR: bottom line is the 25th percentile (Q1), and the top line the 75th percentile (Q3). The whiskers, or lines, range from the 'minimum' to the 'maximum' (Q1–1.5 × IQR or Q3 + 1.5 × IQR). Significant differences between groups are shown: * = p.adj < 0.05, *** = p.adj < 0.001.

Table S3.2. Results of Dunn's test of multiple comparisons between predicted $D_{50s}$ grouped by the $D_{50}$ class of field measurements.

| Group 1 | Group 2 | n1 | n2 | Test statistic | Adjusted p value | p significant |
|---|---|---|---|---|---|---|
| <16 | 16-35 | 5 | 16 | 1.79 | 0.443 | ns |
| <16 | 35-64 | 5 | 42 | 2.76 | 0.0343 | * |
| <16 | >64 | 5 | 86 | 3.84 | 0.0007 | *** |
| 16-35 | 35-64 | 16 | 42 | 1.33 | 1 | ns |
| 16-35 | >64 | 16 | 86 | 3.13 | 0.0104 | * |
| 35-64 | >64 | 42 | 86 | 2.45 | 0.0859 | ns |

Note: Significant differences between groups are as follows: ns= not significant, * = p.adj < 0.05, ** = p.adj < 0.01, *** = p.adj < 0.001. p.adj = Adjusted p value for differences between groups.

# References

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., & Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, *6*(1), 283. https://doi.org/10.1038/s41597-019-0300-6

Natural Resources Canada. (2024). *Product access guide for the Medium Resolution Digital Elevation Model (MRDEM)*. Natural Resources Canada.