

What influences spatial variability in restoration costs?

Econometric cost models for inference and prediction in restoration planning

Abstract

Habitat restoration efforts are often conducted under limited budgets and information. Explicitly incorporating data on project costs can improve the effectiveness of restoration investments but gathering precise data for future projects is often infeasible at the planning stage. We demonstrate econometric and machine learning methods for improving the accessibility of cost information for restoration planning. Using geospatial matching methods, we combine 15 years of spatially explicit project records for over 1,200 culvert fish passage barrier restoration worksites across the Pacific Northwest of the United States with data layers of hypothesized drivers of construction costs. We distinguish between two objectives in analyzing cost data: inference to identify cost drivers and prediction of future project costs. For inference, we multiple linear regression and find that the variables channel bankfull width and slope, road speed class, developed and agricultural land covers, and proximity to privately managed industrial land are associated with culvert restoration costs and may serve as strong proxies in a planning setting. For prediction, we use boosted regression trees to make out-of-sample projections of restoration costs for over 27,000 barrier culverts documented in state inventories. The distribution of these cost projections over space and across jurisdictions reveal higher, and more heterogeneous, culvert restoration costs in the Puget Sound region than in other nearby areas. Our results are directly applicable for resource managers making fish passage restoration decisions and serve as a template for the use of econometrics and machine learning to analyze costs recorded in restoration project databases in other contexts.

Keywords: Restoration costs; Return on investment analysis; Habitat restoration; Machine learning; Cost effectiveness; Conservation planning; Fish passage

1 Introduction

Government agencies, non-profits, and other public and private actors commit vast resources to restoring ecosystems to achieve various objectives, often including the recovery of target species. Restoration activities across biomes contribute 9.5 billion USD of annual economic output to the American economy (BenDor et al. 2015). In the United States, spending on riparian restoration alone is estimated to be over 1 billion USD annually (Bernhardt et al. 2005). Large expenditures are often driven by recovery objectives for specific species. Between 1991 and 2015, over 2.2 billion USD in spending has been documented in a clearinghouse database for salmon habitat restoration projects across the Pacific Northwest, with mean annual expenditures of 130 million USD since 2001 (NMFS 2021).

When allocating resources towards specific objectives, restoration planners rely on decision support frameworks which compare the merits of alternative projects to guide decision making (Roni & Beechie 2013; Boyd et al. 2015; Fonner et al. 2021). Common frameworks include prioritization indices (Roni & Beechie 2013) and return-on-investment (ROI) analysis (Boyd et al. 2015; Fonner et al. 2021), including for example optimization algorithms which consider correlated or interdependent costs and benefits (Newbold & Siikamäki 2015). There has been growing recognition of the value of including costs as well as benefits in conservation planning (Cook et al. 2017). Incorporating costs allows government agencies and conservation organizations to allocate funding to projects based on projected ROI, defined in the simplest terms as the targeted conservation metric (e.g., area of habitat, number of species, etc.) divided by the project cost (Boyd et al. 2015).

Studies have consistently found that incorporating economic costs into planning models can allow for equivalent levels of conservation benefit at much lower costs than models based on benefits alone (Babcock et al. 1997; Naidoo et al. 2006; Perhans et al. 2008; Kujala et al. 2018; Rodewald et al. 2019; Field & Elphick 2019). Beyond their use in planning, studies of conservation and restoration cost data have allowed empirical analyses of management issues, revealing potential conflicts in management priorities (Armsworth et al. 2017) and opportunities to benefit from economies of scale (i.e., when costs increase at a slower rate than output) (Kim et al. 2014; Cho et al. 2017; Armsworth et al. 2018).

Past studies of costs in conservation management have contributed several important results. First, the value of incorporating costs into planning models depends on the distribution, i.e., variability, of costs across projects and in relation to the distribution of benefits (Babcock et al. 1997; Naidoo et al. 2006; Perhans et al. 2008; Rodewald et al. 2019). Second, improved efficiencies from incorporating costs, in terms of benefits achievable on a given budget, are possible even when a priori knowledge of costs is limited (Babcock et al. 1997; Carwardine et al. 2010; Kujala et al. 2018; Field & Elphick 2019). Finally, careful consideration of appropriate proxies for costs of potential conservation actions is needed to ensure that cost measures accurately represent the underlying cost variability and distributions (Armsworth 2014; Burkhalter et al. 2016; Sutton et al. 2016).

However, in applied settings the incorporation of costs in conservation planning studies is dominated by cases where land values, often measured as rents, agricultural productivity, or land acquisition price, serve as valid proxies for economic costs of alternative actions (Armsworth

2014; Sutton et al. 2016; Rodewald et al. 2019). Cost estimates in these cases are commonly used as an input to "reserve site selection" studies of spatial prioritization, as opposed to strategic allocation of resources for restoration for which clear cost proxies are often less readily available (Naidoo et al. 2006; Pienkowski et al. 2021). In this article, we examine methods for determining suitable proxies for and predictions of habitat restoration costs that vary across geography.

For predictions or proxies of project-level restoration costs to have value, they must capture the underlying variability in costs among potential restoration actions. When costs (and often times benefits) for restoration actions are accounted for in large-scale restoration planning studies, researchers most frequently rely on unconditional proxies, heuristics, or expert elicitation (Plummer 2009; Armsworth 2014). In rare cases where empirical methods are applied to historical project records to inform future costs, projects are often divided into simple classes and averages are taken as representative of all projects within that class (e.g., Phillips-Mao et al. 2015). Such methods can obscure the true level of cost variability between projects (see also, Burkhalter et al. 2016; Sacre et al. 2019).

In this article, we demonstrate how econometric and machine learning methods can be used to achieve two distinct but closely related objectives: (1) identifying landscape-level and project-level cost drivers which could be used as potential proxies in planning, and (2) providing predictive estimates that accurately represent the distribution (i.e., variability) of potential costs for future candidate projects. Our empirical findings address a critical and immediate need for policymakers in the Pacific Northwest: the replacement of fish passage barrier culverts at road crossings.

Access to spawning and rearing habitat has long been recognized as a critical factor limiting the recovery of ESA listed Pacific salmon and steelhead (*Oncorhynchus* spp.) across the region (Roni et al. 2002). The most common barriers to fish passage are at stream-road crossings where outdated or deteriorated culverts block access to upstream habitat. At least 401 million USD has been spent on culvert restoration projects within the study region to date (NMFS 2021), while at least 27,450 remaining barrier culverts have been identified across Oregon and Washington. Spending is likely to increase significantly in the coming years across the region as recovery efforts for endangered and co-managed migratory salmon populations accelerate. Recent state and federal legislation have expanded existing and created new funding sources specifically for fish passage barrier correction (Infrastructure Investment and Jobs Act 2021, Move Ahead WA Act 2022).

Following a 2013 federal injunction, Washington state bears responsibility for removing more than 900 state-owned culverts at stream-road crossings that block passage for migratory salmon across the Puget Sound and Washington Coastal basins by 2030 (Hickey 2018). The costs of replacing these state-owned culverts have been estimated at over 3 billion USD. Beyond these state-owned barriers, efforts have accelerated at the county and local level to inventory, prioritize, and remove culvert fish passage barriers (Brown 2019). Both recent state budgets in Washington and legislation under consideration at the federal level have set-aside hundreds of millions of dedicated funding for culvert fish passage improvements. Among all culvert barrier ownership entities in the region, there is an urgent need for data-based characterizations of the relative costs of the thousands of potential barrier improvement projects they might undertake.

To fill this gap, we leverage cost records from over 1,200 georeferenced fish passage restoration worksites from 15 years across the states of Washington and Oregon in the Pacific Northwest of the United States (Figure 1). Historical culvert replacement projects are linked to landscape scale geospatial data for potential biophysical and socioeconomic cost drivers,

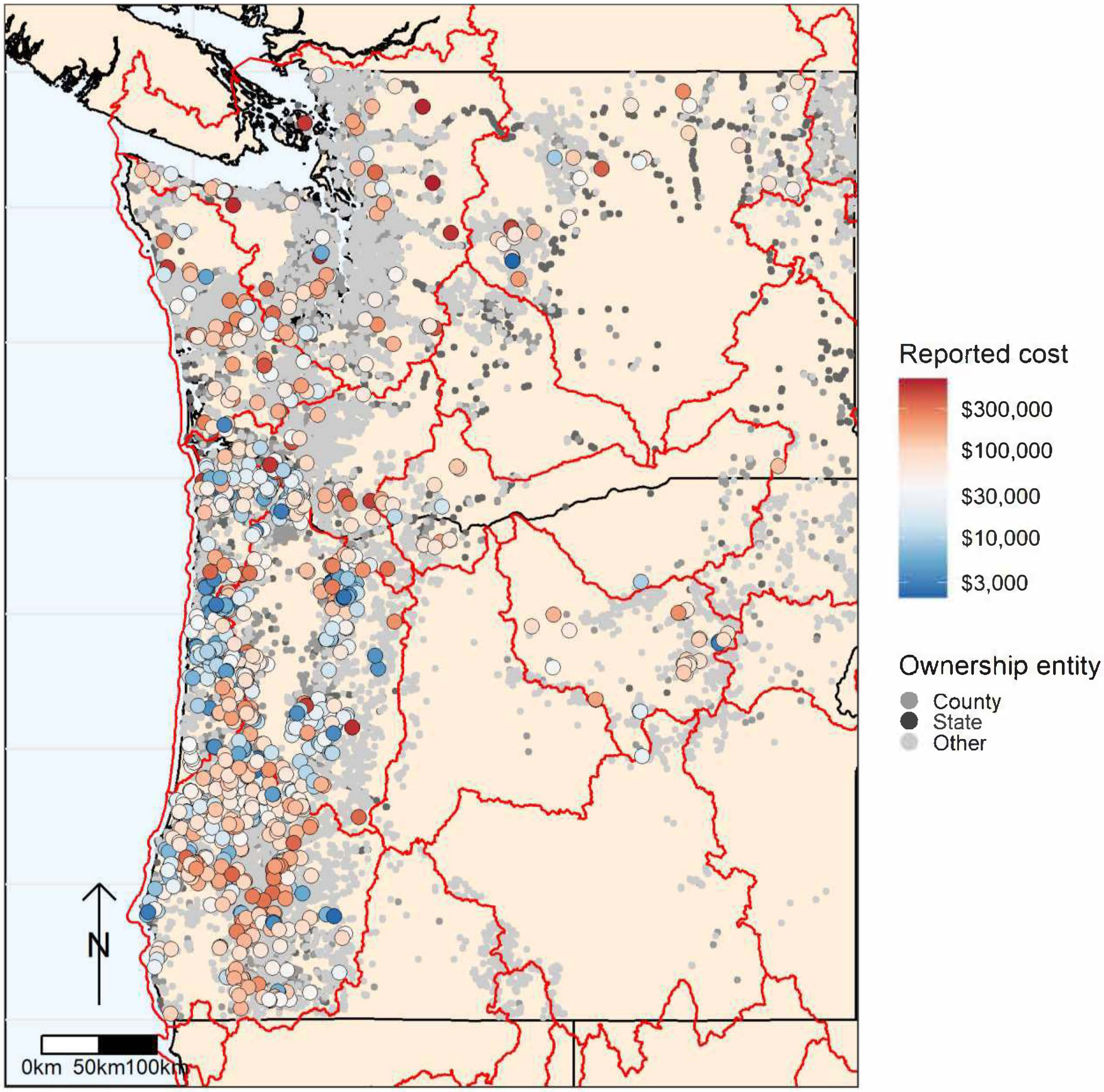

Figure 1. PNSHP culvert restoration worksites by reported cost (color points) and barriers in state inventories by ownership entity class (greyscale points). Costs are converted to 2019 dollars using the CPI. Other includes private landowner, federal, tribal, and city barrier owners. Red borders represent basins (HUC2).

identifying variables strongly associated with fish passage improvement costs which can serve as strong proxies in planning models. Using predictive machine learning methods, we map riverine subbasins where fish passage improvement is likely to be more or less expensive, and where variation between barrier improvement costs is likely to be highest. While these results are of immediate relevance to restoration planners responsible for managing fish passage restoration in the Pacific Northwest, our methods are also of relevance for conservation planners and scientists in other settings seeking tools to empirically inform the distribution of costs for potential restoration actions.

2 Methods

2.1 Data collection and compilation

2.1.1 Historical worksite and cost data

Our analysis of project costs is based on project cost records for 1,236 completed culvert improvement worksites in the Pacific Northwest Salmonid Habitat Project (PNSHP) database (NMFS 2021). PNSHP is a widely cited clearinghouse for habitat restoration projects across four states: Idaho, Montana, Oregon, and Washington (NMFS 2021). PNSHP aggregates habitat restoration project data from over two dozen organizations, referred to as reporting sources, and records are standardized into a consistent format with fields including project cost, restoration actions completed, and worksite coordinates.

PNSHP has a multi-level structure, where one or more restoration actions are grouped into a uniquely identified worksite and one or more worksites are grouped into uniquely identified projects. The unit of analysis for this study is the worksite, while costs are reported at the project level. For each worksite in our sample, we calculate "project average cost" for culvert-related actions by dividing project costs by the number of affected culverts across all associated project worksites. Project average cost is the dependent variable of interest in our analysis and varies widely for culverts across our study region (Figure 1)[1].

[1] A plot of the distribution (kernel density) of the reported cost variable is found in the Supplementary Material as Figure S1.

To ensure sufficient observations in subgroups and remove outliers that may be due to reporting errors or out-of-scope projects (e.g., cases where culverts are replaced with bridges rather than culvert structures), we restrict our analysis to PNSHP records that include only culvert-related habitat improvements, were within Washington state or Oregon, were part of a project initiated between 2001 and 2015, and report project average cost between $2,000 and $750,000 (roughly 5^{th} and 95^{th} percentiles respectively). We further restricted our sample such that each represented basin and reporting source had a minimum of twenty worksite observations[2].

2.1.2 Hypothesized cost drivers

We identified six categories of variables hypothesized to affect culvert restoration costs based on those developed by Washington Department of Fish & Wildlife (WDFW) (Barnard et al. 2013) (Table 1). We compiled existing geospatial data layers from publicly available and restricted access federal sources to identify landscape scale variables associated with the identified categories. The full list of variables collected for each worksite, with descriptions for each, is available in Table S1.

Hydrologic and road variables determine the size, shape, and other design considerations for the replacement culvert design. Channel slope, channel bankfull width, and the length of the culvert are the key variables that determine whether a simple no-slope or more complex and expensive hydraulic design is the appropriate replacement (Barnard et al. 2013). Culvert

[2] The nine specific basins represented are: Southern Oregon Coastal, Northern Oregon Coastal, Washington Coastal, Puget Sound, Lower Columbia, Middle Columbia, Upper Columbia, and John Day. The six reporting sources represented are: Oregon Watershed Restoration Inventory (OWRI), Washington Habitat Work Schedule, Washington Salmon Recovery Funding Board Database (SRFBD), United States Forest Service Region 6 Regional Ecosystem Office (REO), United States Bureau of Land Management (BLM), and Washington Recreation and Conservation Office (WA RCO).

replacements on paved and more heavily trafficked roads are also likely to be more expensive

Table 1. Variable categories with examples and data sources.

Category	Example variables	Data sources
Hydrological	Channel slope (%), bankfull width (m), sinuosity, basin (categorical)	NHDPlus V2, NHDPlus Select Attributes V2.1
Road	Material (categorical), speed limit (categorical), functional class (categorical), distance to road feature (m), road crossing density (crossings per catchment)	HERE Geodatabase Plus, NHDPlus Select Attributes V2.1
Terrain	Terrain slope (degree), elevation (m), land cover (categorical)	NHDPlus Select Attributes V2.1, NLCD, GTOPO30
Population	Housing density (units per km^2), lands managed by private industry (% within 2.5km radius), distance to urban area (m)	BLM Surface Jurisdiction of Lands, US Census Urban and Rural Classifications,
Economic	Construction employment (jobs), density of sand & gravel sales yards	US Census County Business Patterns, HIFLD

Project	Number of worksites, distance between worksites (m), action type (categorical), project year, reporting source (categorical)	PNSHP

due to increased signaling, road diversion, and road reconstruction costs (Washington Department of Fish and Wildlife 2019).

Construction for culvert replacement often requires negotiating with private landowners, working in difficult terrain in remote areas, and is subject to different design standards depending on whether the road is public or private (Barnard et al. 2013), so all of these factors are expected to influence project costs. We used terrain variables as proxies for the ease of access at the worksites, while human population variables represented social factors that affect ease of access, such as efforts to secure staging areas for equipment and supplies, permitting requirements, and travel expenses.

Economic variables proxy for the availability of labor, equipment, and materials to complete a culvert replacement. We expect those worksites with poor access to equipment and materials suppliers and labor are likely to be more expensive. We gathered proxy data for access to the most important materials, equipment, and labor specialties for culvert replacements, including point data for gravel and sand suppliers, concrete suppliers, construction equipment suppliers, and construction and forestry labor measured annually at the county level.

Finally, we include project variables, which represent the scope and scale of the project conducted at the worksite, including the number of other worksites associated with the project, the total distance between those worksites, and the reporting source for the project to the PNSHP database. The number of worksites associated with a project and the distance between them in particular allow us to test conditions where culvert replacement projects could benefit from economies of scale (Armsworth et al. 2011; Kim et al. 2014; Cho et al. 2017).

Economic costs can be divided into two classes: fixed costs that do not depend on the level of output and variable costs that do (Goolsbee et al. 2013). In the case of barrier culvert correction projects, many of the costs incurred, such as costs for mobilizing equipment or negotiating with nearby landowners for access, can be thought of as fixed costs for a project. As the number of barriers corrected under a single project increases, average fixed costs decrease because these costs can be shared across multiple barriers. This phenomena can contribute to economies of scale; sharing costs across multiple barriers reduces the average barrier correction cost for that project. Therefore, accounting for the number of worksites grouped under a single project can identify where economies of scale are present.

On the other hand, at a certain point we expect diseconomies of scale (i.e., when costs grow at a faster rate than quantity) to occur. Grouping worksites under a single project requires more variable costs, including materials and labor hours. There are also limit to how many worksites can share fixed resources; when worksites are distant the costs of moving resources between sites (a variable cost) may outweigh the benefits of sharing these resources under a single project. Therefore, the distance between worksites is also a critical consideration for identifying the limits of scale economies.

Our geospatial data layers include vector (points, lines, and polygons) and raster formats. We relied on several different matching methods to link features from indicator datasets to worksite locations. Methods were selected to accommodate both the feature type and hypothesized mechanisms through which specific variables might influence costs. Detailed descriptions of feature matching methods can be found in the Supplementary Material (Supplementary Material 1).

2.2 Statistical analyses

Our research questions demand both the identification of drivers of costs and the prediction of costs for potential future projects. We approach these two empirical tasks with two separate analyses of our compiled data. Recent methods literature in both ecology and econometrics have suggested that modeling for inference and modeling for prediction are two related but distinct goals, and therefore often require alternative approaches rather than one unified model (Mullainathan & Spiess 2017; Tredennick et al. 2021). We use a linear regression approach with fixed effects for inference and a boosted regression tree model to make predictions of future project costs. All analyses were conducted in R 4.1.2; specific packages used are described in the Supplementary Material.

2.2.1 Identifying cost drivers: Linear regression

To identify the drivers of project costs, we use multiple linear regression, estimated via ordinary least squares. We estimate a parameterized model of mean project average costs, conditional on observable worksite features which allows the comparison of the strength and direction of relationships between features and costs. Formally, the log-linear equation we fit is

$$ln(y_i) = \alpha + HYDRO_i\beta_{HYDRO} + ROAD_i\beta_{ROAD} + TERR_i\beta_{TERR} + POP_i\beta_{POP} + ECON_i\beta_{ECON} + PROJ_i\beta_{PROJ} + \mu_{year(i)} + \mu_{basin(i)} + \mu_{source(i)} + \varepsilon_i \quad (1)$$

where i indexes worksite observations; y_i is project average costs; $HYDRO_i$, $ROAD_i$, $TERR_i$, POP_i, $ECON_i$, $PROJ_i$ are vectors of explanatory variables corresponding to hypothesized driver categories; and ε_i is a zero-mean normally distributed error term. The intercept α; fixed effects $\mu_{year(i)}$, $\mu_{basin(i)}$, and $\mu_{source(i)}$; and the β coefficient vectors are parameters to be estimated. Subscript labels on the fixed effect parameters indicate that the specific fixed effects included depend on the specific worksite observation.

Variables included in explanatory variable vectors include specific indicators for hypothesized mechanisms influencing project costs (Table 2). We include two interaction terms that test for particularly important relationships between explanatory variables and costs. The

Table 2. Descriptive statistics for variables included in linear regression (n = 1,236)

Category	Variable	Mean	Std. dev.	Number of levels
Dependent variable	Cost per culvert ($USD2019)	82,600	93,800	
Hydrological (HYDRO)	Stream slope (%)	4.58	4.06	
	Bankfull width (m)	7.56	5.41	
	Basin			9
Road (ROAD)	Paved road			2
	Road speed class			6
Terrain (TERR)	Land cover class			6
	Terrain slope (deg)	27.3	12	
	Elevation (m)	412.2	317.7	
Population (POP)	Distance to urban area (m)	45,200	31,800	
	Housing density (units per sq. km)	5.96	24.7	
	Private land, individual or company owner (proportion of land within 500m radius)	0.447	0.436	
	Private land, managed by industry (proportion of land within 500m radius)	0.203	0.356	
	Private land, managed by non-industrial owner (proportion of land within 500m radius)	0.0349	0.169	
Economic (ECON)	Construction employment (jobs)	3,010	5,140	
	Ag/forestry employment (jobs)	778	534	
	Density of construction equipment wholesalers (workers per sq. km)	0.0216	0.032	
	Density of brick, stone, and related wholesalers (workers per sq. km)	0.00276	0.00721	
	Density of sand and gravel sales yards (firms per sq. km)	0.0000244	0.00004	
Project (PROJ)	Number of worksites (count)	1.48	0.958	
	Distance between worksites (m)	507	1,360	
	Year			15
	Reporting source			6

first is an interaction between bankfull width and channel slope in the $HYDRO_i$ vector. This interaction is intended to capture situations where channel morphology and hydrological constraints are particularly extreme and require specialized, often more expensive culvert designs (Barnard et al. 2013). We expect this interaction effect to be positive, indicating highest costs when culverts are on streams that are both wide and steep.

The second interaction is between the number of worksites associated with a worksite's project and the total distance between these worksites, both in the $PROJ_i$ vector. This interaction allows for a specific mechanism of diminishing economies of scale. We expect projects with more worksites see lower average costs due to shared fixed costs for organizational resources and mobilization efforts. However, adding additional worksites under a single project also increases project complexity and variable costs due to travel, as captured by the total distance between sites. A negative coefficient on this interaction term, with a positive coefficient on the linear worksite count term, would support this hypothesis, indicating that economies of scale diminish and ultimately become diseconomies as the number of worksites and the distances between them increases.

The fixed effects allow us to condition our estimates on unobserved factors shared between worksites begun a given year, within a specific basin, or reported by the same source. Examples of the types of unobserved factors we envision are captured by these parameters include market or climate conditions affecting the entire Pacific Northwest region (such as the 2008 financial crisis), local government policies or regulations that affect most or all of a basin over the entire study period, and reporting standards that vary between organizations but are consistent over time, respectively. By comparing fixed effects estimates, we can learn where, when, and for whom cost premiums (or discounts) not explained by observed factors are largest or smallest.

We use the individual elements of β to calculate "standardized cost multipliers" for every explanatory variable, allowing for consistent comparison of effect sizes across variables. For continuous variables, we calculate standardized cost multipliers by exponentiating the variable's standard deviation multiplied by its associated β. The resulting value is interpreted as the ratio between the costs of two projects with a standard deviation difference in the explanatory variable, all else equal. For categorical variables, we simply exponentiate the associated β, and the standardized cost multiplier is interpreted as the ratio between the costs of a project in the category and one in the baseline category.

For all standardized cost multipliers reported, we include 95-percent confidence intervals based on heteroskedasticity consistent standard errors clustered at the project-level. Model fit was assessed with adjusted R^2, Akaike information criterion (AIC), and Bayesian information criteria (BIC), with results of alternative specifications and fit on alternative samples presented for comparison (Table S3 and Table S4).

Many of the variables included as hypothesized drivers are potentially correlated, which in a linear regression context can induce multicollinearity leading to inflated standard errors (Wooldridge 2010). To diagnose this potential issue, we computed pairwise Pearson's correlation coefficients for all explanatory variables as well as variance inflation factors (VIFs) for each variable when included in the full specification. Results of these diagnostic procedures can be found in the Supplementary Material and did not reveal pairwise correlations nor VIFs of concerning levels (Table S2, Figure S2).

2.2.2 Predicting future costs: Machine learning

To generate predictions, we considered a suite of machine learning (ML) methods: regression tree, random forest (RF), and boosted regression tree (BRT) (De'ath 2007; Elith et al. 2008; Mullainathan & Spiess 2017; Athey & Imbens 2019; Storm et al. 2020). For all ML methods, we split the culvert project cost data PNSHP sample randomly into two equally proportioned training and testing sets. We trained models on the training set, including fitting the fully specified linear regression model to the training set as a baseline for comparison. The dependent variable was log project average costs for every ML method considered.

We compared root mean squared error (RMSE) (Figure S3), plots of predicted vs. actual values (Figure S4), and plots of the distributions of residuals (Figure S5) for each method and the linear regression as a baseline, with the goal of identifying a preferred method for cost predictions out-of-sample (e.g., on state culvert inventories). We also examined the variables most important to prediction in our ML fits. The Variable Importance metric used is the sum of squared decreases in RMSE at each "node" a variable is associated with across the component models of each fit (Greenwall and Boehmke 2020). After evaluation of these metrics, we selected BRT as the preferred method for out-of-sample predictions. Further discussion of the three methods considered and comparison of prediction performances is found in the Supplementary Material (Supplementary Material 2).

2.3 Predictions for culvert inventories

The purpose of developing predictive models of project costs is threefold: (1) identify relatively high/low-cost barriers, (2) identify regions/barrier ownership entities with relatively high/low-cost barriers, and (3) identify regions/barrier ownership entities with barriers with high variability. Therefore, we need data on the population of potential future fish passage improvement projects in the study region on which we can apply our predictive models.

We applied our preferred predictive model (BRT) to fish passage barriers identified in two inventories maintained by state agencies. Both Washington Department of Fish and Wildlife (WDFW) and Oregon Department of Fish and Wildlife (ODFW) maintain statewide databases of potential and confirmed fish passage barriers, the majority of which are culvert road crossings.

Each observation in these inventories includes coordinates, an evaluation of barrier status (i.e., passability), and ownership information (state, county, private, etc.) (Figure 1). Details on methods and assumptions for gathering data on the explanatory variables are found in the Supplementary Material (Supplementary Material 2).

We mapped predicted cost levels and variability over the study area. Cost levels are presented in relative terms, presented as percentile rank among all inventory barrier culverts in the study region. This step was performed because our predictions are based on historical data rather than forecasts of future trends. As a result, any dollar-amount prediction would require assumptions about the size of the year effect and other time-varying variables like employment levels, land cover, and population density. Therefore, we choose to present only relative costs so that our findings are not improperly interpreted as budgeting forecasts.

Levels are presented both for individual barriers and as mean over hydrological subbasin, specifically the smallest geographic area cataloged in the USGS Hydrological Unit Code (HUC) system (HUC10)[3]. Variability is measured as the coefficient of variation (standard deviation divided by mean) computed over the smallest hydrological subbasin scale (HUC10). We present predictions for all barriers, and by ownership class to highlight barrier ownership entities with more expensive culverts and those who could benefit the most from incorporating costs into planning frameworks.

2.4 Area of applicability and sample selection

[3] The USGS Hydrological Unit Code (HUC) system assigns all US land area into nested units, where subunits are identified by adding two digits for successively smaller areas. HUCs with fewer digits, e.g., HUC6 the largest unit discussed in this paper, subsume HUCs with more digits, e.g., HUC10 the smallest.

When applying predictive models to new settings, predictions are only valid if the prediction space is similar to the data from which the models are derived (Meyer and Pedesma 2022). We assess the similarity of the PNSHP training data to the WDFW and ODFW inventory data using two approaches.

First, we present summary statistics for all variables included in the linear model for both the PNSHP and inventory data, followed by the results of a two-sample comparison of means t-test. These results highlight differences between completed projects and culverts where projects have yet to begin.

Second, we calculated the "area of applicability" for the BRT and RF models, following the methods described in Meyer and Pedesma (2021). We computed the distance in the multivariate predictor space between each point in the inventories and the nearest point in the training data. Each variable is standardized by mean-centering and dividing by the standard deviation, and then weighted by the Variable Importance metric for the relevant model calculated via the *vip* package for R (Greenwall and Boehmke 2020). These distances were then themselves standardized by dividing by the mean pairwise distance between points in the training data; the resulting standardized distances are referred to as dissimilarity index (DI) values. DI values were also computed for each point in the training data and the maximum training set DI value after outliers[4] were removed was established as the "area of applicability" (AOA) threshold, where inventory points with DI values less than the AOA threshold are considered within the applicable range of the predictive model.

Once the AOA is established for the BRT and RF models, we computed the share of inventory barriers that are within this area for each model. For the BRT results, we present maps

[4] Outliers are defined as those greater than the 75-percentile plus 1.5 times the interquartile range.

of both the status of individual barriers within or outside the AOA and the share of barriers in each HUC10 within the AOA. Finally, we calculated the share of barriers within the AOA by categories of management interest (basin, road characteristics, land cover class) to characterize the barriers that fall outside this space where additional data or alternative methods might be needed to improve prediction validity.

Another potential issue that can affect the validity of both prediction and inference in econometric settings is sample selection: when the data used to estimate a model are non-random and therefore potentially non-representative of the population of interest (Wooldridge 2010, Heckman 1979). However, sample selection bias only occurs in a linear regression setting when the selection mechanism is based on the dependent variable, in this case cost (Wooldridge 2010). To determine whether cost was a selecting factor for past culvert correction projects, we examined documents describing prioritization methods for barrier-owning entities in our study region. Of the six prioritization methods we identified, only two included costs as metrics influencing project prioritization. However, in both cases these cost metrics were based entirely on rough measures of the width of the road where the culvert was located and had relatively little influence on the overall priority compared to other metrics. Therefore, we concluded that costs were not a meaningful determinant in the sample selection mechanism for the PNSHP data.

3 Results

3.1 Drivers of culvert restoration costs

Our full linear regression had an adjusted R^2 of 0.404 implying that the model explains around 40% or the variation in the dependent variable (Table S3, “full” column). The specification with no fixed effects (“nofe” column) had an adjusted R-squared of 0.283, indicating that a significant amount of variability in reported costs is explained by the additional explanatory variables. When fixed effect categories are removed, we can check which fixed effects explain the most variation relative to the others. Reporting source accounts for the most variation (adjusted R^2 drops to 0.355 when removed), followed by year (0.386) then basin (0.394) (Table S3, "nosource", "noyear", and "nobasin" columns). We also compared AIC and BIC; the specification with only fixed effects for reporting source had the lowest BIC (3,725.2, vs. 3,790 for the "full" specification), which includes a stronger penalty for the number of estimated parameters, while the full specification had the lowest AIC (3,488.7) (Table S3). When regression is fit only on culverts in the best represented basins (n = 1,091) and sources (n = 779), adjusted R^2 fell significantly (0.337 and 0.272 respectively), suggesting that information from the additional sources and outside the core basins improves the fit (Table S4, "basins_core" and "sources_core" columns). We also compared our worksite-level results to an alternative sample construction at the project level, with explanatory variables averaged across worksites comprising each project (Table S4, "project_level" column). In all cases, key results were qualitatively similar in terms of the signs and relative magnitudes of coefficient estimates.

Bankfull width and channel slope both were positively associated with average project costs, with standard deviation increases associated with 56% and 35% higher average costs respectively when the other is held at its mean (Figure 2). We also generated cost contours over bankfull width and channel slope space, representing combinations of the two variables associated with equivalent cost levels (Figure 3a). As one of the two hydrological variables

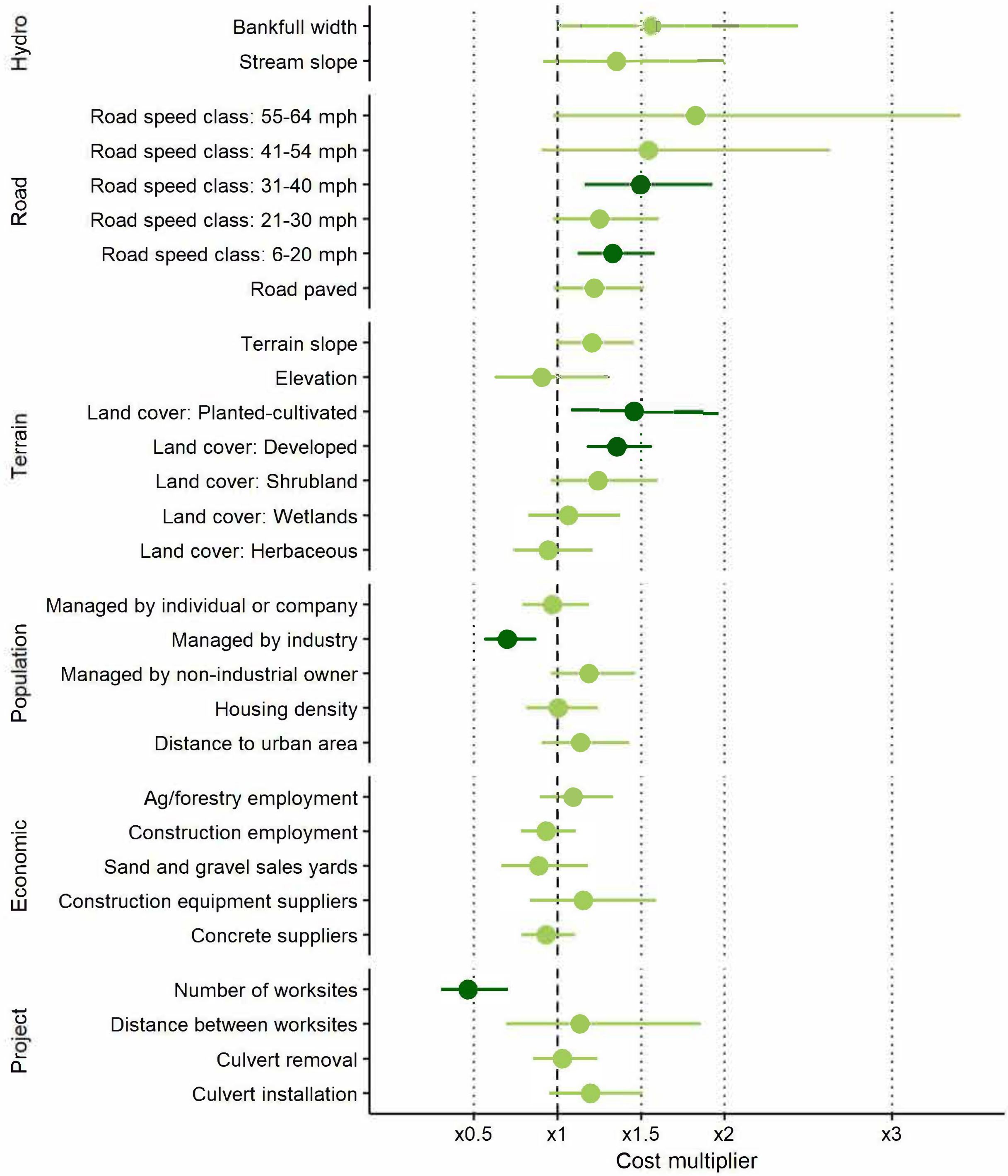

Figure 2. Standardized cost multipliers for hypothesized cost drivers. Multipliers are based on preferred specification of linear regression results ("full" column of Table S3). Horizontal lines represent 95% confidence intervals based on robust standard errors clustered at the project level. Dark indicates when confidence intervals do not include one (a null effect). Multipliers for variables with interactions are calculated with the opposite variable at its mean.

392

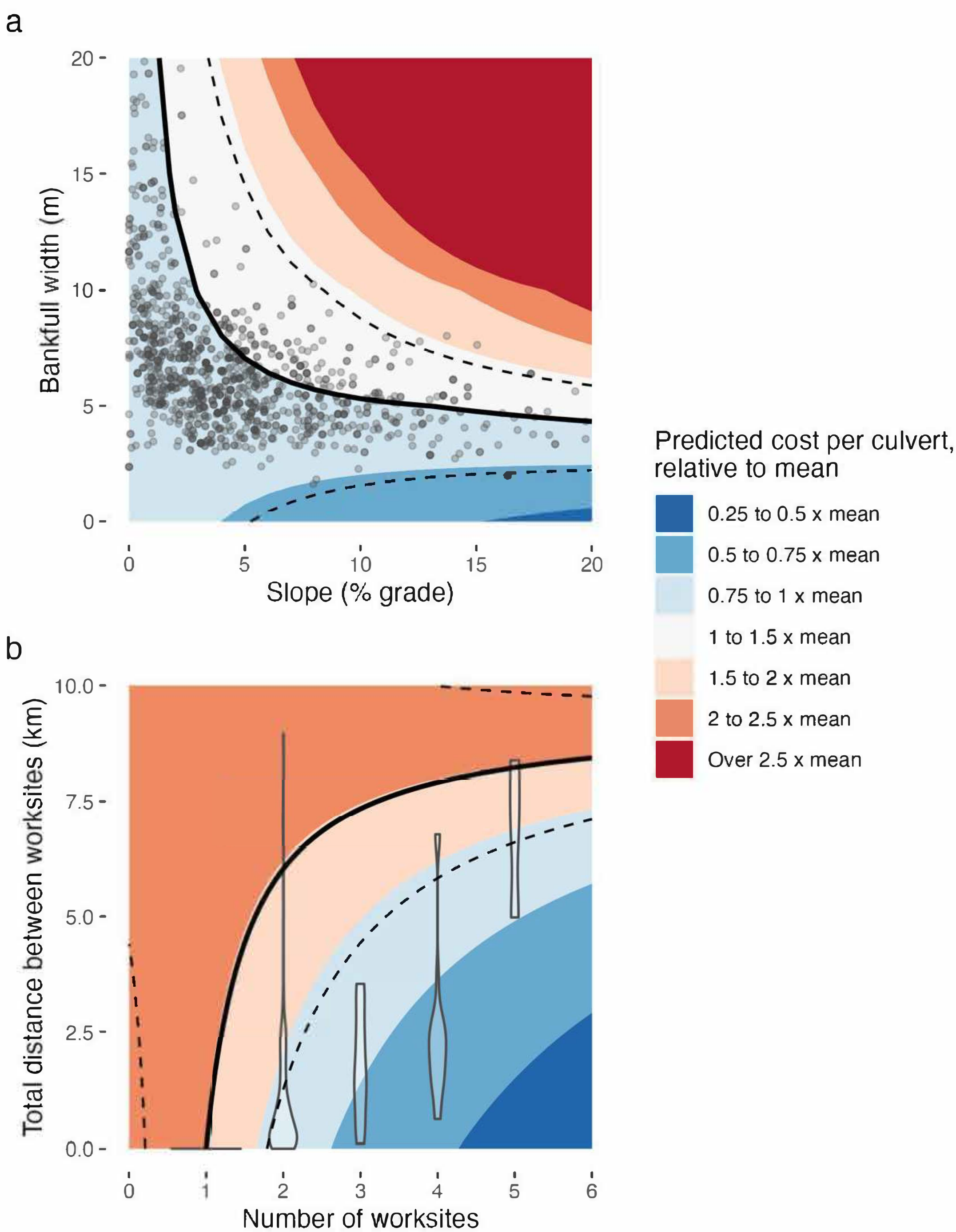

Figure 3. Cost contours for interaction effects of (a) channel slope and bankfull width and (b) number of worksites and total distance between worksites. Colors indicate ranges of predicted costs, based on linear regression, relative to the predicted cost when all explanatory variables are held at their means. Solid line indicates mean cost contour and dashed lines indicate 95% confidence intervals. Points in (a) and violin plots in (b) indicate distributions of underlying data in the sample.

393

increased, the strength of the positive association between the other and costs also increased, suggesting that culverts on streams that are wide and steep are more expensive to restore than those that are only one or the other.

As expected, we found a clear positive relationship between road speed class, which proxies for road width and traffic, and project average costs. Culvert improvements on roads with speed limits in the 6-30mph range were approximately 30% more expensive than worksites on the slowest and smallest speed class (5mph and below). Beyond 30mph, estimated cost multipliers rise to as high as 1.8, indicating costs nearly double on roads with speed limits typical of county highways (55-64mph). However, confidence intervals for the multipliers for the highest speed classes are large, as culvert improvements on these roads have relatively low representation in the PNSHP sample (Figure 2).

A standard deviation increase in the terrain slope of a culvert's catchment is associated with 21% higher project average costs i.e., culverts in hillier or more mountainous areas are more expensive to improve. On the other hand, higher elevations are associated with lower costs, though this relationship is not statistically significant (Figure 2).

Worksites with cultivated cropland and developed land covers had higher project average costs, all else equal, relative to worksites found in areas with forest land cover baseline (Figure 2). Worksites in shrubland land cover were also associated with higher project average costs, though the relationship is not statistically significant (Figure 2). The remaining land covers, wetlands and herbaceous, did not have costs significantly different than the forest land cover baseline.

The cost multipliers for population variables related to urbanization, housing density and distance to urban area, were statistically insignificant in the preferred model ($p > 0.05$) (Figure

2). Private land management by industrial and non-industrial owners had statistically significant cost multipliers, while lands owned by individuals or companies did not (Figure 2). For these variables, the baseline was worksites with no privately managed land within a 500m buffer. Such culverts are most often owned by government agencies. Compared to this baseline, culverts surrounded by more land managed by a non-industrial private owner (i.e., non-profit conservation groups) were associated with higher costs, while culverts surrounded by industrially managed land (i.e., land managed for forestry) were associated with lower costs (Figure 2). This key result indicates efficiencies associated with barrier improvements conducted by large forest landowners.

Worksites closer to suppliers of construction equipment, suppliers of brick and concrete materials, and to gravel and sand sales yards showed no significant associations with project costs (Figure 2). Similarly, the relationships between costs and employment variables were weak. Barriers in counties with more construction jobs were slightly cheaper, while those with more agriculture and forestry jobs were slightly more expensive, though neither association was statistically significant (Figure 2). We consider these estimates to be conservative, because we were only able to link worksites to employment data at the county-level. Counties in Oregon and Washington are large and often exhibit considerable heterogeneity in economic conditions and sectoral composition within their borders.

A standard deviation increase in number of worksites (0.95) was associated with 46% lower project average costs when other variables, most importantly distance between worksites, were held at their means (Figure 2). On the other hand, a standard deviation increase in distance between worksites (1.35 km) is associated with average costs 14% higher under the same conditions (Figure 2). This conflict is at the core of the managerial tradeoffs to consider when grouping worksites under a single project. Our results provide evidence for economies of scale that diminish with increased distance between sites.

The contours of the average project cost surface over number of worksites and the total distance between worksites can be interpreted as the distance limit for which adding an additional worksite to a project is associated with economies or diseconomies of scale (Figure 3b). That these contours tend to be quite steep initially indicates strong potential for economies of scale when opportunities to group nearby worksites under a single project arise. As the number of worksites increase, these contours flatten, indicating that the maximum distance for an efficient additional worksite drops quickly.

There is some evidence from the models that installations are more expensive than improvements (the baseline) and removals, though this association largely washes out when the full suite of fixed effects is included. This may indicate that distinctions between these categories in the data are loosely defined or inconsistently applied across reporting sources or over time.

We also provide standardized cost multipliers for each level of our fixed effects. Figures and discussion for these results can be found in the Supplementary Material (Supplementary Material 3).

3.2 Spatial distribution of predicted culvert restoration costs

We applied our BRT fit to the combined ODFW and WDFW culvert barrier inventories to examine patterns in the predicted costs of improvement for future culverts. To emphasize that these cost predictions are planning-level only and not a substitute for engineering-based design estimates, we only report cost predictions in relative terms (Figure 4a)[5]. Predictions for

[5] We present distributions (kernel density) of these predictions in their original scale in the supplementary material both for the full data and by basin as Figures S11 and S12 respectively.

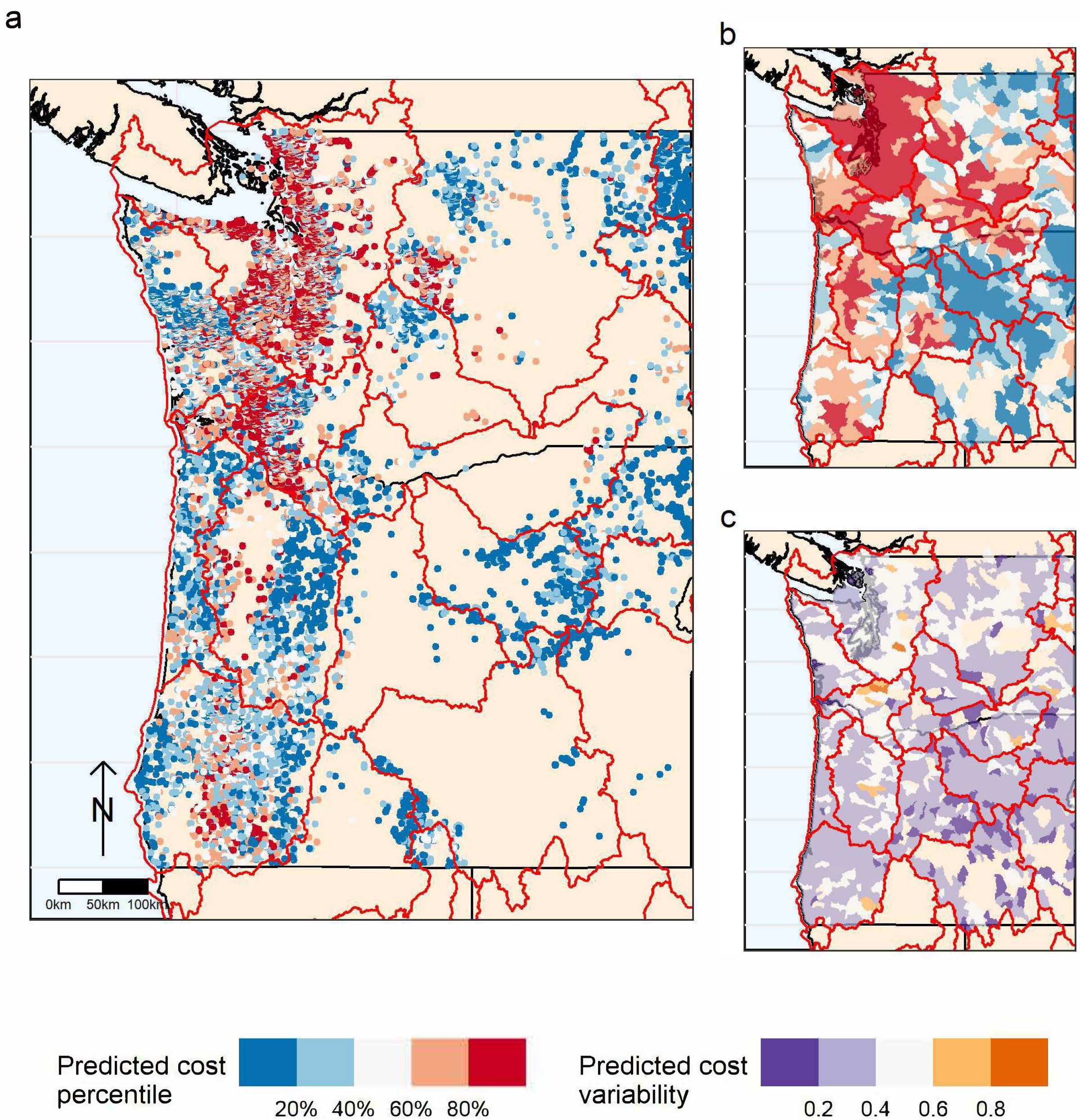

Figure 4. Predicted cost (a) levels for individual culverts, (b) mean levels at watershed-scale (HUC10), and (c) variability (coefficient of variation) at watershed-scale (HUC10). Red borders represent basins (HUC2). Barriers outside the area of applicability removed.

individual barriers are summarized at the HUC10 level both by mean predicted cost as a measure of cost levels and the coefficient of variation as a measure of cost variability.

The Puget Sound, Lower Columbia, and Willamette basins are home to barriers with the highest predicted costs (Figure 4b). These basins contain regions of relatively high development along a major interstate corridor, including the Seattle and Portland metro regions. While culverts in these areas tend to be closer to material suppliers, they also tend to be on larger, more heavily trafficked roads than those in other areas of the study region.

The Upper Columbia, Middle Columbia, Southern Oregon Coastal, and Washington Coastal basins exhibit some areas of high costs as well, though with many watersheds in the middle quantiles or lower (Figure 4b). On the other hand, the Northern Oregon Coastal and basins in Eastern Oregon exhibit consistently lower predicted costs. Barriers in these basins tend to be on smaller, less trafficked roads, on smaller streams, or on privately managed forestland, all of which contribute to lower costs. These regions also lack urban centers, proximity to which is associated with higher costs in the BRT predictions.

Identifying regions with cost variability can reveal where incorporating cost information into decision-making is most likely to improve restoration outcomes when budgets are limited. We mapped the coefficient of variation for cost predictions at the watershed level (Figure 4c). We found higher levels of cost variability in Western Washington watersheds, including those in Puget Sound, Washington Coastal, and Lower Columbia basins. These areas include regions of transition between the heavily urbanized Puget Sound shores and the more foothills of the Cascade and Olympic Ranges. The lowest levels of cost variability were found in Eastern Oregon. This region exhibits relatively consistent stream morphology and population density, while also having lower barrier density.

3.3 Distribution of costs over barrier owners

Incorporating costs into planning can improve agreement between management entities on where to prioritize limited funding, though whether this is possible depends on the co-

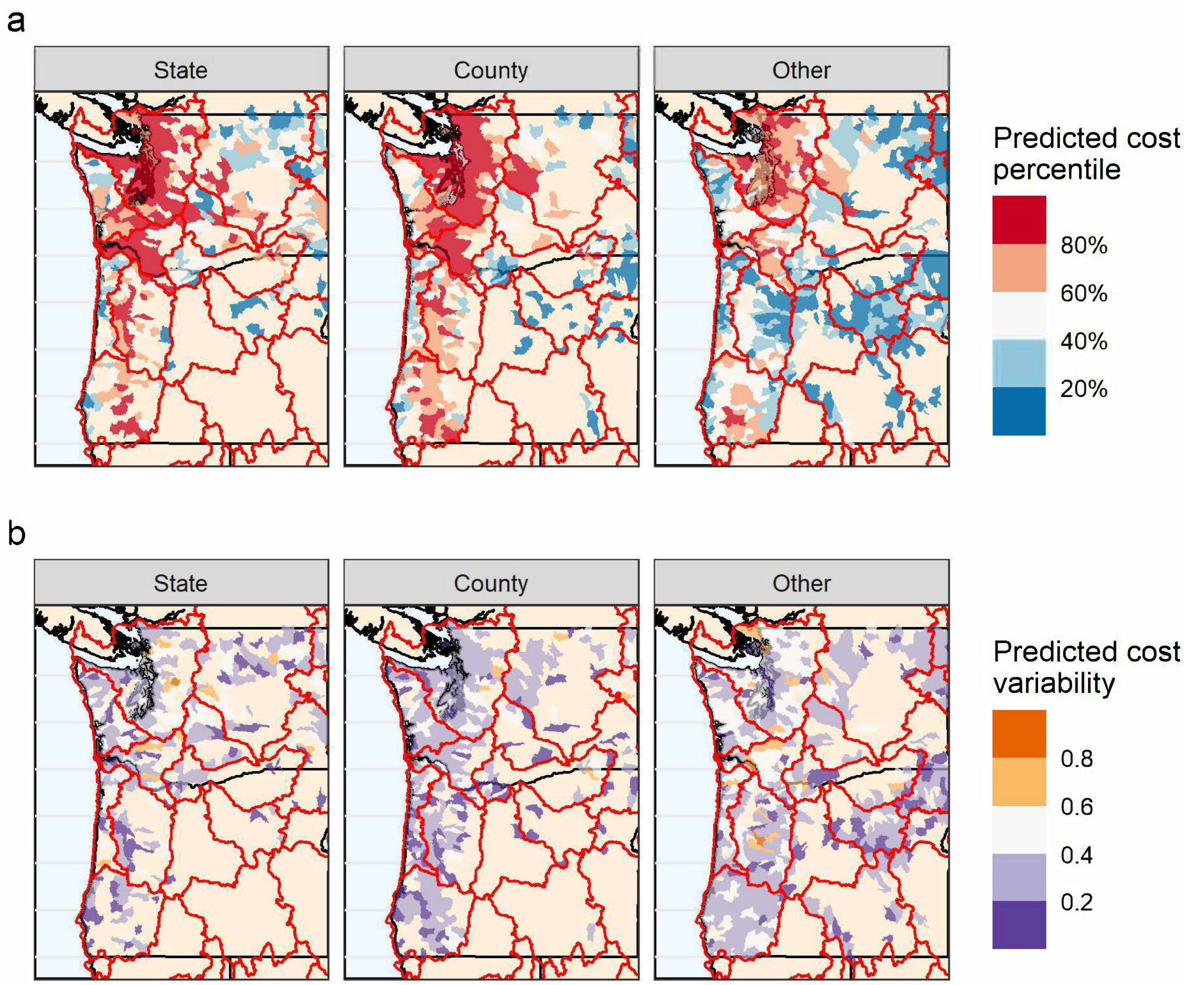

Figure 5. Predicted cost (a) levels for individual culverts, (b) mean levels at watershed-level (HUC10), and (c) variability (coefficient of variation) at watershed-level (HUC10). Red borders represent basins (HUC2). Barriers outside of the area of applicability removed.

distributions of costs and benefits across entities (Armsworth et al. 2017). To examine how cost distributions vary across barrier ownership entities, we calculate relative cost levels and cost variability metrics for three broad classes of entities: state agencies, county governments, and others, a category that includes mainly private landowners (Figure 5).

We find that spatial patterns for cost levels and variability are consistent across ownership entities, with the same regions exhibiting low/high mean costs and cost variability exhibiting the same relative patterns within each ownership group (Figure 5). Barriers and culverts owned by entities in the other category are consistently cheaper to improve than state- or county-owned barriers (Figure 5a), primarily because these barriers are primarily on smaller, privately-owned roads on private forest land. On the other hand, cost variability is consistent across groups, with the same regions exhibiting high-cost variability for all three classes of ownership entities (Figure 5b).

3.4 Area of applicability

The percentages of inventory barriers within the AOA for the BRT and RF models were 93% and 94% respectively. While the means of nearly every variable included in the linear model were statistically different between the training and inventory data (Table S5), the range of the PNSHP data covers the vast majority of the multivariate space occupied by barriers in the inventory data.

The barriers that do fall outside the AOA for our preferred predictive model are not uniformly distributed across the region (Figure 6). Barriers outside the AOA are concentrated in the urbanized and urbanizing watersheds surrounding Seattle, WA, where the share of barriers within the AOA is as low as 25% in some watersheds. Barriers most often outside the AOA are on the smallest roads or located in wetlands or herbaceous lands (Table 3). Barriers found in the Puget Sound basin were also more frequently outside the AOA, as were barriers in the southeastern basins of our study area (Middle Snake – Boise, Clearwater, and Black Rock Desert) (Table 3).

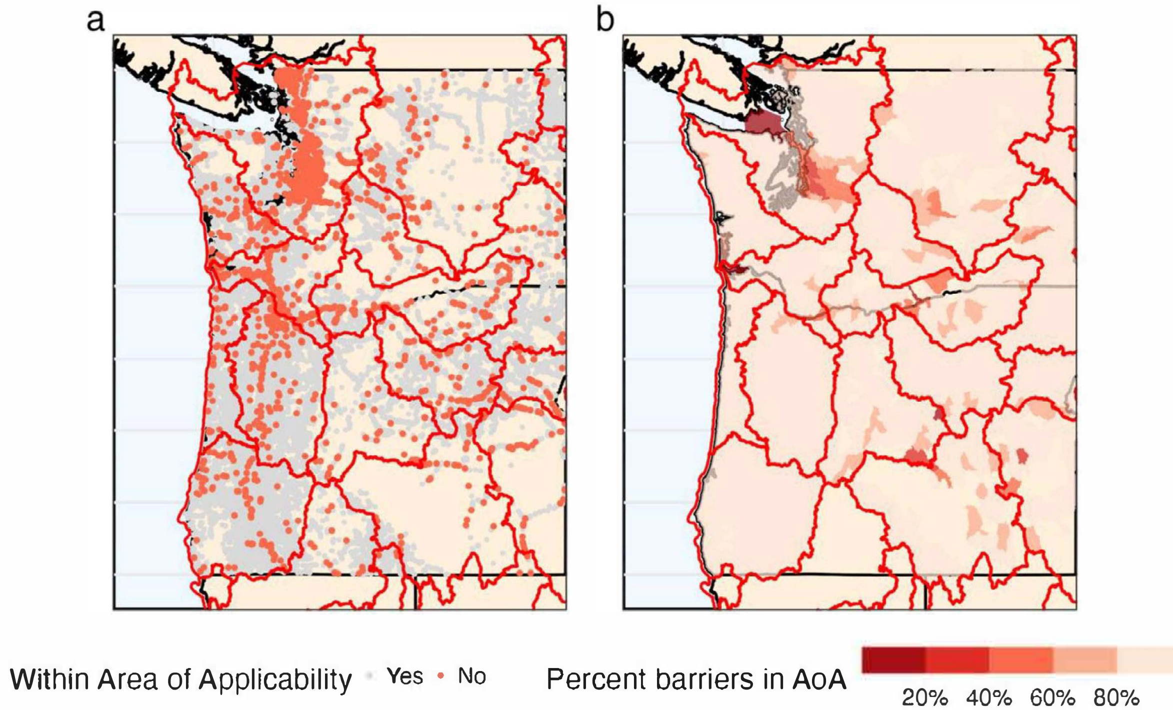

Figure 6. Area of applicability (AoA) for cost predictions presented at (a) individual barrier culvert and (b) watershed-level (HUC10). Red borders represent basins (HUC2).

4 Discussion

In this paper, we demonstrate use of project cost data for restoration planning in an applied setting. In situations like our study area, where planners have large but constrained budgets and hundreds if not thousands of potential projects to consider, it is not feasible to develop engineering estimates for each project. However, cost estimates are still needed to enable the most effective use of limited funds. Using linear regression, we identify cost drivers for culvert barrier removals, which can be used to inform future data gathering efforts or incorporated into planning tools as proxies for high- or low-cost projects. We expand on our linear regression results with predictions based on machine learning methods which provide improved accuracy. Our predictions reveal spatial patterns in cost levels and variation for potential future fish passage barrier removal projects across the Northwest and where

Table 3. Percent of inventory barriers within area of applicability by category.

Speed class	55-64 mph (3)	91.5%
	41-54 mph (4)	90.6%
	31-40 mph (5)	93.0%
	21-30 mph (6)	92.9%
	6-20 mph (7)	97.9%
	5 mph or less (8)	66.7%
Paved class	No	98.2%
	Yes	90.8%
NLCD land cover class	Developed	92.5%
	Forest	96.1%
	Herbaceous	90.0%
	Planted-cultivated	91.8%
	Shrubland	96.5%
	Wetlands	87.3%
Basin	BLACK ROCK DESERT	62.5%
	CLEARWATER	50.0%
	DESCHUTES	94.9%
	JOHN DAY	95.0%
	KLAMATH	97.0%
	LOWER COLUMBIA	92.0%
	LOWER SNAKE	97.4%
	MIDDLE COLUMBIA	91.0%
	MIDDLE SNAKE - BOISE	89.7%
	MIDDLE SNAKE - POWDER	95.5%
	NORTHERN OREGON COASTAL	98.6%
	OREGON CLOSED BASINS	93.5%
	PEND OREILLE	99.3%
	PUGET SOUND	82.8%
	SOUTHERN OREGON COASTAL	98.2%
	SPOKANE	98.2%
	UPPER COLUMBIA	95.4%
	UPPER SACRAMENTO	98.4%
	WASHINGTON COASTAL	98.2%
	WILLAMETTE	95.8%
	YAKIMA	95.6%

incorporating cost information into planning tools is likely to be most valuable. Finally, we provide metrics indicating where and for what barriers our predictions are most likely to be valid based on the similarity of potential worksites to past worksites where project cost data is available.

Our results identify variables that drive culvert improvement costs. With both linear regression and machine learning methods, we find that hydrological variables, especially bankfull width and channel slope, and road variables, especially road speed which serves as a simple proxy for many relevant road features, explain significant variation in culvert improvement costs. These variables and others included in our analysis are readily available for any georeferenced barrier culvert in the study region through either spatial data layers or field surveys. Indeed, existing field survey protocols for culvert evaluation already include methods for measuring many of these variables. Our results describing the area of applicability for our predictive methods highlight the types of barriers where more observations are needed to inform planning, including barriers on the largest and smallest roads, those in wetlands, and those in the Puget Sound basin.

Policymakers and resource managers responsible for planning barrier correction strategies can incorporate our planning-level predictions of relative costs, or measures of identified cost drivers, into decision-making frameworks and tools. There are many examples of tools and models for identifying high-value culverts to replace among a set of candidates. Optimization-based tools, common in the academic literatures, can identify the portfolio of barriers to correct which maximizes some ecological benefit, most often total habitat area or lineal distance, for a given budget while accounting for interdependencies in benefits that occur when multiple barriers are present along a stream network (Neeson et al. 2015; King &

O'Hanley 2016; King et al. 2017). Prioritization index tools, which assign scores to barriers based on a consistent rubric, rank barriers independently and are more commonly used by practitioners because of their consistency and transparency (McKay et al. 2017, 2020; Martin 2019). Both broad classes of prioritization tools can benefit from incorporating results and methods from this study.

Optimization tools most frequently incorporate costs only by assigning a fixed value to all barriers of a certain type, which unrealistically limits the amount of cost variability and fails to account for potential economies of scale achievable by completing nearby barrier corrections simultaneously. These tools can be improved by directly incorporating barrier-specific cost predictions based on models presented here as their cost metric, or by replicating our methods using regionally specific data on past barrier removals. For prioritization index tools, barriers with lower predicted costs, or with values for cost driving variables within certain ranges, can be assigned additional points to upweight projects that are likely to have lower costs. To our knowledge, no barrier prioritization index tool in our study region considers costs explicitly, instead relying entirely on metrics related to ecological benefits.

Our relative cost predictions reveal areas where costs to remove barrier culverts are likely to be high and where costs between projects are likely to be heterogeneous. The cost level results can be used by policymakers to anticipate budgetary needs, noting that all else equal, high-cost regions are likely to require more funding to achieve a targeted level of benefits. While more detailed ex ante cost information can always improve efficiency, the gains from incorporating such information are highest when cost variability is high and when costs are weakly correlated with benefits (Babcock et al. 1997; Naidoo et al. 2006). The cost variability results can be used to identify regions where incorporating cost information into planning is most likely to lead to efficiency gains.

Cost predictions are most useful when paired with a measure of the benefits of a project so that a benefit-cost ratio can be calculated, allowing for a formal return-on-investment analysis (Babcock et al. 1997; Naidoo et al. 2006). Benefit assessment in the fish passage context is a complex problem, arguably more complex than cost assessment. Most typically, benefits for barrier correction are defined in terms of increased fish production through improved habitat access. This can be proxied for through upstream distance or area and can be made more proximate to the true benefits by weighting by habitat quality measures. However, barriers often exist in sequence along branching stream networks, making a benefits assessment of a single barrier's correction futile as correcting one barrier will often affect the benefits for others.

As a "first-order" measure of benefits, we compute the total upstream distance for each barrier based on the NHDPlus V2.1 flow lines and plot this benefit measure against our predicted costs (Figure 7). We emphasize that this is only at best a rough measure of benefits intended to illustrate how costs can be used alongside benefits to assess suitable prioritization frameworks when information and planning is costly. This exercise reveals wide variation in both costs and benefits, as well as in the benefit-cost ratios. Selecting projects based on benefits alone would clearly result in a different set of projects than a selection method based on predicted costs or benefit-cost ratios, demonstrating the value of considering both measures in planning.

Fish passage barrier removal is an ideal first application for our methods due to the discrete point nature of worksites, frequent occurrence of such projects in recent decades, consistent inventory efforts across the region, and availability of spatial data layers for both roads and streams. Beyond our case study application, our methods have the potential to be

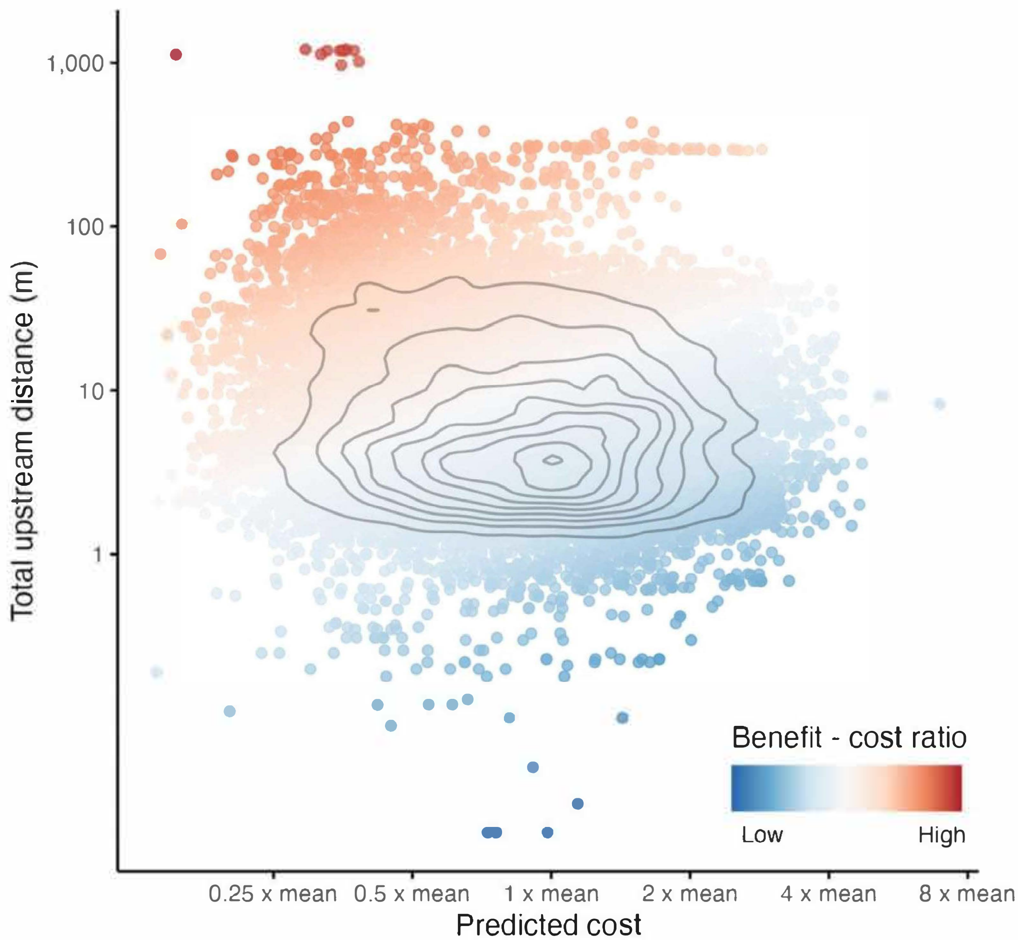

Figure 7. Benefit-cost ratios for inventory barriers. Benefits are defined as total upstream distance. Contours represent density of points. Both axes are on a log-scale. Barriers outside the area of applicability are omitted.

replicated in other restoration settings where ex ante cost data is sparse or expensive to acquire. By relying primarily on readily available spatial data as indicators for cost drivers, we avoid reliance on expensive field surveys of worksite conditions, allowing the comparison of cost estimates for thousands of potential future projects in an affordable and consistent manner. On the other hand, other restoration actions, such as riparian plantings, invasive species control, or the installation of large woody debris, are often conducted over broader areas, and the scope of work and project costs are often poorly documented, which can limit the use of cost data in planning (Cook et al. 2017; Pienkowski et al. 2021).

Comparing culvert removals against other forms of restoration activities is another relevant restoration planning problem where these types of costing models could be useful. Fonner et al. (2021) developed a spatially specific restoration planning model at the HUC12 scale using the average cost of past restoration actions, including culvert removals, to approximate intervention costs. The types of models in this paper can enhance analyses of this type by providing finer granularity in the cost information considered.

In the future, resource managers and scientists should strive to improve the collection and aggregation of consistent data on restoration projects of all types (Cook et al. 2017; Iacona et al. 2018). Doing so will allow for more detailed analysis of restoration costs at broader scales and comparison of costs between action types, allowing for improved restoration planning and the identification of projects that maximize returns on investment. One particularly useful data collection method to improve cost analyses in the future would be the disaggregation of project costs into expenditure categories, such as labor vs. materials or design vs. construction. Breaking down project costs into such categories would allow for the identification of programs to reduce costs in a targeted way, beyond only identifying where projects might be more or less expensive.

Sparse evidence on costs and challenges related to forecasting costs for future projects are frequently cited as a limitation in using return-on-investment or cost-effectiveness as a key decision-making metric (Armsworth 2014; Pienkowski et al. 2021). By utilizing the methods demonstrated in this paper, conservation scientists and practitioners can obtain planning-level cost estimates for future projects when data on past actions are available. Opportunities to apply such methods will increase as more conservation organizations adopt standardized cost reporting procedures (Iacona et al. 2018).

Restoration planners and scientists can improve restoration efficiency by incorporating cost considerations into decision-making. Ignoring costs and basing decisions on ecological benefits alone is likely to result in reduced efficiency, leaving potential gains on the table. The PNSHP database and other clearinghouse databases for restoration projects represent an untapped data resource for analyzing past costs to inform future decisions. Improved recordkeeping and data sharing can allow for better cost data, which in turn will lead to more opportunities to empirically identify cost drivers, areas where costs are likely to be high, and areas with highly variable costs across projects.

Literature cited

Armsworth PR. 2014. Inclusion of costs in conservation planning depends on limited datasets and hopeful assumptions. Annals of the New York Academy of Sciences **1322**:61–76.

Armsworth PR, Cantú-Salazar L, Parnell M, Davies ZG, Stoneman R. 2011. Management costs for small protected areas and economies of scale in habitat conservation. Biological Conservation **144**:423–429.

Armsworth PR, Jackson HB, Cho S-H, Clark M, Fargione JE, Iacona GD, Kim T, Larson ER, Minney T, Sutton NA. 2017. Factoring economic costs into conservation planning may not improve agreement over priorities for protection. Nature Communications **8**:2253. Nature Publishing Group.

Armsworth PR, Jackson HB, Cho S-H, Clark M, Fargione JE, Iacona GD, Kim T, Larson ER, Minney T, Sutton NA. 2018. Is conservation right to go big? Protected area size and conservation return-on-investment. Biological Conservation **225**:229–236.

Athey S, Imbens GW. 2019. Machine Learning Methods That Economists Should Know About. Annual Review of Economics **11**:685–725.

Babcock BA, Lakshminarayan PG, Wu J, Zilberman D. 1997. Targeting Tools for the Purchase of Environmental Amenities. Land Economics **73**:325–339.

Barnard RJ, Johnson J, Brooks P, Bates KM, Heiner B, Klavas JP, Ponder DC, Smith PD, Powers PD. 2013. Water Crossing Design Guidelines. Page 300. Washington Department of Fish and Wildlife, Olympia, Washington. Available from http://wdfw.wa.gov/hab/ahg/culverts.htm.

BenDor T, Lester TW, Livengood A, Davis A, Yonavjak L. 2015. Estimating the Size and Impact of the Ecological Restoration Economy. PLOS ONE **10**:e0128339. Public Library of Science.

Bernhardt ES et al. 2005. Synthesizing U.S. River Restoration Efforts. Science **308**:636–637. American Association for the Advancement of Science.

Boyd J, Epanchin-Niell R, Siikamäki J. 2015. Conservation Planning: A Review of Return on Investment Analysis. Review of Environmental Economics and Policy **9**:23–42. The University of Chicago Press.

Brown A. 2019, August 23. Coming Down the Pipe: Saving Washington's Salmon May Require Replacing Tens of Thousands of Culverts, and Nobody Knows Where the Money Will Come From. The Daily Chronicle. Centralia, Washington. Available from https://www.chronline.com/stories/coming-down-the-pipe-saving-washingtons-salmon-may-require-replacing-tens-of-thousands-of,8033 (accessed May 10, 2021).

Burkhalter JC, Lockwood JL, Maslo B, Fenn KH, Leu K. 2016. Effects of cost metric on cost-effectiveness of protected-area network design in urban landscapes. Conservation Biology **30**:403–412.

Carwardine J, Wilson KA, Hajkowicz SA, Smith RJ, Klein CJ, Watts M, Possingham HP. 2010. Conservation Planning when Costs Are Uncertain. Conservation Biology **24**:1529–1537.

Cho S-H, Kim T, Larson ER, Armsworth PR. 2017. Economies of scale in forestland acquisition costs for nature conservation. Forest Policy and Economics **75**:73–82.

Cook CN, Pullin AS, Sutherland WJ, Stewart GB, Carrasco LR. 2017. Considering cost alongside the effectiveness of management in evidence-based conservation: A systematic reporting protocol. Biological Conservation **209**:508–516.

De'ath G. 2007. Boosted Trees for Ecological Modeling and Prediction. Ecology **88**:243–251.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology **77**:802–813.

Field CR, Elphick CS. 2019. Quantifying the return on investment of social and ecological data for conservation planning. Environmental Research Letters **14**:124081. IOP Publishing.

Fonner R, Honea J, Jorgensen JC, Plummer M, McClure M. 2021. Considering intervention intensity in habitat restoration planning: An application to Pacific salmon. Journal of Environmental Management **299**:113536.

Goolsbee A, Levitt S, Syverson C. 2013. Microeconomics. Worth Publishers, New York, New York.

Greenwall BM, Boehmke BC. 2020. Variable importance plots – An introduction to the vip package. The R Journal **12**: 343-366.

Heckman JJ. 1979. Sample Selection Bias as a Specification Error. Econometrica **47**:153–161.

Hickey R. 2018. Highway Culverts, Salmon Runs, and the Stevens Treaties: A Century of Litigating Pacific Northwest Tribal Fishing Rights. Pub. Land & Resources L. Rev. **39**:253. HeinOnline.

Iacona GD et al. 2018. Standardized reporting of the costs of management interventions for biodiversity conservation. Conservation Biology **32**:979–988.

Infrastructure Investment and Jobs Act, Pub. L. No. 117-58, 135 Stat. 676 (2021).

Kim T, Cho S-H, Larson ER, Armsworth PR. 2014. Protected area acquisition costs show economies of scale with area. Ecological Economics **107**:122–132.

King S, O'Hanley JR. 2016. Optimal Fish Passage Barrier Removal—Revisited. River Research and Applications **32**:418–428.

King S, O'Hanley JR, Newbold LR, Kemp PS, Diebel MW. 2017. A toolkit for optimizing fish passage barrier mitigation actions. Journal of Applied Ecology **54**:599–611.

Kujala H, Lahoz- Monfort JJ, Elith J, Moilanen A. 2018. Not all data are equal: Influence of data type and amount in spatial conservation prioritisation. Methods in Ecology and Evolution **9**:2249–2261.

Martin EH. 2019. Assessing and Prioritizing Barriers to Aquatic Connectivity in the Eastern United States. JAWRA Journal of the American Water Resources Association **55**:401–412.

McKay SK, Cooper AR, Diebel MW, Elkins D, Oldford G, Roghair C, Wieferich D. 2017. Informing Watershed Connectivity Barrier Prioritization Decisions: A Synthesis. River Research and Applications **33**:847–862.

McKay SK, Martin EH, McIntyre PB, Milt AW, Moody AT, Neeson TM. 2020. A comparison of approaches for prioritizing removal and repair of barriers to stream connectivity. River Research and Applications **36**:1754–1761. Wiley Online Library.

Meyer H, Pebesma E. 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. Nature Communications **13**: 2208.

Meyer H, Pebesma E. 2021. Predicting into unkown space? Estimating the area of applicability of spatial prediction models. Methods in Ecology and Evolution **12**: 1620-1633.

Move Ahead WA Act, Washington State Legislature Ch. 182, Laws of 2022 (2022).

Mullainathan S, Spiess J. 2017. Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives **31**:87–106.

Naidoo R, Balmford A, Ferraro PJ, Polasky S, Ricketts TH, Rouget M. 2006. Integrating economic costs into conservation planning. Trends in Ecology & Evolution **21**:681–687.

Neeson TM, Ferris MC, Diebel MW, Doran PJ, O'Hanley JR, McIntyre PB. 2015. Enhancing ecosystem restoration efficiency through spatial and temporal coordination. Proceedings of the National Academy of Sciences **112**:6236–6241. National Academy of Sciences.

Newbold SC, Siikamäki J. 2015. Conservation prioritization using reserve site selection methods. Page in R. Halvorsen and D. Layton, editors. Handbook on the Economics of Natural Resources. Edward Elgar Publishing, Cheltenham, UK.

Pacific Northwest salmon habitat project database. 2021. National Marine Fisheries Service, Northwest Fisheries Science Center. Available from https://www.webapps.nwfsc.noaa.gov/pnshp/.

Perhans K, Kindstrand C, Boman M, Djupström LB, Gustafsson L, Mattsson L, Schroeder LM, Weslien J, Wikberg S. 2008. Conservation Goals and the Relative Importance of Costs and Benefits in Reserve Selection. Conservation Biology **22**:1331–1339.

Phillips-Mao L, Refsland JM, Galatowitsch SM. 2015. Cost-Estimation for Landscape-Scale Restoration Planning in the Upper Midwest, U.S. Ecological Restoration **33**:135–146.

Pienkowski T, Cook C, Verma M, Carrasco LR. 2021. Conservation cost-effectiveness: a review of the evidence base. Conservation Science and Practice **3**:e357.

Rodewald AD, Strimas-Mackey M, Schuster R, Arcese P. 2019. Tradeoffs in the value of biodiversity feature and cost data in conservation prioritization. Scientific Reports **9**:15921.

Roni P, Beechie TJ. 2013. Stream and watershed restoration: a guide to restoring riverine processes and habitats. John Wiley & Sons, Chichester, West Sussex ; Hoboken, NJ.

Roni P, Beechie TJ, Bilby RE, Leonetti FE, Pollock MM, Pess GR. 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.

Sacre E, Pressey RL, Bode M. 2019. Costs are not necessarily correlated with threats in conservation landscapes. Conservation Letters **12**:e12663.

Storm H, Baylis K, Heckelei T. 2020. Machine learning in agricultural and applied economics. European Review of Agricultural Economics **47**:849–892.

Sutton NJ, Cho S, Armsworth PR. 2016. A reliance on agricultural land values in conservation planning alters the spatial distribution of priorities and overestimates the acquisition costs of protected areas. Biological Conservation **194**:2–10.

Tredennick AT, Hooker G, Ellner SP, Adler PB. 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. Ecology **102**:e03336.

Washington Department of Fish and Wildlife. 2019. Fish Passage Inventory, Assessment, and Prioritization Manual. Olympia, Washington. Available from https://wdfw.wa.gov/sites/default/files/publications/02061/Fish%20Passage%20Inventory%2C%20Assessment%2C%20and%20Prioritization%20Manual.pdf.

Wooldridge JW. 2010. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, Massachusetts.

Tables

Table 1. Variable categories with examples and data sources.

Category	Example variables	Data sources
Hydrological	Channel slope (%), bankfull width (m), sinuosity, basin (categorical)	NHDPlus V2, NHDPlus Select Attributes V2.1
Road	Material (categorical), speed limit (categorical), functional class (categorical), distance to road feature (m), road crossing density (crossings per catchment)	HERE Geodatabase Plus, NHDPlus Select Attributes V2.1
Terrain	Terrain slope (degree), elevation (m), land cover (categorical)	NHDPlus Select Attributes V2.1, NLCD, GTOPO30
Population	Housing density (units per km^2), lands managed by private industry (% within 2.5km radius), distance to urban area (m)	BLM Surface Jurisdiction of Lands, US Census Urban and Rural Classifications,

Economic	Construction employment (jobs), density of sand & gravel sales yards	US Census County Business Patterns, HIFLD
Project	Number of worksites, distance between worksites (m), action type (categorical), project year, reporting source (categorical)	PNSHP

Table 2. Descriptive statistics for variables included in linear regression (n = 1,236)

Category	Variable	Mean	Std. dev.	Number of levels
Dependent variable	Cost per culvert ($USD2019)	82,600	93,800	
Hydrological (HYDRO)	Stream slope (%)	4.58	4.06	
	Bankfull width (m)	7.56	5.41	
	Basin			9
Road (ROAD)	Paved road			2
	Road speed class			6
Terrain (TERR)	Land cover class			6
	Terrain slope (deg)	27.3	12	
	Elevation (m)	412.2	317.7	
Population (POP)	Distance to urban area (m)	45,200	31,800	
	Housing density (units per sq. km)	5.96	24.7	
	Private land, individual or company owner (proportion of land within 500m radius)	0.447	0.436	
	Private land, managed by industry (proportion of land within 500m radius)	0.203	0.356	
	Private land, managed by non-industrial owner (proportion of land within 500m radius)	0.0349	0.169	
Economic (ECON)	Construction employment (jobs)	3,010	5,140	
	Ag/forestry employment (jobs)	778	534	
	Density of construction equipment wholesalers (workers per sq. km)	0.0216	0.032	
	Density of brick, stone, and related wholesalers (workers per sq. km)	0.00276	0.00721	
	Density of sand and gravel sales yards (firms per sq. km)	0.0000244	0.00004	
Project (PROJ)	Number of worksites (count)	1.48	0.958	
	Distance between worksites (m)	507	1,360	
	Year			15
	Reporting source			6

Table 3. Percent of inventory barriers within area of applicability by category.

Speed class	55-64 mph (3)	91.5%
	41-54 mph (4)	90.6%
	31-40 mph (5)	93.0%
	21-30 mph (6)	92.9%
	6-20 mph (7)	97.9%
	5 mph or less (8)	66.7%
Paved class	No	98.2%
	Yes	90.8%
NLCD land cover class	Developed	92.5%
	Forest	96.1%
	Herbaceous	90.0%
	Planted-cultivated	91.8%
	Shrubland	96.5%
	Wetlands	87.3%
Basin	BLACK ROCK DESERT	62.5%
	CLEARWATER	50.0%
	DESCHUTES	94.9%
	JOHN DAY	95.0%
	KLAMATH	97.0%
	LOWER COLUMBIA	92.0%
	LOWER SNAKE	97.4%
	MIDDLE COLUMBIA	91.0%
	MIDDLE SNAKE - BOISE	89.7%
	MIDDLE SNAKE - POWDER	95.5%
	NORTHERN OREGON COASTAL	98.6%
	OREGON CLOSED BASINS	93.5%
	PEND OREILLE	99.3%
	PUGET SOUND	82.8%
	SOUTHERN OREGON COASTAL	98.2%
	SPOKANE	98.2%
	UPPER COLUMBIA	95.4%
	UPPER SACRAMENTO	98.4%
	WASHINGTON COASTAL	98.2%
	WILLAMETTE	95.8%
	YAKIMA	95.6%

Table S1. Full variable list with labels and descriptions.

Category	Label	Description	Included in OLS?
Hydrological	basin	Basin (HUC6) containing worksite (categorical)	X
	slope	Channel slope (%) (NHDPlus V2.1)	X
	upst_dist	Upstream routed distance (m) (NHDPlus V2.1)	
	ppt_cat	Annual precipitation in catchment (mm) (NHDPLUS V2.1 Select Attributes data)	
	ppt_acc	… flow accumulated upstream catchments	
	ppt_tot	… all upstream catchments	
	bfi_cat	Base-flow index in catchment (NHDPLUS V2.1 Select Attributes data)	
	bfi_acc	… flow accumulated upstream catchments	
	bfi_tot	… all upstream catchments	
	bankfull_width	Bankfull width (m) (NHDPLUS V2.1 Select Attributes data)	X
	bankfull_depth	Bankfull depth (m) (NHDPLUS V2.1 Select Attributes data)	
	bankfull_xsec_area	Bankfull cross-section area (m2) (NHDPLUS V2.1 Select Attributes data)	
	cat_streamriver	Proportion of catchment waterways classified as stream/river (NHDPLUS V2.1 Select Attributes data)	
	acc_streamriver	… flow accumulated upstream catchments	
	tot_streamriver	… all upstream catchments	
	cat_artificial	Proportion of catchment waterways classified as artificial (NHDPLUS V2.1 Select Attributes data)	
	acc_artificial	… flow accumulated upstream catchments	
	tot_artificial	… all upstream catchments	
	cat_canalditch	Proportion of catchment waterways classified as canal or ditch (NHDPLUS V2.1 Select Attributes data)	
	acc_canalditch	… flow accumulated upstream catchments	
	tot_canalditch	… all upstream catchments	

	cat_connector	Proportion of catchment waterways classified as connector (NHDPLUS V2.1 Select Attributes data)	
	acc_connector	… flow accumulated upstream catchments	
	tot_connector	… all upstream catchments	
	cat_pipeline	Proportion of catchment waterways classified as pipeline(NHDPLUS V2.1 Select Attributes data)	
	acc_pipeline	… flow accumulated upstream catchments	
	tot_pipeline	… all upstream catchments	
	cat_basin_area	Catchment area (km2) (NHDPLUS V2.1 Select Attributes data)	
	acc_basin_area	… flow accumulated upstream catchments	
	tot_basin_area	… all upstream catchments	
	cat_stream_slope	Catchment stream slope, mean (%) (NHDPLUS V2.1 Select Attributes data)	
	acc_stream_slope	… flow accumulated upstream catchments	
	tot_stream_slope	… all upstream catchments	
	cat_stream_length	Catchment main channel length (km) (NHDPLUS V2.1 Select Attributes data)	
	acc_stream_length	… flow accumulated upstream catchments	
	tot_stream_length	… all upstream catchments	
	sinuosity	Catchment sinuosity, or amount of channel meandering (NHDPLUS V2.1 Select Attributes data)	
	cat_strm_dens	Catchment stream density (stream length [m] / catchment area [km2]) (NHDPLUS V2.1 Select Attributes data)	
	acc_strm_dens	… flow accumulated upstream catchments	
	tot_strm_dens	… all upstream catchments	
	hu_type	Type category of 6th field hydrologic unit ("HUC6")	
Road	cat_rdx	Catchment road crossing density (count per catchment) (NHDPLUS V2.1 Select Attributes data)	
	acc_rdx	… flow accumulated upstream catchments	
	tot_rdx	… all upstream catchments	
	here_distm	Distance (m) to nearest HERE (formerly NAVTEQ) road within 5000m	
	here_speed	Speed category of nearest HERE (formerly NAVTEQ) road, worksites	X

		over 150m from match assigned to lowest class	
	here_class	Functional class of nearest HERE (formerly NAVTEQ) road, worksites over 150m from match assigned to lowest class	
	here_paved	Is nearest HERE (formerly NAVTEQ) road paved, worksites over 150m from match assigned to lowest class	X
	here_publi	Is there public access to the nearest HERE (formerly NAVTEQ) road, worksites over 150m from match assigned to lowest class	
	here_class_0	HERE road class with no accuracy check	
	here_class_badmatch	HERE road class with bad matches assigned own class	
	here_speed_0	HERE speed class with no accuracy check	
	here_speed_badmatch	HERE speed class with bad matches assigned own class	
	here_paved_0	HERE paved indicator with no accuracy check	
Terrain	nlcd_barren_cat	NLCD barren land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	
	nlcd_barren_acc	… flow accumulated upstream catchments	
	nlcd_barren_tot	… all upstream catchments	
	nlcd_crop_cat	NLCD cropland land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	
	nlcd_crop_acc	… flow accumulated upstream catchments	
	nlcd_crop_tot	… all upstream catchments	
	nlcd_devhigh_cat	NLCD developed, high-intensity land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	
	nlcd_devhigh_acc	… flow accumulated upstream catchments	
	nlcd_devhigh_tot	… all upstream catchments	
	nlcd_devlow_cat	NLCD developed, low-intensity land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	
	nlcd_devlow_acc	… flow accumulated upstream catchments	
	nlcd_devlow_tot	… all upstream catchments	
	nlcd_devmed_cat	NLCD developed, medium-intensity land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	

nlcd_devmed_acc	… flow accumulated upstream catchments
nlcd_devmed_tot	… all upstream catchments
nlcd_devopen_cat	NLCD developed, open space land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)
nlcd_devopen_acc	… flow accumulated upstream catchments
nlcd_devopen_tot	… all upstream catchments
nlcd_forcon_cat	NLCD forest, coniferous land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)
nlcd_forcon_acc	… flow accumulated upstream catchments
nlcd_forcon_tot	… all upstream catchments
nlcd_fordec_cat	NLCD forest, deciduous land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)
nlcd_fordec_acc	… flow accumulated upstream catchments
nlcd_fordec_tot	… all upstream catchments
nlcd_formix_cat	NLCD forest, mixed land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)
nlcd_formix_acc	… flow accumulated upstream catchments
nlcd_formix_tot	… all upstream catchments
nlcd_grass_cat	NLCD grassland land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)
nlcd_grass_acc	… flow accumulated upstream catchments
nlcd_grass_tot	… all upstream catchments
nlcd_icesnow_cat	NLCD ice/snow land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)
nlcd_icesnow_acc	… flow accumulated upstream catchments
nlcd_icesnow_tot	… all upstream catchments
nlcd_nodat_cat	NLCD no data land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)
nlcd_nodat_acc	… flow accumulated upstream catchments
nlcd_nodat_tot	… all upstream catchments
nlcd_openwater_cat	NLCD open water land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)
nlcd_openwater_acc	… flow accumulated upstream catchments
nlcd_openwater_tot	… all upstream catchments

nlcd_pasture_cat	NLCD pasture land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	
nlcd_pasture_acc	… flow accumulated upstream catchments	
nlcd_pasture_tot	… all upstream catchments	
nlcd_shrub_cat	NLCD shrubland land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	
nlcd_shrub_acc	… flow accumulated upstream catchments	
nlcd_shrub_tot	… all upstream catchments	
nlcd_wetherb_cat	NLCD wetland, herbacous land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	
nlcd_wetherb_acc	… flow accumulated upstream catchments	
nlcd_wetherb_tot	… all upstream catchments	
nlcd_wetwood_cat	NLCD woody wetland land cover in catchment (share) (NHDPlus V2.1 Select Attributes data)	
nlcd_wetwood_acc	… flow accumulated upstream catchments	
nlcd_wetwood_tot	… all upstream catchments	
cat_basin_slope	Catchment slope (%) (NHDPLUS V2.1 Select Attributes data)	X
acc_basin_slope	… flow accumulated upstream catchments	
tot_basin_slope	… all upstream catchments	
cat_elev_mean	Catchment mean elevation (m) (NHDPLUS V2.1 Select Attributes data)	X
acc_elev_mean	… flow accumulated upstream catchments	
tot_elev_mean	… all upstream catchments	
cat_elev_min	Catchment minimum elevation (m) (NHDPLUS V2.1 Select Attributes data)	
acc_elev_min	… flow accumulated upstream catchments	
tot_elev_min	… all upstream catchments	
cat_elev_max	Catchment maximum elevation (m) (NHDPLUS V2.1 Select Attributes data)	
acc_elev_max	… flow accumulated upstream catchments	
tot_elev_max	… all upstream catchments	
slope_deg	Slope (°) of GTOPO30 grid cell that culvert restoration site falls within	
nlcd_current_class	NLCD land cover at worksite point for project year, simplified categories	X

	nlcd_current_fullclass	NLCD land cover at worksite point for project year	
Population	ua_dist	Distance (m) to Census designated “urban area” (50,000 or more people)	X
	uc_dist	Distance (m) to Census designated “urban cluster” (at least 25,000 people and less than 50,000 people) or urban area	
	popdens_cat	Population density (persons per km2) in catchment (NHDPlus V2.1 Select Attributes data)	
	popdens_acc	… flow accumulated upstream catchments	
	popdens_tot	… all upstream catchments	
	hdens_cat	Housing density (units per km2) in catchment (NHDPlus V2.1 Select Attributes data)	X
	hdens_acc	… flow accumulated upstream catchments	
	hdens_tot	… all upstream catchments	
	bia_5km_buff	Proportion of land owned by the Bureau of Indian Affairs within a 2500m radius of the site	
	blm_5km_buff	Proportion of land owned by the Bureau of Land Management within a 2500m radius of the site	
	bpa_5km_buff	Proportion of land owned by the Bonneville Power Administration within a 2500m radius of the site	
	coe_5km_buff	Proportion of land owned by the Corps of Engineers within a 2500m radius of the site	
	fws_5km_buff	Proportion of land owned by the U.S. Fish and Wildlife Service within a 2500m radius of the site	
	gsa_5km_buff	Proportion of land owned by the General Services Administration within a 2500m radius of the site	
	lg_5km_buff	Proportion of land owned by the Local Government within a 2500m radius of the site	
	nps_5km_buff	Proportion of land owned by the National Park Service within a 2500m radius of the site	
	pv_5km_buff	Proportion of land owned by a Private Individual or Company within a 2500m radius of the site	

pvi_5km_buff	Proportion of Lands Managed by Private Industry within a 2500m radius of the site
pvn_5km_buff	Proportion of land owned by Private Non-Industrial Owner within a 2500m radius of the site
pvu_5km_buff	Proportion of land owned by Private Urban Lands within a 2500m radius of the site
st_5km_buff	Proportion of land owned by a State Agency within a 2500m radius of the site
stf_5km_buff	Proportion of land owned by the State Dept. of Forestry within a 2500m radius of the site
stl_5km_buff	Proportion of land owned by the Division of State Lands within a 2500m radius of the site
stp_5km_buff	Proportion of land owned by the State Dept. of Parks and Recreation within a 2500m radius of the site
stw_5km_buff	Proportion of land owned by the State Dept. of Fish and Wildlife within a 2500m radius of the site
und_5km_buff	Proportion of land where ownership is Undetermined within a 2500m radius of the site
usfs_5km_buff	Proportion of land owned by the U.S. Forest Service within a 2500m radius of the site
water_5km_buff	Proportion of land that is Water within a 2500m radius of the site
bia_2km_buff	Proportion of land owned by the Bureau of Indian Affairs within a 1000m radius of the site
blm_2km_buff	Proportion of land owned by the Bureau of Land Management within a 1000m radius of the site
bpa_2km_buff	Proportion of land owned by the Bonneville Power Administration within a 1000m radius of the site
coe_2km_buff	Proportion of land owned by the Corps of Engineers within a 1000m radius of the site

fws_2km_buff	Proportion of land owned by the U.S. Fish and Wildlife Service within a 1000m radius of the site
gsa_2km_buff	Proportion of land owned by the General Services Administration within a 1000m radius of the site
lg_2km_buff	Proportion of land owned by the Local Government within a 1000m radius of the site
nps_2km_buff	Proportion of land owned by the National Park Service within a 1000m radius of the site
pv_2km_buff	Proportion of land owned by a Private Individual or Company within a 1000m radius of the site
pvi_2km_buff	Proportion of Lands Managed by Private Industry within a 1000m radius of the site
pvn_2km_buff	Proportion of land owned by Private Non-Industrial Owner within a 1000m radius of the site
pvu_2km_buff	Proportion of land owned by Private Urban Lands within a 1000m radius of the site
st_2km_buff	Proportion of land owned by a State Agency within a 1000m radius of the site
stf_2km_buff	Proportion of land owned by the State Dept. of Forestry within a 1000m radius of the site
stl_2km_buff	Proportion of land owned by the Division of State Lands within a 1000m radius of the site
stp_2km_buff	Proportion of land owned by the State Dept. of Parks and Recreation within a 1000m radius of the site
stw_2km_buff	Proportion of land owned by the State Dept. of Fish and Wildlife within a 1000m radius of the site
und_2km_buff	Proportion of land where ownership is Undetermined within a 1000m radius of the site
usfs_2km_buff	Proportion of land owned by the U.S. Forest Service within a 1000m radius of the site

water_2km_buff	Proportion of land that is Water within a 1000m radius of the site	
bia_1km_buff	Proportion of land owned by the Bureau of Indian Affairs within a 500m radius of the site	
blm_1km_buff	Proportion of land owned by the Bureau of Land Management within a 500m radius of the site	
bpa_1km_buff	Proportion of land owned by the Bonneville Power Administration within a 500m radius of the site	
coe_1km_buff	Proportion of land owned by the Corps of Engineers within a 500m radius of the site	
fws_1km_buff	Proportion of land owned by the U.S. Fish and Wildlife Service within a 500m radius of the site	
gsa_1km_buff	Proportion of land owned by the General Services Administration within a 500m radius of the site	
lg_1km_buff	Proportion of land owned by the Local Government within a 500m radius of the site	
nps_1km_buff	Proportion of land owned by the National Park Service within a 500m radius of the site	
pv_1km_buff	Proportion of land owned by a Private Individual or Company within a 500m radius of the site	X
pvi_1km_buff	Proportion of Lands Managed by Private Industry within a 500m radius of the site	X
pvn_1km_buff	Proportion of land owned by Private Non-Industrial Owner within a 500m radius of the site	X
pvu_1km_buff	Proportion of land owned by Private Urban Lands within a 500m radius of the site	
st_1km_buff	Proportion of land owned by a State Agency within a 500m radius of the site	
stf_1km_buff	Proportion of land owned by the State Dept. of Forestry within a 500m radius of the site	

	stl_1km_buff	Proportion of land owned by the Division of State Lands within a 500m radius of the site	
	stp_1km_buff	Proportion of land owned by the State Dept. of Parks and Recreation within a 500m radius of the site	
	stw_1km_buff	Proportion of land owned by the State Dept. of Fish and Wildlife within a 500m radius of the site	
	und_1km_buff	Proportion of land where ownership is Undetermined within a 500m radius of the site	
	usfs_1km_buff	Proportion of land owned by the U.S. Forest Service within a 500m radius of the site	
	water_1km_buff	Proportion of land that is Water within a 500m radius of the site	
	pvall_5km_buff	Sum of all private land classes, 2500m radius buffer	
	stall_5km_buff	Sum of all state-managed land classes, 2500m radius buffer	
	fedother_5km_buff	Sum of smaller federally managed land classes, 2500m radius buffer	
	pvall_2km_buff	Sum of all private land classes, 1000m radius buffer	
	stall_2km_buff	Sum of all state-managed land classes, 1000m radius buffer	
	fedother_2km_buff	Sum of smaller federally managed land classes, 1000m radius buffer	
	pvall_1km_buff	Sum of all private land classes, 500m radius buffer	
	stall_1km_buff	Sum of all state-managed land classes, 500m radius buffer	
	fedother_1km_buff	Sum of smaller federally managed land classes, 500m radius buffer	
Economic	emp_agforest	Number of employees in agriculture and forestry sector in county (Census County Business Patterns data)	X
	emp_const	Number of employees in construction sector in county (Census County Business Patterns data)	X
	emp_manuf	Number of employees in manufacturing sector in county (Census County Business Patterns data)	

emp_wholesale	Number of employees in wholesale sector in county (Census County Business Patterns data)	
emp_retail	Number of employees in retail sector in county (Census County Business Patterns data)	
emp_transport	Number of employees in transport sector in county (Census County Business Patterns data)	
emp_info	Number of employees in information sector in county (Census County Business Patterns data)	
emp_finance	Number of employees in finance sector in county (Census County Business Patterns data)	
emp_realestate	Number of employees in real estate sector in county (Census County Business Patterns data)	
emp_profsci	Number of employees in professional and science sector in county (Census County Business Patterns data)	
emp_admin	Number of employees in administrative sector in county (Census County Business Patterns data)	
emp_health	Number of employees in health sector in county (Census County Business Patterns data)	
emp_arts	Number of employees in arts sector in county (Census County Business Patterns data)	
emp_food	Number of employees in food service sector in county (Census County Business Patterns data)	
emp_other	Number of employees in other sectors in county (Census County Business Patterns data)	
brick_coun	Kernel density of "Brick, Stone and Related" category based on sites counts	
brick_totp	Kernel density of "Brick, Stone and Related" category based on total persons/site	X
const_coun	Kernel density of "Construction and Mining" category based on sites counts	

	const_totp	Kernel density of "Construction and Mining" category based on total persons/site	X
	merch_coun	Kernel density of all 3 categories (brick…, construction…, metals…) based on sites counts	
	merch_totp	Kernel density of all 3 categories (brick…, construction…, metals…) based on total persons/site	
	metal_coun	Kernel density of "Metals Service Centers" category based on sites counts	
	metal_totp	Kernel density of "Metals Service Centers" category based on total persons/site	
	sales_coun	Kernel density of "Sand and Gravel Sales Yard" category based on sites counts	X
	sand_count	Kernel density of "Sand and Gravel Operations" based on sites counts	
Project	project_source	Organization reporting project to PNSHP (categorical)	X
	project_year	Year work on project began	X
	n_worksites	Number of worksites associated with project	X
	n_culverts	Number of culverts associated with worksite	
	action_fishpass_culvimp_prj	Indicator if project involves any culvert improvement actions	X
	action_fishpass_culvinst_prj	Indicator if project involves any culvert installation actions	
	action_fishpass_culvrem_prj	Indicator if project involves any culvert removal actions	X
	action_fishpass_culvimp_count_prj	Count of culvert improvement actions across project	
	action_fishpass_culvinst_count_prj	Count of culvert installation actions across project	
	action_fishpass_culvrem_count_prj	Count of culvert removal actions across project	
	action_fishpass_culvimp_wrk	Indicator if worksite involves any culvert improvement actions	
	action_fishpass_culvinst_wrk	Indicator if worksite involves any culvert installation actions	
	action_fishpass_culvrem_wrk	Indicator if worksite involves any culvert removal actions	

action_fishpass_culvimp_count_wrk	Count of culvert improvement actions at worksite	
action_fishpass_culvinst_count_wrk	Count of culvert installation actions at worksite	
action_fishpass_culvrem_count_wrk	Count of culvert removal actions at worksite	
dist_max	Maximum distance between worksites across project (m)	
dist_mean	Mean distance between worksites across project (m)	
tot_dist	Total distance between worksites across project (m)	X

Table S2. Variance inflation factors (VIFs)

Variable	VIF
Bankfull width	1.31
Channel slope	2.27
Terrain slope	2.28
Elevation	4.53
Private land, individual or company (% 500m buffer)	2.64
Private land, managed by industry (% 500m buffer)	2.49
Private land, managed by non-industrial owner (% 500m buffer)	1.85
Housing density	1.44
Distance to urban area	2.92
Construction employment	2.41
Ag/forestry employment	2.31
Density of sand and gravel sales yards	3.71
Density (employee-weighted) of construction equipment suppliers	5.87
Density (employee-weighted) of brick, concrete, and related materials suppliers	2.11
Number of worksites	3.27
Distance between worksites	3.12
Culvert removal (dummy)	1.38
Culvert installation (dummy)	1.58
Road speed class: 4	2.13
Road speed class: 5	9.93
Road speed class: 6	9.24
Road speed class: 7	17.8
Road speed class: 8	12.4
Road paved (dummy)	2.4
Land cover: Forest	1.83
Land cover: Herbaceous	1.15
Land cover: Planted-cultivated	1.23
Land cover: Shrubland	1.25
Land cover: Wetlands	1.29
Basin: JOHN DAY	2.79
Basin: LOWER COLUMBIA	1.87
Basin: MIDDLE COLUMBIA	1.39
Basin: NORTHERN OREGON COASTAL	2.83
Basin: PUGET SOUND	4.93
Basin: UPPER COLUMBIA	2.48
Basin: WASHINGTON COASTAL	3.17
Basin: WILLAMETTE	4.15
Project source: BLM	1.66
Project source: HABITAT WORK SCHEDULE	2.86
Project source: REO	2.35
Project source: SRFBD	1.61

Project source: WA RCO	5.16
Year: 2002	2.22
Year: 2003	2.24
Year: 2004	2.5
Year: 2005	2.14
Year: 2006	1.94
Year: 2007	1.85
Year: 2008	1.5
Year: 2009	1.46
Year: 2010	1.62
Year: 2011	1.54
Year: 2012	1.37
Year: 2013	1.29
Year: 2014	1.18
Year: 2015	1.19

Table S3. Linear regression estimates with alternative fixed effects structures.

Term	full	nofe	nobasin	nosource	noyear	onlybasin	onlysource	onlyyear
Intercept	9.72***	10.8***	9.87***	10.1***	9.98***	10.6***	10.1***	10.3***
	(0.31)	(0.264)	(0.303)	(0.313)	(0.295)	(0.293)	(0.282)	(0.302)
Channel slope	-3.61*	-4.56**	-2.69	-4.53**	-3.47*	-4.29**	-2.44	-4.86**
	(2.03)	(2.13)	(1.99)	(2.1)	(2.06)	(2.13)	(2.03)	(2.06)
Bankfull width	-0.00314	-0.00128	-0.00112	-0.00435	-0.00132	-0.00316	0.00123	-0.00338
	(0.00689)	(0.00773)	(0.00672)	(0.00703)	(0.00705)	(0.00716)	(0.00704)	(0.00734)
Channel slope X	0.968**	1.31***	0.958**	1.19**	0.896*	1.13**	0.841*	1.38***
bankfull width	(0.458)	(0.461)	(0.45)	(0.464)	(0.458)	(0.468)	(0.45)	(0.459)
Road paved	0.198	0.282**	0.183	0.223*	0.266**	0.303**	0.238*	0.226*
(dummy)	(0.12)	(0.126)	(0.127)	(0.119)	(0.12)	(0.119)	(0.126)	(0.126)
Road speed class: 3	0.601*	0.51	0.605*	0.714**	0.532	0.605*	0.555	0.607*
	(0.346)	(0.33)	(0.344)	(0.345)	(0.339)	(0.342)	(0.34)	(0.33)
Road speed class: 4	0.433	0.53*	0.437	0.439	0.359	0.401	0.383	0.503
	(0.301)	(0.313)	(0.289)	(0.335)	(0.304)	(0.337)	(0.287)	(0.329)
Road speed class: 5	0.403***	0.374***	0.418***	0.451***	0.362**	0.416***	0.388***	0.402***
	(0.141)	(0.143)	(0.147)	(0.137)	(0.14)	(0.136)	(0.146)	(0.145)
Road speed class: 6	0.225	0.418***	0.264*	0.239*	0.176	0.192	0.226	0.395***
	(0.142)	(0.148)	(0.144)	(0.139)	(0.143)	(0.14)	(0.145)	(0.145)
Road speed class: 7	0.287***	0.392***	0.307***	0.318***	0.304***	0.348***	0.325***	0.365***
	(0.102)	(0.106)	(0.102)	(0.103)	(0.102)	(0.103)	(0.102)	(0.105)
Terrain slope	0.00775*	0.00524	0.00574	0.00896**	0.00875**	0.0105**	0.00583	0.00597
	(0.00403)	(0.00387)	(0.00377)	(0.00417)	(0.00394)	(0.00415)	(0.00377)	(0.00381)
Elevation	-0.000155	-0.0000567	-0.00027	-0.00000763	-0.000226	-0.0000723	-0.000252	-0.0000977
	(0.000299)	(0.000193)	(0.000198)	(0.000295)	(0.000292)	(0.000286)	(0.000194)	(0.000195)
Land cover:	0.306***	0.361***	0.26***	0.34***	0.313***	0.394***	0.268***	0.296***
Developed	(0.0823)	(0.0906)	(0.081)	(0.0859)	(0.0858)	(0.0913)	(0.0845)	(0.0857)
Land cover:	-0.0585	-0.111	-0.0809	-0.0372	-0.0864	-0.0427	-0.11	-0.118
Herbaceous	(0.155)	(0.184)	(0.158)	(0.159)	(0.157)	(0.165)	(0.159)	(0.176)
Land cover:	0.376**	0.327*	0.37**	0.355**	0.335*	0.32*	0.339**	0.34**
Planted-cultivated	(0.162)	(0.172)	(0.157)	(0.165)	(0.172)	(0.172)	(0.168)	(0.169)
Land cover:	0.215	0.26*	0.174	0.242	0.208	0.245*	0.17	0.238
Shrubland	(0.142)	(0.148)	(0.142)	(0.147)	(0.139)	(0.145)	(0.14)	(0.147)
Land cover:	0.0634	0.119	0.0359	0.0659	0.0479	0.103	0.0284	0.036
Wetlands	(0.157)	(0.162)	(0.165)	(0.157)	(0.147)	(0.148)	(0.159)	(0.171)

Housing density	0.0000902 (0.00193)	0.00198 (0.00184)	0.000292 (0.00188)	0.000428 (0.00198)	0.000127 (0.00185)	0.000668 (0.00195)	0.000222 (0.00178)	0.0018 (0.00192)
Ag/forestry employment	0.0000825 (0.0000955)	0.0000408 (0.0000879)	0.0000397 (0.000085)	0.00017* (0.0000956)	0.0000984 (0.0000932)	0.000182* (0.0000951)	0.0000608 (0.0000849)	0.0000634 (0.0000878)
Construction employment	-0.00000685 (0.00000862)	0.00001 (0.00000837)	0.00000288 (0.0000076)	-0.0000168* (0.00000949)	-0.00000306 (0.0000086)	-0.0000128 (0.00000884)	0.00000549 (0.0000082)	0.00000496 (0.00000834)
Distance to urban area	0.00000204 (0.00000182)	-0.000000137 (0.00000151)	0.00000143 (0.00000143)	-0.000000413 (0.0000018)	0.00000143 (0.0000018)	-0.000000885 (0.00000182)	0.00000117 (0.00000144)	0.0000000773 (0.00000146)
Density (employee-weighted) of construction equipment suppliers	2.23 (2.56)	-0.22 (2.35)	2.15 (2.29)	2.56 (2.66)	2.1 (2.45)	1.84 (2.59)	2.5 (2.26)	0.219 (2.42)
Density (employee-weighted) of brick, concrete, and related materials suppliers	-4.91 (6.02)	6.61 (8.07)	-1 (6.01)	-6.88 (5.01)	-5.79 (4.47)	-7.59* (4.26)	-2.42 (4.83)	7.16 (8.45)
Density of sand and gravel sales yards	-1510 (1830)	-2000 (1560)	-1270 (1570)	-1320 (1910)	-1550 (1830)	-1150 (1890)	-1470 (1590)	-1640 (1610)
Private land, individual or company (% 500m buffer)	-0.0352 (0.119)	-0.253** (0.121)	0.00853 (0.119)	-0.356*** (0.113)	0.00378 (0.12)	-0.371*** (0.115)	0.0473 (0.12)	-0.232** (0.118)
Private land, managed by industry (% 500m buffer)	-0.499*** (0.158)	-0.979*** (0.147)	-0.466*** (0.149)	-0.74*** (0.158)	-0.454*** (0.16)	-0.739*** (0.162)	-0.418*** (0.15)	-0.955*** (0.143)
Private land, managed by non-industrial owner (% 500m buffer)	0.511 (0.316)	-0.186 (0.312)	0.509 (0.32)	0.306 (0.32)	0.383 (0.308)	0.116 (0.312)	0.384 (0.311)	0.0682 (0.325)
Number of worksites	-0.421*** (0.114)	-0.539*** (0.12)	-0.476*** (0.12)	-0.424*** (0.111)	-0.444*** (0.117)	-0.446*** (0.114)	-0.507*** (0.122)	-0.502*** (0.119)
Distance between worksites	-0.00002 (0.000124)	-0.0000512 (0.000127)	-0.0000317 (0.000139)	-0.0000275 (0.000124)	-0.0000236 (0.000106)	-0.0000421 (0.000106)	-0.0000388 (0.000124)	-0.0000282 (0.000145)

Number of worksites X distance	0.0000449 (0.0000299)	0.000065** (0.0000306)	0.0000559 (0.0000354)	0.000047 (0.0000287)	0.0000481* (0.000028)	0.0000511* (0.0000267)	0.0000606* (0.0000348)	0.0000586* (0.0000326)
Culvert installation (dummy)	0.181 (0.149)	0.794*** (0.135)	0.164 (0.148)	0.469*** (0.14)	0.393*** (0.144)	0.672*** (0.136)	0.399*** (0.141)	0.572*** (0.146)
Culvert removal (dummy)	0.0298 (0.118)	-0.0495 (0.117)	0.0471 (0.121)	-0.0732 (0.113)	0.128 (0.115)	0.00362 (0.112)	0.155 (0.117)	-0.123 (0.117)
Project source: BLM	0.885*** (0.144)	–	0.89*** (0.133)	–	0.94*** (0.15)	–	0.939*** (0.14)	–
Project source: HABITAT WORK SCHEDULE	1.1*** (0.326)	–	1.49*** (0.267)	–	0.991*** (0.307)	–	1.5*** (0.259)	–
Project source: REO	0.891*** (0.116)	–	0.919*** (0.11)	–	0.927*** (0.113)	–	0.959*** (0.108)	–
Project source: SRFBD	1*** (0.245)	–	1.25*** (0.221)	–	0.807*** (0.222)	–	1.12*** (0.2)	–
Project source: WA RCO	0.448* (0.234)	–	0.829*** (0.169)	–	0.327 (0.204)	–	0.818*** (0.147)	–
Year: 2002	0.215 (0.15)	–	0.251* (0.149)	0.39*** (0.15)	–	–	–	0.419*** (0.15)
Year: 2003	0.24 (0.158)	–	0.274* (0.159)	0.487*** (0.156)	–	–	–	0.564*** (0.157)
Year: 2004	0.653*** (0.145)	–	0.708*** (0.147)	0.804*** (0.145)	–	–	–	0.887*** (0.151)
Year: 2005	0.314* (0.161)	–	0.335** (0.164)	0.451*** (0.164)	–	–	–	0.514*** (0.17)
Year: 2006	0.391** (0.186)	–	0.409** (0.191)	0.524*** (0.181)	–	–	–	0.564*** (0.187)
Year: 2007	0.543*** (0.17)	–	0.613*** (0.168)	0.73*** (0.169)	–	–	–	0.875*** (0.171)
Year: 2008	0.00225 (0.291)	–	-0.0158 (0.284)	0.48* (0.282)	–	–	–	0.581** (0.288)
Year: 2009	0.449 (0.284)	–	0.443 (0.281)	0.962*** (0.299)	–	–	–	1.06*** (0.275)
Year: 2010	0.151 (0.248)	–	0.18 (0.241)	0.481* (0.252)	–	–	–	0.743*** (0.235)

Year: 2011	0.152 (0.264)	–	0.172 (0.274)	0.589** (0.284)	–	–	–	0.868*** (0.313)
Year: 2012	-0.0957 (0.45)	–	-0.0431 (0.436)	0.18 (0.474)	–	–	–	0.398 (0.445)
Year: 2013	0.0514 (0.293)	–	0.0824 (0.296)	0.143 (0.296)	–	–	–	0.132 (0.31)
Year: 2014	0.172 (0.291)	–	0.234 (0.274)	0.226 (0.289)	–	–	–	0.261 (0.281)
Year: 2015	0.835** (0.33)	–	0.903** (0.36)	1.43*** (0.316)	–	–	–	1.6*** (0.38)
Basin: JOHN DAY	-0.0904 (0.421)	–	–	0.0558 (0.411)	0.0962 (0.414)	0.302 (0.401)	–	–
Basin: LOWER COLUMBIA	0.636*** (0.238)	–	–	0.403* (0.228)	0.731*** (0.236)	0.506** (0.223)	–	–
Basin: MIDDLE COLUMBIA	0.251 (0.304)	–	–	0.206 (0.256)	0.394 (0.297)	0.356 (0.263)	–	–
Basin: NORTHERN OREGON COASTAL	-0.0643 (0.171)	–	–	-0.353** (0.167)	-0.0899 (0.174)	-0.369** (0.172)	–	–
Basin: PUGET SOUND	0.802** (0.312)	–	–	1.13*** (0.303)	0.89*** (0.29)	1.22*** (0.305)	–	–
Basin: UPPER COLUMBIA	0.119 (0.309)	–	–	0.242 (0.282)	0.318 (0.301)	0.415 (0.275)	–	–
Basin: WASHINGTON COASTAL	0.504** (0.239)	–	–	0.591*** (0.196)	0.637*** (0.234)	0.735*** (0.184)	–	–
Basin: WILLAMETTE	0.17 (0.184)	–	–	-0.155 (0.171)	0.198 (0.181)	-0.102 (0.172)	–	–
Adj. R2	0.404	0.283	0.394	0.355	0.386	0.327	0.372	0.323
AIC	3488.7	3690.8	3499.8	3580.3	3512.2	3619.6	3530.6	3633.4
BIC	3795.9	3859.7	3766	3861.9	3747.7	3829.5	3725.2	3874
N	1236	1236	1236	1236	1236	1236	1236	1236

Table S4. Linear regression estimates with alternative samples.

Term	full	basins_core	sources_core	project_level
Intercept	9.72***	10.2***	10.2***	11***
	(0.31)	(0.31)	(0.358)	(0.313)
Channel slope	-3.61*	-7.04***	-2.45	-3.56*
	(2.03)	(2.51)	(3.35)	(2.01)
Bankfull width	-0.00314	-0.00509	0.00194	-0.000173
	(0.00689)	(0.00779)	(0.00883)	(0.00692)
Channel slope X bankfull	0.968**	1.88***	1.09	1.05**
width	(0.458)	(0.593)	(0.756)	(0.418)
Road paved (dummy)	0.198	0.163	0.242	0.243**
	(0.12)	(0.127)	(0.168)	(0.123)
Road speed class: 3	0.601*	0.767**	0.603	0.413
	(0.346)	(0.36)	(0.481)	(0.359)
Road speed class: 4	0.433	0.576	0.389	0.175
	(0.301)	(0.368)	(0.392)	(0.328)
Road speed class: 5	0.403***	0.442***	0.258	0.111
	(0.141)	(0.147)	(0.181)	(0.126)
Road speed class: 6	0.225	0.384**	0.405**	-0.00976
	(0.142)	(0.17)	(0.191)	(0.128)
Road speed class: 7	0.287***	0.328***	0.262**	-0.288***
	(0.102)	(0.116)	(0.128)	(0.103)
Terrain slope	0.00775*	0.00862*	-0.000361	0.00919**
	(0.00403)	(0.0045)	(0.00543)	(0.00393)
Elevation	-0.000155	-0.000271	-0.000129	-0.0001
	(0.000299)	(0.00035)	(0.000268)	(0.000248)
Land cover: Developed	0.306***	0.292***	0.314***	0.302***
	(0.0823)	(0.0971)	(0.111)	(0.0827)
Land cover: Herbaceous	-0.0585	-0.18	-0.061	0.029
	(0.155)	(0.195)	(0.226)	(0.158)
Land cover: Planted-	0.376**	0.334*	0.327*	0.463***
cultivated	(0.162)	(0.174)	(0.191)	(0.151)
Land cover: Shrubland	0.215	0.231	0.144	0.199
	(0.142)	(0.186)	(0.216)	(0.143)
Land cover: Wetlands	0.0634	-0.0587	0.115	0.0664
	(0.157)	(0.185)	(0.189)	(0.155)
Housing density	0.0000902	0.00179	0.000732	0.000709
	(0.00193)	(0.00198)	(0.00395)	(0.00195)
Ag/forestry employment	0.0000825	0.0000796	0.000119	0.0000279
	(0.0000955)	(0.000101)	(0.000118)	(0.0000905)
Construction employment	-0.00000685	0.00000541	0.00000641	-0.00000578
	(0.00000862)	(0.00000885)	(0.0000134)	(0.00000915)

Distance to urban area	0.00000204 (0.00000182)	-0.000000359 (0.00000197)	-0.00000178 (0.00000218)	0.00000214 (0.0000017)
Density (employee-weighted) of construction equipment suppliers	2.23 (2.56)	0.383 (2.86)	-0.0214 (3.61)	2.09 (2.35)
Density (employee-weighted) of brick, concrete, and related materials suppliers	-4.91 (6.02)	7.84 (9)	3.45 (7.53)	-4.02 (5.56)
Density of sand and gravel sales yards	-1510 (1830)	-2120 (1980)	-713 (2370)	-1020 (1570)
Private land, individual or company (% 500m buffer)	-0.0352 (0.119)	-0.161 (0.128)	0.00726 (0.184)	-0.00629 (0.115)
Private land, managed by industry (% 500m buffer)	-0.499*** (0.158)	-0.93*** (0.152)	-0.621*** (0.181)	-0.355** (0.152)
Private land, managed by non-industrial owner (% 500m buffer)	0.511 (0.316)	0.117 (0.343)	0.258 (0.402)	0.654** (0.259)
Number of worksites	-0.421*** (0.114)	-0.486*** (0.124)	-0.53*** (0.145)	-0.47*** (0.104)
Distance between worksites	-0.00002 (0.000124)	-0.0000171 (0.000144)	-0.000058 (0.000151)	-0.000057 (0.000124)
Number of worksites X distance	0.0000449 (0.0000299)	0.0000499 (0.0000367)	0.0000656* (0.0000375)	0.0000603** (0.0000303)
Culvert installation (dummy)	0.181 (0.149)	0.559*** (0.189)	0.177 (0.658)	0.188 (0.139)
Culvert removal (dummy)	0.0298 (0.118)	-0.216* (0.126)	0.117 (0.141)	0.0952 (0.108)
Project source: BLM	0.885*** (0.144)	–	–	0.897*** (0.135)
Project source: HABITAT WORK SCHEDULE	1.1*** (0.326)	–	–	0.114 (0.321)
Project source: REO	0.891*** (0.116)	–	–	0.0559 (0.126)
Project source: SRFBD	1*** (0.245)	–	–	0.0526 (0.265)
Project source: WA RCO	0.448* (0.234)	–	–	-0.363 (0.251)
Year: 2002	0.215 (0.15)	0.467*** (0.154)	0.357** (0.179)	0.156 (0.143)
Year: 2003	0.24 (0.158)	0.595*** (0.161)	0.476** (0.196)	0.0961 (0.132)
Year: 2004	0.653*** (0.145)	0.885*** (0.165)	0.635*** (0.182)	0.439*** (0.122)

Year: 2005	0.314*	0.543***	0.271	0.188
	(0.161)	(0.173)	(0.204)	(0.129)
Year: 2006	0.391**	0.602***	0.408*	0.184
	(0.186)	(0.193)	(0.235)	(0.155)
Year: 2007	0.543***	0.687***	0.703***	0.386***
	(0.17)	(0.181)	(0.21)	(0.15)
Year: 2008	0.00225	0.595*	0.655*	-0.177
	(0.291)	(0.339)	(0.379)	(0.281)
Year: 2009	0.449	1.07***	0.556	0.0623
	(0.284)	(0.323)	(0.441)	(0.234)
Year: 2010	0.151	0.732**	0.512	-0.0665
	(0.248)	(0.342)	(0.348)	(0.243)
Year: 2011	0.152	0.956**	0.424	-0.135
	(0.264)	(0.404)	(0.318)	(0.222)
Year: 2012	-0.0957	0.417	0.587	-0.0873
	(0.45)	(0.518)	(0.487)	(0.262)
Year: 2013	0.0514	0.187	-0.142	-0.105
	(0.293)	(0.317)	(0.358)	(0.285)
Year: 2014	0.172	0.382	0.189	0.0949
	(0.291)	(0.31)	(0.323)	(0.251)
Year: 2015	0.835**	1.65***	–	0.717**
	(0.33)	(0.395)		(0.32)
Basin: JOHN DAY	-0.0904	–	–	0.0245
	(0.421)			(0.358)
Basin: LOWER COLUMBIA	0.636*** (0.238)	–	–	0.463** (0.217)
Basin: MIDDLE COLUMBIA	0.251 (0.304)	–	–	0.173 (0.313)
Basin: NORTHERN OREGON COASTAL	-0.0643 (0.171)	–	–	-0.064 (0.155)
Basin: PUGET SOUND	0.802**	–	–	0.675**
	(0.312)			(0.288)
Basin: UPPER COLUMBIA	0.119 (0.309)	–	–	0.05 (0.309)
Basin: WASHINGTON COASTAL	0.504** (0.239)	–	–	0.448* (0.234)
Basin: WILLAMETTE	0.17	–	–	0.123
	(0.184)			(0.168)
Adj. R2	0.404	0.337	0.272	0.361
AIC	3488.7	3222.6	2301	2905.5
BIC	3795.9	3457.4	2515.3	3201.7
N	1236	1091	779	1028

Table S5. Summary statistics and comparison of means between PNSHP and inventory data for selected variables.

Variable	Mean, PNSHP	Std. Dev., PNSHP	Mean, Inventories	Std. Dev., Inventories	t-value
cost_per_culvert	83,400	96,000	NA	NA	NA
bankfull_width	7.56	5.41	12.2	23.3	-25.8
slope	0.0458	0.0406	0.0361	0.0473	8.31
here_speed_3	0.017	0.129	0.218	0.413	-50.1
here_speed_4	0.017	0.129	0.0535	0.225	-9.69
here_speed_5	0.178	0.383	0.17	0.376	0.745
here_speed_6	0.152	0.359	0.267	0.442	-11.1
here_speed_7	0.409	0.492	0.153	0.36	18.2
here_speed_8	0.227	0.419	0.139	0.346	7.42
here_paved_N	0.606	0.489	0.388	0.487	15.6
here_paved_Y	0.394	0.489	0.612	0.487	-15.6
cat_basin_slope	27.3	12.1	18.8	13.6	24.6
cat_elev_mean	411	317	445	452	-3.76
nlcd_current_class_Developed	0.36	0.48	0.623	0.485	-19.2
nlcd_current_class_Forest	0.456	0.498	0.223	0.417	16.4
nlcd_current_class_Herbaceous	0.0299	0.17	0.0168	0.128	2.72
nlcd_current_class_Planted-cultivated	0.0372	0.189	0.0445	0.206	-1.34
nlcd_current_class_Shrubland	0.0591	0.236	0.0494	0.217	1.44
nlcd_current_class_Wetlands	0.0574	0.233	0.0425	0.202	2.25
pv_1km_buff	0.448	0.436	0.665	0.422	-17.4
pvi_1km_buff	0.203	0.356	0.039	0.164	16.2
pvn_1km_buff	0.0349	0.169	0.0332	0.17	0.347
hdens_cat	6.5	28.1	45	139	-39.6
ua_dist	45,200	31,900	64,800	65,000	-20.9
emp_const	3,010	5,140	6,520	13,500	-22.6
emp_agforest	777	534	582	515	12.8
sales_coun	0.0000244	0.00004	0.0000191	0.0000349	4.63
const_totp	0.0217	0.032	0.0274	0.0341	-6.22
brick_totp	0.0028	0.00725	0.00784	0.0168	-23.4
n_worksites	1.48	0.958	NA	NA	NA
tot_dist	506	1,360	NA	NA	NA
action_fishpass_culvrem_prj	0.199	0.399	NA	NA	NA
action_fishpass_culvinst_prj	0.072	0.259	NA	NA	NA

Figures

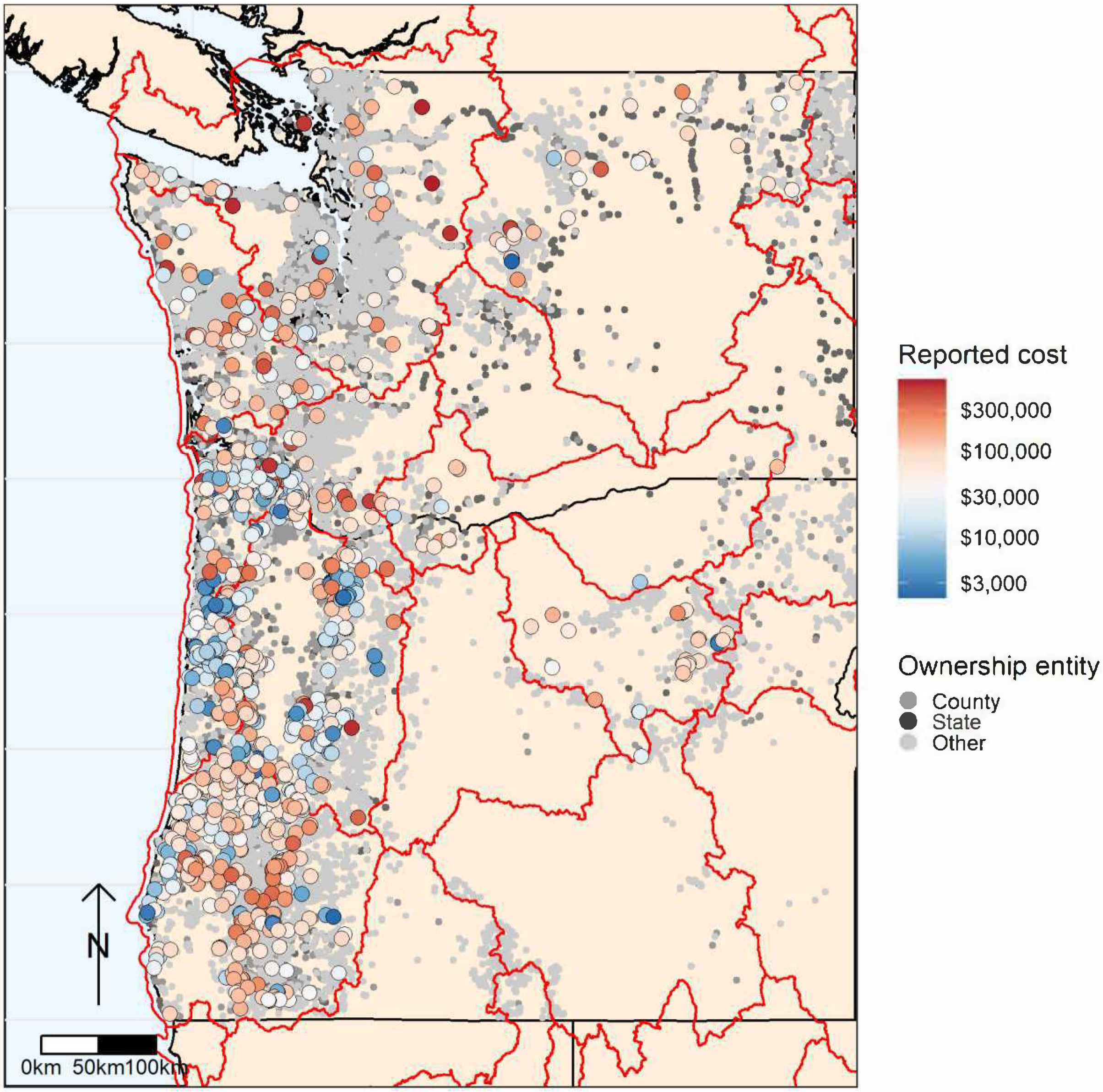

Figure 1. PNSHP culvert restoration worksites by reported cost (color points) and barriers in state inventories by ownership entity class (greyscale points). Costs are converted to 2019 dollars using the CPI. Other includes private landowner, federal, tribal, and city barrier owners. Red borders represent basins (HUC2).

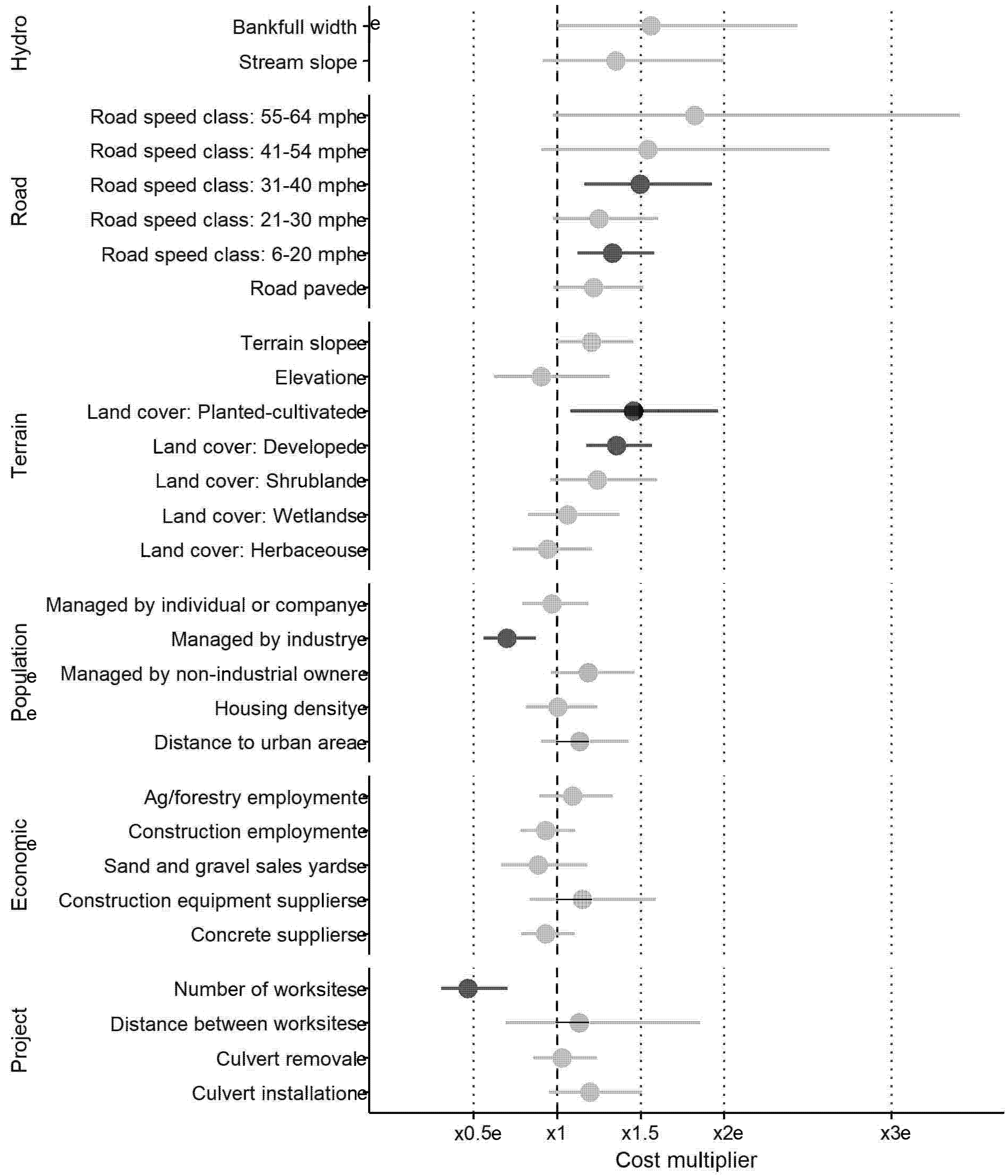

Figure 2. Standardized cost multipliers for hypothesized cost drivers. Multipliers are based on preferred specification of linear regression results ("full" column of Table S3). Horizontal lines represent 95% confidence intervals based on robust standard errors clustered at the project level. Dark indicates when confidence intervals do not include one (a null effect). Multipliers for variables with interactions are calculated with the opposite variable at its mean.

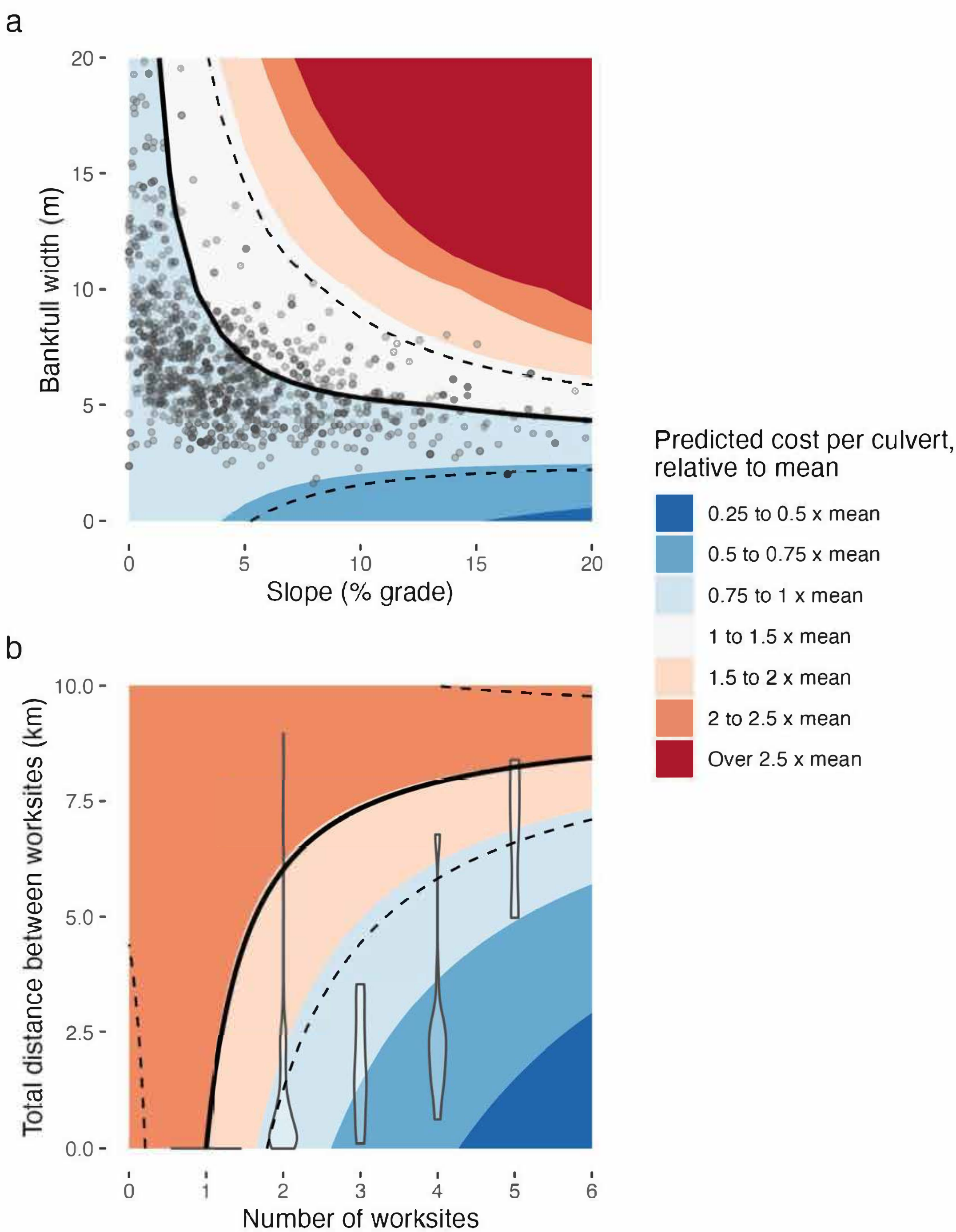

Figure 3. Cost contours for interaction effects of (a) channel slope and bankfull width and (b) number of worksites and total distance between worksites. Colors indicate ranges of predicted costs, based on linear regression, relative to the predicted cost when all explanatory variables are held at their means. Solid line indicates mean cost contour and dashed lines indicate 95% confidence intervals. Points in (a) and violin plots in (b) indicate distributions of underlying data in the sample.

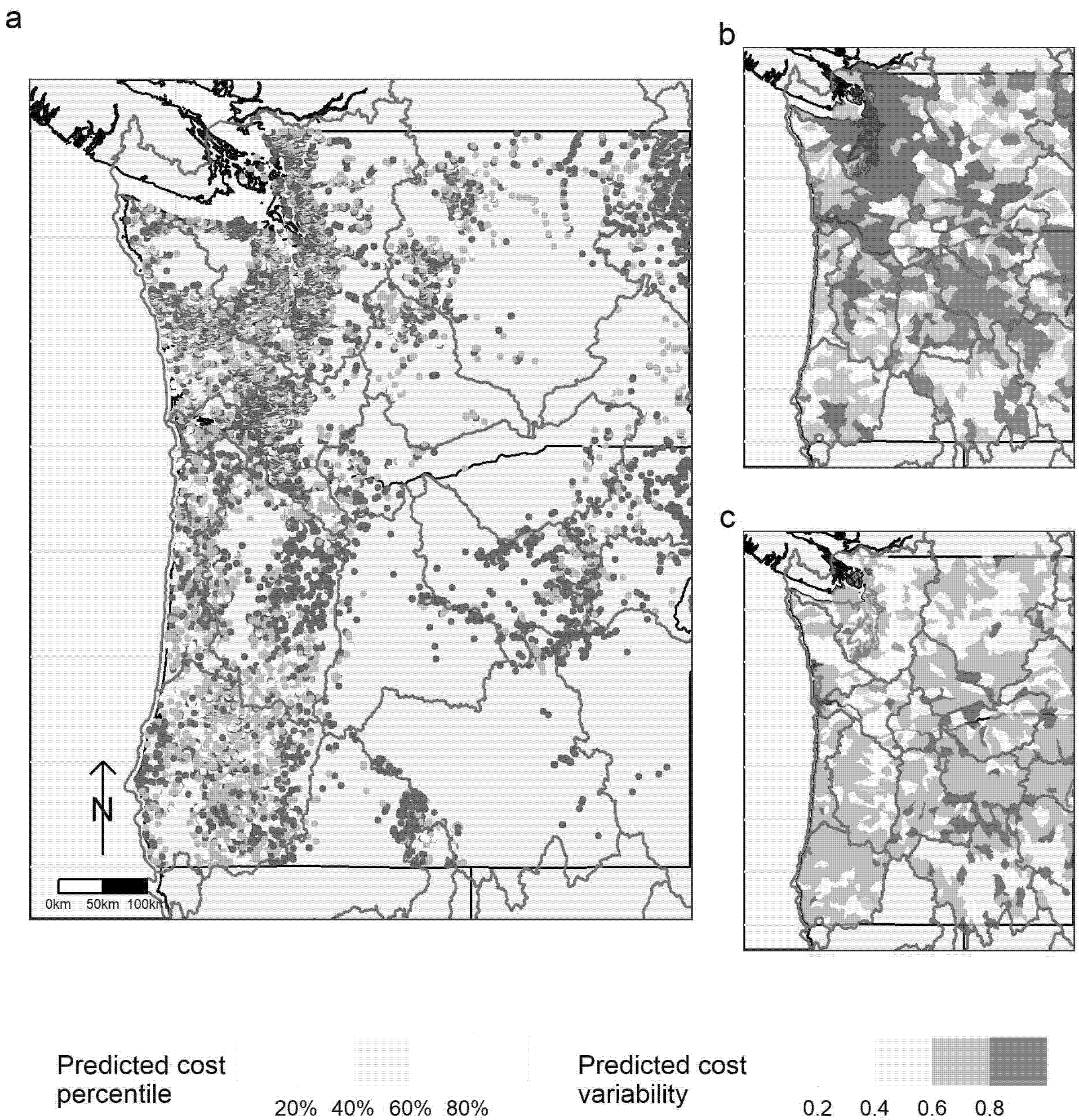

Figure 4. Predicted cost (a) levels for individual culverts, (b) mean levels at watershed-level (HUC10), and (c) variability (coefficient of variation) at watershed-level (HUC10). Red borders represent basins (HUC2). Barriers outside the area of applicability removed.

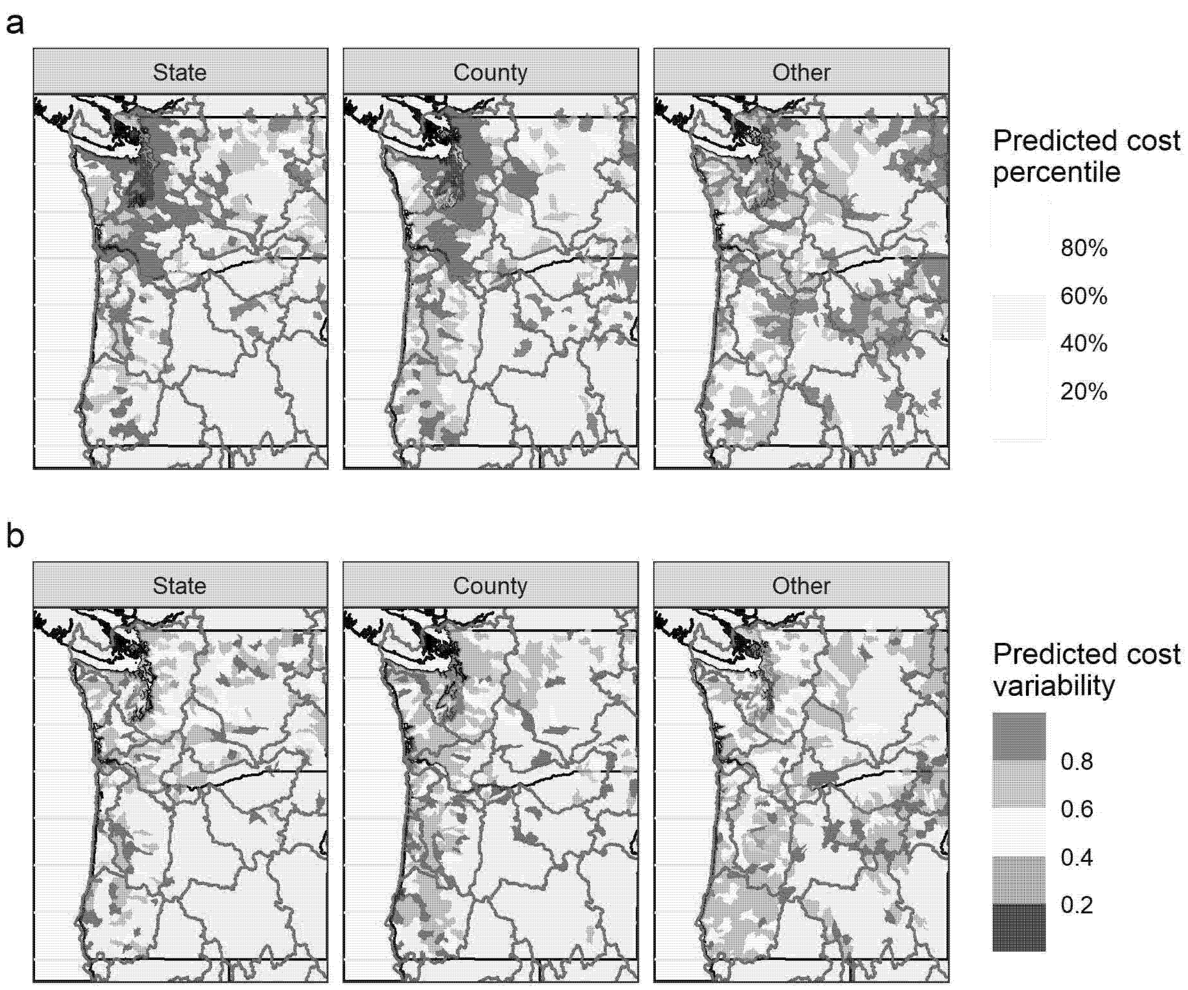

Figure 5. Predicted cost (a) mean levels and (b) variability (coefficient of variation) at the watershed-level (HUC10), separated by ownership entity class. Other includes private landowner, federal, tribal, and city barrier owners. Red borders represent basins (HUC2). Barriers outside the area of applicability removed.

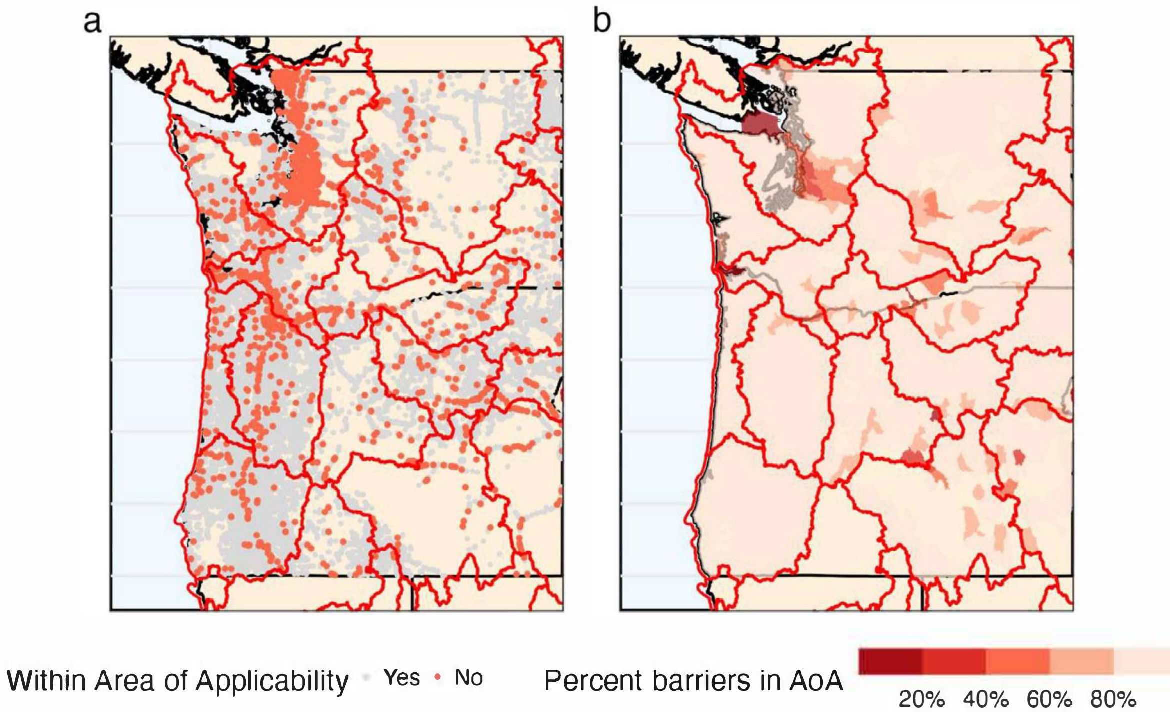

Figure 6. Area of applicability (AoA) for cost predictions presented at (a) individual barrier culvert and (b) watershed-level (HUC10). Red borders represent basins (HUC2).

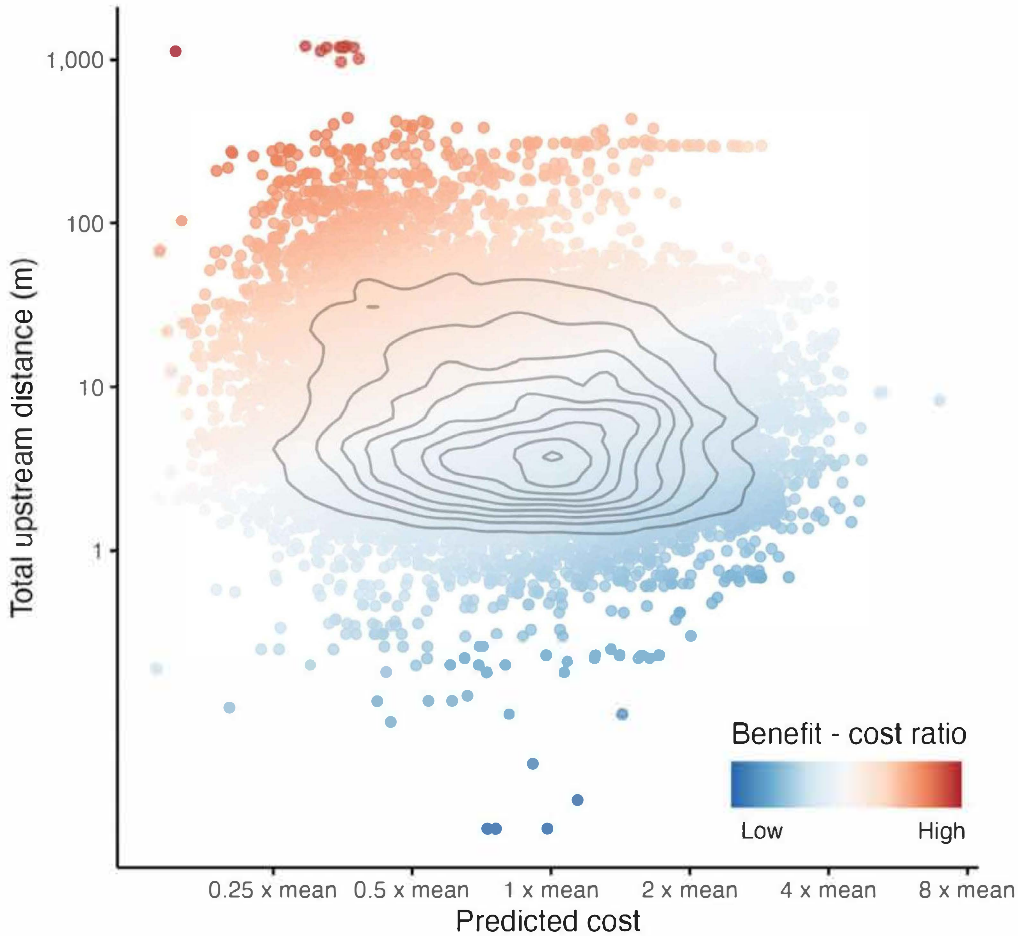

Figure 7. Benefit-cost ratios for inventory barriers. Benefits are defined as total upstream distance. Contours represent density of points. Both axes are on a log-scale. Barriers outside the area of applicability are omitted.

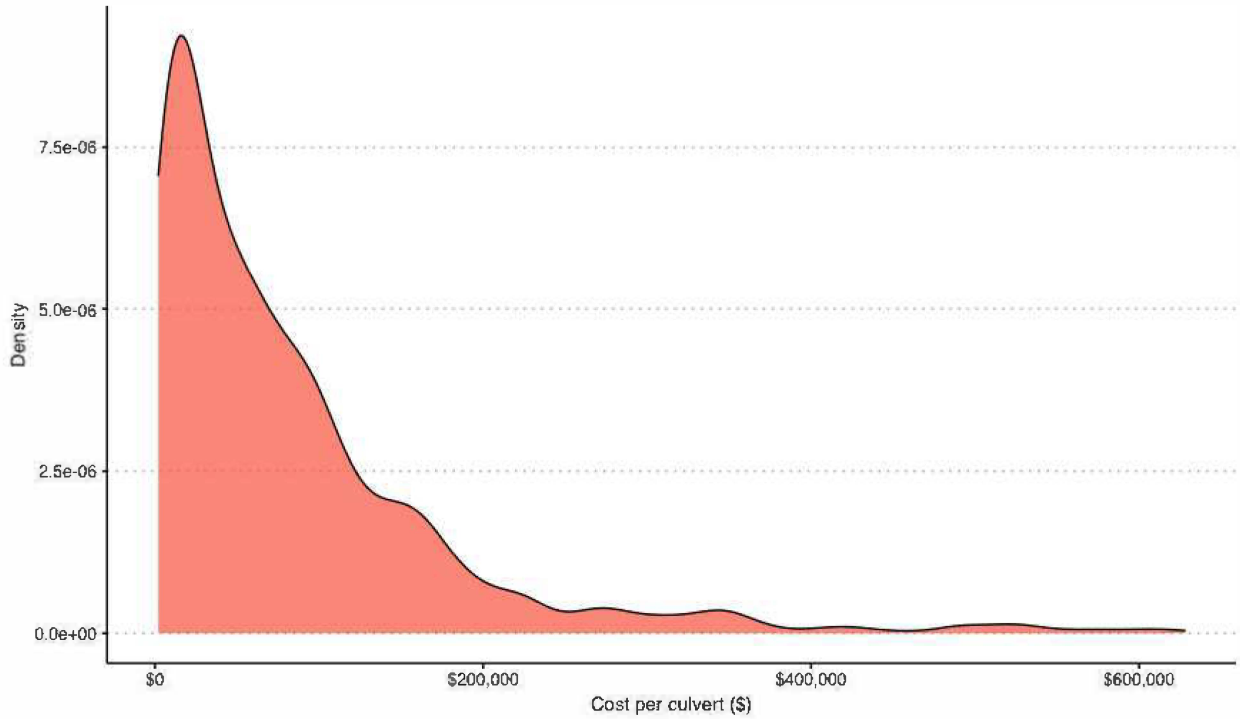

Figure S1. Distribution (kernel density) of reported costs in PNSHP data.

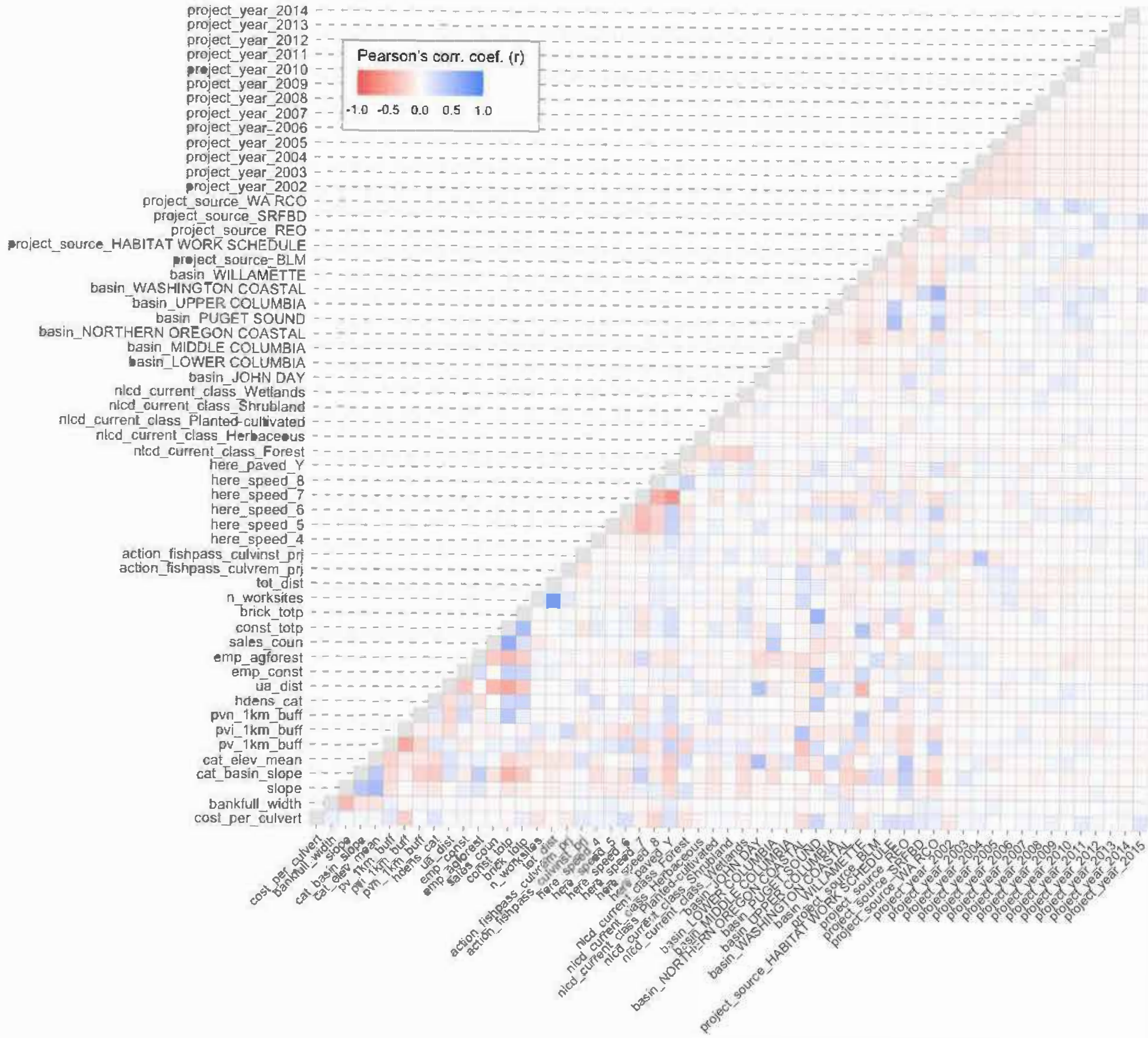

Figure S2. Correlations among covariates included in linear regression. Colors indicate Pearson's correlation coefficient.

Supplementary Material 1 – Feature matching

For point features, used for locations of supplier firms, we estimated weighted kernel densities over a 1km resolution grid with a 100km search radius and sq. km area units. This method was selected in order to better capture effects when worksites are located near multiple suppliers, which may lead to lower costs due to competition between firms. Densities were calculated based on number of firms, and by number of workers as an alternative measure that gives higher weight to larger firms. Worksites were assigned the density value associated with the $1km^2$ grid cell containing the worksite coordinates as a measure of access to suppliers.

For line features, i.e., hydrological and road variables, we matched worksites to the nearest feature (i.e., "snapping"). Euclidean distance to the nearest features was computed using ArcGIS for road line features and the R package *nhdplusTools* for hydrological features. We defined poor matches as worksites further and 150m from any feature. For poor road matches, we assumed the culvert was on an undocumented road and assigned the worksite variable values representing an undeveloped or primitive road (smallest road class, slowest speed class, and unpaved material class). All worksites with poor matches to hydrological features were dropped (25 worksites dropped).

Our indicator datasets feature two distinct types of variables stored in polygon features: county-level workforce data and proximity to urban areas/clusters. For the former we assign county-level variables to all worksites contained within county borders. For the latter we measure the distance to the nearest urban area or cluster borders, with zeros assigned for worksites located within area or cluster boundaries.

For raster features, we use two methods. For property jurisdiction, we calculate the share of land area controlled by each ownership class within 2.5km, 1km, and 0.5km radius buffers around each worksite. For land cover, we assign the land cover value of the raster cell covering the worksite from the NLCD vintage nearest in time to the year of project completion.

In addition to hydrological variables directly available from NHDPlus V2.1, we use the COMID of the nearest stream segment to link to additional variables defined at the catchment level (Wiezorek, Jackson, & Schwarz, 2018). These variables provide additional context on worksite conditions, including measures of bankfull width, elevation, average terrain slope, land cover, housing density, and more, falling into multiple variable categories.

We also include variables derived from PNSHP rather than through feature matching with external data sources. These variables include the number of and total distance between worksites associated with a worksite's project and the scope of culvert modifications (i.e. whether the culvert actions associated with a project include the "culvert removal", "culvert improvement", or "culvert installation" action types).

References

Wieczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2018, Select Attributes for NHDPlus Version 2.1 Reach Catchments and Modified Network Routed Upstream Watersheds for the Conterminous United States (ver. 3.0, January 2021): U.S. Geological Survey data release, https://doi.org/10.5066/F7765D7V.

Supplementary Material 2 – Machine Learning Methods and Prediction Performance

Machine learning methods considered

We compared three ML methods. First, we generated a regression tree, a which partitions the data into groups with similar values of the dependent variable (i.e., log average costs) based on discrete splits in the explanatory variables (i.e., land cover is developed or bankfull width is greater than three meters). This partitioning process continues until a stopping point is reached, based either on some penalization per additional "branch" or a set number of splits. Predictions are made by assigning the median value within each group to all members of that group. We fit trees with the *rpart* package for the statistical software R, fitting two trees: one with the default control parameters and one with the complexity cost parameter set to 0.025, a value associated with the lowest relative error following cross validation exercises.

Second, we considered a bagging technique known as a random forest (RF), a model-averaging procedure based on bootstrapping (bagging is short for "bootstrap aggregating"). This method works by repeatedly randomly selecting a subsample of the training set along with a random selection of the potential explanatory variables and fitting regression trees for each subsample-variables pair. The final prediction is the mean of predictions from each component tree. Our random forest fit is based on 1,000 trees, with 25 explanatory variables randomly selected for each tree and a minimum node size of 5. These parameters performed best when compared to fits with 50 and 100 explanatory variables, minimum node sizes of 1 and 3, and tree numbers of 5,000 and 10,000. Trees were fit with the *randomForest* package for R and all other parameters were left to package defaults.

Third and finally, we considered the boosted regression tree (BRT) method, which works through sequential training. A single regression tree is fit to the data, and then a second regression tree is fit to the residuals of the first tree. The predictions from each model are added together to form an ensemble model, and the residuals from that model are then subsequently fit for an additional tree, and so on and so forth until a stopping point defined as a global minimum in RMSE is reached as assessed with gradient descent. We fit these models with the package *gbm* for R, with a maximum of 1,000 trees, interaction depth of four, and with 5-fold cross validation. All other parameters were set to package defaults.

Prediction performance of machine learning models

We compared RMSE calculated using the testing set (the subsample withheld during fitting) to compare out-of-sample predictive power across methods. Compared to the linear regression baseline, both regression trees performed somewhat worse. On the other hand, both model aggregation methods (RF and BRT) performed significantly better, exhibiting 13% and 11% improvements in RMSE respectively (Figure S3).

We also plotted the predicted cost and actual cost for the testing set to examine how predictive accuracy varies over the range of the dependent variable. These plots show the limitations of smaller regression trees, where predictions are limited to only four values. The aggregating methods RF and BRT performed more similarly to linear regression, providing distinct predictions for each worksite. While RF had a lower RMSE, it tended to over-estimate costs for cheaper projects and under-estimate costs for more expensive ones, which may lead to underestimation of the underlying variability in costs across the landscape (Figure S4).

Finally, we plotted the density of prediction residuals from the testing set across all five methods. BRT exhibited a tighter peak near zero compared to even RF, suggesting that though RF provided a lower out-of-sample RMSE, BRT predictions more consistently hit the mark.

However, for BRT to display this sharp peak of near zero-residuals and while exhibiting a larger RMSE, it must mean that when predictions are inaccurate, it misses by a larger margin, indicated here by longer tails (Figure S5).

Because of the variance-dampening predictions exhibited by RF, we chose to proceed with predictions from BRT, which provided similar improvements in RMSE and smaller residuals over a larger portion of the data. We also checked for patterns of prediction accuracy over space by mapping out-of-sample RMSE from the BRT fit, calculated at the watershed level (Figure S6). We found that the error rate is consistently low across our study area, except for a patch of low accuracy on Jackson Creek near Medford, Oregon.

Variable importance and fit agreement

Many of the variables included in our linear regression appeared among the most important variables in our best performing ML fits (Figure S7). Project source and basin were among the top three most important variables for both RF and BRT, consistent with the adjusted R^2 improvements associated with fixed effects for these factors in our linear regressions. The land cover was the fourth and second most important variable for the RF and BRT fits respectively, which is also consistent with our linear regression findings. In both fits, road features (speed class, distance to road line) were among the important variables as well, consistent with our linear regression results, though speed class played a larger role in BRT than for RF.

Variables measuring the proportion of surround land managed by private industry, measured at all distances, were among the most important variables for the RF fit. Project scale effects like number of worksites, and maximum, mean, and total distance between worksites also played important roles in the RF fit, as did measures of employment by sector and supplier density. For the BRT fit, hydrological features played a relatively more important role, with bankfull width and slope both among the five most important variables.

Overall, the variables we select for the linear regression fit, or close analogs, are among the features that played the most important role in prediction. The BRT fit placed more importance on hydrological and road variables than the RF fit, which leaned more heavily on project and economic variables. The general agreement among both data-driven (i.e., ML) and investigator-guided (i.e., linear regression) methods on which factors drive costs provides strong support for appropriateness of these variables as cost metrics for culvert replacements.

Data and assumptions for out-of-sample predictions

From both inventories, we identified all culverts identified as potential fish passage barriers (i.e., recorded as at least partial or unknown fish passage barrier status) (Figure 1). For all inventory culvert barriers, we gathered indicator variable values using the same feature matching methods used for PNSHP worksites. For variables derived from PNSHP for our observed cost sample (e.g., number of worksites, action type, reporting source, and project year) we made standardizing assumptions to allow direct comparison between barriers. Remedying any barrier was assumed to require only a "culvert improvement" action rather than installation or removal. Land cover classes were based on the most recent available NLCD vintage (2016) and we considered all culverts independently, assigning values of one for the number of worksites variable and zero for all distance variables. Across both inventories, 27,450 potential barrier culverts were identified.

Some inventory culverts were located within basins not observed in the PNSHP sample (n = 2,105); we assigned these barriers to the "Southern Oregon Coastal" basin, the baseline level for the basin factor used in the linear regression estimation. Similarly, some inventory barriers were matched to HERE road segments with speed classifications above those observed in PNSHP (n = 910). We assigned these barriers to the next fastest observed speed class. As Southern Oregon Coastal barriers are among the lowest cost to improve, based on basin fixed effects linear regression parameters, and because higher speed limits are associated with higher cost culvert improvements, these steps mean that predictions for these barriers are conservative.

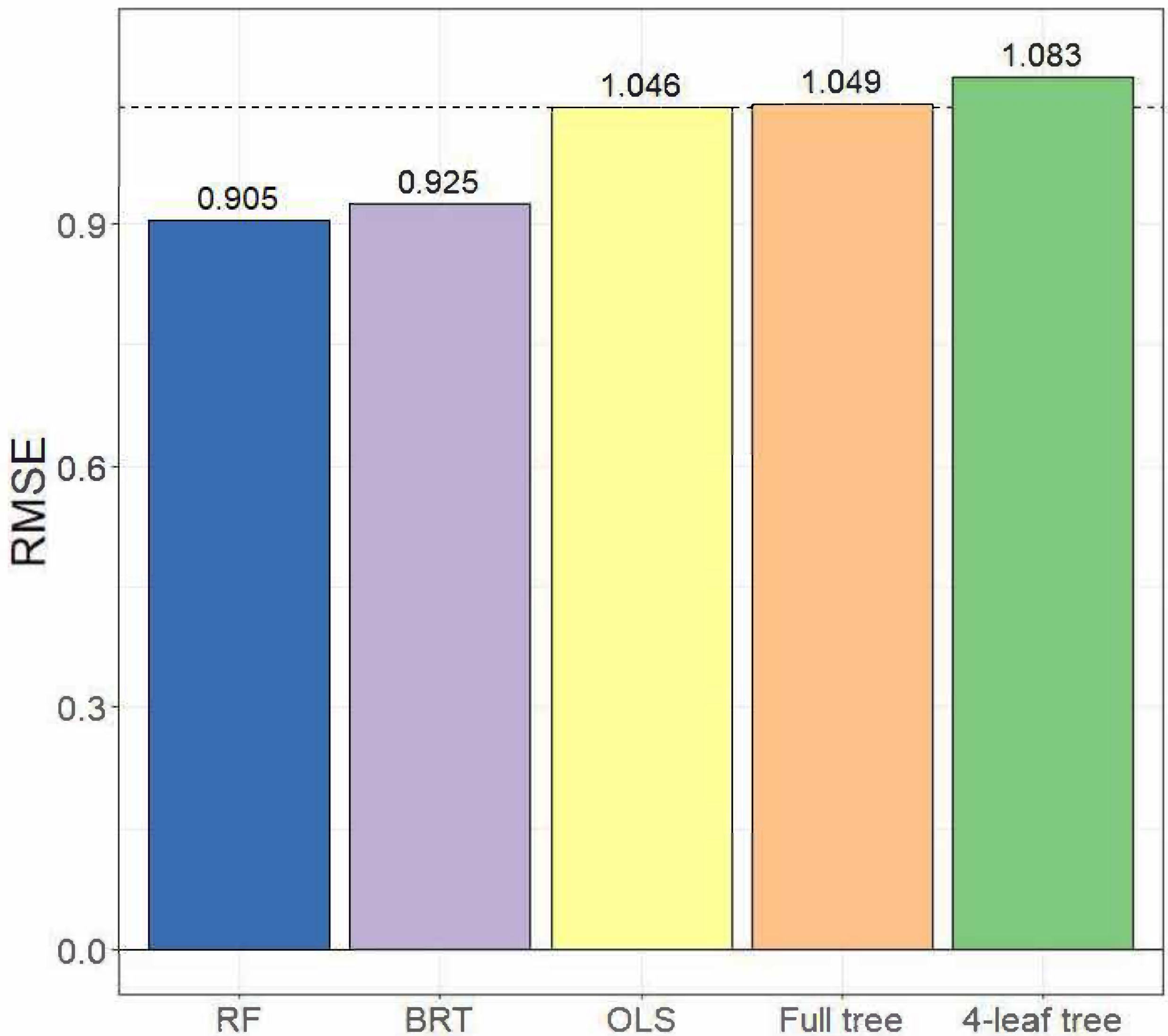

Figure S3. Root mean square error (RMSE) in testing set by machine learning method. RF = random forest, BRT = boosted regression tree, OLS = ordinary least squares (linear regression).

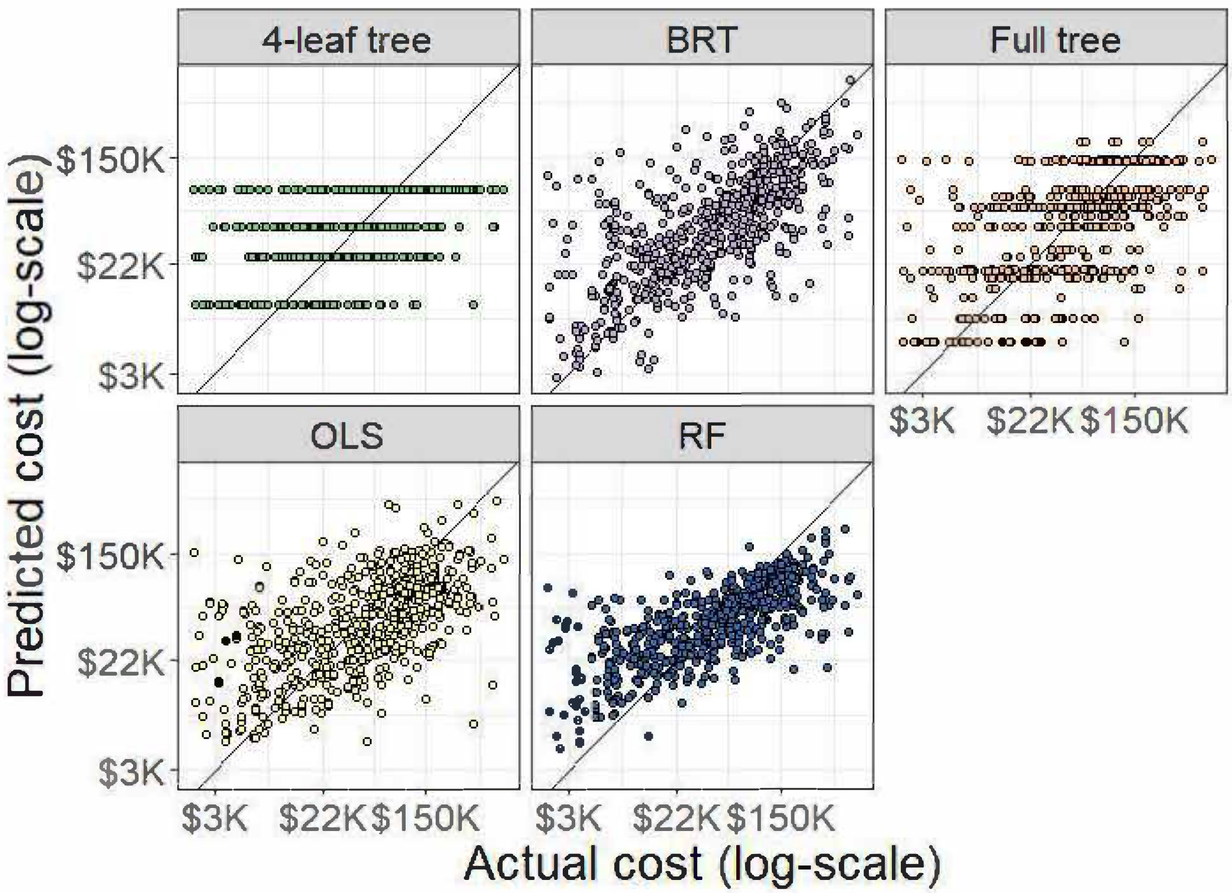

Figure S4. Predicted vs. actual costs in testing set by machine learning method. RF = random forest, BRT = boosted regression tree, OLS = ordinary least squares (linear regression).

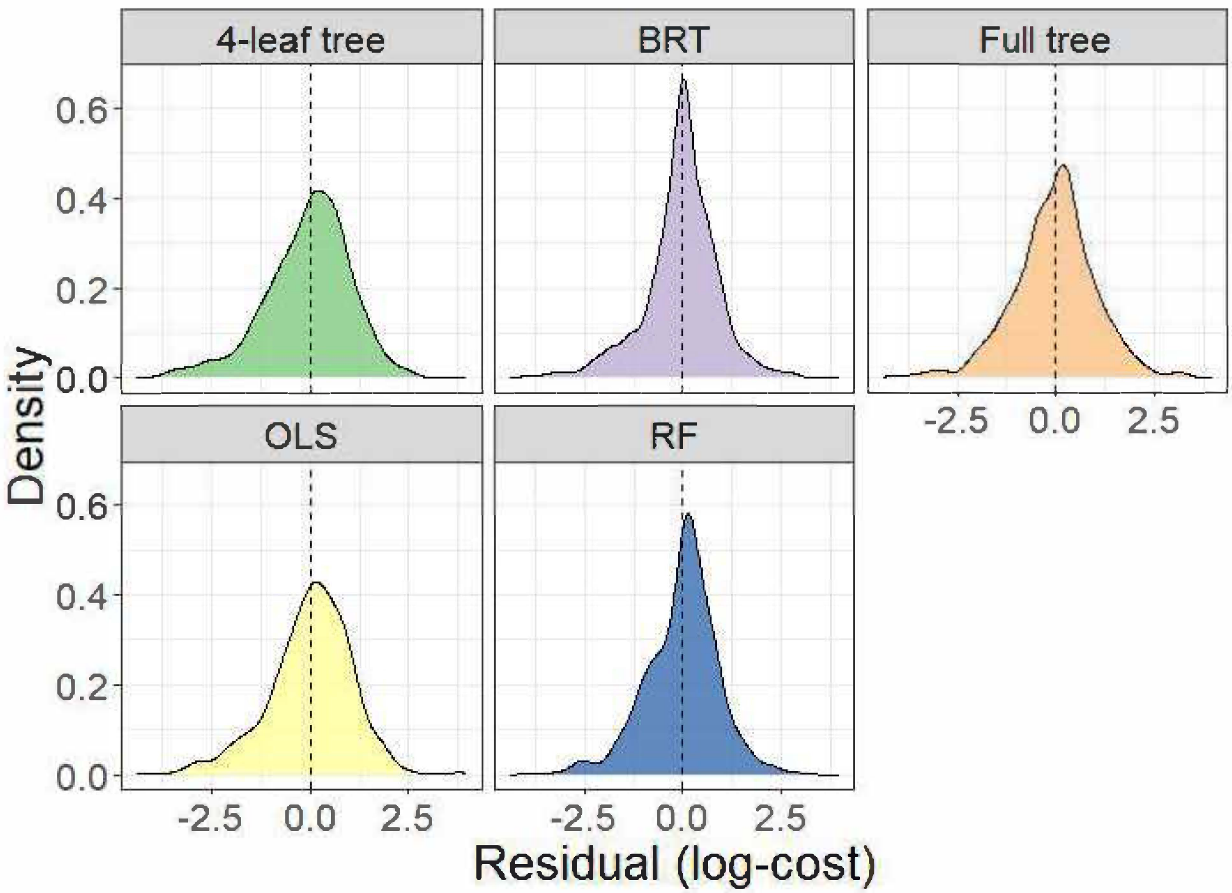

Figure S5. Distribution of residuals in testing set by machine learning method. RF = random forest, BRT = boosted regression tree, OLS = ordinary least squares (linear regression).

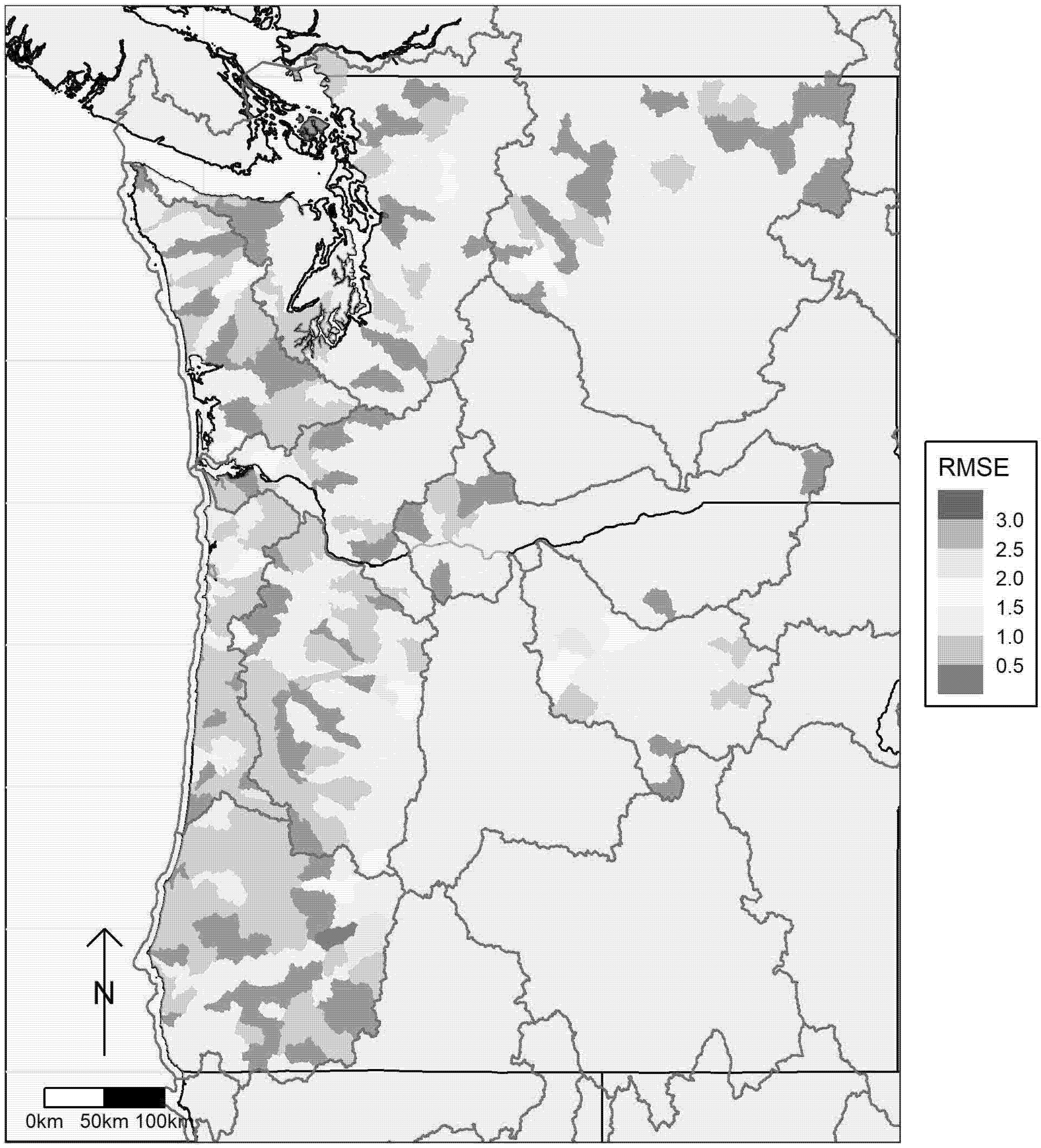

Figure S6. Root mean square error (RMSE) on testing set at HUC10-level for BRT.

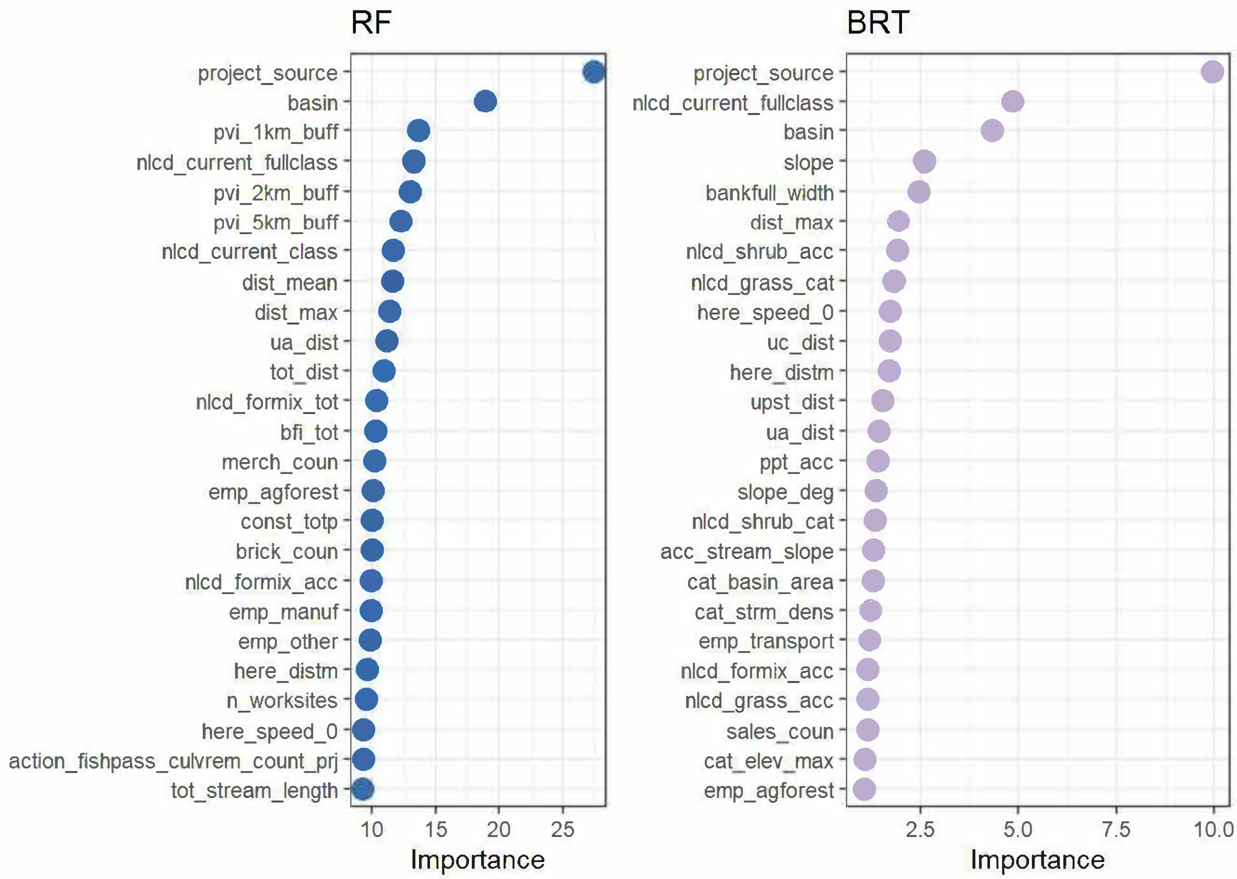

Figure S7. Variable importance plots for random forest (RF) and boosted regression tree (BRT) fits.

Supplementary Material 3 – Fixed effects discussion

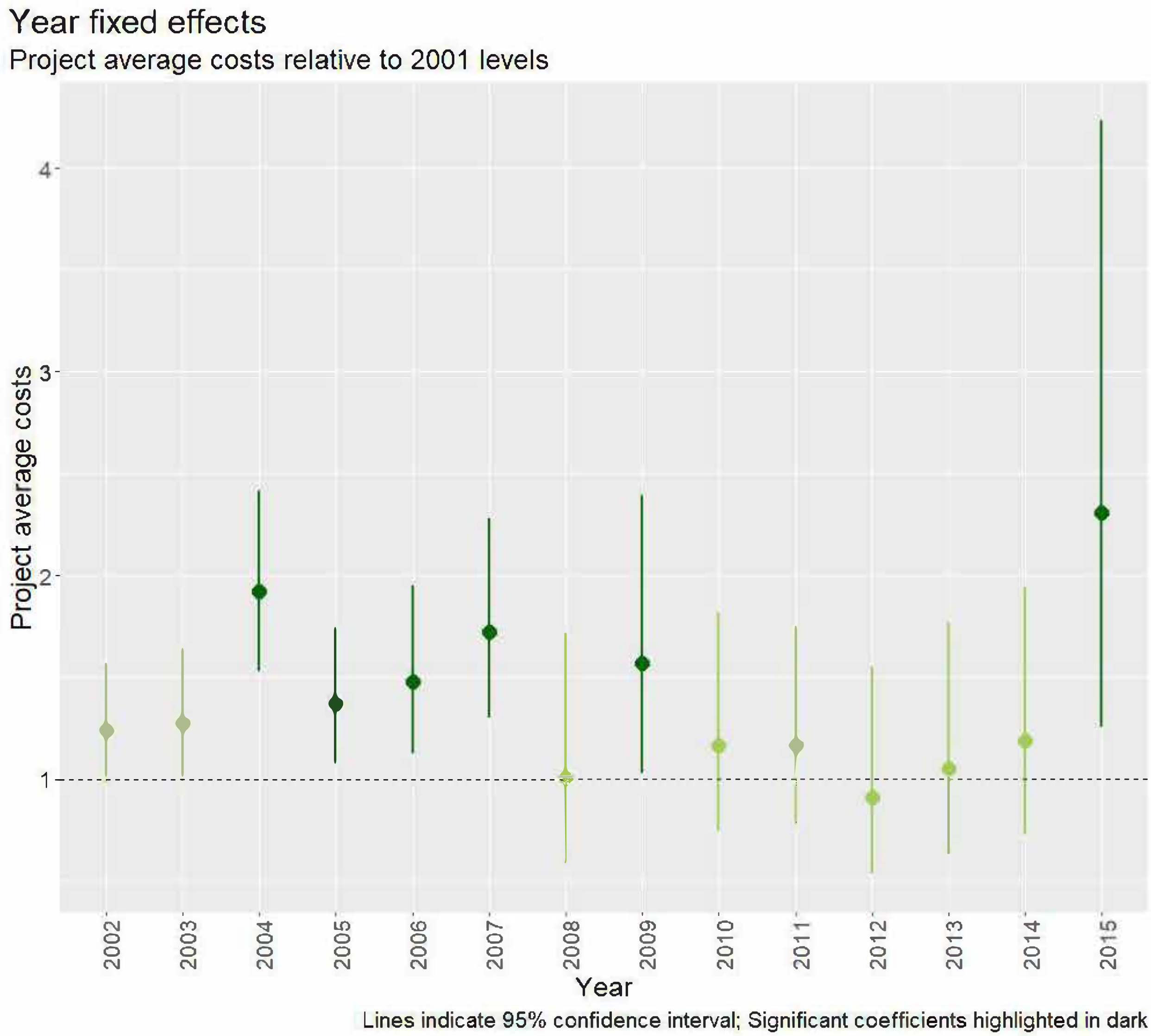

Figure S8. Year fixed effects.

Project average costs are as much as twice as high than 2001 levels between 2002 and 2007, when other factors are accounted for. In years that follow, costs return to around 2001 levels. In 2015, the last year of the sample, costs are nearly two-and-a-half times 2001 average costs, though this effect is estimated with a wide confidence interval and based on only a limited number of observations from this year (n = 6).

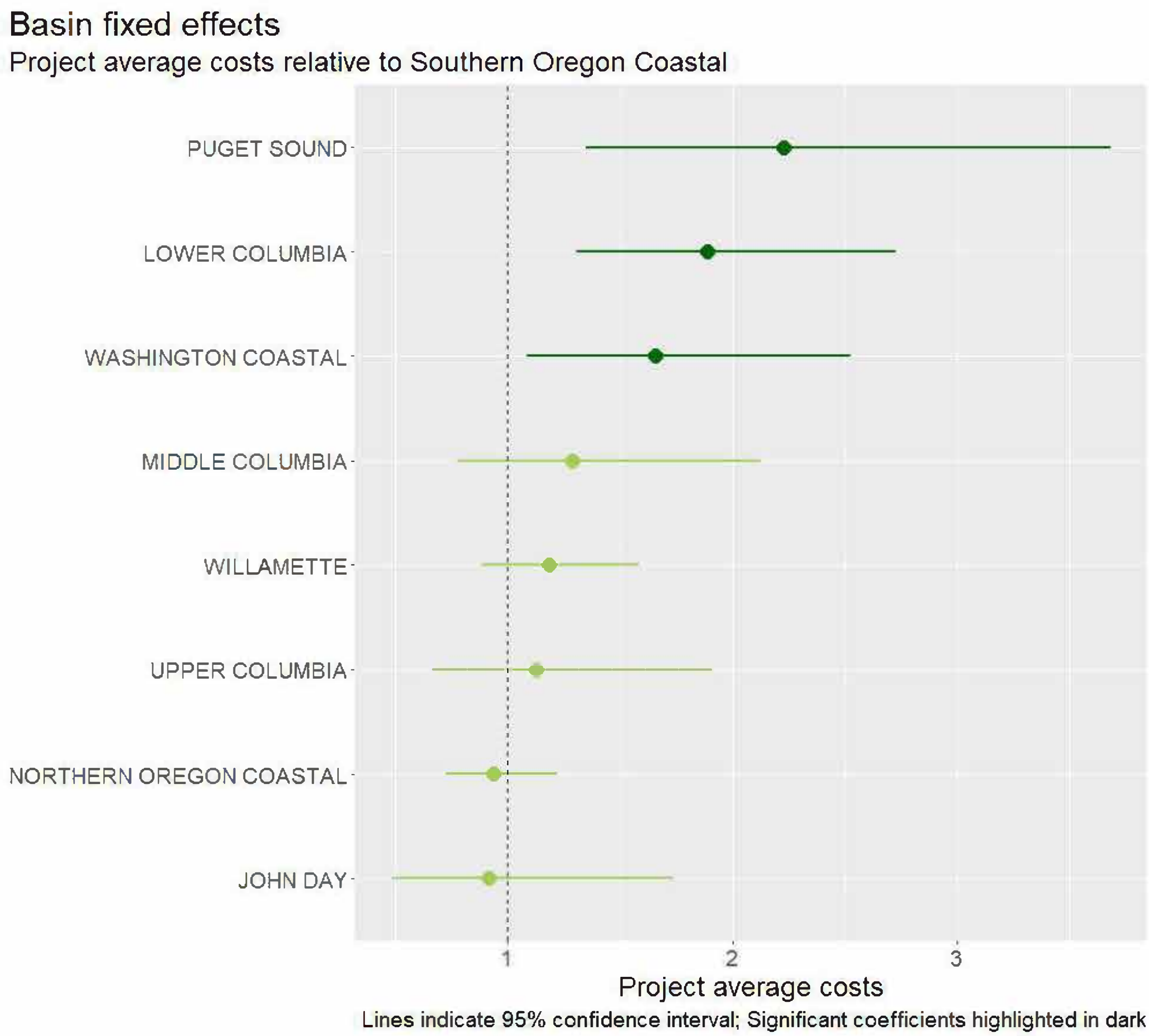

Figure S9. Basin fixed effects.

After accounting for other factors, worksites in the Puget Sound and Lower Columbia basins have the highest project average costs, followed by the Washington Coastal and Middle Columbia basins. The Willamette and both Oregon Coastal basins, including the Southern Oregon Coastal baseline, exhibited cost levels similar to the baseline, as did John Day and the Upper Columbia, both of which are east of the Cascade Range.

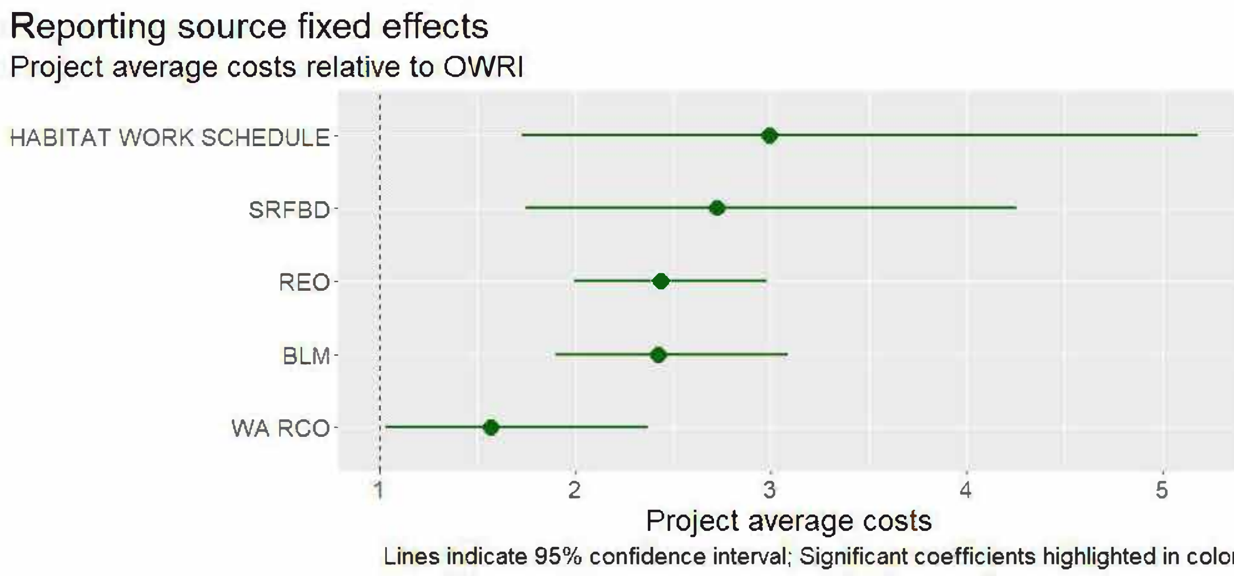

Figure S10. Reporting source fixed effects.

Bureau of Land Management (BLM) and U.S.F.S. Regional Ecosystem Office (REO) reported worksites with costs, two-and-a-half times costs reported by Oregon Watershed Restoration Inventory (OWRI), the baseline. OWRI projects exhibited the lowest average costs, followed by Washington State Recreation and Conservation Office, which reported projects associated with 48% higher costs. Projects reported by Habitat Work Schedule and Salmon Recovery Funding Board had exhibited average costs higher than both REO and BLM, though the confidence intervals for the former two sources fully contain the confidence intervals for each of the latter sources.

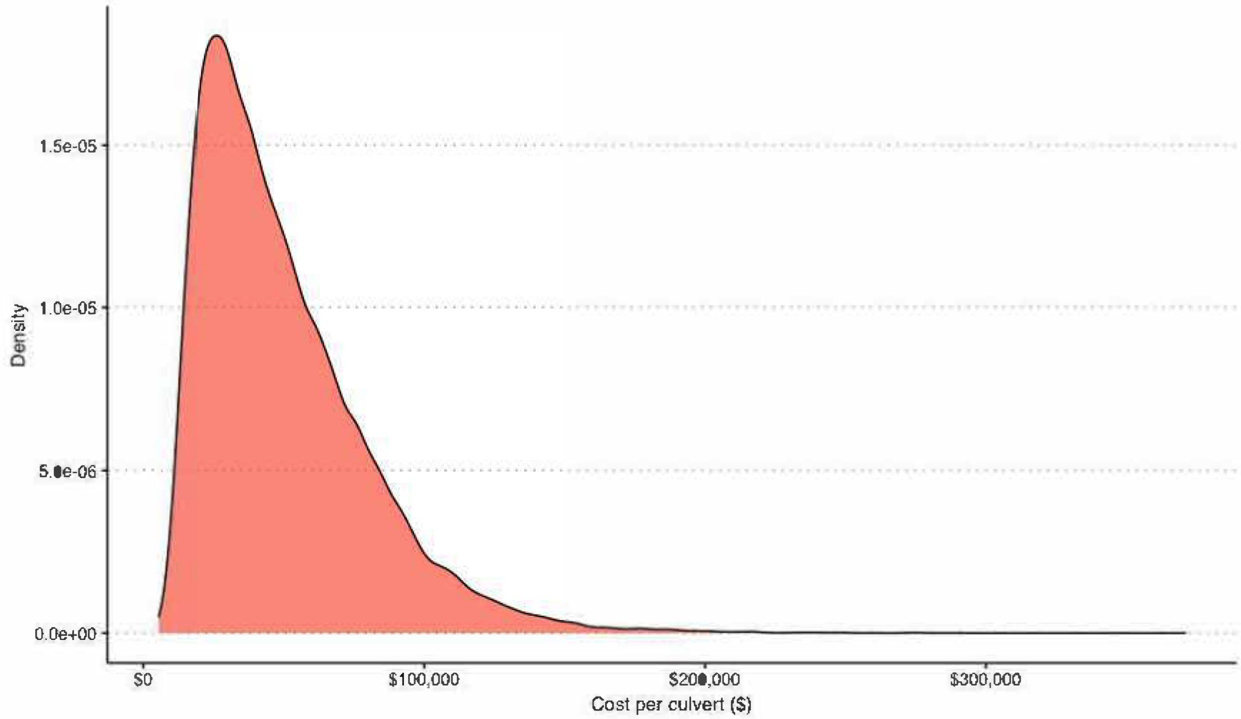

Figure S11. Distribution (kernel density) of predicted costs for inventory data.

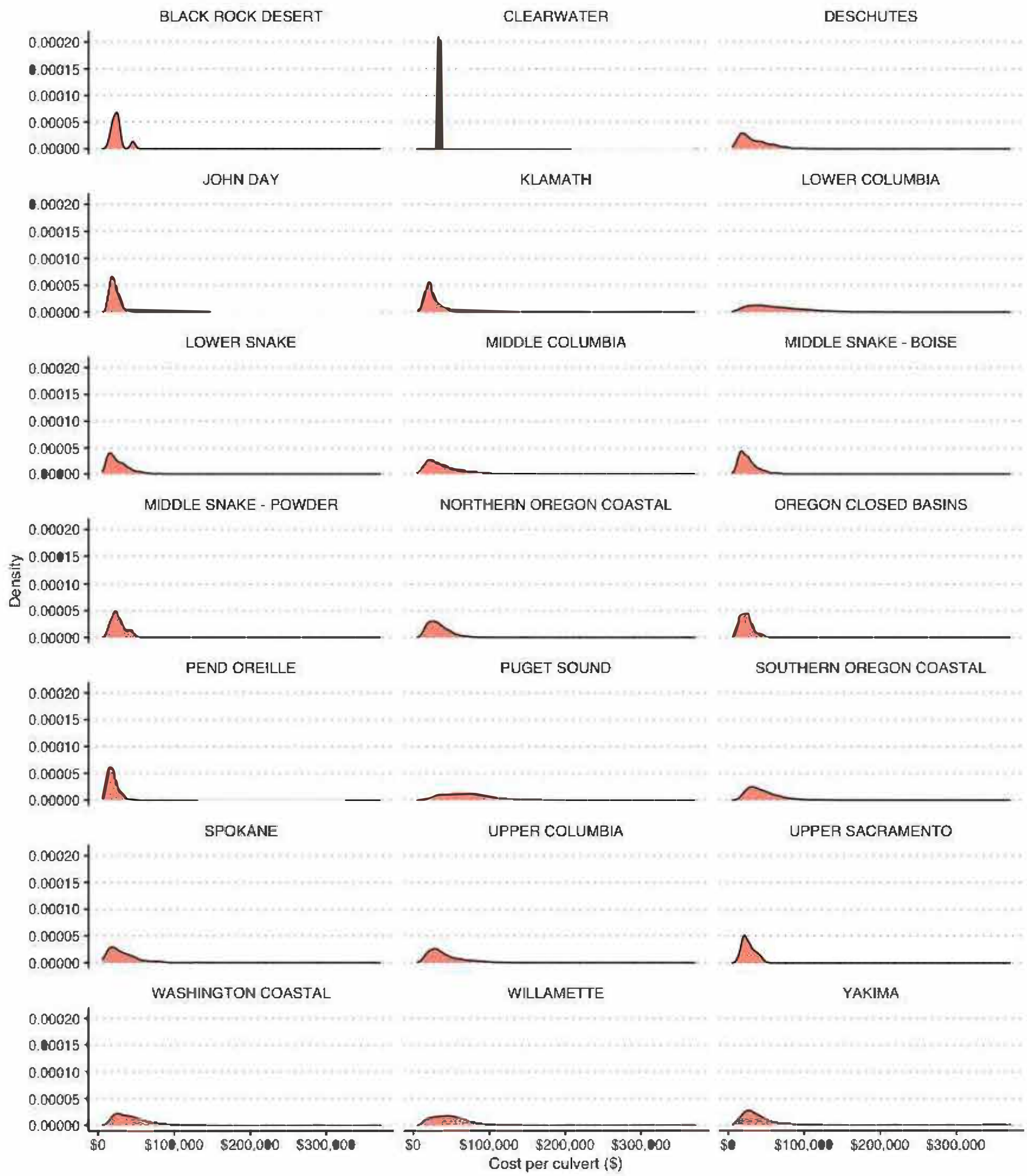

Figure S12. Distribution (kernel density) of predicted costs for inventory data, by basin

What influences spatial variability in restoration costs?

Econometric cost models for inference and prediction in restoration planning

Braeden Van Deynze[a*], Robert Fonner[b], Blake E. Feist[b], Sunny L. Jardine[c], Daniel S. Holland[b]

[a] Washington Department of Fish and Wildlife
1111 Washington St. SE
Olympia, WA 98501
United States

[b] National Marine Fisheries Service, Northwest Fisheries Science Center
2725 Montlake Blvd E
Seattle, WA 98112
United States

[c] University of Washington, School of Marine and Environmental Affairs
3707 Brooklyn Ave NE
Seattle, WA 98105
United States

* Corresponding author: Braeden.VanDeynze@dfw.wa.gov, 1-530-220-5287

Acknowledgements
We thank Katie Barnas and Monica Diaz for curating the restoration project data underlying this analysis. This publication was funded by the Cooperative Institute for Climate, Ocean, & Ecosystem Studies (CIOCES) under NOAA Cooperative Agreement NA20OAR4320271. Opinions and conclusions expressed or implied in this paper are solely those of authors and do not necessarily reflect the views or policy of NOAA Fisheries or Washington Department of Fish and Wildlife.

Author CRediT Statement
Braeden Van Deynze: Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Robert Fonner:** Conceptualization, Methodology, Software, Data Curation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition; **Blake E. Feist:** Methodology, Software, Data Curation, Writing - Review & Editing; **Sunny L. Jardine:** Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition; **Daniel Holland:** Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Conflict of Interest

The authors declare no conflict of interest.

Author CRediT Statement

Braeden Van Deynze: Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization; **Robert Fonner:** Conceptualization, Methodology, Software, Data Curation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition; **Blake E. Feist:** Methodology, Software, Data Curation, Writing - Review & Editing; **Sunny L. Jardine:** Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition; **Daniel Holland:** Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.