Ecological Indicators 50 (2015) 8–23

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Empirical evaluation of the conceptual model underpinning a regional aquatic long-term monitoring program using causal modelling

Kathryn M. Irvine [a,*], Scott W. Miller [b], Robert K. Al-Chokhachy [a], Eric K. Archer [c], Brett B. Roper [c], Jeffrey L. Kershner [a]

[a] U.S. Geological Survey, Northern Rocky Mountain Science Center, 2327 University Way, Suite 2, Bozeman, MT 59715, USA
[b] BLM National Aquatic Monitoring Center, Department of Watershed Sciences, Utah State University, 5210 Old Main Hill, Logan, UT 84322-5210, USA
[c] U.S.D.A. Forest Service, Forest Sciences Lab, 860 North 1200 East, Logan, UT 84321, USA

ARTICLE INFO

Article history:
Received 22 May 2014
Received in revised form 10 October 2014
Accepted 13 October 2014

Keywords:
Adaptive monitoring
Grazing
Heterogeneous landscapes
Long-term monitoring
Mediator
Moderator
Macroinvertebrates
Path analysis
Roads
Sedimentation

ABSTRACT

Conceptual models are an integral facet of long-term monitoring programs. Proposed linkages between drivers, stressors, and ecological indicators are identified within the conceptual model of most mandated programs. We empirically evaluate a conceptual model developed for a regional aquatic and riparian monitoring program using causal models (i.e., Bayesian path analysis). We assess whether data gathered for regional status and trend estimation can also provide insights on why a stream may deviate from reference conditions. We target the hypothesized causal pathways for how anthropogenic drivers of road density, percent grazing, and percent forest within a catchment affect instream biological condition. We found instream temperature and fine sediments in arid sites and only fine sediments in mesic sites accounted for a significant portion of the maximum possible variation explainable in biological condition among managed sites. However, the biological significance of the direct effects of anthropogenic drivers on instream temperature and fine sediments were minimal or not detected. Consequently, there was weak to no biological support for causal pathways related to anthropogenic drivers' impact on biological condition. With weak biological and statistical effect sizes, ignoring environmental contextual variables and covariates that explain natural heterogeneity would have resulted in no evidence of human impacts on biological integrity in some instances. For programs targeting the effects of anthropogenic activities, it is imperative to identify both land use practices and mechanisms that have led to degraded conditions (i.e., moving beyond simple status and trend estimation). Our empirical evaluation of the conceptual model underpinning the long-term monitoring program provided an opportunity for learning and, consequently, we discuss survey design elements that require modification to achieve question driven monitoring, a necessary step in the practice of adaptive monitoring. We suspect our situation is not unique and many programs may suffer from the same inferential disconnect. Commonly, the survey design is optimized for robust estimates of regional status and trend detection and not necessarily to provide statistical inferences on the causal mechanisms outlined in the conceptual model, even though these relationships are typically used to justify and promote the long-term monitoring of a chosen ecological indicator. Our application demonstrates a process for empirical evaluation of conceptual models and exemplifies the need for such interim assessments in order for programs to evolve and persist.

Published by Elsevier Ltd.

1. Introduction

Most regional and national monitoring programs are developed upon the same overarching goal of providing natural resource status and trend assessments beyond the local scale (e.g., Bureau of Land Management's Assessment, Inventory and Monitoring Strategy; Forest Service and BLM's Northwest Forest Plan; National Park Service's Vital Signs Program; and PacFish–InFish Biological Opinions Effectiveness Monitoring Program). A key component in the development of such monitoring programs is the creation of conceptual models that identify core ecosystem processes and factors that may impact them directly or indirectly (Fancy et al., 2009; Noon et al., 1997). Conceptual models can be presented in many different forms; for example, stressors and processes could

* Corresponding author.
E-mail address: kirvine@usgs.gov (K.M. Irvine).

http://dx.doi.org/10.1016/j.ecolind.2014.10.011
1470-160X/Published by Elsevier Ltd.

be listed in tabular form, as flow charts with boxes and arrows connecting the individual components, or as cartoons (e.g., Gross, 2003; Margolius et al., 2009). The utility of framing monitoring programs around conceptual models is the development of scientifically sound monitoring questions, selection of relevant ecological indicators of resource condition, identification of drivers and stressors, and as a communication and information organization tool (Barrows and Allen, 2007; Fancy et al., 2009; Lindenmeyer and Likens, 2009; Noon et al., 1997; Ringold et al., 1999).

Interestingly, although there is emphasis on creating conceptual models during program development, many times the primary measurable objectives are in terms of status and regional trend assessments ("surveillance monitoring", Nichols and Williams, 2006). Consequently, most initial and interim planning focuses on the choice of sampling design (where and how many sample points are selected), temporal revisit design (frequency of data collection over years), or response design (what and how field measurements are collected) that precisely characterize biological condition (status) and maximize trend detection (e.g., Urquhart et al., 1998; Manley et al., 2004; Sims et al., 2006; Reynolds et al., 2011; Levine et al., 2014). For programs targeting the effects of anthropogenic activities, however, it is imperative to identify and measure land use practices and mechanisms that have led to degraded conditions (i.e., moving beyond simple status and trend estimation). Knowledge of such causal linkages is needed for prescribing appropriate changes in management to reduce stressor(s) and improve resource condition. Here we assess whether data collected as part of a regional aquatic monitoring program, PacFish–Infish Biological Opinions Effectiveness Monitoring Program (PIBO) (PACFISH, 1994), can be used to estimate causal pathways as articulated in the original conceptual model. The PIBO program was developed in response to listing of steelhead (*Oncorhynchus mykiss*) and bull trout (*Salvelinus confluentus)* under the Endangered Species Act with a specific focus on evaluating the status and trends of federally-managed headwater streams in the Interior Columbia River Basin and has since expanded to the upper Missouri River Basin (Fig. 1, Kershner et al., 2004).

The PIBO program has been implemented for over a decade, making it ideal for employing causal models to rigorously evaluate whether the original conceptual model of drivers and stressors is supported by the available monitoring data. We describe our process of translating the ecological information within the tabular conceptual model into a causal graph. The linkage structure portrayed in the graph is inherently a complex causal hypothesis and can be considered as a series of working hypotheses for how structure and function of an ecological system responds to various anthropogenic stressors and natural environmental gradients (Grace et al., 2012; Shipley 2009). Some of the advantages of utilizing these models for analyses include tests of mediation (see Burdon et al., 2013; Cubaynes et al., 2012; Gimenez et al., 2012; Riseng et al., 2011), estimation of cumulative and indirect effects (e.g., Clough, 2012), and accounting for context dependency and mitigating factors as commonly encountered with probabilistically sampled data collected across large spatial domains (brief introduction to causal models in

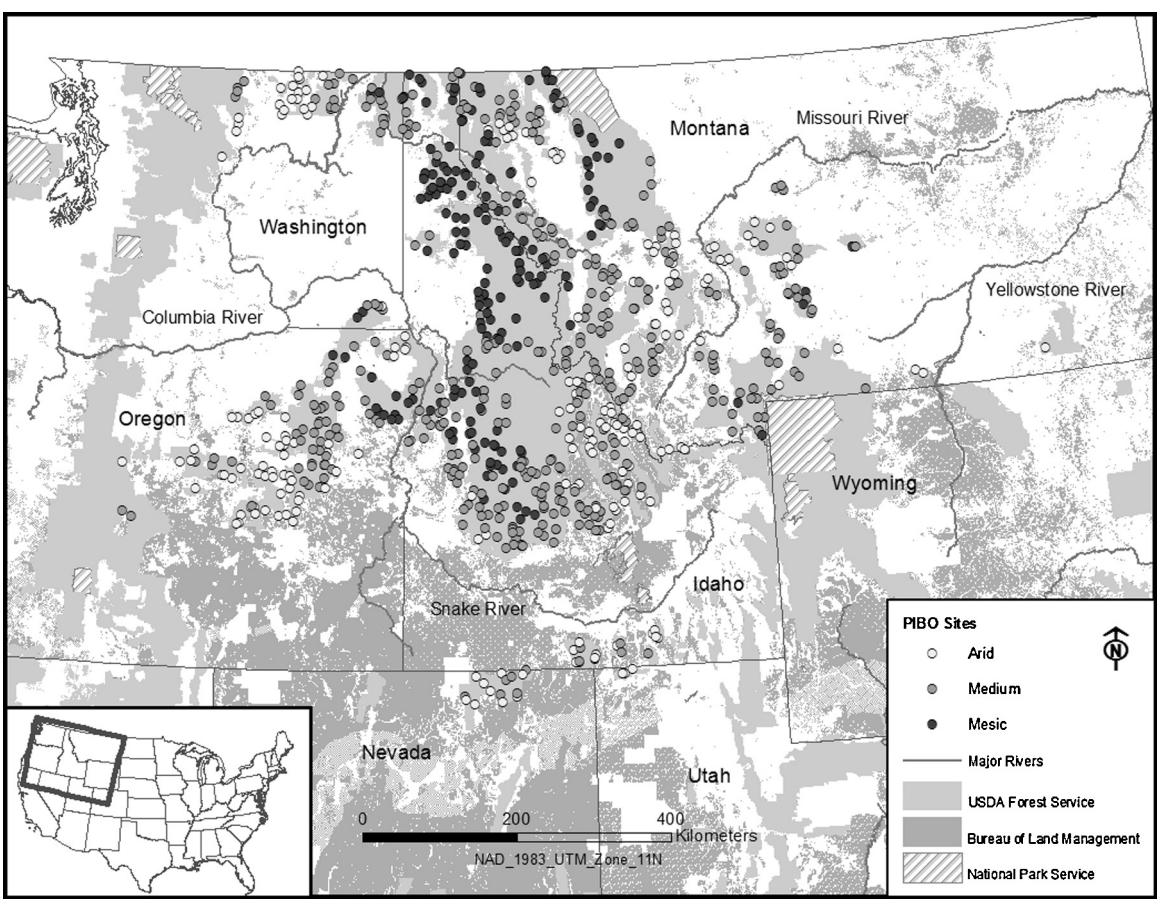

Fig. 1. Spatial distribution of landownership and PIBO sample sites within each of the three precipitation groups (arid [ppt $\leq 0.66\,\mathrm{m\,yr^{-1}}$], medium [$0.66 < \mathrm{ppt} < 1.05\,\mathrm{m\,yr^{-1}}$], and mesic sites [ppt $\geq 1.05\,\mathrm{m\,yr^{-1}}$]) sampled throughout the Interior Columbia and Upper Missouri River Basins.

Appendix A). To our knowledge there have been few published attempts to explicitly evaluate the hypotheses conveyed in the conceptual model underpinning a monitoring program as we do here. Our analytical framework provides an interim assessment of the monitoring program by linking monitoring data back to the original conceptual model and demonstrates the practice of adaptive monitoring at a regional scale (Lindenmeyer and Likens, 2009; Reynolds et al., 2011).

2. Methods

2.1. Study area and sampling design

The study area spans approximately 740,000 km^2 of the Interior Columbia River Basin and the Upper Missouri River Basin (Fig. 1). Elevations vary considerably throughout the study area (450–2460 m) and climatic zones are defined by altitudinal and latitudinal gradients. Ambient air temperatures and precipitation (type and quantity) vary considerably, but the climate is generally characterized by relatively cold winters and hot, dry summers, with monthly average temperatures ranging from −12 to 0 °C in January and from 11 to 23 °C in July. Winter precipitation predominantly falls in the form of snow with mean annual precipitation ranging from 0.24 m yr^{-1} at arid sites to around 2 m yr^{-1} at more mesic sites (2002–08; PRISM Group, Oregon State University, http://www.prismclimate.org).

Sample sites were selected via a spatially balanced random sample design (Stevens and Olsen, 2004) and visited based on a five year rotating panel (a panel is composed of a set of sites with the same revisit schedule, i.e., every five years). Within each randomly-selected sub-watershed, stream conditions were evaluated in the most downstream reach where gradient, measured as the change in the surface water elevations between the bottom and top of the reach divided by the reach length, was <3% and the upstream catchment was primarily (>50%) under federal ownership. Where low gradient (i.e., <3%) reaches were limited within a sub-watershed, we sampled the lowest-gradient stream reaches available and included some with gradients up to 9%; sample reaches were equal to 20× bankfull width or a minimum of 160 m in length and a maximum of 500 m (hereafter, referred to as a site). We focused sampling on wadeable, higher order (3rd/4th) low gradient reaches as these areas are thought to be most sensitive to natural and anthropogenic disturbances (Montgomery and MacDonald, 2002) and fish bearing. We used data collected from 2006 to 2010 for a total of 842 sites, which constituted the second visit for the Upper Columbia sites and first visit for Upper Missouri sites. To minimize measurement error, we selected these years to ensure consistent field protocols across relevant habitat and macroinvertebrate indices.

Under the PIBO monitoring program, sites were selected to represent both reference and managed conditions based on current and historic land-management activities within the upstream catchments. Sites in wilderness areas or in catchments with minimal land management activities were considered 'reference' sites, where minimal land management was defined as (1) no permitted livestock grazing during the last 30 years, (2) minimal timber harvest (<10%), (3) minimal road density (0.5 km/km^2) at the catchment scale, (4) no roads within the proximate (1 km) riparian buffer, and (5) no evidence of historic mining within riparian areas upstream of the site (Kershner et al., 2004). Sites in catchments where any one land management activity exceeded the reference criteria were considered 'managed'. Within our study area livestock grazing and forest roads were two of the most pervasive management activities. Here, we focused our analysis on only 'managed' sites to better isolate the mechanisms by which management activities (road density, timber harvest, and grazing) impact biological condition.

2.2. Translating a conceptual model into a causal graph

The original conceptual model prioritized several key chemical, physical, and biological processes relevant to maintaining or

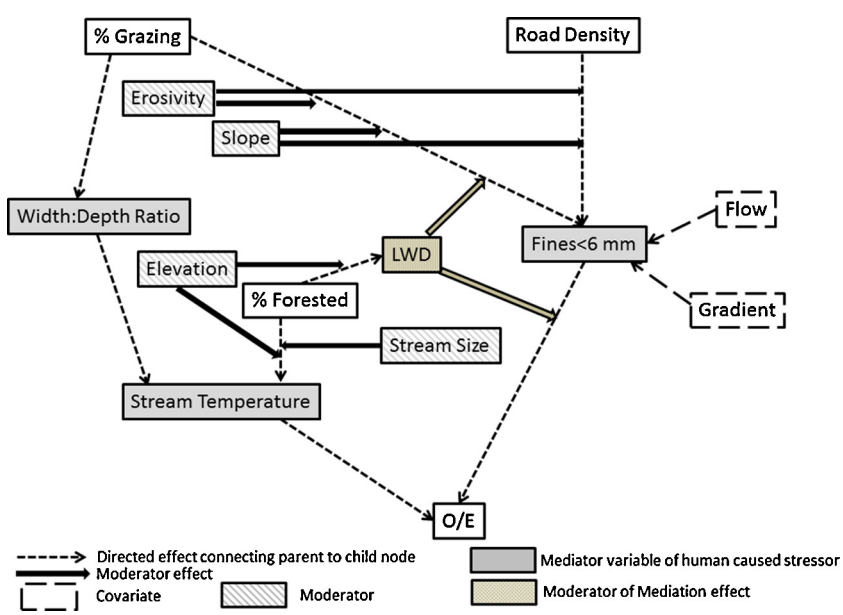

Fig. 2. Causal graph describing the network of causal pathways among anthropogenic drivers in forested streams (% grazing, % forest, and road density in catchments), instream stressors (stream temperature and fine sediments[fines]), and biological condition as moderated by the landscape context. Introduction to causal model terminology provided in Appendix A. In brief, covariates account for natural variation of instream sediments (e.g., Flow and Gradient). Moderator variables (e.g., large woody debris [LWD], Elevation, Erosivity, Slope, Stream Size) are hypothesized to modify the direction and/or strength of the relationship (directed edge) to which the solid black arrow points, e.g., road density may differentially effect fines based on the erosivity of the parent material within a catchment. Mediators (e.g., Fines, Stream Temperature, Width:Depth Ratio) describe mechanisms for how or why anthropogenic drivers affect biological condition (O/E).

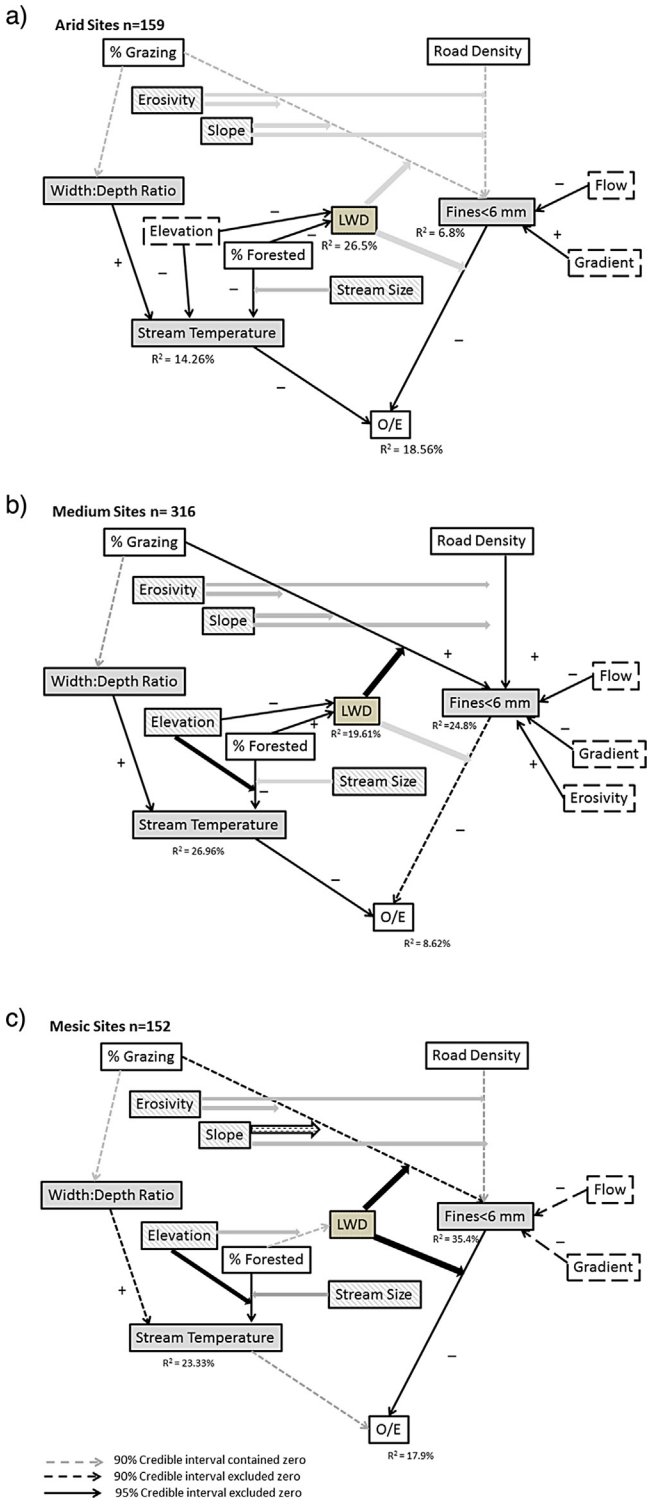

Fig. 3. Fitted causal model to three precipitation groups: arid [PPT ≤ 0.66] (a), medium [0.6 < PPT < 1.05] (b), and mesic [PPT ≥ 1.05] (c). (+) indicates positive effect, (−) indicates negative effect of parent on child node based on reduced graph estimates. Moderation effect estimates when linked to biological condition are reported in Table 4. Sample sizes (n) indicate the number of sites within each precipitation group. R^2 values were calculated based on variance explained in child node based on parent nodes only. See appendix C Table C1 for Bayesian posterior estimates for edge coefficients in reduced graph [excluding edges in gray].

enhancing native fish populations (Figs. 1–3 in Kershner et al., 2004). Stream habitat has been identified as one of four key metrics limiting the success of salmon and steelhead populations in the Upper Columbia River system (habitat, hatcheries, harvest, and hydropower; Good et al., 2007; NMFS, 2000) (www.salmonrecovery.gov). While the PIBO program originated from concerns over fish species, fishes are impractical for long-term monitoring because of the difficulty linking changes in stream habitat to far-ranging species that spend part of their life histories in large migratory river corridors and the ocean, and currently have extremely small populations in some areas.

Aquatic macroinvertebrates, on the other hand, spend most or all of their life cycle in the target streams, are useful indicators of non-point source pollution (grazing, road building, timber harvest), and have numerous advantages as bioindicators of stream conditions (Barbour et al., 1999; Carter et al., 2007; Karr and Chu, 1999). Consequently, the biotic component of the conceptual model assumes that macroinvertebrate assemblages are responsive to habitat alterations, which are known to adversely impact coldwater fish assemblages. Ultimately all of the identified processes, stressors, and consequences were linked to aquatic communities in the conceptual model. As a result, in our causal graph, we used biological condition as our terminal node (see Table A1 definitions, Fig. 2); we quantified biological condition as the observed (O) macroinvertebrate assemblages compared to the expected (E) given the absence of anthropogenic impairment (O/E; Hawkins et al., 2000; Hawkins and Carlisle, 2001). The O/E index is a measure of biodiversity and thus quantifies alterations to aquatic ecosystems as the degree of species loss in relation to least disturbed conditions, with values less than one indicating higher levels of species loss.

The expected assemblages (i.e., least disturbed) were predicted from 201 sites experiencing minimal human disturbance (least disturbed conditions: 174 calibration and 27 randomly selected validation sites) that spanned the physiographic diversity of the study area. The 201 reference sites were grouped into 10 distinct classes based on the similarity of macroinvertebrate assemblage composition among sites following the standard methods of Hawkins et al. (2000). The expected class membership and subsequent reference macroinvertebrate assemblage (E) of managed sites was predicted by linear discriminant function models using watershed area, 30-year average monthly maximum air temperature and the 30-year average precipitation for the 12 months prior to a standardized sample collection date of July, both derived from PRISM estimates. In addition to the O/E index, we also explored measures of assemblage structure as our terminal node (e.g., species diversity, relative abundance of intolerant taxa, functional feeding groups), but the use of different response variables did not change the performance or interpretation of the casual models (data not shown).

Although the tabular conceptual model was useful for developing the foundations for the PIBO program, causal pathways related to O/E were not yet refined with respect to drivers and stressors that were actually collected (for detailed field methods see Appendix B). Our first step was to articulate current scientific knowledge concerning biological and physical relationships for aquatic systems (Table 1). Next, we cross walked these hypothesized relationships with the measured and derived variables in order to estimate pathways within a causal model. The main drivers of interest to the PIBO monitoring program were percent of grazing, density of roads, and percent of forest within a catchment (Fig. 2, Table 1). Our hypothesized mechanisms for how these drivers affect O/E are through sedimentation and streamwater temperature (mediators in the graph, Fig. 2 and Table 1). The next step was to consider mitigating factors at the catchment or reach level that may alter either the driver → stressor or stressor → O/E links through dampening or amplifying the statistical and biological effect sizes (moderators in the graph, Fig. 2, Table 2). Finally, we included additional covariates to account for background environmental variation (Fig. 2).

Table 1
Compiled scientific knowledge concerning biological and physical relationships for aquatic systems as conveyed in causal hypotheses represented in Fig. 2.

Causal path	Causal hypotheses	Citations
% Grazing → fines → O/E	By altering vegetative composition and cover and trampling stream banks, cattle grazing can increase fine sediment loading to streams causing macroinvertebrate assemblages to deviate from reference or expected conditions	(Duff, 1979; McIver and McInnis, 2007; Saunders and Fausch, 2007)
% Grazing → width to depth ratio → stream temperature → O/E	By altering vegetative composition and cover, cattle grazing can increase the amount of incoming solar radiation, while the trampling of stream banks can increase width:depth ratios Collectively these two mechanisms can increase stream temperatures causing observed macroinvertebrate assemblages to deviate from reference or expected conditions	(Kauffman and Krueger, 1984; Kauffman et al., 1983; Knapp and Matthews, 1996; Li et al., 1994; McIver and McInnis, 2007; Platts, 1991; Saunders and Fausch, 2007)
Road density → fines → O/E	Forest road construction can increase fine sediment loading to streams by accelerating mass wasting events and altering the path, timing, and magnitude of runoff events Fine sediment adversely impacts macroinvertebrates, causing O/E to decrease	(Lenat and Crawford, 1994; King, 1983; Wood and Armitage, 1997)
% Forest → large woody debris	The extent of riparian forest influences the amount of large woody debris (LWD) available for recruitment	(Burns, 1972; Lemly and Hilderbrand, 2000; Tschapalinski and Hartman, 1983; WoodSmith and Buffington, 1996)
% Forest → stream temperature → O/E	Stream temperature can be inversely related to percent overhead vegetative cover and increases in stream temperature can cause macroinvertebrate assemblages to deviate from reference or expected conditions	(Davies and Nelson, 1994; Johnson and Jones, 2000; Moore et al., 2005; Tait et al., 1994)

Specifically, measures of streamflow and gradient were included as covariates because we expected fine sediment levels to naturally vary with stream power and we sought to differentiate natural gradients in fine sediment from anthropogenic impacts. The result was a causal graph with both fine sediment within a stream and stream temperature mediating the effect of anthropogenic drivers on O/E (Fig. 2). In addition, watershed attributes that may mitigate the effects of human disturbance were included in the model.

2.3. Bayesian parameter estimation

We hypothesized that precipitation is an overarching driver of the chemical, physical, and biological processes contained within the conceptual model (Feminella, 1996; Monk et al., 2008; Jorgensen et al., 2009). Essentially, we are assuming that precipitation moderates the entire casual graph. To avoid the complexity of incorporating an additional moderator, we subset the data into three groups. Specifically, we post-stratified the sites into three precipitation groups based on 30-year average precipitation (arid [ppt $\leq 0.66\,\mathrm{m\,yr^{-1}}$], medium [$0.66 < \mathrm{ppt} < 1.05\,\mathrm{m\,yr^{-1}}$], and mesic sites [ppt $\geq 1.05\,\mathrm{m\,yr^{-1}}$]) for three separate analyses. The majority of the PIBO sites have a snowpack hydrologic regime with peak flows occurring during May and June. However, the arid sites receive less snowfall resulting in both reduced magnitude and duration of discharge peaks, as well as lower base flows, because less precipitation falls as rain compared to the medium and mesic sites (average bankfull discharge [cfs] = 50.1 in arid sites compared to 125.4 [cfs] in medium and 253.2 [cfs] in mesic sites).

Based on exploratory analysis, we transformed variables to achieve linearity, constant variance, and removed extreme outliers. We used the natural log transformation for fines + 0.05 and LWD + 1 (Keene, 1995). All variables were centered by subtracting the average value within their respective precipitation group. We divided elevation by 1000 and erosivity by 100 to avoid numerical instability in the estimation.

Recently, Grace et al. (2012) provided guidance for a graph-theoretic approach to structural equation modeling and Shipley (2013) provided examples of this approach using path analysis, as we use here (see Appendix A for more details). Based on the conditional independence relationships (implied by Fig. 2), we

Table 2
Description of motivation for including moderation effects for some linkages in the causal graph (Fig. 2). A moderator is a contextual variable that modifies the direction or strength of the directed edge; it clarifies the when, where, or for whom (Wu and Zumbo, 2008; Table 1). LWD is an abbreviation for large woody debris.

Moderation Effects	
LWD for % grazing → fines	Higher amounts of LWD should result in increased fine sediment retention in pools
LWD for fines → O/E	LWD can mitigate the adverse impacts of excessive fine sediment loading on macroinvertebrates by differentially sorting fine sediment loading in depositional (e.g., pools) versus erosional (e.g., riffles) habitats; macroinvertebrate diversity is typically greater in erosional habitats
Elevation with % forest → LWD	Higher elevation forests tend to have sparser density trees resulting in less LWD recruitment compared to lower elevation forests
Elevation with % forest → stream temperature	Elevation can be considered a surrogate for ambient air temperatures, such that reaches at higher elevations are likely thermally insolated compared to lower elevation reaches where shading may have a differential impact on instream temperatures
Stream size with % forest → stream temperature	Larger streams are likely to have dampened effects of forest cover through shading on stream temperature
Erosivity with % grazing → fines	More erosive buffers could amplify the magnitude of grazing impacts on fines deposited in streams
Slope with % grazing → fines	Steeper slopes will increase the transport of grazing derived fine sediment to streams
Erosivity with road density → fines	Similar to % grazing
Slope with road density → fines	Similar to % grazing

modeled the observed data using the following set of equations (bold indicates vector valued quantities):

$$\mathrm{O/E} = \beta_1 \mathrm{fines} + \beta_2 \mathrm{streamtemperature} + \beta_3 \mathrm{LWD} + \beta_4 \mathrm{fines} \times \mathrm{LWD} + \varepsilon_1, \quad (1)$$

$$\mathrm{Fines} = \alpha_1 \%\mathrm{grazing} + \alpha_2 \mathrm{roaddensity} + \alpha_3 \mathrm{LWD} + \alpha_4 \mathrm{flow} + \alpha_5 \mathrm{gradient} + \alpha_6 \mathrm{slope} + \alpha_7 \mathrm{erosivity} + \alpha_8 \%\mathrm{grazing} \times \mathrm{slope} + \alpha_9 \%\mathrm{grazing} \times \mathrm{erosivity} + \alpha_{10} \mathrm{roaddensity} \times \mathrm{slope} + \alpha_{11} \mathrm{roaddensity} \times \mathrm{erosivity} + \alpha_{12} \%\mathrm{grazing} \times \mathrm{LWD} + \varepsilon_2 \quad (2)$$

$$\mathrm{LWD} = \gamma_1 \%\mathrm{forest} + \gamma_2 \%\mathrm{elevation} + \gamma_3 \%\mathrm{forest} \times \mathrm{elevation} + \varepsilon_2 \quad (3)$$

$$\mathrm{Streamtemperature} = \delta_1 \%\mathrm{forest} + \delta_2 \mathrm{elevation} + \delta_3 \mathrm{stream\ size} + \delta_4 \mathrm{width} : \mathrm{depthratio} + \delta_5 \%\mathrm{forest} \times \mathrm{elevation} + \delta_6 \mathrm{forest} \times \mathrm{stream\ size} + \boldsymbol{\varepsilon_4}, \quad (4)$$

and

$$\mathrm{Width} : \mathrm{depth\ ratio} = \theta_1 \%\mathrm{grazing} + \varepsilon_5, \quad (5)$$

where the coefficient parameters (α, β, γ, δ, θ) in Eqs. (1)–(5) correspond to the directed edges in the graph (Fig. 2). Moderator variables are included as interaction terms. We assumed that the error terms were uncorrelated and consistent with a normal distribution and verified this assumption using residual diagnostic plots. Also, because we used centered variables there was no need to specify an intercept term in Eqs. (1)–(5). We estimated the parameters of the causal model using Bayesian estimation implemented in the freeware WinBUGS. Example WinBUGS code used to fit a model and for estimating the cumulative (total) effects of grazing on O/E is provided in Appendix D.

We specified diffuse normal priors on the edge coefficients ($\gamma_1 \sim$ normal[0,100]) normal[0,100]) and diffuse gamma priors for the precisions (e.g., $1/\varepsilon_1 \sim$ gamma[0.5,1.0]) (Lee, 2007). We ran three parallel chains with different randomly selected starting values. We used 15,000 Markov chain Monte Carlo iterations (burn in of 5000) thinned by 10, resulting in 3000 posterior iterations for each parameter. The Gelman-Rubin *R*-hat statistics were all less than 1.1 indicating the chains converged; also trace plots indicated proper mixing.

For model selection, we estimated the 90% credible intervals and reduced the linkage set by removing links whose intervals contained zero. We report the median, 90% and 95% credible intervals for the direct effects (each link between two nodes) based on the reduced graphs (Table C1 Appendix C). We also report a measure of model fit for each child node in the graph as a measure of variance explained by the parent nodes only, R^2.

2.4. Model assessment

A distinction in causal modeling, an inherently multivariate technique, is that posterior predictive distributions can be created for each univariate model, Eqs. (1)–(5) (sub-graphs), to assess the local fit of the model. For these purposes, we randomly selected 25% of the data within each precipitation group as a validation set. Essentially, we simulated 100 realizations of data under the assumed reduced models using the last 100 draws from the posterior distributions for all parameters for comparison to our hold-out dataset, a posterior predictive distribution (Lee, 2007).

Cumulatively, both model and measurement errors introduce noise that can constrain the total amount of variance accounted for in O/E scores as a function of measured stressors. To assess this, we followed Vander Laan et al. (2013) and computed the maximum possible variance explained for each precipitation group model, which was based on the ratio of variance of O/E scores observed at all sites within a precipitation group (i.e., signal) to variance of reference site O/E scores used to validate the model (i.e., noise).

3. Results

Land use differed among the three precipitation groups, with mesic watersheds having the greatest forest cover and least amount of area contained within grazing allotments compared to sites in the medium and arid groups (Table 3). Similarly, the mesic sites were found in larger watersheds (data not shown), occurring at lower elevations that had more erosive lithologies and greater frequencies of LWD on average (Table 3). These physiographic conditions translated to instream conditions characterized by

Table 3
Mean (standard deviation) conditions for landuse drivers, mediators, moderators and the biological response variable compared among precipitation groups. Detailed descriptions of field methods are provided in Appendix B.

		Precipitation groups		
		Arid (n = 212)	Medium (n = 422)	Mesic (n = 204)
Drivers	Road density (km/km^2)	1.4 (0.99)	1.1 (1.04)	1.2 (1.2)
	Grazing (%)	86.7 (27.6)	65.7 (43.2)	27.3 (42.2)
	Forest (%)	66.6 (27.46)	75.4 (18.0)	82.8 (14.1)
Mediators	Fines (%)	36.6 (27.2)	22.6 (23.1)	18.0 (21.7)
	Width:depth (m)	27.5 (18.4)	38.8 (27.9)	45.2 (31.6)
	Stream temperature (°C)	12.6 (2.4)	11.7 (2.3)	11.6 (2.1)
Moderators	Elevation (m)	1500.3 (389.7)	1560.4 (459.7)	1241.6 (403.6)
	Erosivity (unitless)	122.6 (73.3)	142.8 (81.4)	177.4 (80.2)
	LWD (#/100 m)	141.9 (146.6)	219.4 (231.4)	315.7 (263.9)
	Stream size (m)	3.72 (1.68)	5.74 (2.85)	8.09 (3.89)
	Slope (%)	29.85 (10.28)	35.69 (10.91)	39.65 (10.05)
Covariates	Flow (cfs)	50.1 (52.7)	125.4 (156.7)	253.2 (225.8)
	Gradient (%)	2.1 (1.2)	1.9 (1.2)	1.8 (1.0)
Response	O/E (unitless)	0.7 (0.24)	0.87 (0.22)	0.93 (0.21)

higher width:depth ratios and lower fine sediment levels and streamwater temperatures for the mesic sites, as compared to sites in the medium and arid precipitation groups (Table 3).

Thresholds for O/E were developed to create biological condition classes, typically used for biological status assessment reporting for managers. Using the accuracy and precision of the reference site data (mean = 1.03, standard deviation = 0.17), managed sites scoring less than one SD below the mean of reference sites were in "good" biological condition (i.e., comparable to reference conditions); managed sites scoring between one SD and two SD are in "fair" biological condition; and managed sites scoring more than two SD below the mean of reference sites are considered in "poor" biological condition (e.g., Ode et al., 2008). O/E was inversely related to the precipitation groups. The largest percentage of sites in poor condition were in the arid group, 49%, compared to 19% in medium, and only 11% in the mesic group (model building datasets). Conversely, the mesic group had 75% in good condition, and the medium group had 59%, whereas only 33% were in good condition in the arid group (model building datasets).

In general, the signs of the estimated edge coefficients in the causal models were consistent with the a priori hypotheses when detected for all three precipitation groups. However, the linkage structure in the reduced graphs and their associated estimates varied across the three precipitation groups (Fig. 3 and Table C1 Appendix C). Across all three precipitation groups, the maximum possible variance that could be explained in O/E scores by the stressors ranged from 57 to 59%. Thus, considering such constraints, stressors in the arid and mesic models accounted for a significant portion of the variability in biological condition (Fig. 3a and c). The variance explained in streamwater temperature and fine sediment by drivers and covariates was reasonable for medium and mesic sites (from 23% to 35.4%, Fig. 3). However, percent variation explained in fine sediments and streamwater temperature was low in the arid group (6.8% and 14.26%, respectively Fig. 3a).

The supported causal hypotheses related to management impacts were variable across the precipitation groups (Table 4). The causal hypothesis that grazing impacts biological condition by altering fine sediments was supported only in the medium group, although the biological effect size was minimal (Table 4). The effect of grazing on reach width to depth ratio was not detected in any of the precipitation groups (Fig. 3). Although the link between width to depth ratio and stream temperature was maintained in all groups, our quantification of grazing impacts was not the causal driver of these changes through our hypothesized trampling effect (Fig. 3, Table 4). We found evidence that road density within a catchment negatively impacts condition by increasing fine sediments only in the medium precipitation group, but again minimal biological significance (Table 4). The hypothesis that % forest is a measure of potential supply of large woody debris in streams was supported in the medium group; however, the negative effect size for the arid group was contrary to expectations (Table 4). There was weak evidence that increasing amount of forested habitat benefited biological condition through decreasing instream temperature within the arid and medium group (Table 4). In general, the biological effect sizes for the direct effects of grazing, road density, or % forest on fines and stream temperature were minimal or not detected (Appendix C Table C1) leading to the weak indirect effect estimates of management activities on O/E (Table 4).

In terms of only mediator relationships, there was evidence that increasing fine sediments resulted in deviations of macroinvertebrates assemblages from expected conditions, albeit weak effect sizes for all groups (Fig. 3, Table C1 Appendix C). The largest sediment effect was detected in arid sites; we estimated 5% species loss with a doubling of fine sediments within a stream (90% CDI 2.5–7% loss). The link between stream temperature and O/E was maintained in both lower precipitation groups, but not in the wettest group (Fig. 3).

The moderators were important to consider for the medium and mesic causal graphs, but not for the arid graph (Table 4). For example in both reduced graphs (Fig. 3b and c), the direct linkage between grazing and fine sediments would have been removed if the moderators were ignored (Table 4). In the medium graph, although biologically weak, there was support for the hypothesis that increased amounts of LWD in a stream should result in less fine sediment accumulations in erosional habitats and less deviations from reference conditions (Table 4). In the mesic sites there was evidence of LWD moderating the entire pathway from grazing to O/E mediated by fine sediments (Table 4); however, the indirect effect of grazing on O/E's credible interval included zero (90% CDI −0.0039 to 0.005), likely due to the small effect size for the direct link between % grazing to fines. We found weak evidence that for lower elevation sites, increasing forest cover mitigates departures from reference conditions, as hypothesized, in the

Table 4
Summary of statistical support for causal pathways and moderation effects for three different precipitation groups. 90% credible intervals for estimated indirect effects of an increase in the driver nodes (10% increase for grazing and forest; 1 km/km^2 increase for road density) on O/E or LWD. NS indicates no statistical evidence for causal pathway in the graph. The moderator variables included in the reduced model are listed, but if there were none we denote it as none. ^ LWD (large woody debris) moderated the entire pathway from % grazing to O/E in the mesic group; however, the indirect effect was essentially zero. In medium sites, effects for low and high values of moderator variables are reported. Effects on O/E are interpreted in terms of % species loss or gain compared to least disturbed conditions.

Causal pathways	Arid	Medium	Mesic
% Grazing → fines → O/E	NS	LWD (154 pieces per km)[−0.1%, 0.2%] LWD (568 pieces per km) [0.002%, 0.2%]	NS
% Grazing → width to depth ratio → stream temperature → O/E	NS	NS	NS
Road density → fines → O/E	NS	[−1.8%, −0.08%]	NS
% Forest → LWD	[−0.261, −0.032]	[0.012, 0.029]	NS
% Forest → stream temperature → O/E	[0.5%, 1.6%]	Elevation (1120.2 m)[0.7%, 2%] Elevation (1840.1 m)[0.3%, 1%]	NS
Moderation effects			
% Grazing → fines	None	LWD	Slope, LWD^
Fines → O/E	None	None	LWD^
% Forest → LWD	None	None	None
% Forest → stream temperature	None	Elevation	Elevation
Road density → fines	None	None	None

medium sites (Table 4). However, in the arid sites elevation was simply a covariate for stream temperature, not a moderator.

In all three groups, the test of mediation suggested that the proposed stressors retained in the reduced models fully mediated the anthropogenic drivers of % grazing, road density, and/or % forest (Fig. 3). Further there was no evidence of alternative pathways for the drivers to impact biological condition. In other words, there was no need to include an additional link from the drivers directly to O/E in any model (Fig. 3).

3.1. Model assessment

In terms of assessing the local fit of the models, the median prediction errors for O/E varied with over- or under-predicting by approximately 0.30 units (Appendix C Table C2). In terms of classifying a site within a biological condition class based on their posterior predictive distributions, this translated to all sites being classified as good condition for the medium and mesic groups (39% and 33% misclassified, respectively Appendix C Table C3). For the arid sites, the true positive rate for poor condition sites was 90% and good condition sites was 57%, but there was no discrimination of fair sites. These patterns were consistent with the larger R^2 for O/E in the arid group (~18%).

Stream temperature predictions using % forest, elevation, and width to depth ratio compared to the true observed value varied with over- or under- predicting by about 3 °C; on average the prediction error was minimal (median close to zero). Fine sediment prediction performance was similar in the medium and mesic precipitation groups. The median predictive error was 0.09 for the medium group and 0.04 for the mesic sites, on the untransformed scale over-predicting by a factor of 1.09 and 1.04 (Appendix C Table C2). These are reasonable and consistent with the larger R^2 for fine sediments in both models. The arid sites had the largest median difference in fine sediment between the validation set values and the means of their respective posterior predictive distributions resulting in under-predicting by a factor of 0.89 on the un-transformed scale, but this is not surprising given the low R^2 (6.8%) based on just flow and gradient as parents (covariates) (Appendix C Table C2).

4. Discussion

There is an inherent tension within federally mandated monitoring programs such as PIBO to balance the need for robust statistical inferences about status and trends in resource condition at regional and national levels with the desire to understand mechanisms that lead to observed degradation or improvement. Previous analyses with PIBO data have largely focused on habitat metric differences in managed versus reference sites (Al-Chokhachy et al., 2011; Kershner et al., 2004). These correlative statistical techniques are insufficient when trying to understand how or why management intensity influences biological condition. Here, we used causal models to evaluate whether the data gathered for status and trends monitoring can be used to quantify habitat mediated impacts of anthropogenic drivers on biological condition (O/E) within managed sites only. This type of interim assessment that links the monitoring data back to the original causal hypotheses conveyed in the conceptual model could be the most critical step to ensure program longevity and management relevancy.

4.1. Ecological implications and suggestions for survey design adjustments

In our case, the lack of biologically significant results raises several questions: (1) is refinement of the hypotheses underlying the conceptual model needed? (2) Was our quantification of nodes (e.g., response design) in the causal model insufficient for characterizing subtle response gradients related to land use drivers? (3) Was the sampling design for status and trend estimation not suitable for elucidating causal mechanisms? We address these questions and provide suggestions for refinements to the response and sampling design that will help close the apparent inferential disconnect between monitoring status and trends and the need for statistical inferences regarding anthropogenic impacts on biological condition.

Our causal hypotheses were based on decades of both experimental and observational studies (see citations in Table 1). However, these were short-term, small sample size research studies conducted at reach-level over smaller spatial extents, where biologically relevant effect sizes are more easily detected (Clary, 1995; Kauffman et al., 1983; Saunders and Fausch, 2007), not watersheds the size of the interior Columbia River Basin as sampled by PIBO. For example, fine sediment is among the most ubiquitous aquatic stressors in the U.S. (Paulsen et al., 2008) known to have adverse effects on benthic assemblages, and grazing and roads can increase sediment loading to stream systems (reviewed in Wood and Armitage, 1997). Based on our empirical evaluation, the signs of the estimated effects matched our a priori causal hypotheses for this and other attributes, which lend credibility to the conceptual model. We also found evidence of the importance of large woody debris in mitigating the adverse impacts of excessive fine sediment loading associated with grazing on macroinvertebrates. Consequently, we remain confident in the causal hypotheses conveyed within the conceptual model. However, our analyses indicate the importance of setting realistic expectations for monitoring versus research study designs. An omnibus survey design for regional status and trend estimation and explanatory and/or predictive modeling of spatial variability in ecological indicators as related to anthropogenic factors is likely unrealistic, as our results showed.

Substantial effort was placed on developing biological indicators (in our case O/E) and refinement of protocols to minimize measurement errors of stressors (stream temperature and sediments) and riparian habitat conditions (Roper et al., 2010; Whitacre et al., 2007); however, little emphasis or resources were allocated toward quantifying drivers (e.g., grazing). For example, the grazing metric, the percent of a catchment contained within a grazing allotment, was likely quite crude, as it did not quantify the time since an allotment was last grazed nor frequency, timing, duration, or intensity of grazing. Such factors have repeatedly been shown to result in differential rangeland conditions and aquatic responses to grazing (Briske et al., 2008; Sarr, 2002). Similarly, the road node was included to directly quantify impacts associated with roads (e.g., increased sediment supply and transport to streams), but also as a surrogate for the extent of other anthropogenic activities occurring within a catchment (e.g., timber harvest, oil and gas extraction, recreation). We relied on roads as a surrogate of management actions because of the paucity of consistent land use information across the study area. Given the known impacts of these land uses, our analyses indicate the need for an adjustment to our measurement of driver nodes (i.e., grazing and roads), as well as additional consideration of other drivers such as recreation and resource extraction. Better information describing land use intensity will allow managers to understand which land uses are having deleterious effects on aquatic resources, the mechanisms by which degradation occurs, and whether adaptive management practices (e.g., decommissioning roads, regulating recreation, timber harvest and grazing) result in improvements to biological condition across the landscape.

In addition to the quality of data layers used to quantify driver nodes, summarizing road density and percent grazing within a

catchment might have obscured relationships with both stressors and biological responses. Specifically, contemporary forest and rangeland management have heavily emphasized protection of riparian buffers, which can mitigate effects of upslope activities on instream conditions (reviewed in Naiman and Decamps, 1997). For example, when predicting spatial variability in the biological condition of streams throughout the eastern U.S., Carlisle et al. (2009) found estimates of anthropogenic land uses within a riparian buffer to be better predictors than within catchment estimates. We explored whether using estimates of grazing and road density within a 90 m buffer improved our inferences for the arid sites, but there was little evidence of a relationship between density of roads and O/E (correlation = −0.004) and we failed to find a meaningful difference in average O/E between grazed and ungrazed reaches. Therefore, the choice of catchment-level drivers in our models did not appear to affect our findings, so much as how these drivers were quantified.

While higher quality, spatially explicit disturbance information might improve inferences concerning land use activities, we also believe that quantification of stressor and biological response nodes themselves could be improved. The more than 40% of variance unexplainable by any stressor-driver model highlights the need for improvements in the biological response model used to create the O/E metric and/or macroinvertebrate sampling and processing. Additionally, fine sediment, which is one of the most ubiquitous stressors and known to have deleterious effects on benthic organisms (Wood and Armitage, 1997), has a measurement error exceeding 20% among repeated field crew visits (Roper et al., 2010; Whitacre et al., 2007). Our measurement of fine sediments was within pool-tails and macroinvertebrate sampling was within riffle habitats in a reach. This spatial misalignment is likely not an issue in the case of large fine sediment increases; however, in our case, we estimated very subtle changes in fine sediment as a result of altering % grazing within a catchment (Table C1 in Appendix C). This suggests a potential need for co-location of fine sediment and macroinvertebrate sampling within the same habitat and field strategies to minimize controllable sources of error, both of which could maximize the signal between stressors and biological responses.

In fact, an important consideration for interpreting our results relative to previous research studies is the observed gradients of drivers, stressors, and biological condition (Table 3; e.g., Smart et al., 2012). Across the three precipitation groups a range of biological conditions was found, with 49% of arid sites classified as being in 'poor' biological condition, compared to only 19% and 11% in medium and mesic groups, respectively. Thus, it was not surprising that the hypothesized stressors of stream temperature and instream fine sediments were useful for explaining variation in O/E for the arid group. Conversely, average water temperatures were below 12 °C and fine sediment levels averaged less than 23% in the medium and mesic precipitation groups. Such physical conditions are typical of relatively unaltered, forest and rangeland systems and unlikely to elicit significant biological responses (Bryce et al., 2010). Low levels of various stressors and the predominance of intact biological assemblages are likely to be emblematic of many public lands where vast tracks of land are in relatively good condition, most stressors exist as non-point source pollution, and change in resource condition occurs at relatively long timescales (Harding et al., 1998). The subtle gradients in stressors we observed suggest the need to refine the sampling design.

The PIBO sampling design was a spatially balanced design (GRTS) without explicit stratification. Use of such a design over a large geographic region (740,000 km^2) resulted in sampling of streams encompassing substantial environmental heterogeneity that if unaccounted for in a statistical model may mask subtle relationships of interest. We addressed this analytical challenge by subsetting the sites into precipitation groups (post-stratification), and we included covariates to account for the inherent natural environmental heterogeneity in the data. Also, we incorporated moderators to model the context dependency of anthropogenic impacts on biological condition. For example, inclusion of large woody debris as a moderator elucidated context-dependent relationships between grazing and fine sediments in both medium and mesic precipitation groups. Even though we harnessed these statistical modeling options, the fact remains the quality of driver data and possibly the subtle gradients sampled using an unstructured sampling design were insufficient to quantify causal mechanisms for stream degradation.

Sampling designs need to be flexible and allow for precise and unbiased estimates for multiple indicators (Overton and Stehman, 1996; Sergeant et al., 2012). Even though anthropogenic drivers (e.g., grazing or road density) might be of interest to a long-term monitoring program, defining strata by levels of a driver variable could lead to analytical and inference issues because strata membership will likely vary over the course of decades as policy and land ownership change (Johnson, 2012; Mahan et al., 2007; Overton and Stehman, 1996). An alternative, in the absence of the ability to perform experimental manipulations or stratified sampling, is to develop a complementary effort that targets sampling gradients of interest that represent the causal mechanisms of primary concern for resource management (e.g., Burton et al., 2014). As the original sites were selected using GRTS, one tractable option is to reduce the sample size from each panel using the GRTS ordered list and in turn using those resources "saved" to conduct a more targeted and refined assessment. The adjusted panels will still have desirable spatial balance properties and we utilize one of the key benefits to the GRTS sampling design which is the ability to accommodate increases and decreases in sample size (Olsen et al., 2012). Within this supplemental effort the primary objective is to thoroughly investigate the potential impacts of road density and grazing with local spatially explicit characterization of the drivers and co-location of sediment and macroinvertebrates. This approach maintains the current program while adapting to the emerging needs identified by our analysis for monitoring to be management relevant.

5. Conclusions

A key component within the adaptive monitoring paradigm is to learn based on statistical analyses of available monitoring data. There are few, if any, documented examples of an empirical evaluation of a regional monitoring program's conceptual model, as we have done for PIBO. There are examples of retrospective power analyses using data collected for status and trend detection (Mahan et al., 2007; Levine et al., 2014), but these do not address whether the data can be used to understand why a resource is degrading or improving over time. Given the significant financial investment required of long-term, multi-state monitoring programs (e.g., Henderson et al., 2005; Moyer et al., 2003), ensuring that management relevant drivers are characterized sufficiently and can be used in analyses with measured stressors and biological indicators would significantly increase the ecological value of a program and ensure its longevity.

We feel the utility of conceptual models as developed for monitoring programs can be increased by recognizing they inherently provide testable complex causal hypotheses. Causal modeling provides a flexible analytic framework for assessing these proposed direct, indirect, and cumulative effects of drivers and stressors on ecological indicators. Further, if our estimates based on observational data had yielded convincing biological effects this framework could be used for predictions and

assessments of different interventions to which future observations could be compared within an adaptive management context. Arguably, the complexity of the statistical inference methods presented here may hinder the rapid feedback needed in the practice of adaptive monitoring. However, Eqs. (1)–(5) could be estimated using separate multiple linear regression models and graphical techniques could be used to assess the strength of associations between connected components in the causal graph. If no associations are apparent, this could suggest a similar issue may be encountered as in our investigation.

In our case empirical evaluation of the conceptual model allowed us to identify gaps in our monitoring of context-dependent mechanisms of how or why management may impact important conservation targets. We suspect many regionally and nationally mandated monitoring programs would benefit from this type of interim assessment and, as a result, might find similar deficiencies that, if not addressed, would hamper moving from surveillance into more question-based monitoring.

Acknowledgements

We would like to thank Tom Mendenhall and Karl Stein (BLM) for their original ideas and funding and S. Wenger (USFS/Trout Unlimited) for assistance with hydrologic modeling outputs. We appreciate Joel Reynolds' constructive comments and suggestions for reduction of material and James Grace for encouraging the first author to persevere with causal modeling. We thank Tom Rodhouse for his review and suggestions, Siri Wilmoth for her formatting assistance, and two anonymous reviewers for their constructive comments. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Introduction to causal modelling

Recently, Grace et al. (2012) provided guidance for a graph-theoretic approach to structural equation modeling and Shipley (2013) provided examples of this approach using path analysis. A graph is composed of two components: the nodes or random variables (also, referred to as vertices in statistical literature) and the linkages or edge structure among nodes. Relationships between nodes can be described using family tree terminology (e.g., ancestors, parents, children, and descendants). In a causal graph, links are directed (represented by an arrow). A directed edge implies changing the value of a parent node (arrow originates from this node) will change the child node value (arrow points towards this node), a cause and effect (Table A1 provides definitions). For example, in Fig. A1 the parent node of the stressor (the child) is an anthropogenic driver. A path is an ordered sequence of links or edges (such as in Fig. A1 driver → stressor → response). Graphs that contain no cycles (pathways that begin and end at the same node) and only directed edges are directed acyclic graphs (DAGs) or a causal graph. The pathways represented in the DAG are causal hypotheses.

For statistical modeling, the functions or probability distributions used to test the hypotheses conveyed by the graph are specified—a causal model. The complete specification of a causal model is composed of sets of nodes, linkages, and functions (Grace et al., 2012; Pearl, 2009). The casual graph (DAG, Fig. A1) has a local Markov property such that we can use local estimation as opposed to fitting a joint multivariate distribution for the graph [denoted as $P(X)$, typically assumed multivariate Gaussian]. Essentially, we can decompose the joint distribution into the product of conditional distributions as follows $P(X) = \prod_i P_i[X_i|pa(X_i)]$, where P_i is a probability distribution (Guassian, Poisson, etc.), X_i is a node in the graph and $pa(X_i)$ is the set of parents for node X_i (see Grace et al., 2012 for a translation of Pearl, 2009). Rather, the local Markov property implies that a variable (node) is conditionally independent of its non-descendants (excluding its parents), given its parents. In Fig. A1, the joint distribution of {driver, stressor, response, moderator, and covariate}can be decomposed into two conditional distributions: (1) response given stressor, and (2) stressor given driver, moderator, and covariate. A key statistical difference between multiple linear regression and causal models is the assumption of only one response or dependent variable is relaxed in the latter. For example in Fig. A1, stressor and response could be considered dependent variables as they both have a set of parent nodes (Table A1). As a result, complex hypotheses (network of predictors) can be conveyed within a causal model providing more than a simple correlative relationship between two variables. Finally, the parameters in the conditional distributions can be estimated using appropriate statistical models.

One of the important aspects of graphical models is the ability to estimate and operationalize indirect effects via mediator variables (Grace et al., 2012). Mediator variables describe the mechanism or process of how or why an independent variable such as an anthropogenic driver (land use) affects a dependent variable (biological condition) (MacKinnon et al., 2012; Wu and Zumbo, 2008 Wu and Zumbo, 2008; see Table A1 definitions). For motivation, consider potential mediators (e.g., fine sediment concentrations, temperature, or nutrients) for how or why timber harvest and grazing degrade macroinvertebrate biological integrity (Banks et al., 2007; Belsky et al., 1999; Beschta, 1997; Tait et al., 1994). In forested, headwater streams where the predominate human influences are grazing and forest management, surface alterations, such as increased road density, can degrade

Table A1
Definitions for different roles nodes can serve in causal graph. Examples refer to Fig. 3.

Variable type	Definition	Examples
Parent node	An arrow originates from this node	Stream temperature and fines are parent nodes of O/E
Child node	An arrow terminates at this node	O/E is the child node of stream temperature and fines
Root node	No edges directed towards them, only away	% Grazing, road density, and % forest
Terminal node	Only edges directed towards them, and no edges pointing away from them.	O/E
Mediators	Describe the mechanism or process of how or why an independent variable effects a dependent variable	Fines, LWD, stream temperature
Moderators	Contextual variable that modifies the direction or strength of the directed edge; it clarifies the when, where, or for whom	Erosivity, slope
Covariates	A covariate is associated with a dependent variable and accounts for natural variation	Flow, gradient
Confounders	A confounder variable is associated with both a parent and child variable.	Spatial trend in both a parent and child node

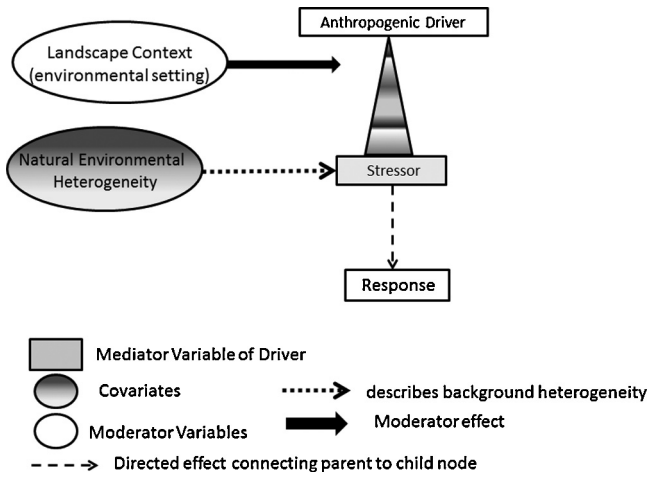

Fig. A1. Illustrative graph of a conceptual model for the indirect effect of a management action (anthropogenic driver) on a response (e.g., measure of biotic integrity). Moderators (landscape context at larger scale) may alter the direction and amplitude of the estimated causal path linking driver to stressor. The causal graph incorporates environmental heterogeneity by including covariates (natural environmental heterogeneity at local scale) on the mediator variable (stressor).

macroinvertebrate assemblages by altering runoff patterns and increasing fine sediment loading. A test of mediation investigates whether there is evidence for an additional direct pathway for an anthropogenic driver (e.g., grazing) to impact biological condition in the causal model. If data support such an edge (link) inclusion, it would suggest an alternative mechanism beyond the proposed mediator for how an anthropogenic driver impacts the response.

A challenge for analyzing data from regional monitoring programs is differentiating between natural and anthropogenic gradients when assessing resource condition. This complication arises due to the myriad of landscape and climate factors operating at different spatial and temporal scales that can mask effects of local management practices (Allan, 2004; Cuffney et al., 2010; Frissell et al., 1986; Larsen et al., 2004). In causal modeling, inherent environmental heterogeneity (geology, climate, topography) can be modeled using moderator variables (Mackinnon et al., 2012; Wu and Zumbo, 2008). A moderator is a contextual variable that modifies the direction or strength of the directed edge; it clarifies the when, where, or for whom (Wu and Zumbo, 2008; Table A1). For monitoring programs, moderator effects are critical to consider as the environmental setting or landscape context could result in differential responses to anthropogenic drivers by either dampening or perhaps amplifying the effect of the driver on the stressor, potential mitigating factors. One approach is to construct homogeneous subsets of data to better detect human impacts on biological condition (macroinvertebrate assemblage; e.g., Riseng et al., 2011). We propose another way to handle this inherent heterogeneity when of ecological interest is to include moderator variables in a causal modeling framework. Additionally, covariates can be included that help describe natural background environmental variation in either response or stressors. We used moderator variables to contextualize potential mechanisms by which anthropogenic drivers (primarily, grazing, timber harvest, and road density) in forested watersheds influence the biological condition of streams as quantified by macroinvertebrate assemblages.

Another attractive property of causal models are the ability to estimate cumulative or total effects of a driver on a response through multiple mediated and possibly moderated causal pathways. For example, in Fig. 3 the total effect of grazing on O/E is found by tracing all paths from the % grazing node to the O/E node. There are two main pathways for grazing to impact O/E, one is % grazing → width:depth ratio → stream temperature → O/E, using Eqs. (1), (4), and (5) this indirect path is estimated by

$$\theta_1 \delta_4 \beta_2 \tag{A.1}$$

The other pathway is mediated by the fine sediment node (fines); the path is % grazing → fines → O/E. However, there are potentially three moderators, erosivity, slope, and LWD on the link between % grazing → fines and LWD also moderates the link between fines → O/E. To estimate the effect of a X-unit change in grazing along this complicated pathway is

$$(\beta_1 + \beta_4 \text{LWD})(\alpha_1 + \alpha_8 \text{slope} + \alpha_9 \text{erosivity} + \alpha_{12} \text{LWD})X, \tag{A.2}$$

given fixed values of the moderators. These estimated effects assume all other variables in the equations are held constant, i.e., road density, flow, and gradient. The total effect is then the sum of these two indirect paths from % grazing to O/E, Eqs. (A.1) + (A.2). Bayesian estimation provides for easy calculation of posterior intervals of these derived parameters.

Causal models, in general, provide a flexible alternative for ecological modeling of heterogeneous data to address hypothesis driven investigations. The local estimation approach is highly versatile in terms of the probability models specified for the conditional distributions of child nodes. For our analysis, we assumed the directed edges were consistent with a linear relationship with the mean in the child (response) node. Based on residual and other diagnostic plots our data appeared consistent with these assumptions and a Gaussian distribution for the error terms. However, our measure of biological condition could have been modeled as a multinomial variable with three condition classes (poor, fair, or good; Carlisle et al., 2009). Also, others have modeled biological response to fine sediment using quantile regression (Bryce et al., 2008). Importantly, these other probability models could be used for conditional distributions within a causal modeling framework (see Grace et al., 2012; Shipley, 2009). Further, if spatial autocorrelation were an issue, these models can be extended to model spatial correlation in both response and parent nodes (Irvine and Gitelman, 2011). This flexibility within the causal modeling framework is a clear benefit for modeling ecological data with the goal of evaluating complex hypotheses.

Appendix B. Field methods

Sample sites were visited during baseflow conditions June–September to minimize sampling error and facilitate safe wading conditions. Sample reach lengths were determined as a function of bankfull width (20×) with a minimum sampling length of 160 m, to increase the likelihood that multiple geomorphic channel units were sampled (Knighton, 1999). At each sample site we collected streamwater temperature data, physical habitat data, and macroinverterbrate samples. We also used GIS data derived for each catchment.

Within stream — we estimated the percent fine sediment in pool tails using a grid-sampling (0.36 m × 0.36 m) approach from the first 10 pools in the reach, if less than 10 pools occurred all pools were sampled. Each grid included 50 equally-spaced intersections, and we collected grid data at three equidistant locations across the wetted width of the lowermost 10% of each pool habitat. We enumerated the number of intersections that overlaid particles with an intermediate axis less than 6 mm, divided by the total number of intersections in the grid; we then calculated mean percent fines for each pool and averaged pool estimates for a site-level estimate of fines sediments ("fines", Fig. 3). Finally, we quantified the amount of large woody debris ("LWD", Fig. 3) in each site. We enumerated all pieces of LWD

>0.1 m in diameter and longer the 1 m in length that were at least partially within the bankfull channel and normalized the counts to pieces/km. Bankfull widths are the average of the measured bankfull widths at each transect ($n \geq 21$; "stream size" in Fig. 3). Our measure for width to depth ratio ("width:depth ratio", Fig. 3) is the average of the wetted width divided by the water depth from 10 cross sections taken at even numbered transects 2–20. If there were <6 suitable measurements no value is reported, these records were not included in the analyses.

We collected streamwater temperatures at 60 min intervals using Hobo data loggers from July 15th to August 31st and calculated the average temperatures at each site for our analyses (degrees celsius; "stream temperature" in Fig. 3).

Benthic macroinvertebrate assemblages were sampled to quantify stream biological integrity following the methods of Peck et al. (2006). Macroinvertebrate samples were preserved and 600 random organisms per sample (Caton, 1991; Vinson and Hawkins, 1996) were identified to the lowest taxonomic resolution possible, generally genus. Prior to analysis, we standardized the taxonomic resolution among all samples by assigning identified macroinvertebrates to unambiguous operational taxonomic units (OTUs) (Cuffney et al., 2007) and counts were standardized to 300 randomly selected individuals per sample.

To compute biological integrity, we constructed an observed/expected (O/E) type predictive model following the procedures outlined in Hawkins et al. (2000) and Hawkins and Carlisle (2001). O/E values compare the actual sampled taxa (O) to the taxa predicted to occur (E) if the site was in reference condition (Wright, 2000). The O/E model was developed using 201 reference sites (174 calibration and 27 validation sites) randomly located throughout the Interior Columbia River Basin and sampled over an eight year timeframe ("O/E" in Fig. 3). The final model had acceptable accuracy and precision (0.95 and 0.16, respectively) (Ostermiller and Hawkins, 2004; Hawkins, 2006) and effectively discriminated between reference and managed sites ($t = -12.57$; $df = 634$; $p < 0.001$).

Table C1
Posterior quantiles for edge coefficients based on reduced graphs [Fig. 3 black colored edges only], excluded edges with 90% credible intervals that included zero. S. Temp. denotes the Stream Temperature node in Fig. 3. Detailed field methods for each variable are provided in Appendix B.

Arid sites	Posterior quantiles					
Parent	Child	2.50%	5%	50%	95%	97.50%
S. Temp.	O/E	−0.053	−0.050	−0.036	−0.021	−0.018
Fines	O/E	−0.106	−0.100	−0.069	−0.037	−0.030
% Forest	S. Temp.	−0.041	−0.039	−0.028	−0.017	−0.014
Elevation	S. Temp.	−2.060	−1.926	−1.128	−0.320	−0.155
Width:depth ratio	S. Temp.	0.011	0.014	0.030	0.046	0.049
Gradient	Fines	0.011	0.012	0.021	0.029	0.031
Flow	Fines	−2.351	−2.231	−1.606	−1.010	−0.884
% Forest	LWD	−0.284	−0.261	−0.146	−0.032	−0.008
Elevation	LWD	−0.011	−0.010	−0.007	−0.004	−0.003
Medium sites	**Posterior quantiles**					
Parent	Child	2.50%	5%	50%	95%	97.50%
S. Temp.	O/E	−0.039	−0.037	−0.028	−0.018	−0.016
Fines	O/E	−0.041	−0.038	−0.020	-0.002	0.001
% Forest	S. Temp.	−0.047	−0.045	−0.033	−0.020	−0.018
Elevation	S. Temp.	−3.031	−2.946	−2.471	−2.017	−1.925
% Forest × elevation	S. Temp.	0.004	0.007	0.029	0.051	0.056
Width:depth ratio	S. Temp.	0.012	0.013	0.020	0.026	0.028
% Forest	LWD	0.010	0.012	0.020	0.029	0.031
Elevation	LWD	−1.470	−1.401	−1.061	−0.716	−0.651
% Grazing	Fines	0.000	0.001	0.003	0.006	0.006
Road Density	Fines	0.049	0.069	0.175	0.274	0.295
Gradient	Fines	−0.346	−0.330	−0.245	−0.163	−0.147
Erosivity	Fines	0.134	0.161	0.290	0.418	0.439
Flow	Fines	−0.004	−0.003	−0.003	−0.002	−0.002
LWD	Fines	−0.047	−0.033	0.037	0.106	0.118
LWD × % grazing	Fines	−0.006	−0.006	−0.004	−0.002	−0.002
Mesic sites	**Posterior quantiles**					
Parent	Child	2.50%	5%	50%	95%	97.50%
Fines	O/E	−0.0720	−0.0687	−0.0468	−0.0249	−0.0213
LWD	O/E	0.0129	0.0191	0.0487	0.0783	0.0846
LWD × fines	O/E	0.0024	0.0066	0.0259	0.0452	0.0491
% Forest	S. Temp.	−0.0570	−0.0510	−0.0230	0.0058	0.0121
Elevation	S. Temp.	−3.3881	−3.2441	−2.4585	−1.6840	−1.5099
% Forest × elevation	S. Temp.	−0.0069	0.0006	0.0429	0.0831	0.0899
Width:depth ratio	S. Temp.	0.0035	0.0050	0.0163	0.0269	0.0290
% Grazing	Fines	−0.0039	−0.0032	0.0006	0.0043	0.0050
Slope	Fines	−0.0535	−0.0499	−0.0334	−0.0169	−0.0140
Gradient	Fines	−0.6099	−0.5843	−0.4272	−0.2773	−0.2432
Flow	Fines	−0.0034	−0.0032	−0.0025	−0.0018	−0.0016
LWD	Fines	0.0576	0.0893	0.2408	0.3862	0.4179
% Grazing × slope	Fines	−0.0011	−0.0010	−0.0006	−0.0003	−0.0002
LWD × % grazing	Fines	−0.0082	−0.0076	−0.0043	−0.0012	−0.0004

Table C2
Quantiles for model based prediction errors for validation dataset by precipitation group ($X_{\text{valid},i} - \hat{X}_{\text{model},i}$),where $\hat{X}_{\text{model},i}$ is the mean of the posterior predictive distribution and $X_{\text{valid},i}$ is the observed value of variable X for site i. The variables to the right of the vertical line enclosed in parenthesis were those used to predict the variable on the left of the vertical line. The different sets of variables correspond to using the measured values of the parents versus predicting the values of the parents using the observed values of the root nodes (refer to Fig. 3). S. Temp. denotes the stream temperature node in Fig. 3. Detailed field methods for each variable are provided in Appendix B.

Arid sites	5%	50%	95%
O/E | (S. Temp., fines)	−0.34	−0.05	0.28
O/E | (flow, gradient, width:depth ratio, elevation, % forest)	−0.40	−0.06	0.34
S. Temp.| (elevation, width:depth ratio, % forest)	−3.87	−0.11	3.99
ln(fines)| (flow, gradient)	−2.18	−0.12	1.28
ln(LWD)| (elevation, % forest)	−4.74	−0.41	2.75

Medium sites	5%	50%	95%
O/E | (fines, S. Temp.)	−0.35	0.07	0.30
O/E |(flow, gradient, erosivity, % grazing, road density,LWD, width:depth ratio, elevation, % forest)	−0.34	0.08	0.33
S. Temp.| (elevation, % forest, width:depth ratio)	−2.58	−0.05	3.09
ln(fines)| (% grazing, LWD, road density, flow, gradient, erosivity)	−1.40	0.09	1.47
ln(fines)| (% grazing, elevation, % forest, road density, flow, gradient, erosivity)	−1.50	0.11	1.61
ln(LWD)| (elevation, % forest)	−2.40	0.12	1.95

Mesic sites	5%	50%	95%
O/E | (fines)	−0.40	−0.04	0.22
O/E | (% grazing, LWD, slope, flow, gradient)	−0.39	−0.03	0.23
S. Temp.| (width:depth ratio, elevation, % forest)	−2.32	0.13	2.65
ln(fines)| (% grazing, LWD, slope, flow, gradient)	−1.11	0.04	1.54

Table C3
Prediction assessment for classifying unobserved sites as being in good (O/E ≥ 0.84), fair (O/E between 0.68 and 0.84), or poor biological condition (O/E < 0.68) based on observed values for parent nodes [O/E| (fines, stream temperature)] versus predictions based on only root nodes [O/E|root of parent nodes]. See Table A1 for terminology of nodes. Sample sizes for hold-out datasets were mesic $n = 52$, medium $n = 106$, and arid $n = 53$. Kendall's tau, a measure of concordance between predicted conditional means and condition class values, is reported in parenthesis (Agresti 2010).

Arid Precipitation Sites

		Predicted with parents (0.54)			Predicted with root variables (0.32)		
		Poor	**Fair**	**Good**	**Poor**	**Fair**	**Good**
Observed	**Poor**	26	0	3	28	0	1
	Fair	7	0	3	10	0	0
	Good	6	0	8	11	0	3

Medium Precipitation Sites

		Predicted with parents (0.25)			Predicted with root variables (0.09)		
		Poor	**Fair**	**Good**	**Poor**	**Fair**	**Good**
Observed	**Poor**	0	0	21	0	0	21
	Fair	0	0	21	0	0	21
	Good	0	0	64	2	0	62

Mesic Precipitation Sites

		Predicted with parents (0.19)			Predicted with root variables (0.14)		
		Poor	**Fair**	**Good**	**Poor**	**Fair**	**Good**
Observed	**Poor**	0	0	8	0	0	8
	Fair	0	0	9	0	0	9
	Good	0	0	35	0	0	35

Reach attributes — we quantified gradient as the change in the elevation of the water surface between the top and bottom of the reach, which we measured using a site level and tripod, and divided this elevation change by site-specific measures of river length ("gradient", Fig. 3). We calculated flow based on the variable infiltration capacity (VIC) model (Wenger et al., 2010). Briefly, the VIC model is largely physically-based and includes empirically-derived relationships for infiltration, runoff, and stream baseflow processes and allows for spatially-continuous streamflow predictions. Given the importance of stream power in controlling local habitat processes (Frissell et al., 1986; Beschta and Platts, 1986; Al-Chokhachy et al., 2011), we included average historic (1915–2006) measures of bankfull discharge at each site in our analyses ("flow" in Fig. 3).

Catchment attributes — we initially delineated the catchment upstream of each sampled site using digital elevation models (10 m DEMs; http://www.ned.usgs.gov) and Arc Macro Language in ArcGIS 9.2 (ESRI 2008). Using these delineations, we quantified differences in soil characteristics by computing a continuous measure of the uniaxial compressive strength of each lithology type ("erosivity" in Fig. 3; Cao et al., 2007); here, increasing values of erosivity positively correspond to the strength of the underlying lithology. Next, we calculated the average catchment slope based on the DEMs ("slope" in Fig. 3). Finally, we quantified the percent of the catchment as forested ("% forest" in Fig. 3, i.e., tree-dominated vegetation; LANDFIRE, 2008). Also, the elevation at the bottom of each stream reach was derived from a 10 m DEM (USGS National Elevation Dataset: http://seamless.usgs.gov/, "elevation" Fig. 3).

Additionally we quantified information regarding the level of anthropogenic disturbances in each catchment. First, we calculated catchment-specific measures of the density of roads ("road density", km/km^2, in Fig. 3.) from the USFS Geodata Clearinghouse (1:24000 scale; svinetfc4.fs.fed.us/clearinghouse/index.html). Next, we compiled grazing allotment boundaries from the USFS and BLM units within our study area and calculated the percentage that contained a grazing allotment ("% grazing" in Fig. 3).

Appendix C. Model estimates and predictive error quantiles

We explored how useful the posterior predictive distributions based on our final graphical models would be for classifying unobserved sites as good (O/E $\geq$ 0.84), fair (O/E between 0.68 and 0.84), or poor biological condition (O/E $<$ 0.68). Let Y_i denote the O/E score and C_i the condition class, where poor = 1, fair = 2, and good = 3 for site i. We calculated the probability for site i being in the biological condition class k, as Prob($C_i = k$) as follows, let $T_0 = -\infty$, $T_1 = 0.68$, $T_2 = 0.84$, and $T_3 = +\infty$, then $\text{Prob}(C_i = k) = \int_{T_{k-1}}^{T_k} f(\tilde{y}_i|y)d\tilde{y}_i$, where $f(\tilde{y}_i|y)$ is the posterior predictive distribution for O/E as a continuous measurement (Gelman et al., 1995). The models displayed poor predictive ability consistent with the weakly estimated linkages in all precipitation groups (Table C3).

Appendix D. Supplement winbugs code

Although this software program has been used by the U.S. Geological Survey (USGS), no warranty, expressed or implied, is made by the USGS or the U.S. Government as to the accuracy and functioning of the program and related program material nor shall the fact of distribution constitute any such warranty, and no responsibility is assumed by the USGS in connection there with.

```
# REDUCED MODEL FOR MESIC SITES
model{

#center variables and perform transformation outside of WinBUGS


            #measurement equation model or likelihood
            for(i in 1:N){

            Bugs[i]~dnorm(mu[i,1],psi[1])  #O/E Scores
            Temp[i]~dnorm(mu[i,2],psi[2])  #streamwater temperature
            fines[i]~dnorm(mu[i,3],psi[4])
            LWD[i]~dnorm(0,psi[3])
            graz[i]~dnorm(0,psi[14]) #grazing
            Grad[i]~dnorm(0,psi[5])  #gradient
            For[i]~dnorm(0,psi[6]) #% Forest
            Slope[i]~dnorm(0,psi[7]) #slope
            Elev[i]~dnorm(0,psi[8]) #elevation
            Rds[i]~dnorm(0,psi[9])  #road density
            Flo[i]~dnorm(0,psi[10]) #flow/ discharge
            Eros[i]~dnorm(0,psi[11])  #erosivity
            bf[i]~dnorm(0,psi[12])    #stream Size
            WDwet[i]~dnorm(0,psi[13])  #Width to Depth ratio

      #these are conditional distrib. on pa(Y_j)

      #assuming intercepts  are all zero since using centered variables


      mu[i,1]<-gam1[1]*fines[i]+gam1[2]*LWD[i]+gam1[3]*LWD[i]*fines[i]

      #O/E| fines, LWD, LWD*fines

      mu[i,2]<-gam2[1]*For[i] +gam2[2]*Elev[i] +
gam2[3]*For[i]*Elev[i]+gam2[4]*WDwet[i]

      #Stream Temp  | % forest Elev %for*Elev WDwet

      mu[i,3]<-gam4[1]*graz[i]+gam4[2]*Slope[i]+gam4[3]*Grad[i]+
             gam4[4]*Flo[i]+gam4[5]*LWD[i]+ gam4[6]*graz[i]*Slope[i]


                  +gam4[7]*LWD[i]*graz[i]

      # fines| grazing,  slope, gradient, flow, LWD, grazing*Slope, LWD*Graz


# RESIDUALS FOR MODEL DIAGNOSTICS

      res.BUGS[i]<-Bugs[i]-mu[i,1]
      res.Fines[i]<-fines[i]-mu[i,3]
      res.ST[i]<-Temp[i]-mu[i,2]



      } #end of i


      #PRIORS on edge/link coefficients

      for(i in 1:3){
      gam1[i]~dnorm(0,.01)}

      for(i in 1:4){
      gam2[i]~dnorm(0,.01)}

      for(i in 1:7){
      gam4[i]~dnorm(0,.01)}


      #PRIORS on precisions
      for(j in 1:14){
      psi[j]~dgamma(0.5, 1.0)
#manifest var. prec; this second parameter is the inverse of rgamma in R
            sgm[j]<-1/psi[j]
      }

#THESE NEED TO BE ON CENTERED SCALE FOR VALUES! Assuming Slope at its mean
value for mesic sites, so term cancels out

#10% change in grazing given high LWD (G1-G2) 10 is a increase (=569 LWDFreq)
#10% change in grazing given low LWD 25% quantile (=155 lwdfreq)
#LWD on log scale - mean value in mesic of 5.27

      Ind.Graz.loEqn7<-(gam1[1]+gam1[3]*-.23)*(gam4[1]+gam4[7]*-.23)*10

      Ind.Graz.hiEqn7<-(gam1[1]+gam1[3]*1.07)*(gam4[1]+gam4[7]*1.07)*10


            }
```

References

Agresti, A., 2010. Analysis of Ordinal and Categorical Data, second edition John Wiley and Sons, Inc., Hoboken, New Jersey.

Al-Chokhachy, R., Roper, B.B., Archer, E.K., Miller, S., 2011. Quantifying the extent of and factors associated with the temporal variability of physical stream habitat in headwater streams in the interior Columbia river basin. Trans. Am. Fish. Soc. 140, 399–414.

Allan, J.D., 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annu. Rev. Ecol. Evol. Syst. 35, 257–284.

Banks, J.L., Li, J., Herlihy, A.T., 2007. Influence of clearcut logging, flow duration, and season on emergent aquatic insects in headwater streams of the Central Oregon Coast Range. J. N. Am. Benthol. Soc. 26, 620–632.

Barbour, M., Gerrtisen, J., Snyder, B.D., Stribling, J.B., 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish, second ed. EPA, Washington, D.C EPA 841-B-99–002.

Barrows, C.W., Allen, M.F., 2007. Biological monitoring and bridging the gap between land management and science. Nat. Areas J. 27, 194–197.

Belsky, A.J., Matzke, A., Uselman, S., 1999. Survey of livestock influences on stream and riparian ecosystems in the western United States. J. Soil Water Conserv. 54, 419–431.

Beschta, R.L., Platts, W.S., 1986. Morphological features of small streams-significance and function. Water Resour. Bull. 22 (3), 369–379.

Beschta, R.L., 1997. Riparian shade and stream temperature: an alternative perspective. Rangelands 19, 25–28.

Briske, D.D., Derner, J.D., Brown, J.R., Fuhlendorf, S.D., Teague, W.R., Havstad, K.M., Gillen, R.L., Ash, A.J., Willms, W.D., 2008. Rotational grazing on rangelands: reconciliation of perception and experimental evidence. Rangeland Ecol. Manage. 61, 3–17.

Bryce, S.A., Lomnicky, G.A., Kaufmann, P.R., McAllister, L.S., Ernst, T.L., 2008. Development of biologically based sediment criteria in Mountain Streams of the Western United States. N. Am. J. Fish. Manage. 28, 1714–1724.
Bryce, S.A., Lomnicky, G.A., Kaufmann, P.R., 2010. Protecting sediment-sensitive aquatic species in mountain streams through the application of biologically based streambed sediment criteria. J. N. Am. Benthol. Soc. 29, 657–672.
Burdon, F.J., McIntosh, A.R., Harding, J.S., 2013. Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. Ecol. Appl. 23, 1036–1047.
Burns, J.W., 1972. Some effects of logging and associated road construction on northern California streams. Trans. Am. Fish. Soc. 101, 1–17.
Burton, A.C., Huggard, E., Bayne, J., Schieck, P., Solymos, T., Farr, D., Boutin, S., 2014. A framework for adaptive monitoring of the cumulative effects of human footprint on biodiversity. Environ. Monit. Assess. 186 (6), 3605–3617.
Cao, Y., Hawkins, C.P., Olson, J., Kosterman, M.A., 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. J. N. Am. Benthol. Soc. 26 (3), 566–585.
Carlisle, D.M., Falcone, J., Meador, M.R., 2009. Predicting the biological condition of streams: use of geospatial indicators of natural and anthropogenic characteristics of watersheds. Environ. Monit. Assess. 151, 143–160.
Carter, J.L., Hannaford, V.H., Myers, M.J., 2007. Macroinvertebrates as biotic indicators of environmental quality. In: Hauer, F.R., Lamberti, G.A. (Eds.), Methods in Stream Ecology. Academic Press, Waltham, Massachusetts, pp. 805–831.
Caton, L., 1991. Improved subsampling methods for the EPA Rapid Bioassessment benthic protocols. J. N. Am. Benthol. Soc. 8, 317–319.
Clary, W.P., 1995. Vegetation and soil responses to grazing simulation on riparian meadows. J. Range Manage. 48, 18–25.
Clough, Y., 2012. A generalized approach to modeling and estimating indirect effects in ecology. Ecology 93, 1809–1815.
Cubaynes, S., Doutrelant, C., Gregoire, A., Perret, P., Faivre, B., Gimenez, O., 2012. Testing hypotheses in evolutionary ecology with imperfect detection: capture-recapture structural equation modeling. Ecology 93, 248–255.
Cuffney, T.E., Bilger, M.D., Haigler, A.M., 2007. Ambiguous taxa: effects on the characterization and interpretation of invertebrate assemblages. J. N. Am. Benthol. Soc. 26 (2), 286–307.
Cuffney, T.F., Brightbill, R.A., May, J.T., Waite, I.R., 2010. Responses of benthic macroinvertebrates to environmental changes associated with urbanization in nine metropolitan areas. Ecol. Appl. 20, 1384–1401.
Davies, P.E., Nelson, M., 1994. Relationships between riparian buffer widths and the effects of logging on stream habitat, invertebrate community composition and fish abundance. Aust. J. Mar. Freshwater Res. 45, 1289–1305.
Duff, D.A., 1979. Riparian habitat recovery on Big Creek, Rich County, Utah. In: Cope, O.B. (Ed.), Method for Analyzing Livestock Impacts on Stream and Riparian Habitat Forum-grazing and Riparian/stream Ecosystems. Trout Unlimited, Denver, Colorado.
ESRI, 2008. Getting to Know ArcView GIS. Environmental Systems Research Institute, Redlands, California.
Fancy, S.G., Gross, J.E., Carter, S.L., 2009. Monitoring the condition of natural resources in US national parks. Environ. Monit. Assess. 151, 161–174.
Feminella, J.W., 1996. Comparison of benthic macroinvertebrate assemblages in small streams along a gradient of flow permanence. J. N. Am. Benthol. Soc. 15, 651–669.
Frissell, C.A., Liss, W.J., Warren, C.E., Hurley, M.D., 1986. A hierarchical framework for stream habitat classification-viewing streams in a watershed context. Environ. Manage. 10, 199–214.
Gelman, A.B., Carlin, J.S., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman & Hall/CRC, Boca Raton, Florida.
Gimenez, O., Anker- Nilssen, T., Grosbois, V., 2012. Exploring causal pathways in demographic parameter variation: path analysis of mark-recapture data. Methods Ecol. Evol. 3, 427–432.
Good, T.P., Beechie, T.J., McElhany, P., McClure, M.M., Ruckelshaus, M.H., 2007. Recovery planning for endangered species act, listed pacific salmon: using science to inform goals and strategies. Fisheries 32, 426–440.
Grace, J.B., Schoolmaster, D.R., Guntenspergen, G.R., Little, A.M., Mitchell, B.R., Miller, K.M., Schweiger, E.W., 2012. Guidelines for a graph-theoretic implementation of structural equation modeling. Ecosphere 3, 1–44.
Gross, J.E., 2003. Developing Conceptual Models for Monitoring Programs. National Park Service, Ft Collins, Colorado, pp. 26.
Harding, J.S., Benfield, E.F., Bolstad, P.V., Helfman, G.S., Jones, E.B.D., 1998. Stream biodiversity: the ghost of land use past. Proc. Natl. Acad. Sci. U. S. A. 95, 14843–14847.
Hawkins, C.P., Carlisle, D.M., 2001. Use of predictive models for assessing the biological integrity of wetlands and other aquatic habitats. In: Rader, R.B., Batzer, D.P. (Eds.), Bioassessment and Management of North American Wetlands. John Wiley & Sons, New York, New York.
Hawkins, C.P., Norris, R.H., Hogue, J.N., Feminella, J.W., 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecol. Appl. 10 (5), 1456–1477.
Hawkins, C.P., 2006. Maintaining and restoring the ecological integrity of freshwater ecosystems: refining biological assessmentsl. Ecol. Appl. 16 (4), 1249–1250.
Henderson, R., Archer, E., Bouwes, B.A., Coles- Ritchie, M.C., Kershner, J.L., 2005. PACFISH/INFISH Biological Opinion (PIBO): Effectiveness Monitoring Program Seven-Year Status Report 1998 Through 2004. RMRS-GTR-162, USDA Forest Service, Fort Collins, Colorado.
Irvine, K.M., Gitelman, A.I., 2011. Graphical spatial models: a new view on interpreting spatial pattern. Environ. Ecol. Stat. 18, 447–469.
Johnson, S.L., Jones, J.A., 2000. Stream temperature responses to forest harvest and debris flows western Cascades, Oregon. Can. J. Fish. Aquat. Sci. 57, 30–39.
Johnson, D., 2012. Monitoring that matters. In: Gitzen, R.A., Millspaugh, J.J., Cooper, A.B., Licht, D.S. (Eds.), Design and Analysis of Long-Term Ecological Monitoring Studies. Cambridge University Press, Cambridge, UK.
Jorgensen, J.C., Honea, J.M., McClure, M.M., Cooney, T.D., Engie, K.I.M., Holzer, D.M., 2009. Linking landscape-level change to habitat quality: an evaluation of restoration actions on the freshwater habitat of spring-run Chinook salmon. Freshwater Biol. 54, 1560–1575.
Karr, J.R., Chu, E.W., 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington, D.C.
Kauffman, J.B., Krueger, W.C., 1984. Livestock impacts on riparian ecosystms and streamside management ipmplications. a review. J. Range Manage. 37, 9.
Kauffman, J.B., Krueger, W.C., Vavra, M., 1983. Impacts of cattle on streambanks in northeastern Oregon. J. Range Manage. 36, 683–685.
Keene, O.N., 1995. The log transformation is special. Stat. Med. 14, 811–819.
Kershner, J.L., Archer, E.K., Coles-Ritchie, M.C., Cowley, E.R., Henderson, K., Quimby, C.M., Turner, D.L., Ulmer, L.C., Vinson, M.R., 2004. Guide to Effective Monitoring of Aquatic and Riparian Resources. RMRS-GTR-121, USDA Forest Service, Ogden, Utah.
King, J.G., 1983. Streamflow and sediment yield responses to forest practices in North Idaho. In: Baumgartner, D.M., Lotan, J.E., Tonn, J.R. (Eds.), Interior Cedar-Hemlock-White Pine Forests: 1993 March 2–4; Spokane, WA. Washington State University Cooperative Extension, Pullman, Washington, pp. 213–220.
Knapp, R.A., Matthews, K.R., 1996. Livestock grazing, golden trout, and streams in the golden trout wilderness, California: impacts and management implications. N. Am. J. Fish. Manage. 16, 805–820.
Knighton, A.D., 1999. Downstream variation in stream power. Geomorphology 29, 293–306.
LANDFIRE, 2008. Landscape, fire, and resource management planning tools. U.S. Geol. Surv.
Larsen, D.P., Kaufmann, P.R., Kincaid, T.M., Urquhart, N.S., 2004. Detecting persistent change in the habitat of salmon-bearing streams in the Pacific Northwest. Can. J. Fish. Aquat. Sci. 61, 283–291.
Lee, S., 2007. Structural Equation Modeling: A Bayesian Approach. John Wiley & Sons Ltd., West Sussex, England.
Lemly, A.D., Hilderbrand, R.H., 2000. Influence of large woody debris on stream insect communities and benthic detritus. Hydrobiologia 421, 179–185.
Lenat, D.R., Crawford, J.K., 1994. Effects of land-use on water-quality and aquatic biota of 3 North-Carolina Piedmont streams. Hydrobiologia 294, 185–199.
Levine, C.R., Yanai, R.D., Lampman, G.G., Burns, D.A., Driscoll, C.T., Lawrence, G.B., Lynch, J.A., Schoch, N., 2014. Evaluating the efficiency of environmental monitoring programs. Ecol. Ind. 39, 94–101.
Li, H.W., Lamberti, G.A., Pearsons, T.N., Tait, C.K., Li, J.L., Buckhouse, J.C., 1994. Cumulative effects of riparian disturbances along high desert trout streams of the John Day Basin, Oregon. Trans. Am. Fish. Soc. 123, 627–640.
Lindenmeyer, D.B., Likens, G.E., 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. Trends Ecol. Evol. 24, 482–486.
Mackinnon, D.P., Coxe, S., Baraldi, A.N., 2012. Guidelines for the investigation of mediating variables in business research. J. Bus. Psychol. 27, 1–14.
Mahan, C.G., Diefenbach, D.R., Cass, W.B., 2007. Evaluating and revising a long-term monitoring program for vascular plants: lessons from Shenandoah National Park. Nat. Areas J. 27, 16–24.
Manley, P.N., Zielinski, W.J., Schlesinger, M.D., Mori, S.R., 2004. Evaluation of a multiple-species approach to monitoring species at the ecoregional scale. Ecol. Appl. 14, 296–310.
Margolius, R., Stem, C., Salafsky, N., Brown, M., 2009. Using conceptual models as a planning and evaluation tool in conservation. Eval. Program Plann. 32.
McIver, J.D., McInnis, M.L., 2007. Cattle grazing effects on macroinvertebrates in an Oregon mountain stream. Rangeland Ecol. Manage. 60, 293–303.
Monk, W.A., Wood, P.J., Hannah, D.M., Wilson, D.A., 2008. Macroinvertebrate community response to inter-annual and regional river flow regime dynamics. River Res. Appl. 24, 988–1001.
Montgomery, D.R., MacDonald, L.H., 2002. Diagnostic approach to stream channel assessment and monitoring. J. Am. Water Resour. Assoc. 38, 1–16.
Moore, R.D., Spittlehouse, D.L., Story, A., 2005. Riparian microclimate and stream temperature response to forest harvesting: a review. J. Am. Water Resour. Assoc. 41, 813–834.
Moyer, C., Gallo, K., Lanigan, S.H., 2003. Interagency Regional Monitoring, Northwest Forest Plan Aquatic and Riparian Effectiveness Monitoring Program. USDA Forest Service, Portland, Oregon.
NMFS, 2000. Endangered Species Act-Section 7 Consultation, Biological Opinion, Reinitiation of Consultation on Operation of the Federal Columbia River Power System, Including the Juvenile Fish Transportation Program, and 19 Bureau of Reclamation Projects in the Columbia Basin. National Marine Fisheries Service, Portland, Oregon.
Naiman, R.J., Decamps, H., 1997. The ecology of interfaces: riparian zones. Annu. Rev. Ecol. Syst. 28, 621–658.
Nichols, J.D., Williams, B.K., 2006. Monitoring for conservation. Trends Ecol. Evol. 21, 668–673.
Noon, B.R., Raphael, M.G., Spies, T.A., Olsen, A.R., 1997. Conceptual basis for designing and effectiveness monitoring program. In: Mulder, B.S., Noon, B.R., Olsen, A.R., Palmer, C.J., Raphael, M.G., Reeves, G.H., Spies, T.A. (Eds.), The

Strategy and Designing of the Effectiveness Monitoring Program for the Northwest Forest Plan. USDA Forest Service, Portland, OR.
Ode, P.R., Hawkins, C.P., Mazor, R.D., 2008. Comparability of biological assessments derived from predictive models and multimetric indices of increasing geographic scope. J. N. Am. Benthol. Soc. 27, 967–985.
Olsen, A.R., Kincaid, T.M., Payton, Q., 2012. Spatially balanced survey designs for natural resources. In: Gitzen, R.A., Millspaugh, J.J., Cooper, A.B., Licht, D.S. (Eds.), Design and Analysis of Long-Term Ecological Monitoring Studies. Cambridge University Press, Cambridge, UK.
Ostermiller, J.D., Hawkins, C.P., 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. J. N. Am. Benthol. Soc. 23 (2), 363–382.
Overton, W.S., Stehman, S.V., 1996. Desirable design characteristics for long-term monitoring of ecological variables. Environ. Ecol. Stat. 3, 349–361.
PACFISH, 1994. Environmental Assessment for the Implementation of Interim Strategies for Managing Anadromous Fish-producing Watersheds in Eastern Oregon, Washington, Idaho, and Portions of California. USDA Forest Service, Washington, D.C.
Paulsen, S.G., Mayio, A., Peck, D.V., Stoddard, J.L., Tarquinio, E., Holdsworth, S.M., Van Sickle, J., Yuan, L.L., Hawkins, C.P., Herlihy, A.T., Kaufmann, P.R., Barbour, M.T., Larsen, D.P., Olsen, A.R., 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. J. N. Am. Benthol. Soc. 27, 812–821.
Pearl, J., 2009. Causality: Models, Reasoning and Inference, second edition Cambridge University Press, New York, New York.
Peck, D.V., Herlihy, A.T., Hill, B.H., Hughes, R.M., Kaufmann, P.R., Klemm, D.J., Lazorchak, J.M., McCormick, F.H., Peterson, S.A., Ringold, S.A., Magee, T.K., Cappaert, M., 2006. Environmental Monitoring and Assessment Program-surface Waters Western Pilot Study: Field Operations Manual for Wadeable Streams. U.S. Environmental Protection Agency, EPA/620/R-06/003, Corvallis, Oregon.
Platts, W.S., 1991. Livestock grazing. In: Meehan, W.R. (Ed.), Influences of Forest and Rangeland Management on Salmonid Fishes and Their Habitats. American Fisheries Society, Bethesda, Maryland.
Reynolds, J.H., Thompson, W.L., Russell, B., 2011. Planning for success: identifying effective and efficient survey designs for monitoring. Biol. Conserv. 144, 1278–1284.
Ringold, P.L., Mulder, B., Alegria, J., Czaplewski, R.L., Tolle, T., Burnett, K., 1999. Establishing a regional monitoring strategy: the Pacific Northwest Forest Plan. Environ. Manage. 23, 179–192.
Riseng, C.M., Wiley, M.J., Black, R.W., Munn, M.D., 2011. Impacts of agricultural land use on biological integrity: a causal analysis. Ecol. Appl. 21, 3128–3146.
Roper, B.B., Buffington, J.M., Bennett, S., Lanigan, S.H., Archer, E., Downie, S.T., Faustini, J., Hillman, T.W., Hubler, S., Jones, K., Jordan, C., Kaufmann, P.R., Merritt, G., Moyer, C., Pleus, A., 2010. A comparison of the performance and compatibility of protocols used by seven monitoring groups to measure stream habitat in the Pacific Northwest. N. Am. J. Fish. Manag. 30, 565–587.
Sarr, D.A., 2002. Riparian livestock exclosure research in the western United States: a critique and some recommendations. Environ. Manage. 30, 516–526.
Saunders, W.C., Fausch, K.D., 2007. Improved grazing management increases terrestrial invertebrate inputs that feed trout in Wyoming rangeland streams. Trans. Am. Fish. Soc. 136, 1216–1230.
Sergeant, C.J., Moynahan, B.J., Johnson, W.F., 2012. Practical advice for implementing long-term ecosystem monitoring. J. Appl. Ecol. 49, 969–973.
Shipley, B., 2009. Confirmatory path analysis in a generalized multilevel context. Ecology 90, 363–368.
Shipley, B., 2013. The AIC model selection method applied to path analytic models compared using a d-separation test. Ecology 94, 560–564.
Sims, M., Wanless, S., Harris, M.P., Mitchell, P.I., Elston, D.A., 2006. Evaluating the power of long-term trends in the numbers of common guillemots. J. Appl. Ecol. 43, 537–546.
Smart, S.M., Henrys, P.A., Purse, B.V., Murphy, J.M., Bailey, M.J., Marrs, R.H., 2012. Clarity or confusion? – Problems in attributing large-scale ecological changes to anthropgenic drivers. Ecol. Ind. 20, 51–55.
Stevens, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. J. Am. Stat. Assoc. 99, 262–278.
Tait, C.K., Li, J.L., Lamberti, G.A., Pearsons, T.N., Li, H.W., 1994. Relationships between riparian cover and the community structure of high desert streams. J. N. Am. Benthol. Soc. 13, 45–56.
Tschapalinski, P.J., Hartman, G.F., 1983. Winter distribution of juvenile coho salmon (*Oncoryhnchus kisutch*) before and after logging in Carnation Creek, British Columbia, and some implications for overwinter survival. Can. J. Fish. Aquat. Sci. 40, 452–461.
Urquhart, N.S., Paulsen, S.G., Larsen, D.P., 1998. Monitoring for policy-relevant regional trends over time. Ecol. Appl. 8, 246–257.
Vander Laan, J.J., Hawkins, C.P., Olson, J.R., Hill, R.A., 2013. Linking land use, in-stream stressors, and biological condition to infer causes of regional ecological impairment in streams. Freshwater Sci. 32, 801–820.
Vinson, M.R., Hawkins, C.P., 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. J. N. Am. Benthol. Soc. 15 (3), 392–399.
Wenger, S.J., Luce, C.H., Hamlet, A.F., Isaak, D.J., Neville, H.M., 2010. Macroscale hydrologic modeling of ecologically relevant flow metrics. Water Resour. Res. 46.
Whitacre, H.W., Roper, B.B., Kershner, J.L., 2007. A comparison of protocols and observer precision for measuring physical stream attributes. J. Am. Water Resour. Assoc. 43, 923–937.
Wood, P.J., Armitage, P.D., 1997. Biological effects of fine sediment in the lotic environment. Environ. Manage. 21, 203–217.
WoodSmith, R.D., Buffington, J.M., 1996. Multivariate geomorphic analysis of forest streams: implications for assessment of land use impacts on channel condition. Earth Surf. Process. Landforms 21, 377–393.
Wright, J.E., 2000. An introduction to RIVPACS. In: Wright, J.E., Sutcliffe, D.W., Furse, M.T. (Eds.), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ambleside, UK.
Wu, A.D., Zumbo, B.D., 2008. Understanding and using mediators and moderators. Sociol. Indic. Res. 87, 367–392.