Assessing impact of exogenous features on biotic phenomena in the presence of strong spatial dependence: A lake sturgeon case study in natural stream settings

Andrew O. Finley [ID]1*, **Patrick S. Forsythe**2, **James A. Crossman**3, **Edward A. Baker**4, **Kim T. Scribner**5,6

1 Department of Forestry, Michigan State University, East Lansing, Michigan, United States of America, **2** Department of Natural and Applied Sciences, University of Wisconsin, Green Bay, Wisconsin, United States of America, **3** BC Hydro, Water License Requirements, Castlegar, British Columbia, Canada, **4** Michigan Department of Natural Resources, Fisheries Division, Marquette, Michigan, United States of America, **5** Department of Zoology, Michigan State University, East Lansing, Michigan, United States of America, **6** Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, United States of America

* finleya@msu.edu

Abstract

Modeling spatially explicit data provides a powerful approach to identify the effects of exogenous features associated with biological processes, including recruitment of stream fishes. However, the complex spatial and temporal dynamics of the stream and the species' reproductive and early life stage behaviors present challenges to drawing valid inference using traditional regression models. In these settings it is often difficult to ensure the spatial independence among model residuals—a key assumption that must be met to ensure valid inference. We present statistical models capable of capturing complex residual anisotropic patterns through the addition of spatial random effects within an inferential framework that acknowledges uncertainty in the data and parameters. Proposed models are used to explore the impact of environmental variables on Lake sturgeon (*Acipenser fulvescens*) reproduction, particularly questions about patterns in egg deposition. Our results demonstrate the need to apply valid statistical methods to identify relationships between response variables, e.g., egg counts, across locations, and environmental covariates in the presence of strong and anisotropic autocorrelation in stream systems. The models may be applied to other settings where gamete distribution or, more generally, other biotic phenomena may be associated with spatially dynamic and anisotropic processes.

Introduction

Ecology, hydrology, statistics, and interrelated subdisciplines in aquatic sciences are increasingly integrated to study complex relationships between stream physical features and the life histories of stream fishes at multiple spatial scales [1–3] that have important implications for

management and conservation [2]. Research must be conducted in a spatially explicit fashion because stream hydrogeomorphological features are highly heterogeneous. Data are currently available for many fluvial systems (e.g., [4]) allowing studies of associations between aquatic physical data and biological phenomena (e.g., [3, 5]). However, practitioners face challenges developing and applying models capable of adequately accommodating non-Gaussian response variables and spatially complex autocorrelation, e.g., directional dependence, of physical and biotic variables within and among different watersheds and stream segments [1, 3, 6].

Spatial heterogeneity in physical stream characteristics associated with spawning means that eggs, during and after spawning event, will experience a range of environmental conditions that affect deposition and ultimately survival to subsequent development stages. From a management perspective, egg count is commonly used as a measure of year class strength; hence, there is a need to understand the relationships between the complex spatial and temporal dynamics of stream characteristics and reproductive output. It is, however, often challenging to develop valid statistical models to draw inferences about processes of interest from such complex systems—let alone assessing the extent to which findings can be extended to novel domains. For example, many standard parametric tests or model frameworks require data that is independent across space [7]. Failing to account for spatial correlation may result in inaccurate parameter estimates and erroneous conclusions regarding associations between environmental covariates and the response variable(s) of interest. Geostatisical methods attempt to account for spatially correlated residuals by including model components that estimate the characteristics of the underlying spatial process (see, e.g., [8, 9]).

Most available geostatisical models assume the relationship between point-referenced observations can be described using a function of Euclidean distance. Specific examples of such methods within the fisheries literature appear in classical kriging [10] and model-based survey design [11]. [12] extend these ideas to specify model random effects that accommodate both space and time dependence structures for modeling fish population dynamics.

Here, we define spatial regression models to quantify the effects of stream environmental covariates and the spatial dimensions of these effects associated with egg deposition within and among stream segments. We worked within a geostatistical setting, given that egg counts (the response) are point-referenced. Because spawning is confined to small stream segments, modeling spatial dependence using Euclidean distance is sensible. If observations were recorded along longer stream segments, or entire networks, then we would need to consider modeling approaches that are not strictly based on Euclidean distance to describe spatial dependence, see, e.g., [13–16]. Although we do not deal with network distance metrics, care is needed to specify models capable of accommodating complex residual patterns, e.g., anisotropy, through the addition of spatial random effects, and the use of inferential frameworks that acknowledge uncertainty in the data and parameters. The models proposed in this paper were developed to explore relationships between environmental variables and the deposition of gametes within and among spawning sites.

Many aspects of stream fishes reproductive ecology and physiology are tied to environmental variables associated with stream habitats [17]. Fishes often rely on exogenous cues to decide on the location and timing of reproduction [18–21]. Many fish species exhibit behaviors whereby parental decisions result in gametes being placed into a specific location (e.g., nest) characterized by a selected suite of environmental features [22, 23]. Other species, such as sturgeon (family *Acipenserformes*) are broadcast-spawners. For example, lake sturgeon (*Acipenser fulvescens*) are characterized by a promiscuous and aggregate mating system (Bruch and Binkowski 2002). Females are highly fecund (~11,000 eggs/kg; [24]) and have adhesive eggs (2.7-3.8 mm diameter). Negatively buoyant eggs and sperm are released by groups of females and males into the water column. Eggs are distributed over broad areas and exposed to spatially

heterogeneous environmental variables without post-ovulatory parental care. Reliance on environmental cues suggests that environmental variables that are associated with time and location of reproduction may be consistent across multiple spatial and temporal scales.

The lake sturgeon mating system and broad distribution of gametes creates a natural experimental setting to develop and illustrate models to investigate the effects of environmental variables on egg distribution. Because eggs from a single female are distributed passively over large expanses of stream habitat, micro-spatial heterogeneity in physical stream characteristics associated with the location of egg deposition means that eggs will experience a range of environmental conditions. Spawning time and location within and across years are characterized by differences in several environmental variables including temperature and river flow [20], which can be high during incubation [25]. High rates of egg mortality in other sturgeon species have been attributed to developmental arrest, predation, and physical stream processes that dislodge eggs [26, 27]. Features of stream habitat shown to be associated with egg loss include substrate [28, 29], ground and surface water exchange [30–32], and water velocity [29, 33].

The goals of this study were to provide insight into the: *i*) relationship between stream covariates and lake sturgeon egg deposition; *ii*) extent to which these relationships can be generalized to other locales and systems; and *iii*) level of model sophistication needed to draw valid inference about these relationships. Regarding the third point, we specifically wanted to know if failure to account for spatial structure, beyond that described by covariates, results in inaccurate characterization about the effect of covariates on egg deposition. To explore these questions, we developed spatial regression models that characterize egg deposition within different stream sections and allow us to quantify the effects of covariates that might affect the location of egg deposition.

Materials and methods

Study location

The models developed were applied to data from a well-studied lake sturgeon population that spawn in the Upper Black River (UBR), the primary tributary of Black Lake, MI (Fig 1). The UBR is relatively narrow and shallow with low average discharge. Spawning occurs within a shallow (<1.5 m) and wadable section (1.5 km) of the river. Within this section, individuals spawn during the spring (depending on the year between late April and early June) within six primary areas over a period of 28-45 days [21]. Seasonal patterns in water temperature and discharge are well characterized and have been shown to predict the timing and location of spawning activities [20, 21]. The stream hydrology at spawning sites allow unrestricted access to spawning adults, resulting in detailed observations of spawning behavior (e.g., spawning location, number of spawners, sex ratios, and spawning duration). Further, accessibility in the study site allows for the majority of the spawning population to be observed and captured at the spawning grounds. Enumeration of adult spawning aggregations also identifies the location of early life stage rearing habitats. Deposited eggs can be easily collected from stream substrates using sampling gear that is well suited to these stream conditions though not widely uses in larger rivers. Access to the stream also allows measurement of environmental covariates that might contribute to egg placement.

Field sampling

Surveys for spawning adults were conducted daily, starting upstream and moving downstream through the 1.5 km section of river. Upon capture, adults were tagged uniquely by sex with external floy tags (details in [27]). The location of each spawning aggregation was recorded using a handheld GPS and permanent stakes were placed along the stream bank. Data were
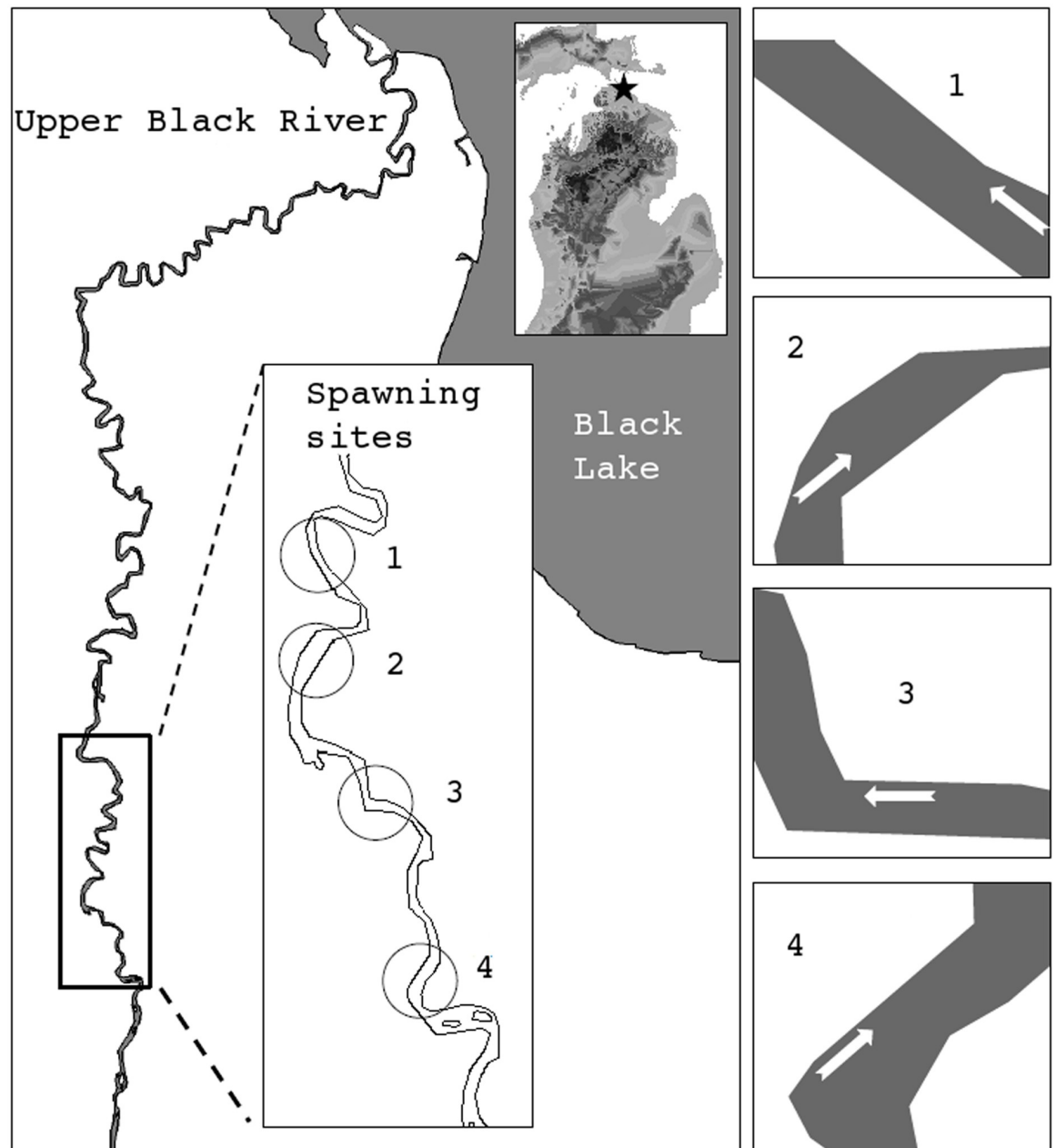
Fig 1. Left panel: star symbol on the inset map of Michigan indicates the Upper Black River study location. The inset box of the Upper Black River identifies the four spawning site locations with circle symbols. Right panels: provide a characterization of each spawning site with flow direction indicated with an arrow. River segment lengths over which measurements were taken are approximately 45, 25, 60, and 30 m, for Sites 1-4.

https://doi.org/10.1371/journal.pone.0204150.g001

collected at four spawning sites selected by spawning lake sturgeon in the UBR. Site 1 was sampled on 5/19/03, Sites 2 and 3 were sampled on 5/10/05 and 5/12/05, and Site 4 was sampled on 4/25/06, respectively (Fig 1). Transects were conducted across the stream channel starting ~1 m below the observed spawning group and were repeated every 5 m downstream to cover the area of egg deposition. To ensure spawning had ceased and no new eggs were deposited, transects were conducted one day after spawning was completed and adults had left
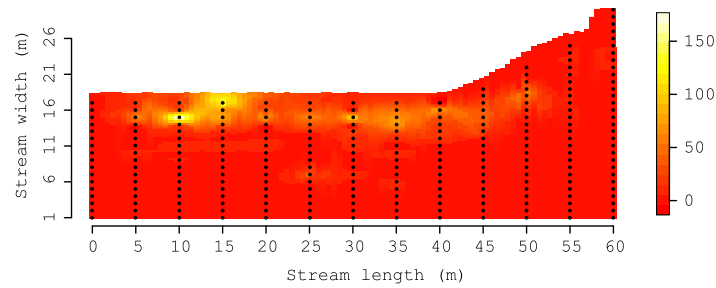
(a) Egg count
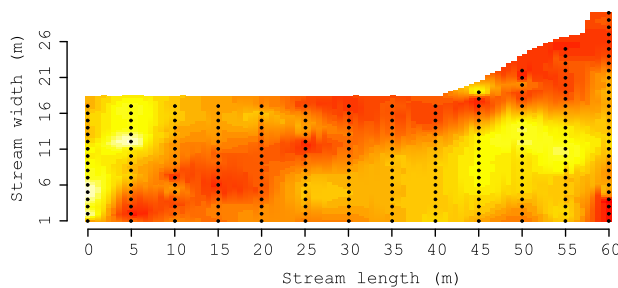
(b) Water depth (m)
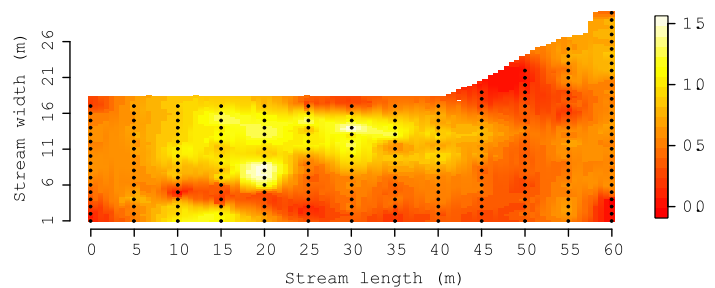
(c) Water velocity (m^3/s)
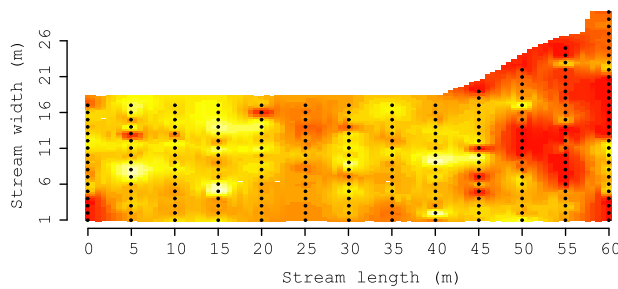
(d) Average substrate (mm)
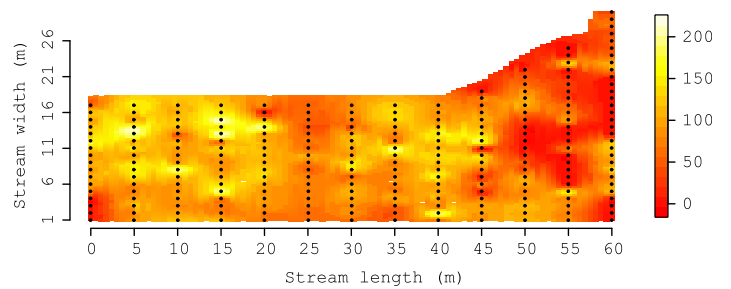
(e) MaxSub

Fig 2. Point symbols indicate the locations where egg count and environmental covariates were measured at site 3. Underlying surfaces were generated by passing the given point values through an interpolator.

the site. Counts of eggs and measurements of environmental covariates were collected at a 1 m interval along each transect (see, e.g., Fig 2) using kick nets. Kick nets are a widely used approach to sample stream benthos including sturgeon [34] and have been used successfully to collect eggs in other studies (e.g., [35, 36]). Substrate was thoroughly disturbed for 5 seconds in the area (~0.125 m^2) immediately upstream of a small triangular net (305×305×305 mm opening). The kick net was held immediately down stream of the agitated area to ensure

displaced materials were swept into the net. We assume the egg counts are proportional to the actual eggs present across the range of habitats encountered (deep vs shallow, coarse substrate vs fine). Materials collected by the kick net were emptied into white bottom trays and all contents were sorted and the number and status (live versus dead) of collected eggs were recorded.

Data on environmental covariates (water depth, water velocity, and substrate size) were collected concurrent with surveys of egg deposition. Stream depth (*Depth*) was recorded in meters using a stadia rod. Water velocity (*Velocity*) at the stream bed was recorded in cubic meters per second using an electromagnetic flow meter (Marsh-McBirney, Inc.). Average substrate size (*AvgSub*; mm length along the longest axis) was estimated using four stones collected from either the kick net sample or from the corresponding point on the transect. Maximum substrate size (*MaxSub*; mm) was estimated by measuring the largest stone collected within the transect point. Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.tg706dv.

Statistical analyses

Egg deposition. A Poisson regression model was used to relate egg count at sampling points along the transects and coinciding measures of environmental covariates. This model is adequate if spatial correlation in the response is completely explained by the covariates. Because lake sturgeon are broadcast spawners and eggs are extruded into a spatially complex fluvial environment, egg counts are expected to exhibit strong spatial dependence along the down-stream and across-stream sampling axes (i.e., similar egg counts in the proximate locations). Unmeasured covariates or covariates measured at the *wrong* scale can result in spatial dependence among the residuals, which violates the model assumptions. Therefore, to accommodate any lingering spatial structure, we include a spatially structured random effect in the model intensity. Given observations arise over a set of locations, say $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, where $\mathbf{s}_i$ is a sample location's spatial coordinates, we specify the log-link model as:

$$y(\mathbf{s}_i) \sim Poisson(\mu(\mathbf{s}_i)); \quad \log(\mu(\mathbf{s}_i)) = \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + w(\mathbf{s}_i), \tag{1}$$

where $y(\mathbf{s}_i)$ is the i-th sample location egg count, which is assumed to be distributed ("$\sim$") as a Poisson variable with location specific conditional expectation $\mu(\mathbf{s}_i)$. This conditional expectation comprises vectors of covariates and associated slope parameters, $\mathbf{x}(\mathbf{s}_i)$ and $\boldsymbol{\beta}$ respectively, and $w(\mathbf{s}_i)$ which provides local adjustment (with structured dependence) and captures the effect of unmeasured or unobserved spatial covariates.

We assume the random effects follow a zero mean Gaussian Process (GP), $w(\mathbf{s}) \sim GP(0, \sigma^2 \rho(\cdot; \boldsymbol{\theta}))$, with variance σ^2, and correlation function $\rho(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})$, where $\mathbf{s}$ and $\mathbf{s}^*$ are two generic locations. The vector of parameters $\boldsymbol{\theta}$ control the correlation function's behavior. For a collection of n locations the $n \times 1$ vector of random effects follows a Multivariate Normal distribution, $\mathbf{w} \sim MVN(\mathbf{0}, \sigma^2 R(\boldsymbol{\theta}))$, where $R(\theta) = [\rho(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})]_{i,j=1}^n$ is the correlation matrix. Spatial correlation functions of varying complexity are available for defining $\rho(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta})$, see, e.g., [37]. A common choice for the correlation function is an isotropic exponential, $\rho(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}) = \exp(-\phi\|\mathbf{s} - \mathbf{s}^*\|)$, where ϕ is the spatial range parameter. Here it is assumed the spatial dependence is the same in all directions. Given the dynamics of the stream's flow, isotropy is a rather unrealistic assumption. To explore the possibility of a more complex spatial structure we use an anisotropic form $\rho(\mathbf{s}, \mathbf{s}^*; \boldsymbol{\theta}) = \exp[-(\mathbf{s} - \mathbf{s}^*)' [G(\psi)\Lambda^2 G'(\psi)]^{-1}(\mathbf{s} - \mathbf{s}^*)]$. Here, the rotation matrix $G(\psi)$ controls directional dependence given the angle parameter ψ and the diagonal matrix Λ defines the rate of spatial decay along perpendicular axes. Specifically, the positive diagonal elements λ_0 and λ_1 corresponding to spatial decay parameters for the major and minor axes, respectively [38, 39], which in our situation corresponds to down stream vs cross stream, respectively.

When describing a spatial process it is useful to report the distance at which spatial dependence is negligible. This distance is referred to as the effective spatial range, and is defined here as the distance at which the spatial correlation drops to 0.05. For the isotropic exponential correlation function this is $-\log(0.05)/\phi$ and for the anisotropic form this is $-\log(0.05)\lambda$, where λ is a diagonal value in Λ. Note, there are two effective spatial ranges for each anisotropic process, one oriented with ψ, and the other perpendicular to ψ.

Implementation and model selection. The proposed models were fit following a Bayesian paradigm [40, 41]. As such, a prior distribution must be assigned to each parameter to complete the model specification. For all models the $\boldsymbol{\beta}$'s received *flat* prior distributions and the spatial variance parameters were assigned inverse-Gamma (*IG*) priors with hyperparameters $IG(2, \cdot)$. This is considered a non-informative prior; with a shape value of 2, the *IG* distribution has infinite variance and is centered on the scale value, which is data set specific. The spatial range parameters λ's and ϕ's followed Uniform priors, which are chosen to support a spatial range from 0 to the maximum inter-location distance in the data set. The rotation parameters ψ's also received a data set specific Uniform prior providing rotation estimates consistent with possible stream flow direction.

To compare several alternative models, we used the Deviance Information Criterion (DIC) [42]. Letting Ω be the generic set of parameters being estimated for each model (including random effects), we computed the expected posterior deviance $\overline{D(\Omega)} = E_{\Omega|Y}[-2\log L(Data|\,\Omega)]$, where $L(Data|\Omega)$ is the first stage likelihood from the respective model and the effective number of parameters (as a penalty) as $p_D = \overline{D(\Omega)} - D(\bar{\Omega})$, where $\bar{\Omega}$ is the posterior mean of the model parameters. The DIC is then given by $\overline{D(\Omega)} + p_D$ and is easily computed from the posterior samples with lower values indicating better models.

For all models, three MCMC chains, with unique starting values, were run for 50,000 iterations. The CODA package in R (www.r-project.org) was used to diagnose convergence by monitoring mixing using Gelman-Rubin diagnostics and autocorrelations (see, e.g., [40]). For all analyses, acceptable convergence was diagnosed within 10,000 iterations (which were discarded as burn-in). The sampler was coded in C++ and Fortran and leveraged Intel's Math Kernel Library threaded BLAS and LAPACK routines for matrix computations.

Results

Summary of spawning groups and environmental covariates

As illustrated in Fig 3, the four sampling sites exhibited similar ranges in environmental covariate characteristics including substrate sizes (1-200 mm), maximum substrate size (1-300 mm), water depth (0.1-1.6 m), and water velocity (0.1-1.6 m³/s). However, the distribution of values for each environmental covariate indicate that the physical features of areas where eggs were deposited varied considerably among sites, Fig 3. These distributions suggest Sites 1 and 3 have many sample locations with average and maximum substrate smaller than Sites 2 and 4. Also, Site 3 has several locations with water velocity greater than the other sites. Fig 2b–2e) offers a spatial representation of the distribution of these covariates for Site 3. Corresponding figures for Sites 1, 2, and 4, as well as figures for subsequent analysis results are provided as supplemental material.

Embryo deposition and environmental covariates

A total of 9,426 eggs were collected across all spawning locations from 651 kick net samples. Egg counts varied across locations (Site 1 = 2,509, Site 2 = 564, Site 3 = 2,111, Site 4 = 4,242). Within each site, egg deposition exhibited non-random and spatially-dependent patterns, see,
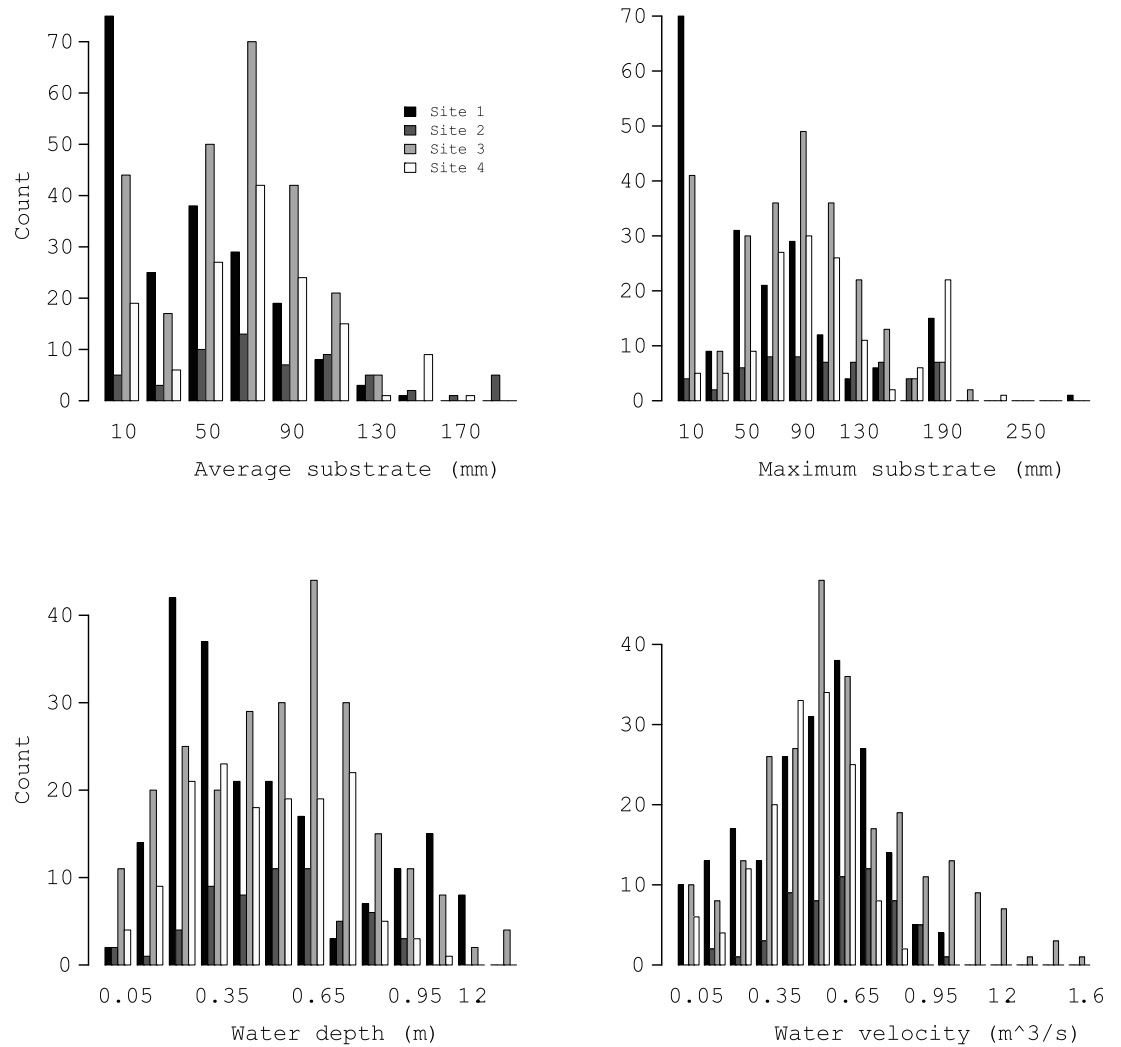
Fig 3. Frequency histograms for average substrate size, maximum substrate size, water depth, and water velocity for sites 1-4.

e.g., Fig 2(a), often occurring in discrete patches. Eggs were primarily deposited along the downstream axis on the side of the river where spawning took place and from 0 to 60 m downstream from spawning locations.

Both the non-spatial and spatial versions of Model (1) were fit using observations from the 198, 60, 249, and 144 sampling locations surveyed at Sites 1-4, respectively. Table 1 provides candidate models' parameter estimates for each site. Looking first at the spatial model's effective spatial range parameter estimates, λ_0 and λ_1, we see there is relatively strong but variable spatial patterns among the residuals. For example, Site 2 exhibits the shortest effective spatial range of $\sim$5 m, whereas Site 3 shows the longest range of $\sim$54 m. For Sites 1, 3, and 4, the disparity between λ_0 and λ_1 suggests there are strong anisotropic patterns in spatial dependence and hence an isotropic spatial covariance function would not be appropriate. In each of these sites ψ describes the orientation of the major spatial range parameter axis λ_0. Here we present ψ in both radians and degrees, where in both cases zero is perpendicular to the west river bank. Not surprisingly, parameter estimates of λ_1 suggest a shorter range of spatial dependence

Table 1. Summary of egg deposition non-spatial and spatial models' parameter estimates for each site. Parameter posterior credible intervals, 50 (2.5 97.5) percentiles. Those β parameters that are significant at the 0.05 level are bolded.

Parameters	Site 1		Site 2	
	Non-spatial	Spatial	Non-spatial	Spatial
β_0	**1.86 (1.71, 1.99)**	-1.94 (-4.61, 0.43)	**1.42 (0.90, 1.93)**	-0.27 (-2.54, 2.04)
β_{Depth}	**-0.44 (-0.61, -0.28)**	**1.89 (0.11, 3.84)**	**0.96 (0.33, 1.39)**	0.87 (-2.18, 3.44)
$\beta_{Velocity}$	-0.04 (-0.19, 0.15)	1.18 (-0.21, 2.60)	**-0.79 (-1.29, -0.35)**	-0.20 (-2.84, 2.25)
β_{AvgSub}	**0.014 (0.012, 0.015)**	0.004 (-0.005, 0.013)	**-0.012 (-0.014, -0.009)**	-0.015 (-0.034, 0.001)
β_{MaxSub}	**0.003 (0.002, 0.004)**	**0.007 (0.002, 0.012)**	**0.015 (0.012, 0.018)**	**0.022 (0.008, 0.039)**
σ^2	–	5.19 (2.96, 9.56)	–	2.42 (1.36, 4.93)
ψ degree	–	10.78 (-36.31, 39.46)	–	-14.49 (-43.19, 43.08)
λ_0	–	10.96 (5.00, 16.10)	–	2.23 (0.37, 9.51)
λ_1	–	5.66 (2.94, 11.65)	–	1.83 (0.39, 9.29)
Eff. range $(-\log(0.05)\lambda_0)$	–	32.85 (14.98, 48.25)	–	6.68 (1.13, 28.50)
Eff. range $(-\log(0.05)\lambda_1)$	–	16.97 (8.81, 34.91)	–	5.48 (1.17, 27.85)
DIC	-8710	-14115	-1548	-2139
p_D	5	117.4	5.7	48.5

| Parameters	Site 3		Site 4	
	Non-spatial	Spatial	Non-spatial	Spatial
β_0	0.02 (-0.12, 0.25)	**-3.49 (-5.79, -0.41)**	**0.87 (0.72, 1.01)**	**-4.20 (-7.02, -0.81)**
β_{Depth}	**-0.34 (-0.58, -0.18)**	0.68 (-1.36, 2.82)	**0.81 (0.62, 0.93)**	**7.26 (2.05, 9.16)**
$\beta_{Velocity}$	**1.71 (1.57, 1.87)**	0.22 (-1.02, 1.50)	**1.78 (1.58, 1.96)**	**2.64 (0.42, 4.42)**
β_{AvgSub}	**0.016 (0.013, 0.019)**	**0.019 (0.005, 0.035)**	**0.009 (0.008, 0.010)**	0.002 (-0.004, 0.009)
β_{MaxSub}	-0.001 (-0.003, 0.001)	-0.009 (-0.020, 0.002)	**0.003 (0.003, 0.005)**	-0.002 (-0.007, 0.003)
σ^2	–	10.43 (6.33, 17.35)	–	5.84 (3.54, 10.51)
ψ degree	–	5.88 (-7.44, 19.98)	–	16.94 (-8.51, 64.34)
λ_0	–	18.12 (10.84, 22.05)	–	10.33 (5.87, 12.14)
λ_1	–	6.30 (3.71, 10.40)	–	5.045 (2.83, 10.31)
Eff. range $(-\log(0.05)\lambda_0)$	–	54.29 (32.48, 66.07)	–	30.97 (17.60, 36.37)
Eff. range $(-\log(0.05)\lambda_1)$	–	18.89 (11.13, 31.18)	–	15.11 (8.47, 30.89)
DIC	-6163	-11080	-22051	-28620
p_D	5	117.1	5.1	95.4

perpendicular to water flow. Sites with the maximum egg dispersal distance, e.g., Site 3, also had the highest average water velocity.

The relatively long effective spatial ranges as well as the magnitude of the spatial variance parameter values, i.e., σ^2, suggest that even after accounting for the covariates there is substantial unexplained spatial pattern in the number of eggs within the spawning sites. The presence of this residual spatial autocorrelation violates an assumption of the non-spatial equivalent to Model (1). As noted in Section 1, this violation can result in erroneous estimates of the regression coefficients associated with the covariates.

Comparison between the non-spatial and spatial model regression coefficients, i.e., β's, in Table 1 shows that ignoring the residual autocorrelation does in fact result in very different conclusions about the association between the covariates and egg deposition. For example, in Site 1 the non-spatial model estimates the median and 95% credible interval (CI) for β_{Depth} at -0.44 (-0.61, -0.28), which excludes zero and can hence be considered statistically significant at a 0.05 level. This estimate suggests that water depth is negatively associated with the number of

eggs deposited. The relationship between depth and number of eggs changes to positive and significant after model assumptions are satisfied through the addition of the spatial random effects. This is presumably a more accurate portrayal of the system. The non-spatial model identifies more covariates as significant than the spatial model across all sites. Further, once the residual space dependence is accommodated, there is little consistency in regression coefficient across the sites. For instance, the spatial model parameter estimates suggest water depth is significant and positive for Sites 1 and 4, but not in Sites 2 and 3. Maximum substrate size is significant in Sites 1 and 2 suggesting that larger or coarser substrate is positively associated with number of eggs deposited—a relationship that does not hold for Sites 3 and 4. In addition to meeting model assumptions, and presumably providing a more accurate view of the relationship between deposition and the environmental covariates, the spatial random effects in Model (1) improves fit to the data as reflected in the consistently lower DIC values given in Table 1. The larger effective number of parameters, p_D, for the spatial models versus the non-spatial model, is due to the addition of the random effects. The fit of the non-spatial and spatial models is illustrated in Fig 4(a) and 4(b), respectively. These surfaces can be compared to the observed values illustrated in Fig 2(a) and shows the spatial model does, in fact, more closely approximate the distribution of eggs across the stream segment. Fig 4(c) provides the spatial random effects surface that illustrates where local adjustment to $\mu(\mathbf{s})$ is needed to fit the observed data after accounting for the covariates.

(a) Non-spatial fitted

(b) Spatial fitted

(c) Spatial random effects
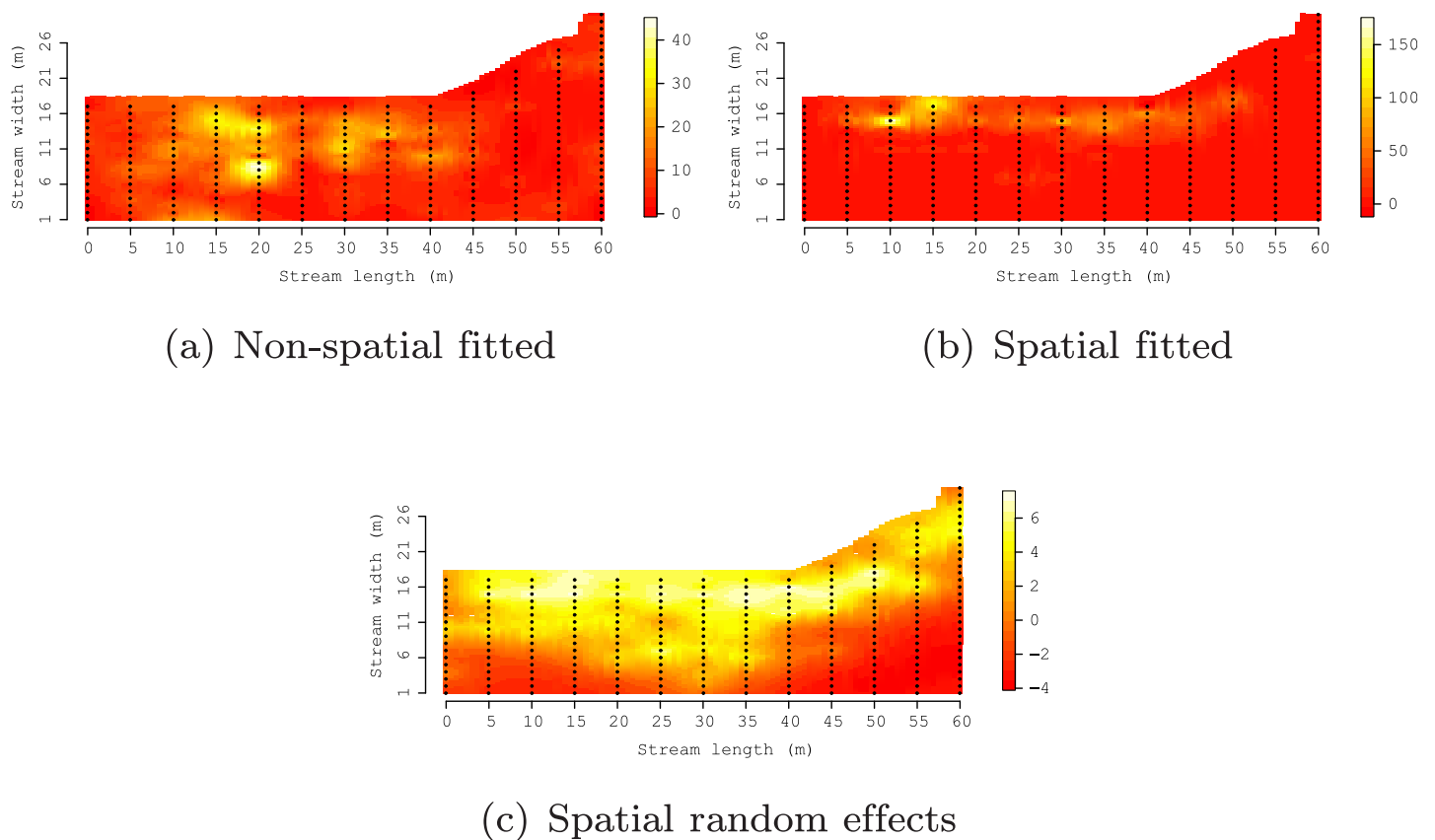
Fig 4. Interpolated surfaces of fitted values from the non-spatial and spatial models (a) and (b), respectfully, and spatial model random effects (c) for site 3. These fitted values can be compared to the observed egg counts illustrated in Fig 2(a).

https://doi.org/10.1371/journal.pone.0204150.g004

Discussion

Following from Section 1, the non-spatial models' parameter estimates provided misleading conclusions about the relationship between deposition and the covariates. Of 16 environmental covariates measured across sites (i.e., 4 covariates times 4 sites), 14 were found to be significant predictors of egg deposition based on the non-spatial model, compared with only six when using the more statistically appropriate spatial models, Table 1. Several lines of evidence including measures of model fit (DIC, pD; Table 1) and the magnitude of spatial range parameters demonstrate that without accounting for spatial dependencies, predictions of egg deposition, and subsequent uses of these predictions, would be inaccurate and misleading from a biological or management perspective.

The spatial models showed that lake sturgeon egg depositional profiles were spatially autocorrelated and highly anisotropic across all spawning sites. Comparison of spatial range parameters revealed that spatial dependence was always greater in the downstream than across the stream direction. Environmental covariates did not show consistent significant impact across sites and generally did not explain a substantial portion of responses' variability. Rather, the spatial random effects explained the majority of spatial variability in egg distribution. Importantly, results underscores the need to work within modeling frameworks capable of capturing residual spatial dependence if we wish to draw valid inference about those covariates that are included in the model. Examples of egg deposition studies that have failed to use spatial models are common (see, e.g., review by [43]). Failure to incorporate spatial models can mis-direct efforts to use direct estimates of total egg counts as a measure of recruitment potential [44]. Comparisons of the actual egg depositional surface (Fig 2(a)) to spatially and non-spatially fitted egg surfaces in Fig 4 clearly show that if egg count estimates are based on non-spatial models, results should be interpreted with caution. In dynamic stream environments, estimates based on a sample of available habitat should not be extrapolated to larger areas to estimate total reproductive effort (total numbers of eggs).

Despite the specificity of spawning habitats typically selected by lake sturgeon [45], the known importance of "high quality" habitats [17], the spatial model results suggested considerable heterogeneity across sites in the relative magnitude and direction of effects of stream environmental variates on egg depositional patterns. Assumptions that adult reliance on environmental cues to time and select locations for reproduction imply that effects of environmental covariates would be comparable across multiple spatial and temporal scales. Adult behavioral plasticity in spawning site selection (e.g., four sites sampled in this study) resulted in heterogeneous egg depositional profiles across sites dictated by site-specific environmental covariates.

Spatial dependence is likely due to the species' spawning behavior and the physical system [45, 46]. Eggs are released and subjected to site-specific fluvial dynamics near the stream bed that non-randomly determine the direction of egg drift and the distance traveled. Model results suggest some environmental covariates are useful for explaining patterns in egg deposition, supporting our prediction that hydro-geomophological environmental stream variables will explain some variability in egg deposition within a site. Specifically, egg deposition was significantly associated with substrate maximum and average size. Water velocity and depth were also found to be significant predictors of egg deposition. Our results are concordant with other studies that found stream variables including water depth, water velocity, and substrate size are predictive of egg deposition for migratory fishes [47–49]. The magnitude and direction of association between numbers of lake sturgeon eggs deposited with substrate size, depth, and velocity reported in other studies (e.g., [50, 51]) might differ if spatially explicit models were considered. The effect of flow to generate "hot spots" has been widely described in other

aquatic organisms (e.g., macroinvertebrates) and is frequently observed in river systems [52]. Our results are consistent with other studies of behaviors such as nest site selection in birds and in migratory fishes which exhibit clustered or patchy distributions due to abiotic (i.e., habitat) specificity related to the species reproductive ecology [53, 54]. Our proposed models, can be used in future studies to quantify the impact of additional spawning behavior and physical system covariates that might help explain patterns seen in the spatial random effects. Potentially useful covariates that influence where eggs ultimately settle include flow at the substrate subsurface, hyporheic discharge, and other microhabitat characteristics (e.g., interstitial spacing or substrate embeddedness).

Summary

Properly linking components of the physical and biotic environment to currencies of reproductive performance is important to understand population, species, and community responses to anthropogenic change and abilities of natural systems to deliver ecosystem services [55–57]. Our study demonstrates the need to employ appropriate statistical methods to properly elucidate relationships between response and environmental covariates in the presence of strong and anisotropic autocorrelation in complex stream systems. As demonstrated here, ignoring spatial dependence can result in falsely precise estimates of regression coefficients associated with environmental covariates and erroneous predictions of egg deposition. If estimates are based on non-spatial models, results should be interpreted with caution. In dynamic stream environments, estimates based on a sample of available habitat should not be extrapolated to larger areas to estimate total reproductive effort (total numbers of eggs).

Most commonly applied regression models require the residuals, i.e., after accounting for covariates, to be independent and identically distributed. In many settings the biotic phenomena of interest exhibit spatial dependence that cannot be completely explained by covariates. Exploratory analyses of non-spatial model residuals using surface plots, e.g., Fig 4, or variograms (if the response is Gaussian) will typically reveal if there is lingering spatial structure and hence if a spatial random effect is needed. As shown in this study, residual spatial patterns can exhibit directional dependence, in which cases an anisotropic spatial correlation function should used. Failing to account for spatial correlation can result in inaccurate parameter estimates and erroneous conclusions regarding associations between environmental covariates and the response variable(s) of interest.

Supporting information

S1 File. Data and results for sites 1, 2, and 4.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Andrew O. Finley, Edward A. Baker, Kim T. Scribner.

Data curation: Patrick S. Forsythe, James A. Crossman, Edward A. Baker, Kim T. Scribner.

Formal analysis: Andrew O. Finley.

Funding acquisition: Kim T. Scribner.

Investigation: Andrew O. Finley, Patrick S. Forsythe.

Methodology: Andrew O. Finley, Patrick S. Forsythe.

Resources: Andrew O. Finley.

Software: Andrew O. Finley.

Validation: Andrew O. Finley.

Visualization: Andrew O. Finley.

Writing – original draft: Andrew O. Finley, Patrick S. Forsythe, James A. Crossman, Edward A. Baker, Kim T. Scribner.

Writing – review & editing: Andrew O. Finley, Patrick S. Forsythe, James A. Crossman.

References

1. Fausch KD, Torgersen CE, Baxter CV, Li HW. Landscapes to Riverscapes: Bridging the Gap between Research and Conservation of Stream Fishes: A Continuous View of the River is Needed to Understand How Processes Interacting among Scales Set the Context for Stream Fishes and Their Habitat. BioScience. 2002; 52(6):483–498. https://doi.org/10.1641/0006-3568(2002)052%5B0483:LTRBTG%5D2.0.CO;2

2. Schlosser IJ, Angermeier P. Spatial variation in demographic processes of Lotic fishes: conceptual models, empirical evidence, and implications for conservation. American Fisheries Society Symposium. 1995; 17:392–401.

3. Allan JD. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology, Evolution, and Systematics. 2004; 35:257–284. https://doi.org/10.1146/annurev.ecolsys.35.120202.110122

4. Esselman PC, Infante DM, Wang L, Wu D, Cooper AR, Taylor WW. An index of cumulative disturbance to river fish habitats of the conterminous United States from landscape anthropogenic activities. Ecological Restoration. 2011; 29(1-2):133–151. https://doi.org/10.3368/er.29.1-2.133

5. Poff NL. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. Journal of the north american Benthological society. 1997; 16(2):391–409. https://doi.org/10.2307/1468026

6. Gergel SE, Turner MG, Miller JR, Melack JM, Stanley EH. Landscape indicators of human impacts to riverine systems. Aquatic Sciences. 2002; 64(2):118–128. https://doi.org/10.1007/s00027-002-8060-2

7. Hoeting JA. The importance of accounting for spatial and temporal correlation in analyses of ecological data. Ecological Applications. 2009; 19(3):574–577. https://doi.org/10.1890/08-0836.1 PMID: 19425418

8. Schabenberger O, Gotway CA. Statistical methods for spatial data analysis. CRC press; 2004.

9. Cressie N, Wikle CK. Statistics for spatio-temporal data. John Wiley & Sons; 2015.

10. Petitgas P. Geostatistics in fisheries survey design and stock assessment: models, variances and applications. Fish and Fisheries. 2001; 2(3):231–249. https://doi.org/10.1046/j.1467-2960.2001.00047.x

11. Kimura DK, Somerton DA. Review of statistical aspects of survey sampling for marine fisheries. Reviews in Fisheries Science. 2006; 14(3):245–283. https://doi.org/10.1080/10641260600621761

12. Kristensen K, Thygesen UH, Andersen KH, Beyer JE. Estimating spatio-temporal dynamics of size-structured populations. Canadian Journal of Fisheries and Aquatic Sciences. 2013; 71(2):326–336. https://doi.org/10.1139/cjfas-2013-0151

13. Ver Hoef JM, Peterson E, Theobald D. Spatial statistical models that use flow and stream distance. Environmental and Ecological statistics. 2006; 13(4):449–464. https://doi.org/10.1007/s10651-006-0022-8

14. Curriero FC. On the use of non-Euclidean distance measures in geostatistics. Mathematical Geology. 2006; 38(8):907–926. https://doi.org/10.1007/s11004-006-9055-7

15. Cressie N, Frey J, Harch B, Smith M. Spatial prediction on a river network. Journal of Agricultural, Biological, and Environmental Statistics. 2006; 11(2):127. https://doi.org/10.1198/108571106X110649

16. Peterson EE, Ver Hoef JM, Isaak DJ, Falke JA, Fortin MJ, Jordan CE, et al. Modelling dendritic ecological networks in space: an integrated network perspective. Ecology Letters. 2013; 16(5):707–719. https://doi.org/10.1111/ele.12084 PMID: 23458322

17. Kamler E. Ontogeny of yolk-feeding fish: an ecological perspective. Reviews in Fish Biology and Fisheries. 2002; 12(1):79–103. https://doi.org/10.1023/A:1022603204337

18. Workman RD, Hayes DB, Coon TG. A model of steelhead movement in relation to water temperature in two Lake Michigan tributaries. Transactions of the American Fisheries Society. 2002; 131(3):463–475. https://doi.org/10.1577/1548-8659(2002)131%3C0463:AMOSMI%3E2.0.CO;2

19. Keefer ML, Moser ML, Boggs CT, Daigle WR, Peery CA. Variability in migration timing of adult Pacific lamprey (Lampetra tridentata) in the Columbia River, USA. Environmental biology of fishes. 2009; 85 (3):253–264. https://doi.org/10.1007/s10641-009-9490-7

20. Forsythe PS, Crossman JA, Bello NM, Baker EA, Scribner KT. Individual-based analyses reveal high repeatability in timing and location of reproduction in lake sturgeon (Acipenser fulvescens). Canadian Journal of Fisheries and Aquatic Sciences. 2012; 69(1):60–72. https://doi.org/10.1139/f2011-132

21. Forsythe P, Scribner K, Crossman J, Ragavendran A, Baker E, Davis C, et al. Environmental and lunar cues are predictive of the timing of river entry and spawning-site arrival in lake sturgeon Acipenser fulvescens. Journal of Fish Biology. 2012; 81(1):35–53. https://doi.org/10.1111/j.1095-8649.2012.03308.x PMID: 22747803

22. Trippel EA, Kjesbu OS, Solemdal P. Effects of adult age and size structure on reproductive output in marine fishes. In: Early life history and recruitment in fish populations. Springer; 1997. p. 31–62.

23. Jørgensen C, Dunlop ES, Opdal AF, Fiksen Ø. The evolution of spawning migrations: state dependence and fishing-induced changes. Ecology. 2008; 89(12):3436–3448. https://doi.org/10.1890/07-1469.1 PMID: 19137949

24. Bruch RM, Miller G, Hansen MJ. Fecundity of Lake Sturgeon (Acipenser fulvescens, Rafinesque) in Lake Winnebago, Wisconsin, USA. Journal of Applied Ichthyology. 2006; 22:116. https://doi.org/10.1111/j.1439-0426.2007.00938.x

25. Johnson J, LaPan S, Klindt R, Schiavone A. Lake sturgeon spawning on artificial habitat in the St Lawrence River. Journal of Applied Ichthyology. 2006; 22(6):465–470. https://doi.org/10.1111/j.1439-0426.2006.00812.x

26. Parsley MJ, Anders PJ, Miller AI, Beckman LG, McCabe GTJ. Recovery of white sturgeon populations through natural production: Understanding the influence of abiotic and biotic factors on spawning and subsequent recruitment. American Fisheries Society Symposium. 2002; 28:55–66.

27. Forsythe P, Scribner K, Crossman J, Ragavendran A, Baker E. Experimental assessment of the magnitude and sources of lake sturgeon egg mortality. Transactions of the American Fisheries Society. 2013; 142(4):1005–1011. https://doi.org/10.1080/00028487.2013.790847

28. Bernier-Bourgault I, Magnan P. Factors affecting redd site selection, hatching, and emergence of brook charr, Salvelinus fontinalis, in an artificially enhanced site. Environmental Biology of Fishes. 2002; p. 333–341. https://doi.org/10.1023/A:1016006303854

29. Smith SJ, Marsden JE. Factors affecting sea lamprey egg survival. North American Journal of Fisheries Management. 2009; 29(4):859–868. https://doi.org/10.1577/M07-196.1

30. Malcolm IA, Youngson AF, Soulsby C. Survival of salmonid eggs in a degraded gravel-bed stream: effects of groundwater–surface water interactions. River Research and Applications. 2003; 19(4): 303–316. https://doi.org/10.1002/rra.706

31. Malcolm I, Soulsby C, Youngson A, Hannah D, McLaren I, Thorne A. Hydrological influences on hyporheic water quality: implications for salmon egg survival. Hydrological Processes. 2004; 18(9): 1543–1560. https://doi.org/10.1002/hyp.1405

32. Youngson A, Malcolm I, Thorley J, Bacon P, Soulsby C. Long-residence groundwater effects on incubating salmonid eggs: low hyporheic oxygen impairs embryo development. Canadian Journal of Fisheries and Aquatic Sciences. 2004; 61(12):2278–2287. https://doi.org/10.1139/f04-217

33. Curry RA, MacNeill WS. Population-level responses to sediment during early life inbrook trout. Journal of the North American Benthological Society. 2004; 23(1):140–150. https://doi.org/10.1899/0887-3593(2004)023%3C0140:PRTSDE%3E2.0.CO;2

34. van der Leeuw BK, Parsley MJ, Wright CD, Kofoot EE. Validation of a critical assumption of the riparian habitat hypothesis for white sturgeon. US Geological Survey Scientific Investigations Report. 2006; 5225:20.

35. Gönczi AP. A study of physical parameters at the spawning sites of the European grayling (Thymallus thymallus L.). Regulated Rivers: Research & Management. 1989; 3(1):221–224. https://doi.org/10. 1002/rrr.3450030121

36. Katt JD, Koupal KD, Schoenebeck CW, Hoback WW. Assessment of a New Gear to Sample Walleye Eggs. North American Journal of Fisheries Management. 2012; 32(1):44–48. https://doi.org/10.1080/ 02755947.2012.655842

37. Cressie NA. Statistics for spatial data: Wiley Series in Probability and Mathematical Statistics. Find this article online. 1993;.

38. Finley AO. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. Methods in Ecology and Evolution. 2011; 2(2):143–154. https://doi.org/10.1111/j.2041-210X.2010.00060.x

39. Paciorek CJ, Schervish MJ. Spatial modelling using a new class of nonstationary covariance functions. Environmetrics. 2006; 17(5):483–506. https://doi.org/10.1002/env.785 PMID: 18163157

40. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. vol. 2. Chapman & Hall/CRC Boca Raton, FL, USA; 2014.

41. Banerjee S, Carlin BP, Gelfand AE. Hierarchical Modeling and Analysis for Spatial Data, Second Edition. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press; 2014.

42. Spiegelhalter D, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B. 2002; 64(4):583–639. https://doi.org/10.1111/1467-9868.00353

43. Bunn N, Fox CJ, Webb T. A literature review of studies on fish egg mortality: implications for the estimation of spawning stock biomass by the annual egg production method. Ministry of Agriculture, Fisheries and Food, Centre for Environment, Fisheries and Aquaculture Science; 2000.

44. Fitzsimons J, Jonas J, Claramunt R, Williston B, Williston G, Marsden J, et al. Influence of egg predation and physical disturbance on lake trout Salvelinus namaycush egg mortality and implications for life-history theory. Journal of Fish Biology. 2007; 71(1):1–16. https://doi.org/10.1111/j.1095-8649.2007. 01437.x

45. Kempinger JJ. Spawning and early life history of lake sturgeon in the Lake Winnebago system, Wisconsin. American Fisheries Society Symposium. 1988; 5:110–122.

46. Bruch R, Binkowski F. Spawning behavior of lake sturgeon (Acipenser fulvescens). Journal of Applied Ichthyology. 2002; 18(4-6):570–579. https://doi.org/10.1046/j.1439-0426.2002.00421.x

47. Mull KE, Wilzbach MA. Selection of spawning sites by coho salmon in a northern California stream. North American Journal of Fisheries Management. 2007; 27(4):1343–1354. https://doi.org/10.1577/ M06-054.1

48. Knapp RA, Preisler HK. Is it possible to predict habitat use by spawning salmonids? A test using California golden trout (Oncorhynchus mykiss aguabonita). Canadian Journal of Fisheries and Aquatic Sciences. 1999; 56(9):1576–1584. https://doi.org/10.1139/f99-081

49. Thurow RF, King JG. Attributes of Yellowstone cutthroat trout redds in a tributary of the Snake River, Idaho. Transactions of the American Fisheries Society. 1994; 123(1):37–50. https://doi.org/10.1577/ 1548-8659(1994)123%3C0037:AOYCTR%3E2.3.CO;2

50. LaHaye M, Branchaud A, Gendron M, Verdon R, Fortin R. Reproduction, early life history, and characteristics of the spawning grounds of the lake sturgeon (Acipenser fulvescens) in Des Prairies and L'Assomption rivers, near Montreal, Quebec. Canadian Journal of Zoology. 1992; 70(9):1681–1689. https://doi.org/10.1139/z92-234

51. Chiotti JA, Holtgren JM, Auer NA, Ogren SA. Lake sturgeon spawning habitat in the Big Manistee River, Michigan. North American Journal of Fisheries Management. 2008; 28(4):1009–1019. https://doi.org/ 10.1577/M07-051.1

52. Bond NR, Downes BJ. The independent and interactive effects of fine sediment and flow on benthic invertebrate communities characteristic of small upland streams. Freshwater Biology. 2003; 48(3): 455–465. https://doi.org/10.1046/j.1365-2427.2003.01016.x

53. Melles S, Badzinski D, Fortin MJ, Csillag F, Lindsay K. Disentangling habitat and social drivers of nesting patterns in songbirds. Landscape Ecology. 2009; 24(4):519. https://doi.org/10.1007/s10980-009-9329-9

54. Hendry AP, Berg OK, Quinn TP. Breeding location choice in salmon: causes (habitat, competition, body size, energy stores) and consequences (life span, energy stores). Oikos. 2001; 93(3):407–418. https:// doi.org/10.1034/j.1600-0706.2001.930306.x

55. Post DM, Palkovacs EP. Eco-evolutionary feedbacks in community and ecosystem ecology: interactions between the ecological theatre and the evolutionary play. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2009; 364(1523):1629–1640. https://doi.org/10.1098/rstb.2009.0012 PMID: 19414476

56. Schoener TW. The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. science. 2011; 331(6016):426–429. https://doi.org/10.1126/science.1193954 PMID: 21273479

57. Frimpong EA, Angermeier PL. Trait-based approaches in the analysis of stream fish communities. In: Community ecology of stream fishes: concepts, approaches, and techniques. American Fisheries Society, Symposium. vol. 73; 2010. p. 109–136.