RESEARCH ARTICLE

# New methods for quantifying the effects of catchment spatial patterns on aquatic responses

**Donald E. Weller** · **Matthew E. Baker** · **Ryan S. King**

**Abstract**

*Context* Previous studies developed distance-weighted regression to describe how land use effects on aquatic systems attenuate with arrangement of source areas within catchments.

*Objectives* We clarify and extend the conceptual foundations of this approach, enhance the spatial and statistical methods, and provide new tools to interpret the results.

*Methods* We derive the framework from first principles to resolve conceptual issues with how weighting is applied to source area versus total area, and we formalize the requirements for an ideal weighting function. We quantify the spatial distributions of land areas in a way that integrates with model fitting. We adapt non-linear optimization to simultaneously fit regression and weighting parameters. We quantify the spatial distribution of source effects with arrangement and document how different weighting functions alter that distribution.

*Results* To verify their utility, we applied these methods to a published analysis relating polychlorinated biphenyls in fish to developed land use in catchments. We identified a stronger distance-weighted model and more completely characterized the effects of weighting on where aquatic impacts originate.

*Conclusions* Our methods enable more comprehensive analyses of the effects of spatial arrangement to better inform a wide range of scientific investigations and applications. Our methods can relate almost any spatially distributed source or driver to an integrated response at a point or along a boundary; and alternate hypotheses about the effects of pattern or proximity on processes can be tested with alternative weighting functions. New applications will generate additional weighting functions that enhance the general approach.

**Keywords** Distance weighting · Land use regression · Catchment pattern · Spatial arrangement · Aquatic impact

D. E. Weller (✉)
Smithsonian Environmental Research Center, 647 Contees Wharf Rd., Edgewater, MD 21037-3702, USA
e-mail: wellerd@si.edu

M. E. Baker
Department of Geography and Environmental Systems, University of Maryland Baltimore County, 211 Sondheim Hall, 1000 Hilltop Circle, Baltimore, MD 21250, USA

R. S. King
Department of Biology, Baylor University, One Bear Place #97388, Waco, TX 76798-7388, USA

## Introduction

Freshwater and estuarine ecosystems are intimately connected to their catchments (Hynes 1975; Johnson

et al. 1997). Hydrologic behavior (Rinaldo et al. 1995), biophysical processes, and ecological functions (Black 1997; Johnson et al. 1997; Benda et al. 2004) are all related to the geomorphology of the collection basin. Given these connections, land use within catchments can have strong impacts on aquatic ecosystems.

Many studies have estimated land use impacts by relating aquatic responses to catchment land use proportions (Osborne and Wiley 1988; Allan et al. 1997; Herlihy et al. 1998; Liu et al. 2000; King et al. 2005, 2007; Goetz and Fiske 2008; Jordan et al. 2018), often by applying a linear regression across a sample of catchments:

$$C = \beta_0 + \beta_1 F \tag{1}$$

where $C$ is an aquatic response, such as the concentration (mass per unit volume) of a material in the receiving water, $F$ is the fraction (proportion) of a catchment occupied by the land use that is the source of the material, and $\beta_0$ and $\beta_1$ are fitted regression coefficients. Both regression coefficients have units of the response (mass/volume for a concentration response), and the land proportion $F$ is dimensionless. For simplicity, we focus on the case where the effect originates from a single land use, here called the source land.

For a concentration response, the regression parameter $\beta_1$ can be interpreted by defining the mass $M$ of material released from source land in a catchment as $M = mS$, where $S$ is the area of source land and $m$ is mass release per unit area. The entire catchment delivers a volume of water $V$ to the receiving water as described by $V = vA$, where $A$ is the total area of the catchment and $v$ is the water yield (volume per unit area). Then,

$$C = \frac{M}{V} = \frac{mS}{vA} = q\frac{S}{A} = qF \tag{2}$$

where $q = m/v$ is the concentration loading rate (in mass volume$^{-1}$ area$^{-1}$), and $F$ is the proportion of source land relative to total catchment area. Regression parameter $\beta_1$ in Eq. (1) provides a statistical estimate of the conceptual loading rate $q$ in Eq. (2).

The regression approach has been modified to evaluate how aquatic responses to land use are moderated by spatial factors, such as transport distance or the arrangement of source land relative to sink ecosystems like riparian buffers, floodplains, or stream corridors (Bohlke and Denver 1995; Lowrance et al. 1997; O'Neill et al. 1997; Omernik et al. 1981; Sponseller et al. 2001; Weller et al. 1998, 2011; Weller and Baker 2014).

To account for transport distance, many studies have replaced land use or impervious surface proportions with "effective" proportions derived from weighting by distance (Comeleo et al. 1996; Van Sickle and Johnson 2008; Walsh and Kunapo 2009; Peterson et al. 2011; Peterson and Pearse 2017; Walsh and Webb 2014). Some studies implemented an attenuation function within a geographic information system (GIS) calculation (Hunsaker and Levine 1995; Soranno et al. 1996; Zhang 2010). Other studies separated spatial analysis from weighting by dividing catchments into distance increments, tabulating total and source areas within each band, and fitting different weighting models using that tabular summary (King et al. 2004, 2005; Zhang 2010). With this enhancement, GIS calculations need not be repeated for each weighting considered, but using just a few bands can sacrifice distance resolution and reduce the accuracy of the analysis.

Inverse power distance weighting ($1/d$, where $d$ is distance) is the most widely used weighting function (Comeleo et al. 1996; Goetz and Fiske 2008; Peterson et al. 2011), but step, exponential and linear weighting functions have also been used (e.g., Van Sickle and Johnson 2008; Walsh and Webb 2014). Far more studies have implicitly used distance weighting by only considering land use within a fixed distance from a response point (Osborne and Wiley 1988; Johnson et al. 1997; Tufford et al. 1998; Jones et al. 2001; Sponseller et al. 2001; Snyder et al. 2003), often implemented with the "buffer" function of a GIS package.

Most studies have used a weighting calculation that applies distance weights to total area as well to as source area (King et al. 2004, 2005; Van Sickle and Johnson 2008; Peterson et al. 2011; Sheldon et al. 2012; Walsh and Webb 2014), but some have left total area unweighted (e.g., Comeleo et al. 1996; Walsh and Kunapo 2009). The implications of how total area is weighted have not been explored.

Weighting function parameters have not been estimated with statistical rigor. King et al. (2004) tested three inverse power functions and chose one that yielded the greatest improvement in model $R^2$. Others

sought optimum weighting function parameters by trying a set of values, creating contours of model skill versus parameter value, and visually identifying the value maximizing model skill (Hunsaker and Levine 1995; Soranno et al. 1996; Van Sickle and Johnson 2008; Zhang 2010; Walsh and Webb 2014). No studies have optimized the weighting function parameters simultaneously with the slope and intercept in land use regression.

Past studies have not effectively quantified how weighting changes inferences about where land use impacts originate within catchments. Most interpretations have been based on plots of the weighting function versus distance (King et al. 2004, 2005) or on the "half distance" at which an effect is reduced by 50% (Van Sickle and Johnson 2008; Walsh and Webb 2014). These approaches do not account for the effects of weighting total land as well as source land, nor do they consider how weighting interacts with the distance distributions of source and total land.

In this paper, we clarify and extend the conceptual foundation of distance-weighted land use regression, enhance the spatial and statistical methods, and provide new tools for understanding spatial patterns. We apply these new methods to a published analysis to illustrate the improved understanding they provide.

## Methods

### Conceptual foundations of distance weighting

#### *Derivation of weighted land use regression*

Distance-weighted regression is very general and can be applied to many problems where an effect propagates from sources to impact points. To simplify the derivation, we focus on a material of interest released from a single source land use type and transported hydrologically to an aquatic measurement point. The derivation accounts for non-conservative transport and dilution while assuming that the system is at steady state so that transport time need not be considered explicitly. We provide a list of mathematical symbols (Table S1) to help interpret the equations.

To account for arrangement effects on an aquatic response $C$, we divide catchments into bands, each defined by its distance $d$ from the receiving water. The number of bands is $B$ while $s_b$ and $a_b$ are the

areas of source and total land in band $b$. Summing across the bands recovers the catchment total source area $S$ and catchment area $A$.

$$C = \frac{\sum_{b=1}^{B} ms_b}{\sum_{b=1}^{B} va_b} = q\frac{\sum_{b=1}^{B} s_b}{\sum_{b=1}^{B} a_b} = q\frac{S}{A} = qF \qquad (3)$$

Distance weighting hypothesizes that source land further from the receiving water delivers less material than closer source land. We implement this by multiplying the source area $s_b$ in the numerator by a distance weighting function $w(d)$ that declines with distance $d$ from the receiving water. Distant land may also deliver less water than closer land, so we multiply the denominator by another function that declines with distance $x(d)$. The expression for concentration in the receiving water becomes

$$C = \frac{\sum_{b=1}^{B} ms_b w(d_b)}{\sum_{b=1}^{B} va_b x(d_b)} = q\frac{\sum_{b=1}^{B} s_b w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)} = qG \qquad (4)$$

The quantity $G$ is the "distance-weighted source proportion" or "effective source proportion"—the proportion of catchment area occupied by source land after adjusting for the effects of distance on the delivery of material and water to the receiving water. We refer to the use of different numerator and denominator weighting functions [$w(d)$ and $x(d)$] as "separate weighting." Conceptually, the numerator accounts for attenuating, non-conservative transmission of the material or effect from source land to the receiving water, while the denominator accounts for non-conservative transmission of water or other diluting effects from all the land.

Incorporating Eq. (4) into regression Eq. (1) yields the weighted of concentration or aquatic response to land use

$$C = \beta_0 + \beta_1 \frac{\sum_{b=1}^{B} s_b w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)} = \beta_0 + \beta_1 G \qquad (5)$$

where the regression coefficient $\beta_1$ provides a statistical estimate of the conceptual loading rate $q$.

Two published weighting approaches are special cases of Eq. (5). If water delivery attenuates with distance the same way as material delivery, then the denominator weighting function is the same as the numerator weighting function [$x(d) = w(d)$] and Eq. (5) simplifies to

$$C = \beta_0 + \beta_1 \frac{\sum_{b=1}^{B} s_b w(d_b)}{\sum_{b=1}^{B} a_b w(d_b)} = \beta_0 + \beta_1 G \qquad (6)$$

This represents the "double weighting" method that has been most widely applied (King et al. 2004, 2005; Van Sickle and Johnson 2008; Peterson et al. 2011; Sheldon et al. 2012; Walsh and Webb 2014). If the denominator weighting function $x(d)$ has a constant value of 1 rather than declining with distance, then Eq. (5) simplifies to

$$C = \beta_0 + \beta_1 \frac{\sum_{b=1}^{B} s_b w(d_b)}{\sum_{b=1}^{B} a_b} = \beta_0 + \beta_1 \frac{\sum_{b=1}^{B} s_b w(d_b)}{A} \qquad (7)$$

This equation incorporates the weighting method applied by Comeleo et al. (1996) and Walsh and Kunapo (2009), which implicitly assumes that water delivery or other diluting effects do not attenuate with distance ("single weighting"). Finally, if $w(d)$ also equals 1, then Eq. 7 reduces to the unweighted relationship (Eq. 2) and $G = F$.

No assumption about the attenuation of water with distance can be universally correct. One can avoid both assumptions by statistically fitting separate functions for the numerator and denominator (Eq. 4), allowing the data to indicate how much the diluent is attenuated.

*Weighting functions*

Four weighting functions have been applied (King et al. 2004, 2005; Van Sickle and Johnson 2008; Walsh and Kunapo 2009; Walsh and Webb 2014):

Step function with threshold $t_s$

$$w(d) = \begin{cases} 1, & d < t_s \\ 0, & d \geq t_s \end{cases} \qquad (8)$$

Linear function with threshold $t_l$

$$w(d) = \begin{cases} 1 - d/t_l, & d < t_l \\ 0, & d \geq t_l \end{cases} \qquad (9)$$

Exponential function with constant $k$

$$w(d) = e^{-kd} \qquad (10)$$

Power function with power $p$

$$w(d) = \frac{1}{d^p} = d^{-p} \qquad (11)$$

Each of these has a single fitted parameter ($t_s$, $t_l$, $k$, or $p$). We also tested a generalized threshold function that can represent concave, convex, linear, and step functions (see Supplemental Information).

An ideal proximity weighting function has three characteristics: declines with distance $d$, approaching 0 as $d$ gets large; is 1 (no distance effect) at $d = 0$, and yields the same model results regardless of the units of distance measurement (scale invariance). The last condition can be met if changing units does not change the weighted source proportions $G$ or if changing units multiplies the weighted proportion of every catchment by a constant $\gamma$ so that regression with the rescaled data yields the same intercept and fit statistics as the unscaled data, but with rescaled slope $\beta_1/\gamma$. The step, linear, and exponential functions meet all three criteria. The inverse distance function $d^{-p}$ meets two criteria but is discontinuous at $d = 0$. A modified version, $(d + 1)^{-p}$ (Van Sickle and Johnson 2008; Peterson et al. 2011; Sheldon et al. 2012; Peterson and Pearse 2017) is 1 at $d = 0$ but can yield different results in statistical models if the distance measurement units are changed. We also tested two other modified power functions that do meet all three criteria (see Supporting Information).

*Spatial distributions of weighted source land and delivered materials*

The proportion of source land in distance band $b$ expressed *as a fraction of the total catchment area* is

$$f_b = \frac{s_b}{A} \qquad (12)$$

Without distance weighting, the contribution of band $b$ to material concentration at the response point is

$$c_b = q f_b = q \frac{s_b}{A} \qquad (13)$$

The sum across bands $\sum_{b=1}^{B} f_b$ is the overall fraction of source land in the catchment ($F$) and the sum $\sum_{b=1}^{B} c_b$ is the concentration in the receiving water ($C$).

The proportion of distance-weighted source land $g_b$ in band $b$ *as a fraction of distance-weighted total catchment area* is

$$g_b = \frac{s_b w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)} \tag{14}$$

And the weighted contribution $h_b$ of the band to delivered concentration $H$ in the receiving water is

$$h_b = \beta_1 g_b = \beta_1 \frac{s_b w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)} \tag{15}$$

The sum across bands $\sum_{b=1}^{B} g_b$ is the weighted catchment source proportion ($G$) and the sum $\sum_{b=1}^{B} h_b$ (plus $\beta_0$) is the weighted concentration delivered to the receiving water ($H$). These equations are conceptually related to the "width function" describing the distribution of basin area with distance from the outlet (Rinaldo et al. 1995). In the regression model, $\beta_1$ provides a statistical estimate of the loading rate $q$ so $h_b = \beta_1 g_b$ can be used to interpret the spatial distribution of material source for the regression model.

When applied across the distance bands of a catchment, Eq. (14) captures the combined effect of the spatial distributions of source and total land, weighting of source land, and weighting of total land on the spatial distribution of effective source land. Equation (15) provides the same integration to capture the combined effects on the spatial distribution of material delivery.

The relationships in Eqs. (14) and (15) help visualize and summarize the spatial distributions of source land and delivered materials in a catchment as well as to quantify how those distributions shift in a weighted analysis compared to an unweighted one. Plotting weighted source land proportion minus actual source land proportion against distance reveals where weighting increases or decreases effective source proportion relative to actual source proportion. Similarly, plotting $h_b - c_b$ reveals how weighting shifts the distances from which materials are delivered to the receiving water.

Weighted source land proportion minus actual source land proportion is

$$g_b - f_b = \frac{s_b w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)} - \frac{s_b}{A} = s_b \left[ \frac{w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)} - \frac{1}{A} \right] \tag{16}$$

At high distances, $w(d_b)$ approaches 0, so $g_b - f_b$ approaches a negative value ($-s_b/A$). At the distance where $w(d_b) = \left[ \sum_{b=1}^{B} a_b x(d_b) \right]/A$, the term in brackets is 0 so that $g_b = f_b$ and their difference is zero. Below that distance, the weighted source proportion $g_b$ is greater than unweighted proportion $f_b$, while above that distance the weighted proportion $g_b$ is less than unweighted source proportion $f_b$.

If the total land area in the denominator is not weighted (single weighting), then Eq. (16) simplifies to

$$g_b - f_b = \frac{s_b w(d_b)}{A} - \frac{s_b}{A} = s_b \left[ \frac{w(d_b) - 1}{A} \right] \tag{17}$$

Because $w(d_b)$ is a fraction, the term in brackets is always negative, so that with single weighting, the weighted proportion $g_b$ is less than the unweighted $f_b$ at all distances.

The same patterns apply to the difference in material delivery between weighted and unweighted models because material delivery is simply the constant $q$ times weighted proportion (Eq. 16).

$$\begin{aligned} h_b - c_b &= q s_b \left| \frac{w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)} - \frac{1}{A} \right| \\ &= \beta_1 s_b \left[ \frac{w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)} - \frac{1}{A} \right] \end{aligned} \tag{18}$$

*Realized weighting function*

We define the "Realized weighting" $y_b$ for a band $b$ as the ratio of weighted source proportion $g_b$ for that band (Eq. 14) to the unweighted source proportion $f_b$ for that band (Eq. 13)

$$y_b = \frac{g_b}{f_b} = \frac{\frac{s_b w(d_b)}{\sum_{b=1}^{B} a_b x(d_b)}}{\frac{s_b}{A}} = w(d_b) \frac{A}{\sum_{b=1}^{B} a_b x(d_b)} = w(d_b) J \tag{19}$$

The realized weighting function $y_b$ is simply a constant ($J$) times the distance weighting function $w(d_b)$, but that constant differs among catchments. The term $A / \sum_{b=1}^{B} a_b x(d_b)$ is the ratio of unweighted to weighted total catchment area. For each catchment, this ratio takes a unique value ($J$) that depends on weighting function $x(d_b)$, catchment area $A$, and the distance distribution of total area within that catchment (the width function, Rinaldo et al. 1995). With an ideal weighting function, weighted total area is less than total area, so that the term is greater than one and amplifies the effect of the weighting function. The function $y_b$ takes the value 1 at the distance where $w(d_b) = 1/J$. This point is helpful for interpreting the behavior of $y_b$. At smaller distances, realized weighting $y_b$ is $> 1$, while $y_b < 1$ at distances beyond this point.

### New tools for implementing weighted regression

#### *Geographic methods for quantifying spatial distributions*

We quantified the distance distributions of source and total land for catchments using a GIS script that integrates digital maps of land use or land cover, water bodies, and elevation to calculate the distance from each cell of the raster land use or land cover map to the receiving water body. The script can calculate Cartesian distance or distances along downhill flow paths. Raw distance values were divided by 10 and converted to integers to produce compact tabular summaries, which were exported to files with columns for catchment identifier, distance increment, total land, and source land. This method delivers distance distributions resolved to the precision of the land cover raster instead of lumping distances into a few bands defined by arbitrary, a priori ranges (as in King et al. 2004, 2005; Zhang 2011).

#### *Improved regression methods*

We implemented non-linear least squares regression for estimating weighting function parameters simultaneously with the regression slope and intercept (Eq. 1). We used the R *optimx* package (Nash and Varadhan 2011), which includes algorithms that work well with discontinuous functions (e.g., Eqs. 8 or 9). *optimx* requires a user-supplied function to calculate the model residual sum of squares (RSS) for any values of the weighting and regression parameters. We wrote an R function that uses the tabular summaries of distances from the spatial analysis. It applies the weighting equation with the parameters supplied by *optimx* to all distance values, source areas, and total areas for each catchment. Those weighted areas are summed for each catchment to produce a weighted source proportion $G$ for each catchment. Then, the values of regression parameters $\beta_0$ and $\beta_1$ provided by *optimx* are applied to each weighted source proportion $G$ to calculate a predicted response. Finally, observed values are subtracted from the predictions, squared, and summed for return to *optimx*.

We performed model comparisons based on information theory to contrast model performance for different weighting functions and to test hypotheses about weighting (Akiake Information Criterion, AICc, Burnham and Anderson 2002). We also calculated model probabilities (Akaike model weights) as a weight of evidence for the best model within a model set (Burnham and Anderson 2002).

### Case study

#### *Test data*

To test our methods, we reanalyzed data from a published study of how the distance from developed land to the shoreline affected the concentration of polychlorinated biphenyls (PCBs) in the tissues of fish caught in 14 subestuaries of Chesapeake Bay (King et al. 2004). In that study, the distance between each land cover cell and the shoreline was calculated. Cartesian distance was used even though downhill flow path distance provides a better description of hydrologic transport in most landscapes (King et al. 2005; Baker et al. 2006, 2007; Weller et al. 2011; Weller and Baker 2014), because flow path algorithms did not perform effectively for the very flat catchments of the subestuaries. For four possible measures of developed land cover

(total developed, high intensity residential plus commercial, commercial, and impervious surface) and for all land, the counts of pixels were aggregated into seven distance classes. Subestuary average PCB concentration in fish was related to developed land covers using inverse distance-weighted regression. Percent commercial land, weighted by $d^{-1}$, was the best predictor of PCBs in fish, and weighting increased $R^2$ to 99% compared to 55% for unweighted commercial land. The study demonstrated a strong spatial effect but left opportunities to improve the analysis and refine its interpretation. King et al. (2004) provide more details and results.

*Catchment characteristics*

We used the watershed boundaries and land cover data presented by King et al. (2004) but applied our new spatial analysis (above) to obtain more precise tabulations of the spatial distributions of the Cartesian distances from commercial land and all land to the subestuary. To visualize the distributions, we plotted histograms of commercial land area and total land area versus distance to the subestuary using 200 m distance bins.

*Model comparisons*

We fit 15 weighted models to test hypotheses about distance weighting and to identify the most effective weighting functions. These 15 models reflect three ways of weighting the total area in the weighting equation (single, double, and separate) and four distance weighting functions (step, linear, exponential, and power) plus the unweighted model. For the power function, we used $(d+1)^{-p}$ instead of $d^{-p}$. We quantified model performance with four metrics (RMSE, $R^2$, AICc, and Akaike model weight). We compared models within three subsets that were selected to answer three questions:

- Is there evidence for an arrangement effect?
- How should total land be weighted?
- What is the best weighting function?

*How does weighting affect apparent land use proportions?*

For each weighting function, we quantified how weighting changed the catchment proportion of weighted source land ($G$) relative to the unweighted proportion ($F$) as well as weighted material concentration ($H$) relative to the unweighted concentration ($C$). We also plotted predicted material concentration versus weighted source proportion ($G$) for the 14 subestuaries along with the fitted regression line to show how weighting affected the strength of association and the magnitude of residuals around the fitted line.

*Where in the system do materials originate?*

For models with four weighting functions as well as the unweighted model, we visualized the spatial distributions of how weighting shifts source land proportions ($g_b - f_b$, Eq. 16) and contribution to material concentration ($h_b - f_b$, Eq. 18) using histograms with 200 m distance bins. We examined aggregate distance distributions for all the catchments together to focus on the differences among models, and we examined them for individual catchments to reveal how weighting functions interact with differences in catchment land use patterns.

*Evaluation of extreme observations*

Like King et al. (2004), we repeated the analysis after omitting two systems with distinctly high PCB to verify that key results do not depend on extreme values (see Supporting Information).

## Results of the case study

Distance distributions of source land and total land

The catchment areas of the 14 subestuaries ranged from 46 to 662 km$^2$ and summed to 2720 km$^2$. The proportion of commercial land ranged from 0.25 to 13.9% among the catchments and was 4.33% of the summed study area. GIS analysis of distances to shoreline yielded 4715 unique increments ranging from 30 to 47,180 m as well as the areas of commercial and total land at each distance for each catchment. The distributions of total and commercial land areas versus distance from the
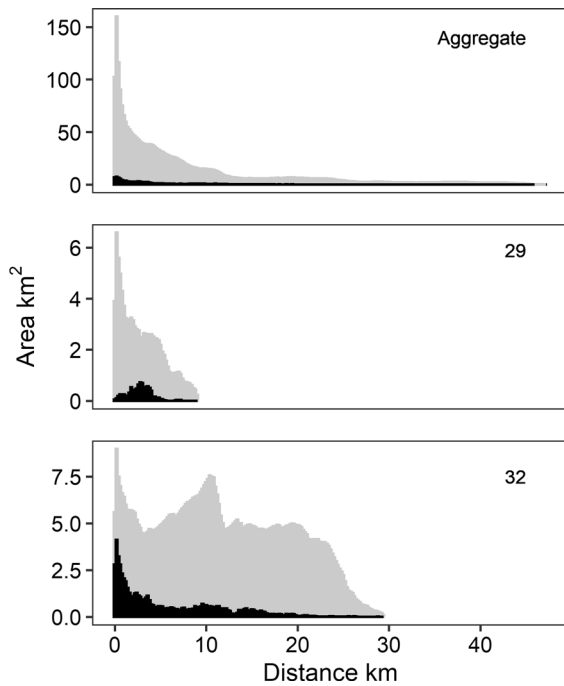
**Fig. 1** Distributions of watershed area (gray) and commercial land area (black) versus distance from the subestuary for the watersheds of two numbered subestuaries and for the aggregate of all study watersheds

subestuary for two example catchments illustrate differences among catchments in size, shape, and the spatial distributions of source and total area (Fig. 1). Watershed 32 was larger than catchment 29, and the distribution of catchment area declined much less strongly with distance in watershed 32. Commercial land in watershed 32 was concentrated very close to the subestuary, while the peak in commercial land area for watershed 29 occurred at some distance from the subestuary.

## Model comparisons

AICc-based model comparisons within subsets of models addressed the specific hypotheses or questions represented in the headings and tables presented below.

### *Is there an arrangement effect?*

There was very strong evidence that spatial arrangement is important for the test data. Among double-weighted models with four different weighting functions, even the poorest lowered AICc by 43.0 compared to the unweighted model (Table 1), translating to a model probability of 1.0 for the weighted model compared to 0.0 for the unweighted. Any of the models was overwhelmingly more likely than the unweighted model, providing rigorous, probabilistic evidence that spatial arrangement is important and should be included in relating land use to the observed PCB concentrations in fish.

### *How should total land be weighted?*

For the test data, there was strong support for double weighting (Eq. 6) and virtually no support for single weighting (Eq. 7). With double weighting, any of the weighting functions (Eqs. 8–11) provided a far better explanation of the data than the unweighted model (Table 2) improving $R^2$ from 51% to at least 98% and lowering AICc by at least 43.0. None of the weighting functions applied with single weighting provided any increase in $R^2$ and all had higher AICc values than the unweighted model (see also Supporting Information).

The test data did not support separate weighting parameters for the numerator and denominator of the weighting function. Across all weighting

**Table 1** Performance metrics and model parameters for the unweighted model and for double-weighted models of the test data with five different weighting functions

| Weighting function | $R^2$ | AICc | $\Delta$AICc[a] | $w_i$ | $\beta_0$ | SE $\beta_0$ | $\beta_1$ | SE $\beta_1$ | $t$ | SE $t$ | $k$ | SE $k$ | $p$ | SE $p$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unweighted | 0.511 | 181.3 | 55.0 | 0.000 | $-8.11$ | 44.09 | 2720 | 768 | | | | | | |
| Step | 0.993 | 126.3 | 0 | 0.912 | $-0.46$ | 5.07 | 1670 | 50 | 1606 | 94 | | | | |
| Linear | 0.990 | 131.3 | 5.0 | 0.075 | $-1.04$ | 6.27 | 1662 | 73 | 2405 | 350 | | | | |
| Exponential | 0.986 | 135.2 | 8.9 | 0.011 | $-1.95$ | 7.20 | 1665 | 86 | | | 0.0010 | 0.0002 | | |
| Power | 0.983 | 138.4 | 12.1 | 0.002 | 25.33 | 8.38 | 2120 | 143 | | | | | 0.798 | 0.052 |

[a]$\Delta$AICc relative to the best model, positive indicates poorer model

**Table 2** Comparison of models of the test data using single or double distance weighting

| Weighting function | $R^2$ | AICc | $\Delta$AICc[a] | $\beta_0$ | $\beta_1$ | $t$ | $k$ | $p$ |
|---|---|---|---|---|---|---|---|---|
| Unweighted model | | | | | | | | |
| Unweighted | 0.511 | 181.3 | 0.00 | −8.11 | 2720 | | | |
| Single-weighted models | | | | | | | | |
| Step | 0.511 | 185.4 | 4.04 | −8.11 | 2720 | 50,000 | | |
| Linear | 0.511 | 185.4 | 4.04 | −8.10 | 2720 | 1.5E+10 | | |
| Exponential | 0.511 | 185.4 | 4.04 | −8.11 | 2720 | | 0 | |
| Power | 0.511 | 185.4 | 4.04 | −8.11 | 2720 | | | 0 |
| Double-weighted models | | | | | | | | |
| Step | 0.993 | 126.3 | −55.04 | −0.46 | 1670 | 1606 | | |
| Linear | 0.990 | 131.3 | −50.04 | −1.04 | 1662 | 2405 | | |
| Exponential | 0.986 | 135.2 | −46.11 | −1.95 | 1665 | | 0.00103 | |
| Power | 0.983 | 138.4 | −42.97 | −25.33 | 2120 | | | 0.798 |

[a] $\Delta$AICc relative to the unweighted model, negative for better models and positive for poorer

**Table 3** Comparison of models for the test data using double or separate distance weighting

| Weighting type | $R^2$ | AICc | $\Delta$AICc[a] | $w_i$ | $\beta_0$ | $\beta_1$ | $t$ | $t_d$ | $k$ | $k_d$ | $p$ | $p_d$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Step function | | | | | | | | | | | | |
| Double | 0.99281 | 126.3 | 0 | 0.87 | −0.46 | 1670 | 1606 | | | | | |
| Separate | 0.99344 | 130.1 | 3.8 | 0.13 | −2.68 | 1891 | 1641 | 1918 | | | | |
| Linear function | | | | | | | | | | | | |
| Double | 0.98973 | 131.3 | 0 | 0.64 | −1.04 | 1662 | 2405 | | | | | |
| Separate | 0.99221 | 132.5 | 1.2 | 0.36 | −2.90 | 2340 | 2039 | 3301 | | | | |
| Exponential function | | | | | | | | | | | | |
| Double | 0.98641 | 135.2 | 0 | 0.65 | −1.95 | 1665 | | | 0.00103 | | | |
| Separate | 0.98965 | 136.5 | 1.2 | 0.35 | −5.13 | 2742 | | | 0.00129 | 0.00066 | | |
| Power function | | | | | | | | | | | | |
| Double | 0.98298 | 138.4 | 0 | 0.92 | −25.33 | 2120 | | | | | 0.798 | |
| Separate | 0.98303 | 143.4 | 5.0 | 0.08 | −24.33 | 1863 | | | | | 0.784 | 0.806 |

[a] $\Delta$AICc relative to double weighting, positive if separate weighting is poorer

functions, fitting separate weighting functions (which added a parameter to any weighted regression model) increased $R^2$ at most by 0.0032 and raised rather than lowered AICc, which increased between 1.2 and 5 above the unweighted model (Table 3). Also, fitted parameters of the separate weighting functions in the numerator and denominator were similar.

*What is the best weighting function?*

With double weighting, the step function provided the best model among the four candidate functions (Table 1). The AICc for the step function was 5.0 lower than for the next best fit, the linear function. Relative probabilities were 0.91 for the step function and 0.08 for the linear function. Exponential and power weighting functions performed more poorly, with $\Delta$AICc values of 8.9 and 12.1 and model probabilities of 0.01 and 0.002, respectively. When the two subestuaries with the highest PCB levels were removed, the step function still provided the strongest model, but its AICc differences with the linear, exponential, and power functions were reduced ($\Delta$AICc 1.3, 1.9, and 4.2, respectively, Table S7).

How does weighting affect apparent land use proportions?

Different weighting functions applied in double-weighted models reduced effective source and total areas by differing degrees. Compared to the unweighted total area of 2720 km$^2$, weighted total areas for step, linear, exponential, and power functions were reduced to 750, 594, 486, and 116 km$^2$, respectively, while commercial land area was reduced from the unweighted 118 km$^2$ to 42.3, 33.4, 27.4, and 0.71 km$^2$, respectively. Smaller and smaller fractions of the catchment appeared to contribute to material delivery as the strength of weighting increased, and the contributing fraction of catchment was especially small with inverse distance weighting.

Weighted commercial land cover proportions ($G$) were higher than unweighted proportions ($F$) for some catchments and lower for others (Figs. 2, S4; Table S8). All four weighting functions increased $G$ relative to $F$ for watershed 32 and did the opposite for watershed 29. Watershed 3 and 4 showed smaller changes. The remaining watersheds had lower commercial land proportions that were not shifted much by weighting.

Shifts in $G$ relative to $F$ increased the correlation of the dependent variable (PCB concentration) with $G$ compared to the correlation of PCB with $F$, thus increasing the explanatory power ($R^2$) of weighted compared to unweighted regressions (Fig. 2; Table 1). The increase in $G$ relative to $F$ for watershed 32 had the strongest leverage in increasing the correlation of PCB versus $G$ compared to the correlation of PCB versus $F$ and in yielding a stronger weighted regression (higher $R^2$, lower AICc) with narrower confidence limits compared to the unweighted regression. The decrease in $G$ relative to $F$ for watershed 29 had the next strongest effect, followed by watersheds 4 and 3.

Where in the system do materials originate?

Distance weighting changed where PCBs appeared to originate within the study catchments. The four weighting functions had different effects, and the effects of each function differed among catchments.

All weighting functions increased weighted source proportions ($g_b$) relative to simple proportions ($f_b$) for small distances while decreasing weighted
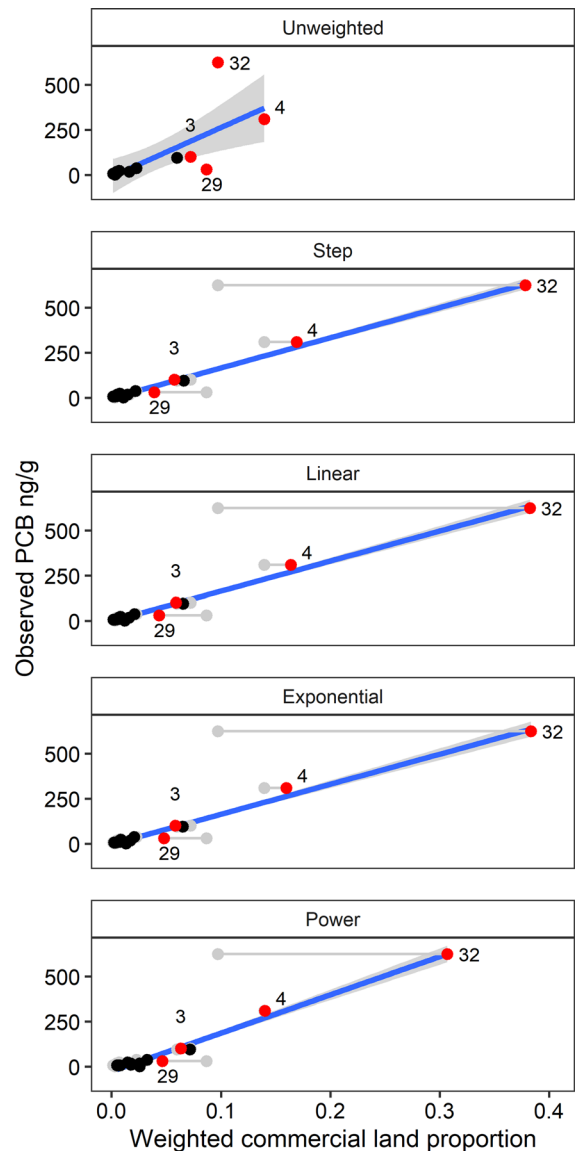


**Fig. 2** Regressions of PCB concentration versus commercial land proportion (blue lines with gray confidence intervals) for unweighted proportions and for weighted proportions from four different distance weighting functions applied with double weighting. Four numbered watersheds are highlighted in red. Grey lines show how distance weighting shifts the commercial land proportions from their unweighted values (gray dots) to distance-weighted values (black or red dots)

proportions relative to simple proportions at larger distances (Fig. 3). The power function generated the greatest increases in effective proportion at small distances, followed in order by the exponential, linear, and step functions. The same patterns were seen
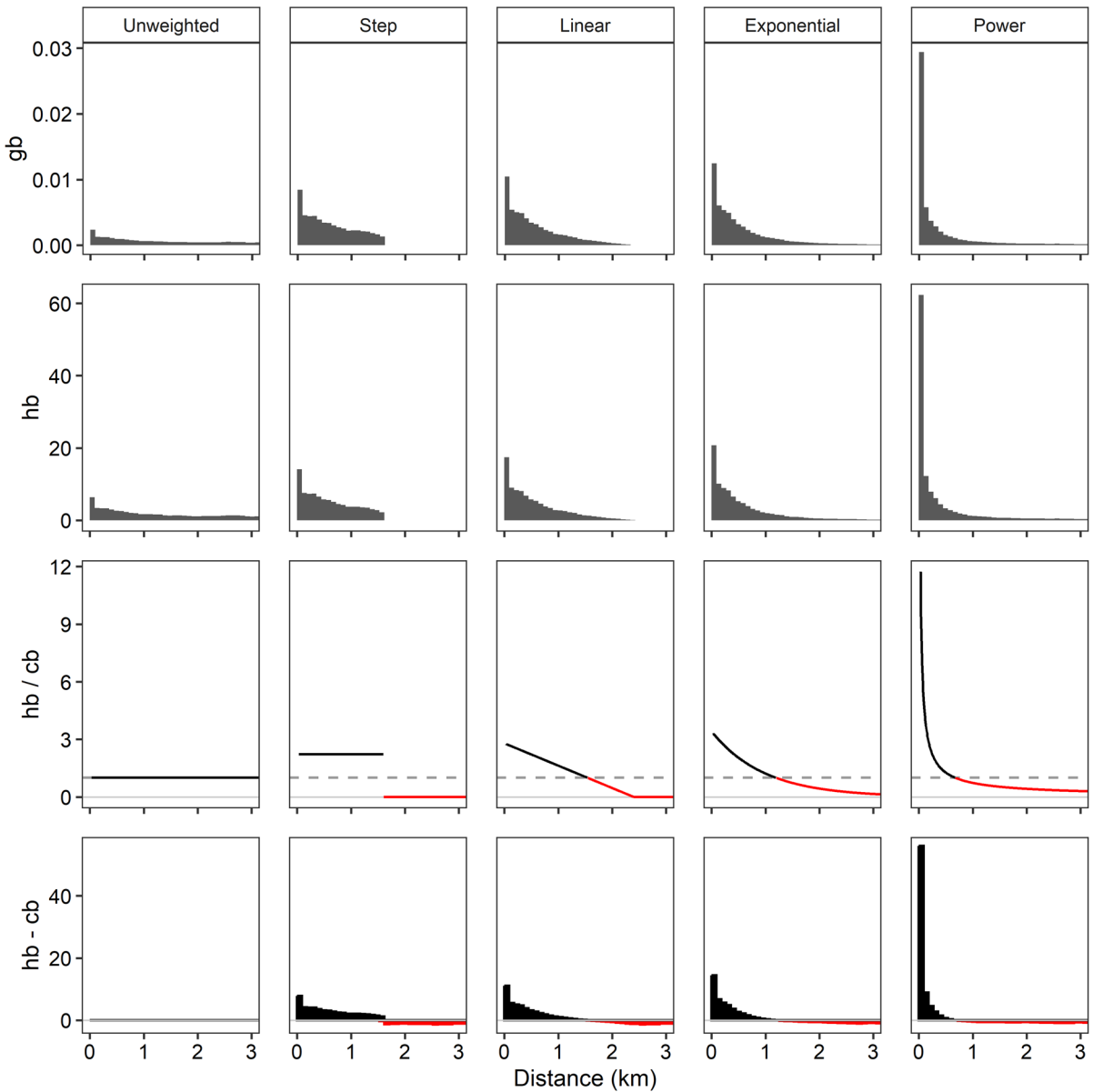
**Fig. 3** Distance distributions of four measures of material source strength for the aggregate of all study watersheds. The measures are (top) distance-weighted commercial land proportion ($g_b$), (second row) distance-weighted material source strength ($h_b$), (third row) the ratio $h_b/c_b$ of weighted material delivery to unweighted material delivery (the realized weighting), and (bottom) the difference between distance-weighted material delivery and unweighted material delivery ($h_b - c_b$). In the lower two panels, data for distances where weighted delivery is greater than or equal to unweighted delivery ($h_b \geq c_b$) are shown in black, while data for distances where weighted delivery is less than unweighted delivery ($h_b < c_b$) are shown in red

in other measures: material delivery ($h_b$) (Fig. 3), the minimum distance needed to supply certain percentages of material delivery (Table S9), and the cumulative distribution of delivery (Fig. 4). In the unweighted model, land less than 2811 m from the subestuary supplied half of the material delivery across the entire study area, while the step function model indicated land less than 516 m supplied half of material delivery. The distance decreased further to
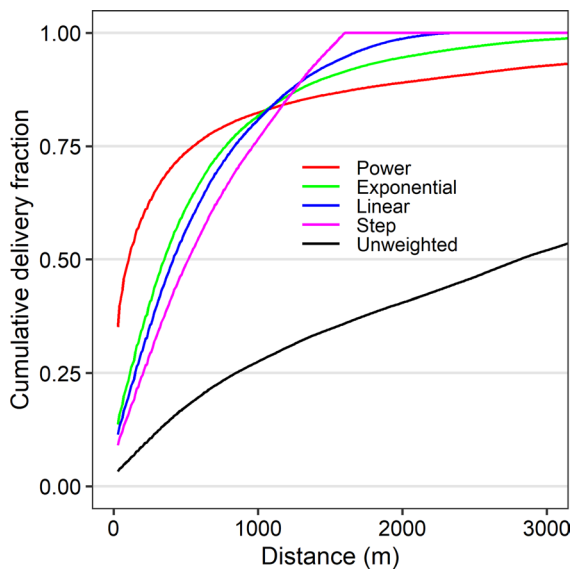
**Fig. 4** Cumulative fraction of total material delivery versus distance from a subestuary for the aggregate of all study watersheds. Cumulative distributions are shown for double-weighted models with four distance weighting functions as well as for the unweighted model

413, 351, and 101 m for the linear, exponential, and power functions, respectively (Fig. 4, Table S9).

The distributions of material delivery relative to the unweighted model ($h_b - c_b$, Eq. 18, Fig. 3) for individual catchments (Fig. 4) also explain how weighting interacted with the distributions of source land and total area to produce shifts in $G$ relative to $F$ (Sect. 3.3). The greatest change with all four weighting functions was a more than threefold increase in $G$ relative to $F$ for watershed 32 (Figs. 2, S4; Table S8). The subestuary of watershed 32 had the highest measured PCB concentration and its distributions of land were unusual (Fig. 1). The distribution of source (commercial) land $f_b$ versus distance had a strong peak at small distance (around 300 m), reflecting higher prevalence of source land near the subestuary shorelines than in the aggregate of all the catchments. In addition, the distribution of catchment area in watershed 32 fell off more slowly with distance than in the other watersheds because a greater fraction of total catchment area was further from the subestuary than in the aggregate. Weighting functions that increased effective proportion $g_b$ relative to simple proportion $f_b$ at small distances strongly increased overall effective proportion $G$ in this catchment where source land was more prevalent near the subestuary. Watershed 32 had very high values of

$g_b - f_b$ and $h_b - c_b$ at small distances with any weighting function (Fig. 5).

Watershed 29 showed the next greatest change in $G$ relative to $F$, but in the opposite direction: $G$ was reduced to roughly half of $F$ by any of the weighting functions (Fig. 2, Figure S4, Table S8). In watershed 29, the peak of the distribution of commercial land relative to distance was at larger distances, peaking around 2900 m (Fig. 1). Compared to watershed 32, less commercial land occurred close to the subestuary where weighting increased effective proportions $g_b$ relative to $f_b$, whereas more occurred farther away where weighting reduced $g_b$ relative to $f_b$ (Fig. 5). Watershed 32 showed big gains in $g_b$ and $h_b$ at small distances and minor changes in $g_b$ and $h_b$ at larger distances. This led to the large increase in $G$ relative to $F$. In contrast, low $f_b$ values close to the subestuary in watershed 29 provided little opportunity for weighting to yield high $g_b$ values, but weighting did lead to large reductions in $g_b$ relative to $f_b$ at larger distances where much commercial land was located (Fig. 5).

## Realized weighting

Realized weighting for any of the weighting functions (black lines, Figure S7) emphasized nearby land relative to distant land even more strongly than suggested by the weighting function alone (red lines, Figure S7). Exaggeration of realized weighting relative to the weighting function was greatest for the power function and less for the exponential, linear, and step functions, in that order. Realized weighting also differed among catchments because the realized effect depends on the spatial distributions of both source and total land (Eq. 19), which differed among catchments. The ratio of realized weighting to the weighting function was a distance-independent constant that differed among catchments (factor $J$ in Eq. 19, Table S8).

## Discussion

We have presented a comprehensive set of methodological improvements for fitting and interpreting spatially weighted regressions relating land use to aquatic responses. We tested our methods by reanalyzing a published data set to focus on the methodological enhancements rather than on introducing new data. King et al.
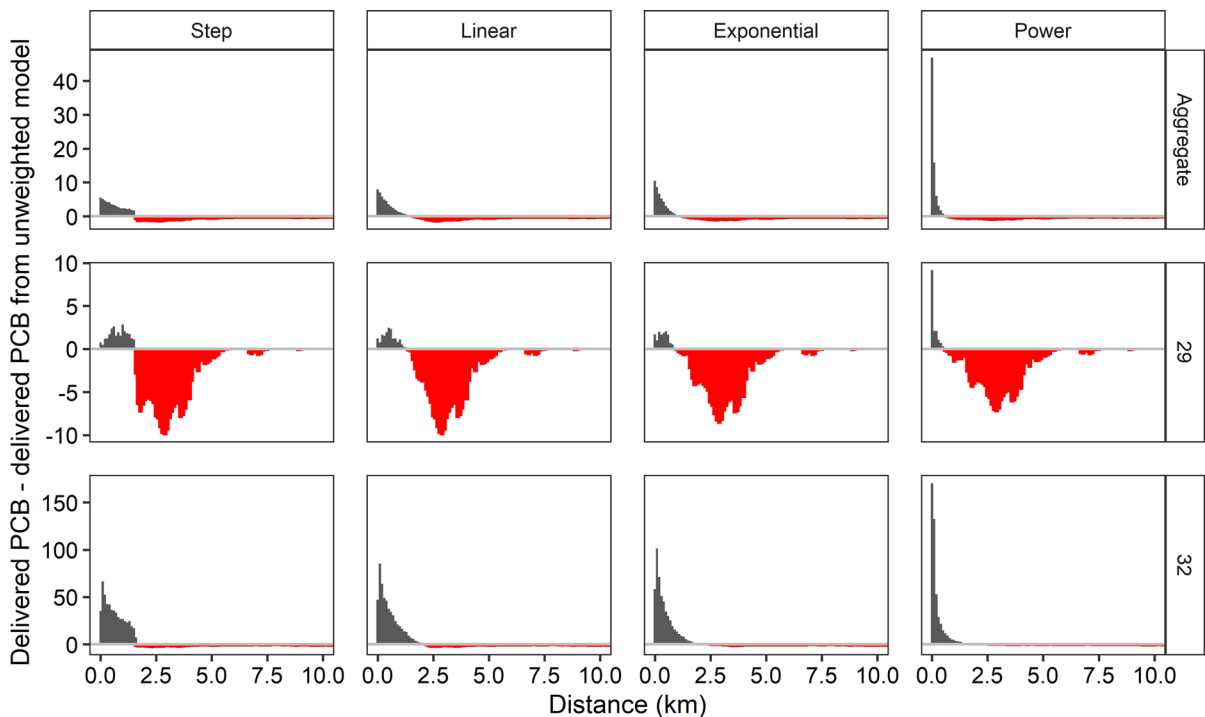
**Fig. 5** Changes in PCB delivery with distance in double-weighted models with four distance weighting functions relative to the unweighted model, calculated as weighted delivery minus unweighted delivery ($h_b - c_b$) at each distance. Black bars show how much weighted delivery exceeds unweighted delivery, while red bars show how much weighting reduces delivery relative to the unweighted model. Distributions are shown for the watersheds of two subsestuaries and for the aggregate of all the watersheds

(2004) had already documented that distance weighting yields strong increases in explanatory power for the test data. Our new methods were able to enhance an already strong published analysis by providing more rigorous answers to the following four questions:

Is there an effect of arrangement?

The first challenge for models incorporating landscape pattern is demonstrating an arrangement effect. Model comparison with AICc demonstrated that even the poorest of the four double-weighted models for the test data has a model probability of 1.0 compared to an unweighted model–rigorous, probabilistic evidence that spatial arrangement is important and should be included in relating land use to the observed PCB concentrations in fish. The previous study reported that distance-weighted models increased $R^2$ relative to unweighted models (King et al. 2004), but that measure alone provides much weaker evidence for an arrangement effect.

What is the best weighting function?

Specifying the weighting function is also an important step in spatially weighted regression. Most studies have just used inverse distance weighting ($1/d$). King et al. (2004) tested inverse power functions ($d^{-p}$) with three different $p$ values. Other studies considered additional mathematical functions and estimated weighting parameters with response surfaces created from small sets of possible values (Van Sickle and Johnson 2008; Walsh and Webb 2014). We tested several functions, optimized weighting and regression parameters simultaneously, and compared models with AICc. For the test data, we identified an inverse distance power function with $p = 0.80$ that gave a stronger weighted model (model probability = 0.92, Table S3) than originally reported with $p = 1$ (King et al. 2004). Furthermore, exponential, step, and linear weighting functions all produced even stronger models.

A step function gave the best distance-weighted model for the test data. Considering only land use within a fixed distance is a simple idea that is easy to implement with the GIS "buffer" function, making it the most common way to represent a greater impact of nearby land than distant land. However, our results do *not* support this method because our implementation of the step function has important differences from applying a fixed width buffer. We treated the threshold distance for the step function as a parameter to be estimated from the data rather than an arbitrarily, a priori value; and the step function was judged best only when it excelled compared to other functions. This approach brings the idea of a fixed zone of impact under the more general and rigorous umbrella of spatially weighted regression.

The best weighting function will likely differ among data sets that represent different variables and processes. For some problems, a priori considerations may suggest a particular weighting function. For example, the exponential weighting function can represent a distance effect arising from a first order uptake process while inverse distance models are often used for processes like seed dispersal (Canham and Uriarte 2006). We recommend comparing models with different weighting functions. Better performing models for other published data sets might have been found if more weighting functions and parameter values had been considered, and comparing models can yield insight into likely mechanisms. For future analyses, our general framework for fitting and comparing multiple models should become a standard approach for catchment-scale assessment of land use impacts on aquatic responses. Applying our methods to other data sets will test the generality of our findings with the test data.

How should total land be weighted?

The question of how to weight total land in the denominator of the weighting equation was not thoroughly discussed in previous papers. For the test data, there was very strong support for weighting total land the same way as source land (double weighting) and virtually none for leaving source land unweighted (Table 2). Double weighting also performed better than using different weighting parameters for source and total land (separate weighting) although $R^2$ and AICc values for this

comparison were similar. Conceptually, the superiority of double weighting suggests that PCBs from source land and water from all land decline with distance from the estuary in a similar way. The test data results support the common choice of double weighting (as in King et al. 2005, 2004; Van Sickle and Johnson 2008; Peterson et al. 2011; Sheldon et al. 2012; Walsh and Webb 2014). We expect that the best choice will differ among problems and data sets. More studies need to explicitly compare single, double, and separately weighted models to assess the generality of our findings for other problems and processes. Larger data sets could also support fitting more weighting parameters and so provide more scope for separate weighting to yield different results from double weighting. There may also be cases where weighting of numerator and denominator should use different mathematical functions, not just the same function with a different parameter.

Where in the system do materials originate?

Our new interpretative tools reveal how weighting affects the attribution of where effects originate. Comparing the spatial distributions of weighted source proportion (Eq. 14) and weighted material delivery (Eq. 15) to the corresponding distributions for an unweighted model reveals how weighting affects the relative contributions of land near to and more distant from receiving waters (Figs. 3, 5; Table S9). Such comparisons enable us to visualize, quantify, and interpret the impacts of different weighting functions. These analyses also reveal how weighting effects differ among catchments (Figs. 5, S5, S6; Tables S8, S9) because of differences in spatial distributions of source land and total land (Fig. 1). This is a key strength of our approach. Previous studies with distance-weighted land use regressions could not offer such interpretations because they relied only on plots of the weighting function versus distance (e.g., King et al. 2004) or on half decay distances (Van Sickle and Johnson 2008; Walsh and Webb 2014). To correctly quantify the weighting effect, one must integrate effects of applying a weighting function in both the numerator and denominator of the weighting equation as well as the actual distributions of source and total land, as achieved here. Accomplishing this provides

enhanced guidance for environmental management. In the test case, the best model suggests that half the PCBs originate within just 0.516 m of a subestuary (Fig. 4, Table S9), so that concentrating management actions to reduce sources within that specific zone would have the maximum benefit.

Generality and applicability

The methodological enhancements provided here make it easier to conceptualize and operationalize models linking landscape pattern to biophysical processes. For simplicity and clarity, our language has focused on the context of materials released from source areas and transported hydrologically to response points in aquatic systems. Studies in this vein have applied distance weighting to understand both water quality and biological responses (e.g., Frimpong et al. 2005; King et al. 2005; Van Sickle and Johnson 2008; Peterson et al. 2011). However, the underlying mathematics are completely general and can be applied to other contexts where patterns affect processes, such as understanding the impacts of proximate physiography on ground water delivery, land cover on animal occupancy or movement within landscape patches, or the influence of surrounding land cover on urban heat (e.g., Baker et al. 2003; Cushman et al. 2008; Cunningham and Johnson 2016; Dixon et al. 2020; Alonzo et al. 2021).

Different transport or movement processes can be represented by choosing different measures of distance. In the test case, we measured distance as the shortest path from each land pixel to the receiving water (Cartesian distance). Other ways of expressing can represent different levels of specificity about effect transmission. Many studies have measured distance along downhill flow paths (e.g., King et al. 2005, 2012; Baker et al. 2006; Walsh and Webb 2014; Weller et al. 2011, 1998; Weller and Baker 2014) or along in-stream flow paths through a stream network (e.g., King et al. 2005; Sheldon et al. 2012; Walsh and Webb 2014). Distance measures can also integrate other hydrological factors, such as local slope, flow path gradient, upslope contributing area (Peterson et al. 2011), or other spatial indices of hydrologic function. Because models are fit by numerical optimization rather than linear regression, our method can also be applied to problems where the underling response model is non-linear.

Our methods can be further extended to questions arising from many disciplines, including ecology, evolution, geography, geomorphology, epidemiology, public policy, and others. For problems in which spatial patterns can affect processes, the impacts of patterns on overall function can now be explicitly tested and quantified. Our methods can relate almost any spatially distributed source or driver to an integrated response at a point or along a boundary; and alternate hypotheses about the effects of pattern or proximity on processes can be tested with alternative weighting functions. As new applications are implemented, additional weighting functions will likely emerge and enhance our general approach. The evolving framework can be applied to inform understanding in a wide range of scientific investigations and applied contexts.

**Declarations**

# References

Allan JD, Erickson DL, Fay J (1997) The influence of catchment land use on stream integrity across multiple spatial scales. Fresh Biol 37(1):149–161

Alonzo M, Baker ME, Gao Y, Shandas V (2021) Spatial configuration and time of day impact the magnitude of urban tree canopy cooling. Environ Res Lett 16(8):084028

Baker ME, Wiley MJ, Carlson ML, Seelbach PW (2003) A GIS model of subsurface water potential for aquatic resource inventory, assessment, and environmental management. Environ Manag 32(6):706–719

Baker ME, Weller DE, Jordan TE (2006) Improved methods for quantifying potential nutrient interception by riparian buffers. Landsc Ecol 21(8):1327–1345

Baker ME, Weller DE, Jordan TE (2007) Effects of stream map resolution on measures of riparian buffer distribution and nutrient retention potential. Landsc Ecol 22(7):973–992

Benda L, Poff NL, Miller D et al (2004) The network dynamics hypothesis: how channel networks structure riverine habitats. Bioscience 54(5):413–427

Black PE (1997) Watershed functions. J Am Water Res Assoc 33(1):1–11

Bohlke JK, Denver JM (1995) Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, Atlantic coastal plain, Maryland. Water Resour Res 31(9):2319–2339

Burnham KP, Anderson DR (2002) Model selection and multimodel inference, 2nd edn. Springer, New York

Canham CD, Uriarte M (2006) Analysis of neighborhood dynamics of forest ecosystems using likelihood methods and modeling. Ecol Appl 16(1):62–73

Comeleo RL, Paul JF, August PV et al (1996) Relationships between watershed stressors and sediment contamination in Chesapeake Bay estuaries. Landsc Ecol 11(5):307–319

Cunningham MA, Johnson DH (2016) What you find depends on where you look: responses to proximate habitat vary with landscape context. Avian Conserv Ecol. https://doi.org/10.5751/ACE-00865-110201

Cushman SA, McGarigal K, Neel MC (2008) Parsimony in landscape metrics: strength, universality, and consistency. Ecol Indic 8(5):691–703

Dixon AP, Baker ME, Ellis EC (2020) Agricultural landscape composition linked with acoustic measures of avian diversity. Land 9(5):145

Frimpong EA, Sutton TM, Lim KJ et al (2005) Determination of optimal riparian forest buffer dimensions for stream biota-landscape association models using multimetric and multivariate responses. Can J Fish Aquat Sci 62(1):1–6

Goetz S, Fiske G (2008) Linking the diversity and abundance of stream biota to landscapes in the mid-Atlantic USA. Remote Sens Environ 112(11):4075–4085

Herlihy AT, Stoddard JL, Johnson CB (1998) The relationship between stream chemistry and watershed land cover data in the Mid-Atlantic region, U.S. Water Air Soil Pollut 105:377–386

Hunsaker CT, Levine DA (1995) Hierarchical approaches to the study of water-quality in rivers. Bioscience 45(3):193–203

Hynes HBN (1975) The stream and its valley. SIL Proc 1922-2010 19(1):1–15

Johnson LB, Richards C, Host GE, Arthur JW (1997) Landscape influences on water chemistry in Midwestern stream ecosystems. Fresh Biol 37(1):193

Jones KB, Neale AC, Nash MS et al (2001) Predicting nutrient and sediment loadings to streams from landscape metrics: a multiple watershed study from the United States Mid-Atlantic region. Landsc Ecol 16:301–312

Jordan TE, Weller DE, Pelc CE (2018) Effects of local watershed land use on water quality in mid-Atlantic coastal bays and subestuaries of the Chesapeake Bay. Estuaries Coasts 41(S1):S38–S53

King RS, Beaman JR, Whigham DF, Hines AH, Baker ME, Weller DE (2004) Watershed land use is strongly linked to PCBs in white perch in Chesapeake Bay subestuaries. Environ Sci Technol 38(24):6546–6552

King RS, Baker ME, Whigham DF et al (2005) Spatial considerations for linking watershed landcover to ecological indicators in streams. Ecol Appl 15(1):137–153

King RS, Deluca WV, Whigham DF, Marra PP (2007) Threshold effects of coastal urbanization on *Phragmites australis* (common reed) abundance and foliar nitrogen in Chesapeake Bay. Estuaries Coasts 30(3):469–481

King RS, Walker CM, Whigham DF, Baird SJ, Back JA (2012) Catchment topography and wetland geomorphology drive macroinvertebrate community structure and juvenile salmonid distributions in south-central Alaska headwater streams. Freshw Sci 31(2):341–364

Liu Z-J, Weller DE, Correll DL, Jordan TE (2000) Effects of land cover and geology on stream chemistry in watersheds of Chesapeake Bay. J Am Water Resour Assoc 36(6):1349–1365

Lowrance R, Altier LS, Newbold JD et al (1997) Water quality functions of riparian forest buffer systems in the Chesapeake Bay watershed. Environ Manag 21(5):687–712

Nash JC, Varadhan R (2011) Unifying optimization algorithms to aid software system users: optimx for R. J Stat Softw 43(9):1–14

Omernik JM, Abernathy AR, Male LM (1981) Stream nutrient levels and proximity of agricultural and forest land to streams: Some relationships. J Soil Water Conserv 36(4):227–231

O'Neill RV, Hunsaker CT, Jones KB et al (1997) Monitoring environmental quality at the landscape scale. Bioscience 47(8):513–519

Osborne LL, Wiley MJ (1988) Empirical relationships between land use/cover and stream water quality in an agricultural watershed. J Environ Manag 26:9–27

Peterson EE, Pearse AR (2017) IDW-Plus: an ArcGIS toolset for calculating spatially explicit watershed attributes for survey sites. JAWRA J Am Water Resour Assoc 53(5):1241–1249

Peterson EE, Sheldon F, Darnell R, Bunn SE, Harch BD (2011) A comparison of spatially explicit landscape representation methods and their relationship to stream condition. Freshw Biol 56(3):590–610

Rinaldo A, Vogel GK, Rigon R, Rodriguez-Iturbe I (1995) Can one gauge the shape of a basin? Water Resour Res 31(4):1119–1127

Sheldon F, Peterson EE, Boone EL, Sippel S, Bunn SE, Harch BD (2012) Identifying the spatial scale of land use that most strongly influences overall river ecosystem health score. Ecol Appl 22(8):2188–2203

Snyder CD, Young JA, Villella R, Lemarie DP (2003) Influences of upland and riparian land use patterns on stream biotic integrity. Landsc Ecol 18(7):647–664

Soranno PA, Hubler SL, Carpenter SR, Lathrop RC (1996) Phosphorus loads to surface waters: a simple model to account for spatial pattern of land use. Ecol Appl 6(3):865–878

Sponseller RA, Benfield EF, Vallet HM (2001) Relationships between land use, spatial scale and stream macroinvertebrate communities. Freshw Biol 46:1409–1424

Tufford DL, McKellar HN, Hussey JR (1998) In-stream nonpoint source nutrient prediction with land-use proximity and seasonality. J Environ Qual 27(1):100–111

Van Sickle J, Johnson CB (2008) Parametric distance weighting of landscape influence on streams. Landsc Ecol 23(4):427–438

Walsh CJ, Kunapo J (2009) The importance of upland flow paths in determining urban effects on stream ecosystems. J N Am Benthol Soc 28(4):977–990

Walsh CJ, Webb JA (2014) Spatial weighting of land use and temporal weighting of antecedent discharge improves prediction of stream condition. Landsc Ecol 29(7):1171–1185

Weller DE, Baker ME (2014) Cropland riparian buffers throughout Chesapeake Bay watershed: spatial patterns and effects on nitrate loads delivered to streams. J Am Water Resour Assoc 50(3):696–712

Weller DE, Jordan TE, Correll DL (1998) Heuristic models for material discharge from landscapes with riparian buffers. Ecol Appl 8(4):1156–1169

Weller DE, Baker ME, Jordan TE (2011) Effects of riparian buffers on nitrate concentrations in watershed discharges: new models and management implications. Ecol Appl 21(5):1679–1695

Zhang T (2010) A spatially explicit model for estimating annual average loads of nonpoint source nutrient at the watershed scale. Environ Model Assess 15(6):569–581

Zhang T (2011) Distance-decay patterns of nutrient loading at watershed scale: regression modeling with a special spatial aggregation strategy. J Hydrol 402(3–4):239–249