Environmetrics Research Article

Received: 27 September 2013, Revised: 29 March 2014, Accepted: 21 April 2014, Published online in Wiley Online Library: 21 May 2014

(wileyonlinelibrary.com) DOI: 10.1002/env.2284

Spatial sampling on streams: principles for inference on aquatic networks

Nicholas A. Som[a]*, Pascal Monestiez[b], Jay M. Ver Hoef[c], Dale L. Zimmerman[d] and Erin E. Peterson[e]

For ecological and environmental data, prior inquiries into spatial sampling designs have considered two-dimensional domains and have shown that design optimality depends on the characteristics of the target spatial domain and intended inference. The structure and water-driven continuity of streams prompted the development of spatial autocovariance models for stream networks. The unique properties of stream networks, and their spatial processes, warrant evaluation of sampling design characteristics in comparison with their two-dimensional counterparts. Common inference scenarios in stream networks include spatial prediction, estimation of fixed effects parameters, and estimation of autocovariance parameters, with prediction and fixed effects estimation most commonly coupled with autocovariance parameter estimation. We consider these inference scenarios under a suite of network characteristics and stream-network spatial processes. Our results demonstrate, for parameter estimation and prediction, the importance of collecting samples from specific network locations. Additionally, our results mirror aspects from the prior two-dimensional sampling design inquiries, namely, the importance of collecting some samples within clusters when autocovariance parameter estimation is required. These results can be applied to help refine sample site selection for future studies and further showcase that understanding the characteristics of the targeted spatial domain is essential for sampling design planning. Published 2014. This article has been contributed to by US Government employees and their work is in the public domain in the USA.

Keywords: sampling design; spatial statistics; stream networks

1. INTRODUCTION

Efficient management and monitoring of natural resources requires informative data (Mahjouri and Kerachian, 2011), with inferences for spatial data affected substantially by the spatial configuration of sampled sites (Zimmerman, 2006); hence, a common issue in spatial statistics is how to choose an "optimal" set of sample locations (Zhu and Stein, 2006). Optimality for spatial sampling designs depends on the characteristics of the target spatial domain and intended inference. Regardless of the inference goal, estimation of spatial autocorrelation is among the essential steps in any geostatistical analysis (Gascuel-Odoux and Boivin, 1994), and poorly constructed sampling designs can lead to data lacking sufficient practical value (Strobl *et al.*, 2006). This is as true for stream and river networks as it is for the two-dimensional domain usually considered by geostatistics. Recently, new classes of geostatistical models for stream networks have been developed (summarized by Ver Hoef and Peterson, 2010), but the development of optimal stream-network designs under these models has received little attention. The overall goal of this manuscript is to study optimal spatial-sampling designs for the class of stream network models described in Ver Hoef and Peterson (2010).

There is a substantial and coherent body of work evaluating sampling designs for geostatistical data, which are commonly used for monitoring population status and trend, estimating relationships between ecological phenomena, and predicting values at unobserved locations. Sampling design optimality varies by these inference scenarios (Zhu and Stein, 2005), prior knowledge of the covariance parameters (Zimmerman, 2006), and the strength of spatial autocorrelation, among other factors. Characteristics of optimal designs for estimation of a constant mean with known covariance parameters include some samples uniformly spaced throughout the domain and the rest focused near domain boundaries (Xia *et al.*, 2006). To estimate spatial trend, clusters are needed in areas of high leverage, as well as samples near domain boundaries (Müller, 2005). When covariance parameters are also estimated, high densities of samples near the boundary are replaced with

* *Correspondence to: Nicholas A. Som, US Fish & Wildlife Service, Arcata FWO, Arcata, CA, U.S.A. E-mail: nicholas_som@fws.gov*

a *US Fish & Wildlife Service, Arcata FWO, Arcata, CA, U.S.A.*

b *Biostatistique et Processus Spatiaux, INRA, Avignon, France*

c *National Marine Mammal Laboratory, NOAA Fisheries, Seattle, WA, U.S.A.*

d *Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA, U.S.A.*

e *CSIRO Computational Informatics, Dutton Park, QLD, Australia*

samples distributed throughout the domain with some clusters, and fewer samples at high leverage locations (Müller and Stehlk, 2010). Similarly, the best designs for prediction with known covariance parameters and a constant mean are uniformly distributed across the domain, while locations are allocated to high leverage areas when trend is estimated. These cases contrast with optimal designs developed specifically for the estimation of covariance parameters, which include a distribution of locations with many small and large separation distances (Zimmerman, 2006; Zhu and Zhang, 2006; Irvine *et al.*, 2007).

The results described in the preceding paragraph rely on "model-based" inference, which assumes that the data are generated by a spatial stochastic process. In contrast, "design-based" inference assumes that the data are fixed, and all inference relies on random properties of the design (Ver Hoef, 2002; Stevens Jr., 2006). Sample designs for monitoring and inference in aquatic networks have been commonly chosen with design-based inference in mind, where the inference goals include estimation of population means or totals, and designs commonly include spatially balanced samples that broadly cover the domain (Al-Chokhachy *et al.*, 2011). Samples of this type result in more efficient samples (under the design-based paradigm) through less redundant information (Stevens Jr. and Jensen, 2007) and maintain randomness to avoid biases (Herlihy *et al.*, 2000). They are broadly similar to those outlined for model-based inference, except where fine-scale clustering is indicated to better estimate spatial autocorrelation.

The aforementioned results are satisfactory for estimation and prediction in the traditional geostatistical, two-dimensional setting. They may not, however, be sufficient for spatial processes on stream networks. Each stream network consists of a set of linear segments that are connected by flowing water. The direction of water flow and the branching structure of the network strongly influence the spatial relationships within a stream network. This branching structure forms distinctly different network shapes, from linear trellis networks to oval-shaped, pear-shaped, or heart-shaped dendritic networks, depending on the topography and geology of the region (Benda, 2008). The shape and complexity of the network controls the spatial configuration of confluences (i.e., locations where stream segments converge), which strongly influences the heterogeneity of chemical, physical, and biological processes throughout the network (Fagan, 2002; Benda *et al.*, 2004; Labonne *et al.*, 2008). For example, the diversity of instream habitat such as flow characteristics, woody debris, or substrate size is expected to be higher near confluences within a dendritic network compared with a trellis network because segments of similar size converge causing disturbance and heterogeneity in the environment (Benda, 2008).

These traits have led to further developments in the study of covariance models for stream networks. Ver Hoef *et al.* (2006) utilized moving-average constructions to develop spatial covariance models for stream networks based on stream distance. These models also account for the relative flow contribution between segments at confluences and are discussed in more detail in the succeeding text. Stream-network moving-average models allow for the modeling of a broad suite of stream network-based spatial processes. Thus, their unique properties warrant an evaluation of sampling design characteristics in comparison with their traditional geostatistical counterparts.

In this work, we consider optimal sampling designs for stream networks in the context of common inference scenarios, more specifically, fixed effects estimation and prediction (each with known and unknown covariance parameter values), and the estimation of only covariance parameters. We pay particular attention to the role that confluences, flow connectivity, and directionality might play in regards to sample site selection. In Section 2, we begin with an introduction to spatial moving-average models for stream networks. In Section 3, we first define our sampling design criterion, and then use a small "toy" stream network to find locally optimal sampling designs under a suite of spatial stream-network covariance processes, spatial autocorrelation strengths, fixed effects structures, and segment weighting schemes. In Section 4, results from the toy networks are then pursued in larger simulated networks to explore the relative merits of sampling design choice and the potential impacts of contrasting stream-network shapes. We conclude with a discussion of the results and sampling design recommendations for the most common inference scenarios.

2. MOVING AVERAGE AUTOCOVARIANCE MODELS FOR STREAM NETWORKS

Mathematically, we envision a stream network as a set of lines (stream segments) that branch upstream from the most downstream segment on the network (outlet segment) to the most upstream segments on the network (headwater segments; Figure 1). We assume that the branching is binary (i.e., three or more segments never branch upstream from the same confluence). Observations are represented by points on the network, which have two coordinate systems (Peterson *et al.*, 2013); one is the usual two-dimensional coordinate system and the other is based on the network topology (i.e., branching structure and connectivity of segments). Note that separation distance between two locations along the network is the shortest distance between them when movement is confined to the network (e.g., stream distance; Dent and Grimm, 1999). If water flows from an upstream location to a downstream location, we refer to these locations as flow-connected (Figure 1, A and C, B and C) and refer to two locations within the same network not connected by flowing water as flow-unconnected (Figure 1, A and B).

Some new models for stream networks, based on moving average constructions, were initially described by Ver Hoef *et al.* (2006) and Cressie *et al.* (2006). In these models, the observed values measured at points are assumed to come from one realization of a continuous spatial random field on the network (i.e., samples could be collected continuously along the network). The models summarized in Ver Hoef and Peterson (2010) use a spatial moving-average approach to construct Gaussian random fields based on the network topology, rather than the usual two-dimensional coordinate system commonly used in geostatistics. This approach yields random processes that are similar to typical geostatistical models; they can be described by a mean function that depends on the location within the network, and a second-order, stationary covariance function that depends on the separation distance between two locations.

The fact that streams are branching and have flow creates a rich set of models not seen in either time series analysis or spatial statistics. Using the moving average constructions, if a moving average function starts at some location and is non-zero only upstream of that location, it is called a "tail-up" model. The function must split at confluences as it goes upstream to maintain stationarity of variances, so some weighting of segments must occur. If a moving average function starts at some location and is non-zero only downstream of that location, it is called a "tail-down" model. Consider two pairs of sites that have the same stream distance between them, but one pair is flow-connected and the other pair is flow-unconnected; in general, the amount of autocorrelation will be different between them. For the following development,

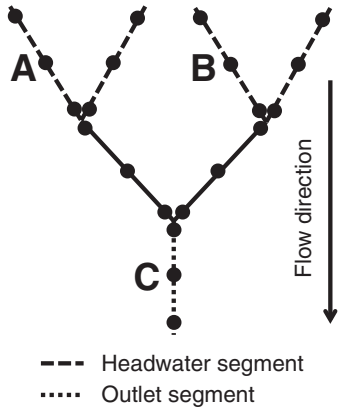

Figure 1. Toy network with seven stream segments and 21 potential sampling locations. Dashed lines represent headwater segments, the dotted line represents the outlet segment, and solid lines represent segments in the network between the outlet and headwater segments. Black circles represent sampling locations at 0.01, 0.50, and 0.99 units along each segment. The arrow shows the direction of flowing water within the network. A, B, and C represent three network locations. In this network, C is flow-connected to both A and B, but A and B are flow-unconnected

let r_i and s_j denote two locations on a stream network, and let h be the stream distance between them. Then, the following models have been developed to describe different forms of covariance of the response at locations r_i and s_j.

The moving average construction for tail-up models, as described by Ver Hoef *et al.* (2006), is

$$C_u(r_i, s_j|\boldsymbol{\theta}_u) = \begin{cases} \pi_{i,j} C_t(h|\boldsymbol{\theta}_u) & \text{if } r_i \text{ and } s_j \text{ are flow-connected} \\ 0 & \text{if } r_i \text{ and } s_j \text{ are flow-unconnected} \end{cases} \quad (1)$$

where $C_u(r_i, s_j|\boldsymbol{\theta}_u)$ is the spatial autocovariance between r_i and s_j, u denotes a tail-up model, $\boldsymbol{\theta}_u$ is the set of covariance function parameters, $C_t(h|\boldsymbol{\theta}_u)$ is the value of a covariance function based on h, $\boldsymbol{\theta}_u$, and a selected covariance model (e.g., exponential), and $\pi_{i,j}$ are weights to account for the branching characteristics of the stream and maintain variance stationarity.

For the weights, $\pi_{i,j}$, consider the following. First, associate a value of 1 with each headwater segment in a stream network. When two segments join at a confluence, the value for the segment downstream of the confluence is the sum of the two upstream values, and this summation continues downstream until the network outlet is reached. Segment values formed in this manner are known as Shreve's stream order (Shreve, 1967) and are an example of an additive function. These additive values are a property of whole stream segments and are applied to any point along a particular segment. If we denote the value of the additive function as $\Omega(x)$ for some point x on the stream network, then the weight for two flow-connected points where r_i is downstream from s_j is

$$\pi_{i,j} = \sqrt{\frac{\Omega(s_j)}{\Omega(r_i)}}; \pi_{i,j} \in (0, 1)$$

The weights reflect the relative shared flow among locations, and again, serve to maintain stationary variances. More details can be found in Ver Hoef and Peterson (2010), including ways to create an additive function from values associated with stream segments, such as flow volume, a proxy for flow volume (e.g., basin area), or any other ecologically relevant variable.

Herein, we feature the exponential stream-network covariance function because its geostatistical counterpart is frequently applied by practitioners and has been featured in previous optimal sampling design work in the two-dimensional setting (e.g., Zimmerman (2006) and Xia *et al.* (2006)). For the exponential stream-network covariance function, $C_t(h|\boldsymbol{\theta}_u)$ has the following form (Ver Hoef *et al.*, 2006):

$$C_t(h|\boldsymbol{\theta}_u) = \sigma_u^2 \exp(-3h/\alpha_u) \quad (2)$$

where $\sigma_u^2 > 0$ is an overall variance parameter (also known as the partial sill), $\alpha_u > 0$ is the range parameter, and $\boldsymbol{\theta}_u = \left(\sigma_u^2, \alpha_u\right)'$. Via (1), spatial autocorrelation is only permitted between flow-connected locations in the tail-up model.

For tail-down models, spatial autocorrelation is permitted between both flow-connected and flow-unconnected locations, but we generally distinguish between the two cases. When two sites are flow-unconnected, there will always be at least one common confluence (i.e., a downstream confluence that receives water from each of the two upstream sites). Let b denote the longer of the two distances to the closest common downstream confluence, and a denote the shorter of the two distances. If two sites are flow-connected, again use h to denote their stream distance. Again, the only tail-down model we consider is the exponential, defined as follows:

$$C_d(a, b, h|\boldsymbol{\theta}_d) = \begin{cases} \sigma_d^2 \exp(-3h/\alpha_d) & \text{if flow-connected} \\ \sigma_d^2 \exp(-3(a+b)/\alpha_d) & \text{if flow-unconnected} \end{cases} \quad (3)$$

where $C_d(a, b, h|\boldsymbol{\theta}_d)$ is the spatial autocovariance between r_i and $s_j, \sigma_d^2 > 0$ is an overall variance parameter, $\alpha_d > 0$ is the range parameter, $\boldsymbol{\theta}_d = \left(\sigma_d^2, \alpha_d\right)'$, and d denotes a tail-down model. We note, for the exponential model, that when $a + b = h$, the flow-connected and flow-unconnected models are equivalent, and stress is a unique property of the exponential form of tail-down covariance models (Garreta *et al.*, 2010). A full development and more detail regarding the suite of stream-network moving-average models can be found in Ver Hoef and Peterson (2010).

A mixed linear model combining tail-up and tail-down components is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}_u + \mathbf{z}_d + \boldsymbol{\epsilon} \tag{4}$$

where $\mathbf{Y}$ is the vector of random variables for an observable stream attribute at sampled locations, $\mathbf{X}$ is a design matrix of fixed effects, $\boldsymbol{\beta}$ contains fixed effects parameters, $\mathbf{z}_u$ contains spatially autocorrelated random variables with a tail-up autocovariance (e.g., (2)), with $\text{var}(\mathbf{z}_u) = \sigma_u^2\mathbf{R}(\alpha_u)$, and $\mathbf{R}(\alpha_u)$ is a correlation matrix that depends on the range parameter α_u; $\mathbf{z}_d$ contains spatially autocorrelated random variables with a tail-down autocovariance (e.g., (3)) such that $\text{var}(\mathbf{z}_d) = \sigma_d^2\mathbf{R}(\alpha_d)$; and $\boldsymbol{\epsilon}$ contains independent random variables with $\text{var}(\boldsymbol{\epsilon}) = \sigma_0^2\mathbf{I}$. When used for spatial prediction, this model is referred to as "universal" kriging (Le and Zidek, 2006, p. 107), with "ordinary" kriging being the special case where the design matrix $\mathbf{X}$ is a single column of ones (Cressie, 1993, p. 119). This yields a covariance matrix of the form

$$\text{var}(\mathbf{Y}) = \boldsymbol{\Sigma} = \sigma_u^2\mathbf{R}(\alpha_u) + \sigma_d^2\mathbf{R}(\alpha_d) + \sigma_0^2\mathbf{I} \tag{5}$$

The mixed model in (4) is quite flexible. For example, including both tail-up and tail-down models may be useful for modeling stream-network attributes of active organisms that also respond to in-stream habitat (e.g., cutthroat trout abundance (Ganio *et al.*, 2005)), where both variance components might better model the spatial dependence than either model alone.

3. TOY NETWORKS

We next consider the inference scenarios of spatial prediction and fixed effects estimation, each with known and estimated covariance parameters, and estimation of only the covariance parameters, for stream-network spatial processes. We begin with a small, toy stream network. This dendritic-shaped network consists of seven unit-length stream segments, each with three potential sampling locations placed at stream distances of 0.01, 0.50, and 0.99 units along each segment (Figure 1). The network has four headwater segments and a single outlet segment that drains the entire network. The set W consists of the seven stream segments, and the set S the finite set of 21 points on W. We wish to find the six-point subset $\{s_1, \ldots, s_6\} \subset S$ that optimizes the unique design criterion for each inference scenario. We chose a sample size of six because it was large enough to reveal differences, from clustered to balanced designs, but small enough so that we could computationally evaluate the optimality criteria for all 54,264 designs. For each inference scenario, we applied tail-up, tail-down, and tail-up + tail-down exponential stream-network spatial-covariance functions and assumed that the underlying spatial process in the mixed model (4) was Gaussian. Segment weights for the tail-up model were apportioned in two ways: equally at each confluence or with the upstream-left segment receiving a weight of 0.8 and the upstream-right segment receiving 0.2. The range parameters (α_u and α_d) were also set in two ways: at approximately half the maximum stream distance or near the maximum stream distance. We did not alter the values of σ_u^2 and σ_d^2, setting each equal to 1, and set $\sigma_0^2 = 0$ (i.e., not incorporating a "nugget" effect). The nugget effect combines micro-scale variation (i.e., variation at distances smaller than observed), with measurement error (Schabenberger and Gotway, 2005, p. 139), and can result from unresolved scales of variability (Cressie and Wikle, 2011, p. 6). No-nugget models were selected to allow, to the maximum extent, the differences in design criteria to depend on the spatial effects, and note an increasing nugget component approaches the case of independence among sampled locations. We also note, that for the no-nugget model, changes to σ_u^2 and σ_d^2 simply scale the criteria values described in the succeeding text, and do not effect the rankings among designs. Finally, the fixed effects structure was either a constant mean, where $\mathbf{X}$ is a single column of ones, or a spatial trend representing a linear function of stream distance from the outlet, where $\mathbf{X}$ is a two-column matrix with the second column containing distances to the network outlet (Table 1). The vectors of estimated parameters depend on the fixed effects structure and spatial-covariance function. For the constant-mean case $\beta = (\beta_0)$ and for the spatial-trend case $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. Likewise, $\boldsymbol{\theta} = \left(\sigma_u^2, \alpha_u\right)'$ for tail-up covariance function, $\boldsymbol{\theta} = \left(\sigma_d^2, \alpha_d\right)'$ for the tail-down covariance function, and $\boldsymbol{\theta} = \left(\sigma_u^2, \alpha_u, \sigma_d^2, \alpha_d\right)'$ for the tail-up + tail-down covariance function. In general, the optimal design for a given criterion may depend on the covariance parameters; hence, we will use the term "locally optimal" to indicate optimality under a model having particular values of those parameters.

3.1. Covariance parameter estimation

We begin with the case of covariance parameter estimation. We assume that the covariance function is known only up to a parameter vector $\boldsymbol{\theta}$, and that the elements of $\boldsymbol{\theta}$ are unknown and estimated from the observed data. We sought a scalar summary of design quality that encapsulated the estimation of (potentially) multiple covariance parameters. We followed Zhu and Stein (2005) and Zimmerman (2006) by considering the inverse of the information matrix, $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$, as a reasonable approximation to the mean squared error matrix

$$\mathbf{M}(\boldsymbol{\theta}) = E\left[\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)'\right] \tag{6}$$

and our measure of design quality was the determinant of the inverse information matrix

$$\text{CP} = |[\mathbf{I}(\boldsymbol{\theta})]^{-1}| \tag{7}$$

Environmetrics 2014; **25**: 306–323 Published 2014. This article has been contributed to by US Government employees and their work is in the public domain in the USA. wileyonlinelibrary.com/journal/environmetrics

Table 1. Combinations of stream network covariance functions, segment weighting at each confluence ($\pi_{i,j}$), strength of autocorrelation (i.e., range parameter(s) α_u and/or α_d), and fixed effects mean structure (mean) for each inference scenario

Covariance function	$\pi_{i,j}$	$\alpha_{u,d}$	Mean
Tail-up	0.5/0.5	2	Constant
			Trend
		4	Constant
			Trend
	0.8/0.2	2	Constant
			Trend
		4	Constant
			Trend
Tail-down	-	2	Constant
			Trend
		4	Constant
			Trend
Tail-up + tail-down	0.5/0.5	2, 2	Constant
			Trend
		4, 4	Constant
			Trend

Zhu and Stein (2005) demonstrated that although the approximation of $\mathbf{M}(\boldsymbol{\theta})$ by $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$ is poor for small samples sizes, the relationship between their determinants is approximately monotonic and hence (7) is a suitable design criterion. In our toy examples, we used restricted maximum likelihood (REML) estimation, so the ijth element of $\mathbf{I}_{REML}(\boldsymbol{\theta})$ is

$$\frac{1}{2}\mathrm{tr}\left(\mathbf{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_i}\mathbf{Q}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_j}\right) \tag{8}$$

where tr represents the matrix trace operator and $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$. We thus defined the locally CP-optimal six-point design as that which minimized $|[\mathbf{I}_{REML}(\boldsymbol{\theta})]^{-1}|$ among all six-point designs.

The results for all locally CP-optimal designs contained at least a portion of samples within "clusters," defined here as either adjacent site pairs/triads on the same segment or the triad of sites within stream distance 0.02 of each other at a stream confluence (Figure 2; and hereafter, the filled black circles represent each locally optimal six-point design). Which of these cluster types occurred, and what their distribution was within the network, depended considerably on the spatial process, to a lesser extent on the mean function, and little on the strength of spatial autocorrelation. For the tail-up process, all clusters were adjacent site pairs or triads that occurred on stream segments that were flow-unconnected yielding highly correlated observations within, but completely uncorrelated observations across, clusters. The tail-up spatial-trend case favored triads, while the constant-mean case favored pairs, and there was no effect of the segment weighting scheme. Locally CP-optimal designs for the tail-down process featured one confluence-based cluster plus single samples in headwater segments, of which at least two were separated from the cluster by distances at or near the maximum distance. In addition, differences due to mean function appeared to be small. The tail-up + tail-down scenario exhibited attributes of each model: pairs/triads of samples were found within segments, in some cases accompanied by single samples in locations separated from the clusters by the maximum stream distance (or nearly so).

3.2. Prediction with known covariance parameters

We next consider spatial prediction, commonly referred to as kriging, and begin by assuming that the covariance function and parameter vector $\boldsymbol{\theta}$ are known. Under model (4), the best linear unbiased predictor (BLUP) of an unobserved value of Y at a stream network location $s_0 \in S$ is

$$\hat{Y}(s_0) = [\mathbf{c} + \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{x}_0 - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c})]'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \tag{9}$$

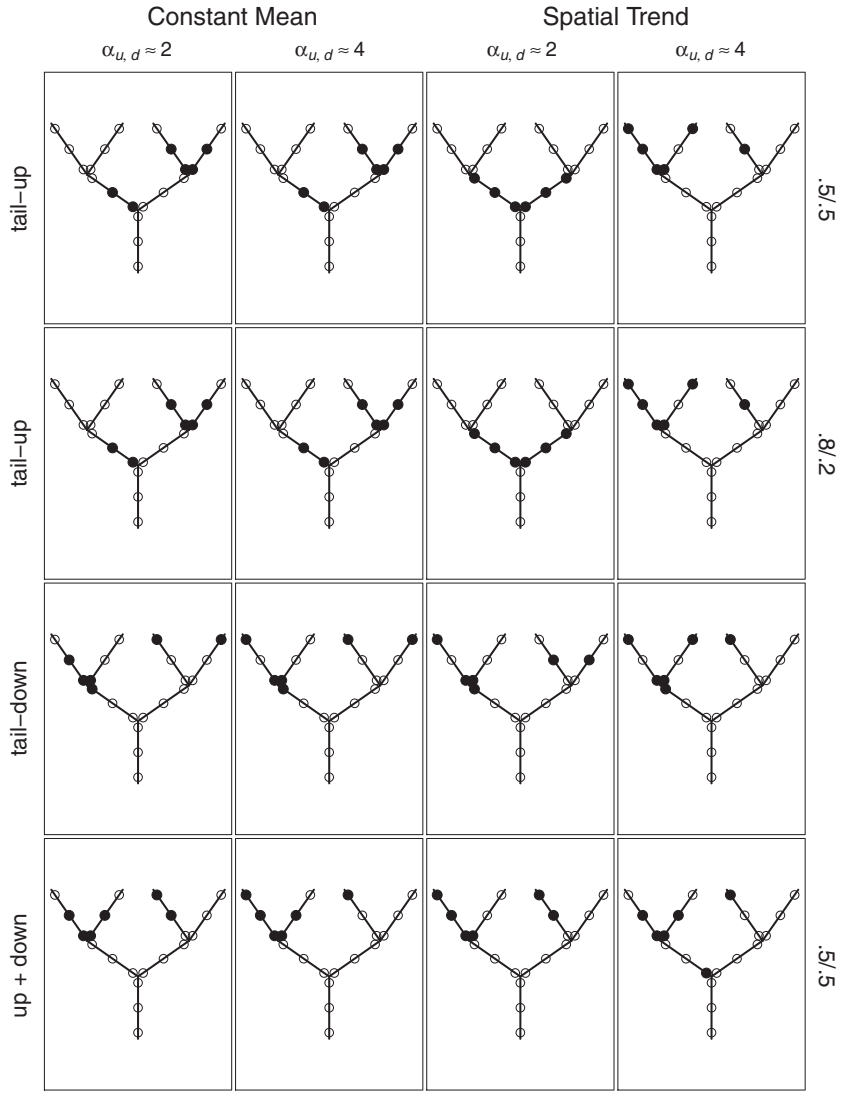

Figure 2. Locally CP-optimal toy network designs for estimation of covariance parameters. All circles represent potential sampling locations, and filled black circles are locations included in the locally optimal six-point designs. Rows, from top to bottom, correspond to tail-up covariance process with equal segment weighting (0.5/0.5), tail-up covariance process with unequal segment weighting (0.8/0.2), tail-down covariance process, and tail-up + tail-down covariance process with equal segment weighting for the tail-up component. Columns, from left to right, correspond to constant mean with $\boldsymbol{\alpha} \approx 2$, constant mean with $\boldsymbol{\alpha} \approx 4$, spatial trend mean structure with $\boldsymbol{\alpha} \approx 2$, and spatial trend mean structure with $\boldsymbol{\alpha} \approx 4$

where $\mathbf{c}$ is the vector of covariances between $\mathbf{Y}$ and $Y(s_0)$, $\mathbf{x_0}$ is the design vector of fixed effects values at s_0, and $\mathbf{X}$ and $\boldsymbol{\Sigma}$ are as in (4) and (5), respectively. The kriging variance, $\text{var}\left[\hat{Y}(s_0) - Y(s_0)\right]$, is

$$\sigma_K^2(s_0) = C - \mathbf{c}'\boldsymbol{\Sigma}^{-1}\mathbf{c} + \left(\mathbf{x} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c}\right)' \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1} \left(\mathbf{x} - \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{c}\right) \tag{10}$$

where C is the var(s_0) and the K subscript denotes kriging. Given our finite set of points, an appropriate design criterion was the maximum kriging variance (Zimmerman, 2006), given by

$$\text{K} = \max_{s \in S} \sigma_K^2(s) \tag{11}$$

although the average (over S) kriging variance would also be an acceptable design criterion (Zhu and Zhang, 2006). We defined the locally K-optimal six-point stream-network design as that which minimized (11) among all six-point designs.

All results for locally K-optimal designs exhibited spatially balanced patterns, with no clustering (in the sense defined in the previous section) whatsoever (Figure 3). There was no effect of mean function and only minor effects of segment weighting scheme and spatial

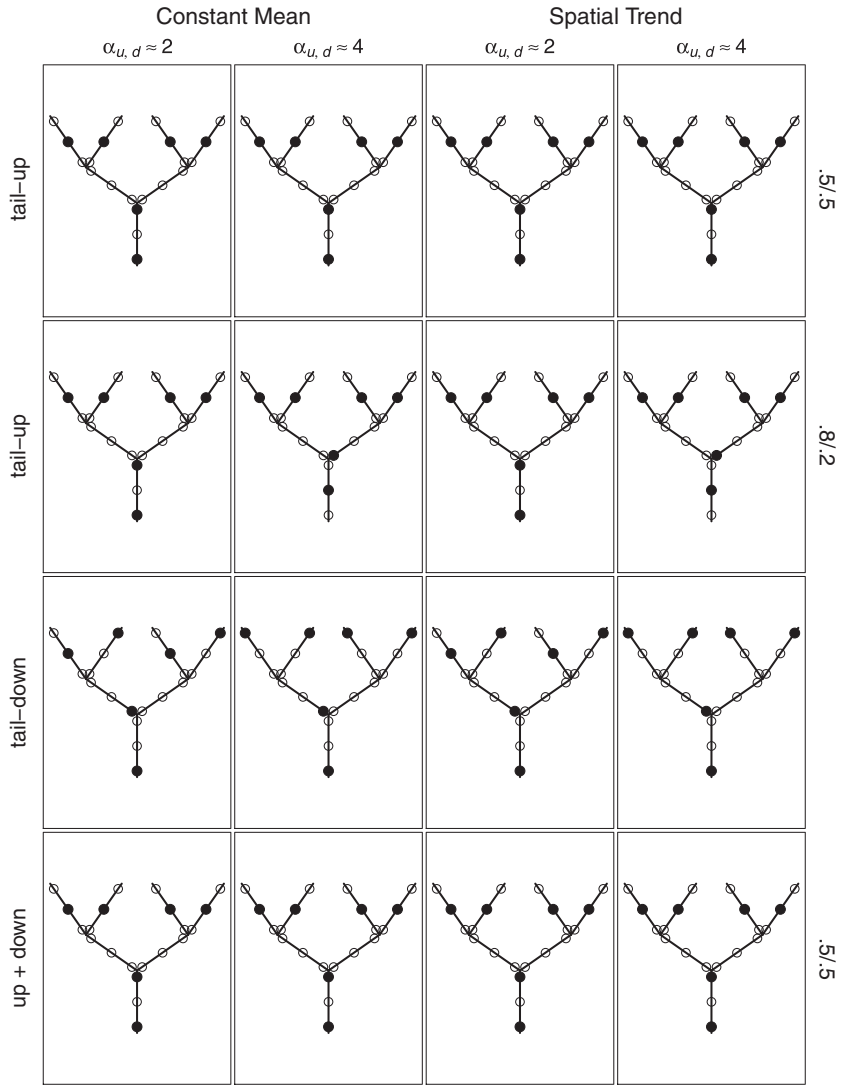

Figure 3. Locally K-optimal toy network designs for prediction with known covariance parameters. All circles represent potential sampling locations, and filled black circles are locations included in the locally optimal six-point designs. Rows, from top to bottom, correspond to tail-up covariance process with equal segment weighting (0.5/0.5), tail-up covariance process with unequal segment weighting (0.8/0.2), tail-down covariance process, and tail-up + tail-down covariance process with equal segment weighting for the tail-up component. Columns, from left to right, correspond to constant mean with $\boldsymbol{\alpha} \approx 2$, constant mean with $\boldsymbol{\alpha} \approx 4$, spatial trend mean structure with $\boldsymbol{\alpha} \approx 2$, and spatial trend mean structure with $\boldsymbol{\alpha} \approx 4$

autocorrelation strength. For the tail-up process, nearly all samples occurred in the outlet and headwater segments. For the tail-down process, stronger spatial autocorrelation led to more points at uppermost locations on headwater segments. Results for the tail-up + tail-down processes matched those for the tail-up processes exactly. In all cases, at least one sample was located in the outlet segment, and in nearly all cases, a sample was located at the most downstream location in the outlet segment.

3.3. Prediction with estimated covariance parameters

We continue with kriging but under the more common scenario of estimating both the fixed effects and covariance function parameters. The predictor of $Y(s_0)$ depended on $\hat{\boldsymbol{\theta}}$, a vector of estimated covariance parameters, and is expressed similarly to (9), but with $\boldsymbol{\Sigma}$ evaluated at $\hat{\boldsymbol{\theta}}$. This is known as the empirical best linear unbiased predictor (E-BLUP) of $Y(s_0)$. Although the E-BLUP remains unbiased under a weak set of regularity conditions met by both ML and REML, exact expressions for other moments of the E-BLUP's prediction error are not known. Nevertheless, reasonable approximations to the E-BLUP's prediction error variance exist. Building on the prior work of Kackar and Harville (1984), Harville and Jeske (1992) and Zimmerman and Cressie (1992) proposed

$$\sigma_K^2\left(s_0;\hat{\boldsymbol{\theta}}\right) \doteq \sigma_K^2(s_0) + \operatorname{tr}\left\{\mathbf{A}(\boldsymbol{\theta})[\mathbf{I}(\boldsymbol{\theta})]^{-1}\right\} \tag{12}$$

as an approximation to the prediction error variance, where

$$\mathbf{A}(\boldsymbol{\theta}) = \text{var}\left[\frac{\partial \hat{Y}(s_0)}{\partial \boldsymbol{\theta}}\right]$$

As for our criterion for kriging with known $\boldsymbol{\theta}$, we relied on the maximum prediction-error variance among all stream network locations in S to rank the sampling designs,

$$\text{EK} = \max_{s \in S}\left(\sigma_K^2(s) + \text{tr}\left\{\mathbf{A}(\boldsymbol{\theta})[\mathbf{I}(\boldsymbol{\theta})]^{-1}\right\}\right) \tag{13}$$

and defined the locally empirically K-optimal (EK-optimal) six-point stream-network design as that which minimized (13) among all six-point designs.

The results for locally EK-optimal designs were "hybrids," of a sort, of the corresponding locally CP-optimal and K-optimal designs, exhibiting some characteristics of each. As such, the result figure can be found in the online appendix (Figure A.1). A portion of the samples occurred in pairs at adjacent sites within the same segment, and these occurred predominantly in headwater and outlet segments. For the tail-up process, stronger correlation and unequal segment weighting favored regular spacing among samples in lieu of any clustering, and there was a preference for lower-weighted segments. Differences in mean function had little impact on overall design characteristics. For the tail-down process, stronger autocorrelation resulted in samples separated by the maximum stream distance, in contrast to the weak autocorrelation scenario, which produced a design with a broader distribution of inter-point distances due to the presence of some within-segment pairs. Interestingly, all samples lay within the headwater and outlet segments. The tail-up + tail-down process, like each on its own, had more samples within clusters when the autocorrelation was weak, all of which were found within headwater and outlet segments.

3.4. Fixed effects estimation with known covariance parameters

Next, we consider estimation of $\boldsymbol{\beta}$ from (4) given observed $\mathbf{X}$, $\mathbf{Y}$, and known covariance parameters. The best linear unbiased estimator (BLUE) (e.g., Schabenberger and Gotway, 2005, p. 134) of $\boldsymbol{\beta}$ is the generalized least squares estimator

$$\hat{\boldsymbol{\beta}}_{gls} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \tag{14}$$

where $\boldsymbol{\Sigma}$ is defined by (5), and $\hat{\boldsymbol{\beta}}_{gls}$ has covariance matrix

$$\text{var}\left(\hat{\boldsymbol{\beta}}_{gls}\right) = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1} \tag{15}$$

We employed a standard measure of a design's quality for estimating $\hat{\boldsymbol{\beta}}_{gls}$, known as D-optimality (Kiefer, 1958) and defined as

$$\text{D} = \left|\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\right| \tag{16}$$

which is the determinant of var $\left(\hat{\boldsymbol{\beta}}_{gls}\right)$. A locally D-optimal six-point stream-network design was one that minimized (16) among the set of all six-point designs.

The locally D-optimal results varied by spatial process and mean function, but less so by spatial autocorrelation strength and segment weighting (Figure 4). For the tail-up process, all locally optimal designs for the constant mean scenarios had a striking pattern in which all samples occurred in two confluence-based clusters, with sites in different clusters flow-unconnected. For the spatial-trend case, locally optimal designs usually contained a single confluence-based cluster at the upstream terminus of the outlet segment. This cluster was accompanied by single samples at locations of highest leverage under weak spatial autocorrelation but not at highest leverage locations under stronger spatial autocorrelation. For weaker spatial autocorrelation with spatial trend, the results suggested that samples need to be located in segments across the spectrum of segment weightings. For the tail-down process, samples were allocated to areas of high leverage and greatest stream distance, regardless of spatial autocorrelation strength or mean function. The tail-up + tail-down results aligned very closely with those of the tail-down process.

3.5. Fixed effects and covariance parameters estimation

Finally, we extend the inference scenario of Section 3.4 to that in which both $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are unknown and estimated from the observed data, probably the most common scenario for scientists and managers. Again, our scalar measure of design quality needed to incorporate both parameter vectors and does so by extending the development of our criterion from Section 3.3. When covariance parameters are estimated, the empirical-best linear-unbiased estimator for $\boldsymbol{\theta}$ is the estimated generalized least-squares estimator, obtained by evaluating (14) at the estimated parameter vector $\hat{\boldsymbol{\theta}}$,

$$\hat{\boldsymbol{\beta}}_{egls} = \left(\mathbf{X}'\left[\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)\right]^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\left[\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)\right]^{-1}\mathbf{Y} \tag{17}$$

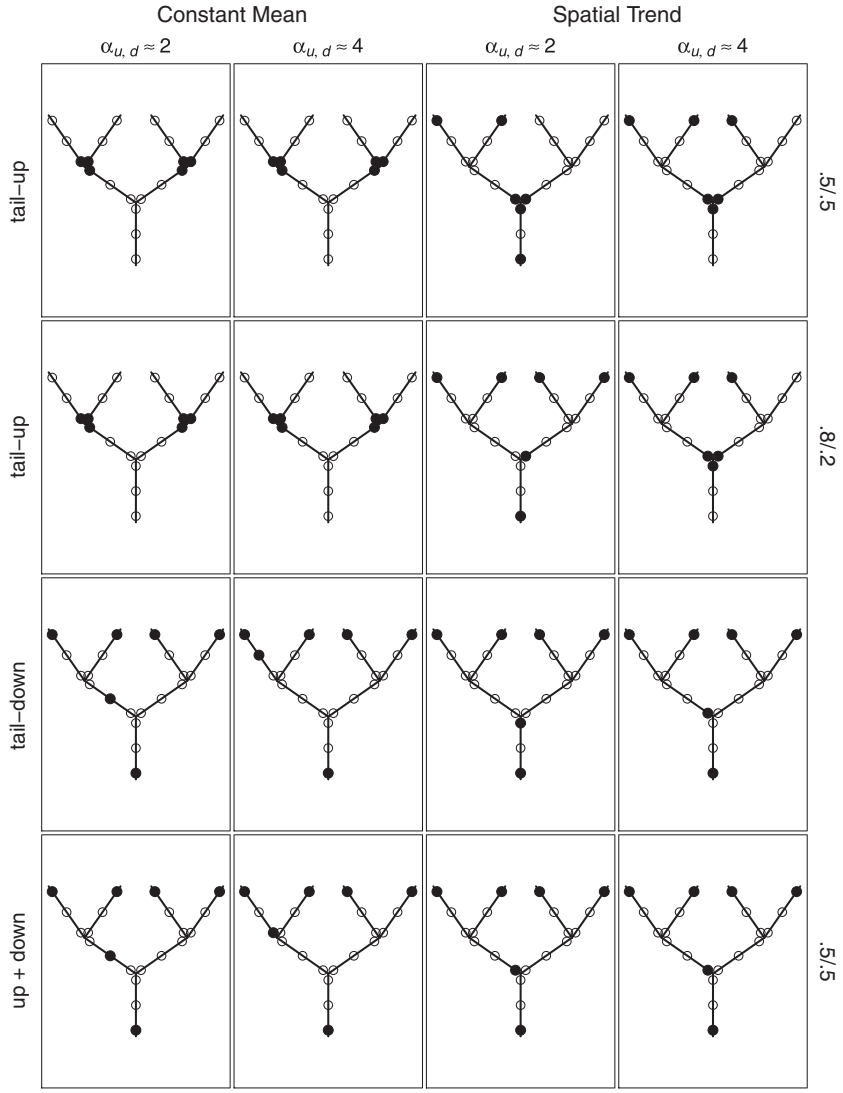

Figure 4. Locally D-optimal toy network designs for fixed effects estimation with known covariance parameters. All circles represent potential sampling locations, and filled black circles are locations included in the locally optimal six-point designs. Rows, from top to bottom, correspond to tail-up covariance process with equal segment weighting (0.5/0.5), tail-up covariance process with unequal segment weighting (0.8/0.2), tail-down covariance process, and tail-up + tail-down covariance process with equal segment weighting for the tail-up component. Columns, from left to right, correspond to constant mean with $\alpha \approx 2$, constant mean with $\alpha \approx 4$, spatial trend mean structure with $\alpha \approx 2$, and spatial trend mean structure with $\alpha \approx 4$

where the elements of $\hat{\boldsymbol{\theta}}$ are estimated using REML or ML, and other terms are as in (14). Incorporating $\hat{\boldsymbol{\theta}}$ to (14) again complicates the sampling distribution of $\hat{\boldsymbol{\beta}}_{egls}$, and we thus turned to an approximation derived in a similar fashion to that for (12). Specifically, we approximated the variance of $\hat{\boldsymbol{\beta}}_{egls}$ by

$$\text{var}\left(\hat{\boldsymbol{\beta}}_{egls}\right) \doteq \left(\mathbf{X}'\left[\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\theta}}\right)\right]^{-1}\mathbf{X}\right)^{-1} + \begin{pmatrix} \text{tr}\left\{\mathbf{A}_{11}(\boldsymbol{\theta})[\mathbf{I}(\boldsymbol{\theta})]^{-1}\right\} & \dots & \text{tr}\left\{\mathbf{A}_{1m}(\boldsymbol{\theta})[\mathbf{I}(\boldsymbol{\theta})]^{-1}\right\} \\ \vdots & \ddots & \vdots \\ \text{tr}\left\{\mathbf{A}_{m1}(\boldsymbol{\theta})[\mathbf{I}(\boldsymbol{\theta})]^{-1}\right\} & \dots & \text{tr}\left\{\mathbf{A}_{mm}(\boldsymbol{\theta})[\mathbf{I}(\boldsymbol{\theta})]^{-1}\right\} \end{pmatrix} \tag{18}$$

where

$$\mathbf{A}_{kj}(\boldsymbol{\theta}) = \text{cov}\left(\frac{\partial\hat{\beta}_k(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}, \frac{\partial\hat{\beta}_j(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)$$

$\hat{\beta}_k(\boldsymbol{\theta})$ is the kth element of $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$, m indexes the length of $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$, $k = 1, \dots, m$, and $j = 1, \dots, m$. We again relied on the determinant of (18) as our design criterion,

$$\text{ED} = \left|\text{var}\left(\hat{\boldsymbol{\beta}}_{egls}\right)\right| \tag{19}$$

and defined a six-point stream-network design as locally empirically D-optimal (ED-optimal) if it minimized (19).

For the locally ED-optimal results, the general distribution of samples was often similar to those of locally D-optimal designs, including clusters and an indication that samples from the spectrum of segment weights was important. Given the similarity to the D-optimal designs, the results figure can be found in the online appendix (Figure A.2). For the tail-up process with constant mean and stronger spatial autocorrelation, the best designs consisted of confluence-based clusters. All other conditions produced optimal designs without confluence-based clusters and samples at highest leverage locations for the spatial-trend case. The results from the tail-down and tail-up + tail-down processes were quite similar. All locally optimal designs for these latter processes included samples at the headwater and outlet segments of the network, and all contained an adjacent pair in either a headwater or outlet segment.

3.6. Conclusions from toy networks

The toy network results matched patterns for optimal designs from the traditional geostatistical realm and highlighted the unique aspects of sampling spatial processes in stream networks. Similarities included the importance of sampling at areas of high leverage when spatial trend is present, general spatial balance among samples for prediction with known covariance parameters, designs consisting predominantly of clusters for only estimating covariance parameters, and some samples within clusters for fixed effects estimation and prediction when covariance parameters are also estimated. Most striking was that for each inference scenario, the locally optimal designs often differed between the tail-up and tail-down processes. The toy results also highlighted the importance of sampling the outlet segment for both prediction (with and without estimation of covariance parameters) and tail-down spatial processes (alone or in conjunction with tail-up), regardless of mean function. Finally, several of the locally optimal designs included confluence-focused clusters, which occurred more commonly for the tail-up spatial processes.

4. SIMULATED STREAM NETWORKS

We next explore the prevalent design themes from the toy networks with larger, simulated networks. We chose two contrasting network shapes: dendritic and trellis. Dendritic-shaped networks have a tree-like appearance and are more space-filling. Trellis-shaped networks are more linear and have a higher proportion of headwater segments. To create the simulated networks, we developed a function using R statistical software (R Development Core Team, 2011) that builds and branches the networks. Differing network shapes are created by altering the probabilities with which new upstream segments are added to existing segments (e.g., a higher probability of adding a new segment to the most recently created segment will lead to trellis-shaped networks). Examples of the simulated network shapes can be found in the subsequent sections describing the results of the inference scenarios, and the function used to simulate the networks is available from the authors upon request. For both shapes, 1000 stream networks were simulated with 101 segments each, a number of segments sufficient to create ample confluences and headwater segments for evaluating the merits of sampling from these network locations. In each simulated network, every segment's length was drawn from a uniform(0.5, 2.5) distribution to further induce variation in the shape and structure of the networks. A dense grid of potential sampling locations was placed along each simulated network. The two network shapes produced differences in the maximum distances among flow-connected and flow-unconnected locations. For example, the maximum flow-unconnected distances among all simulations averaged 31 units for dendritic networks (maximum = 36 units) and 55 units for trellis networks (maximum = 63 units). Values of σ_u^2 and σ_d^2 were the same as in the toy examples (noting that changes to σ_u^2 and σ_d^2 simply scale the criteria scores), and we again set $\sigma_0^2 = 0$. Two range parameter values (α_u, α_d) of 18 and 36 units were used to represent relatively weak and strong spatial autocorrelation. For the segment weights, we again applied two differing methods, an equal segment weighting above each confluence and Shreve stream order. These weighting schemes were selected because they were applicable to our simulated network topologies, in contrast to simulating flow volumes or catchment areas.

For the larger networks and sample sizes, an exhaustive search for the best design is not possible (as it was for the toy networks), and the number of sampling designs to consider becomes enormous. Instead, we used what we learned from the toy networks to create sampling designs for the larger networks that preserved important characteristics observed from the toy-network results. These designs, and the toy-network designs that inspired them, are given in Table 2. In addition to the design patterns that emerged as locally optimal for the toy networks, we created several hybrid designs by combining the beneficial characteristics among multiple designs to evaluate their potential merits. For each simulated network, we applied an example of each sampling design with a sample size (n) of 40, which is a sample size consistent with those applied in previous spatial sampling design work (e.g., Zhu and Stein (2005, $n = 30$); Zimmerman (2006, $n = 50$); Müller and Stehlk (2010, $n = 36$)). For presentation ease, we assigned a name abbreviation to each class of sampling design characteristic (Table 2) where "H" indicates designs with headwater and outlet segment clusters, "U" refers to sample clusters from mutually flow-unconnected stream segments, "C" indicates designs containing confluence-based clusters, "G" indicates spatially balanced designs, and "SRS" denotes simple random sampling. The numerals that follow each letter simply label designs within each class.

To implement the network-based sampling designs from Table 2, we first selected segments based on the intended network characteristics, and then selected sampling locations within the segments. For all sampling designs that included a single sample from the outlet segment, we selected the location at the very bottom of the network. To select segments for same-segment clusters, we first found the set of all segments that matched the necessary characteristics (e.g., all headwater segments), and then randomly selected the required number. We next added size-2 clusters to these segments by first choosing a random location in the segment. We then drew a distance value from a uniform (0, 0.5) distribution (noting 0.5 units was the minimum segment length) and selected a clustered neighbor randomly from the set of locations within the drawn stream distance value. For the confluence-based clusters of size-3, we randomly selected the necessary number of confluences from each network and selected the location from each of the three adjoining stream segments closest to the confluence. Spatially balanced samples were generated using generalized random tessellation stratified (GRTS) sampling. GRTS samples rely on quadrant-recursive addressing and hierarchical randomization to choose samples that are generally spatially balanced (Stevens Jr. and Olsen, 2004) and are commonly applied in

Environmetrics 2014; **25**: 306–323 Published 2014. This article has been contributed to by US Government employees and their work is in the public domain in the USA. wileyonlinelibrary.com/journal/environmetrics

Table 2. The suite of sampling designs applied to all simulated networks

Clusters			Singles			
Source	Number	Size	Source	Number	Toy	Abbrv.
Outlet	1	2	Headwater	36	ED, EK	H1
Headwater	1	2				
Outlet	1	2	Headwater	20	ED, EK	H2
Headwater	9	2				
Outlet	1	2	Headwater	2	ED, EK	H3
Headwater	18	2				
Outlet	1	2	Middle	36	EK	H4
Headwater	9	2				
Outlet	1	2	Middle	2	EK	H5
Headwater	18	2				
Unconnected	20	2	—	—	CP	U1
Unconnected	10	4	—	—	CP	U2
Confluence	7	3	Headwater	19	D, ED	C1
Confluence	7	3	Headwater	18	D	C2
			Outlet	1		
Confluence	13	3	Headwater	1	D, ED, CP	C3
Confluence	13	3	Outlet	1	Hybrid	C4
—	—	—	SRS	40		SRS
—	—	—	GRTS	40	K, EK	G1
—	—	—	GRTS	39	Hybrid	G2
—	—	—	Outlet	1		
GRTS	20	2	—	—	Hybrid	G3
GRTS	5	2	GRTS	30	Hybrid	G4

The "Toy" column indicates which toy-network inference-scenario results prompted a design's inclusion, and a "hybrid" label denotes that prevailing characteristics from multiple toy network results were combined. The abbreviations column ("Abbrv.") provides a condensed label for reference ease in figures, tables, and text: "H" indicates designs with headwater and outlet segment clusters; "U" refers to sample clusters from mutually flow-unconnected stream segments; "C" indicates designs containing confluence-based clusters; "G" indicates spatially balanced designs; and "SRS" denotes simple random sampling. For designs that include clustered samples, the "Number" column indicates how many clusters the design contains, and the "Size" column indicates the number of samples in each cluster. "Source" indicates from which network features the samples were selected. Total sample size for all designs is 40.

aquatic systems. We selected GRTS samples via the R package *spsurvey* (Kincaid *et al.*, 2008). To obtain samples weighted toward the middle portions of networks, a weighted random sample was selected with weights assigned inversely proportional to each location's distances from the outlet and most upstream location. We also selected SRS samples as a relative baseline regarding design choice. The SRS samples were selected by randomly selecting 40 locations from the dense grid of potential sampling locations on each simulated network. To efficiently create samples with stream network-based characteristics (i.e., simulated random variables with appropriate covariance properties), we relied on functions from the R package *SSN* (Ver Hoef *et al.*, 2014).

Design performance was evaluated in two ways. First, we ranked the criteria scores of each sampling design for each simulated network and calculated the proportion of times a design ranked best (i.e., had the best criteria score among all sampling designs), or among the two best scoring designs. Next, we computed the mean and standard deviation of the criteria scores for each sampling design across all simulated networks of each shape. The ranking of designs within each simulation aided in determining best designs when there was overlap in the distributions of criteria scores over the 1000 simulations for each network shape. The mean criteria scores revealed the relative performance

Table 3. Summaries of criteria scores from the simulated networks

Inference scenario	Spatial process	Mean function	Range	Weights	Shape	Design	Best	Top two	Mean	SD
EK	Tail-up	Constant	18	Equal	Dendritic	H1	0.000	0.000	1.028	9.6e-4
						C1	0.000	0.000	1.023	5.9e-4
						C3	0.454	0.999	1.018	7.0e-4
						C4	0.545	1.00	1.018	1.0e-3
						SRS	0.000	0.000	1.046	3.6e-3
EK	Tail-up	Trend	36	Shreve	Trellis	H1	0.730	0.923	1.097	0.017
						H2	0.058	0.325	1.130	0.023
						G1	0.040	0.102	1.174	0.050
						SRS	0.022	0.062	1.197	0.061
EK	Tail-down	Constant	36	—	Trellis	H1	0.199	0.419	0.445	0.072
						G1	0.320	0.540	0.427	0.063
						G2	0.297	0.542	0.425	0.061
						SRS	0.021	0.067	0.554	0.110
EK	Tail-down	Constant	36	—	Dendritic	H1	0.323	0.520	0.461	0.087
						G1	0.258	0.493	0.462	0.072
						G2	0.224	0.451	0.473	0.087
						SRS	0.027	0.070	0.587	0.103
ED	Tail-up	Constant	18	Shreve	Dendritic	H1	0.000	0.000	0.029	3.1e-4
						C3	0.517	0.916	0.023	1.7e-3
						C4	0.449	0.909	0.023	1.7e-3
						SRS	0.000	0.000	0.048	3.9e-3
ED	Tail-up	Constant	18	Shreve	Trellis	H1	0.882	0.977	0.027	1.4e-4
						H2	0.000	0.038	0.034	1.7e-4
						C1	0.048	0.454	0.031	0.002
						SRS	0.000	0.000	0.051	0.005
ED	Tail-up	Trend	18	Shreve	Trellis	H1	0.963	0.998	3.2e-6	8.7e-7
						H2	0.000	0.290	5.0e-6	1.4e-6
						C1	0.022	0.419	4.9e-6	1.4e-6
						SRS	0.000	0.000	1.1e-5	3.8e-6
ED	Tail-down	Constant	18	—	Trellis	H1	0.773	0.971	0.137	0.007
						H2	0.202	0.792	0.141	0.007
						SRS	0.001	0.017	0.155	0.009
ED	Tail-down	Trend	36	—	Trellis	H1	0.323	0.593	2.1e-4	5.8e-5
						H2	0.312	0.551	2.2e-4	6.0e-5
						H3	0.171	0.312	2.3e-4	6.8e-5
						SRS	0.009	0.037	2.6e-4	8.0e-5

Summaries include the proportion of 1000 simulations where each sampling design had the best criteria score ("Best"), were within the best two scoring designs ("Top two"), the mean value of a design's criteria scores ("Mean"), and the standard deviation of a design's criteria scores ("SD"). Inference scenarios include prediction (EK) and fixed effects estimation (ED), both with estimated covariance parameters. The simple random sampling (SRS) design is included for reference along with best-performing designs, and poorer preforming designs are omitted. Table 2 contains the sampling design abbreviations.

of designs, and the potential to discover designs that might commonly be near-optimal among the set of inference scenarios. They also allowed for clear "winners", where a sampling design's criteria scores were markedly lower than other designs despite the added randomness injected via the simulated networks and specific placement of sample points as described previously.

It is most realistic to assume that covariance parameter values are not known *a priori* (Gilmour and Trinca, 2012). To save space, details are provided in an online appendix for the scenarios with known covariance parameters (D and K) and only covariance parameter estimation (CP). However, we summarize those results in the succeeding text, and we also provide detailed tabular and figure results for the scenarios with estimated covariance parameters (ED and EK). Finally, we remind readers that the design name abbreviations can be found in Table 2.

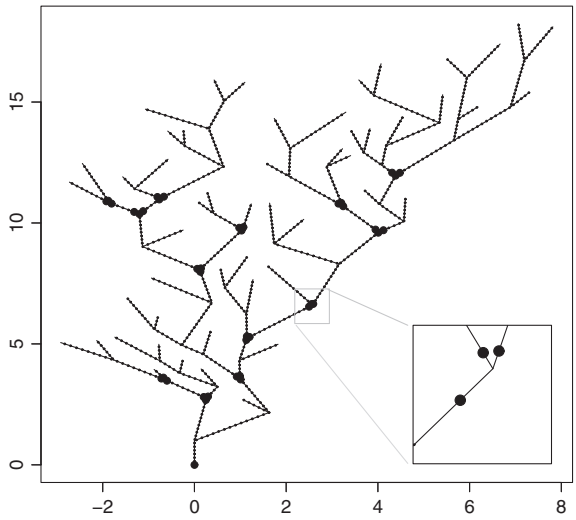

Figure 5. The best scoring design for the prediction with estimated covariance parameters (EK) inference scenario for a tail-up spatial process, dendritic-shaped networks, effective range = 18, equal segment weighting, and constant mean structure. This design, labeled as C4 in Table 2, contains 39 samples within size-3 clusters at confluences, and a single sample from the outlet segment. The inset figure highlights a size-3 cluster from one of the confluences with clustered samples

4.1. Prediction with estimated covariance parameters

For the tail-up spatial process, effects of mean function were pronounced, but strength of spatial autocorrelation, segment weighting, and network shape had minor effects on EK-criteria scores. For the case of the constant mean, the confluence-based designs were clearly best (Table 3, and Figure 5 demonstrates a best overall scoring design). Note that although the H1 design was never among the two best scoring designs, its criteria scores were close to those of the confluence-based designs (Table 3). Results not presented in Table 3 include, for the trellis-shaped networks under weaker spatial autocorrelation and Shreve-based weights, the C1 design scored slightly better than C3 or C4, and for both network shapes, especially with stronger spatial autocorrelation under equal segment weighting, several GRTS-based designs performed as well as the confluence-based designs.

For the spatial trend scenarios, scores among designs were generally more homogeneous than the constant-mean case. The best scoring designs included the H1, G1, and G2 designs with comparable scores. The exception was for the trellis-shaped networks with Shreve-based weighting where the H1 design ranked better than any of the GRTS-based designs (Table 3).

The results for the tail-down and tail-up + tail-down processes were quite similar, and differences in mean function, spatial autocorrelation strength, and network shape did not generally affect the set of best designs. The preponderance of the EK scores were best for the H1 and GRTS-based designs (Table 3). The GRTS designs ranked best moderately more frequently for most of the scenarios of trellis-shaped networks (Table 3), with the exception of a spatial trend and tail-up + tail-down process, and the H1 design ranked best moderately more frequently in all scenarios for the dendritic-shaped networks (Table 3, and Figure 6 demonstrates an example of a best overall scoring design).

4.2. Estimation of fixed effects and covariance parameters

For the tail-up spatial process, the best ED-scoring designs in all cases included clusters. For all dendritic networks and trellis networks with equal segment weighting, the confluence-based clusters, particularly C3 and C4, were generally best regardless of mean function or spatial autocorrelation strength (Table 3, and Figure 7 demonstrates a best overall scoring design). The only deviation from this pattern (and hence not presented in Table 3) was for the spatial-trend case with $\alpha = 18$, where the C1 design scored better than the C3 design. For the trellis networks with Shreve weighting, the H1 design scored at least as well as the confluence-based designs, and consistently ranked best (Table 3, and Figure 8 demonstrates a best overall scoring design). The amount of improvement in ED scores for the H1 design over the confluence-based designs was greater in the spatial-trend case than the constant-mean case (Table 3), but this effect was reduced with stronger spatial autocorrelation.

For the tail-down spatial process, the ED criteria scores were generally more homogeneous among designs than they were for the tail-up process, and the C3 design was among the worst. The ranking of criteria scores revealed that designs with outlet and headwater clusters scored best for both network shapes (Table 3). Stronger spatial autocorrelation or a spatial trend did not alter the rankings of best designs relative to others but suggested the proportion of headwater and outlet samples within clusters mattered less (Table 3). Results for the tail-up + tail-down process were similar to those of the tail-down (hence, tabular summaries are not presented, but Figure 9 demonstrates a best overall scoring design), with one exception; for the case of a spatial trend and stronger spatial autocorrelation, the C2 design scored nearly as well as the H1 and H2 designs.

The next three paragraphs provide summaries of the results for the less common scenarios; the interested reader can find associated tabular and figure details in the online appendix.

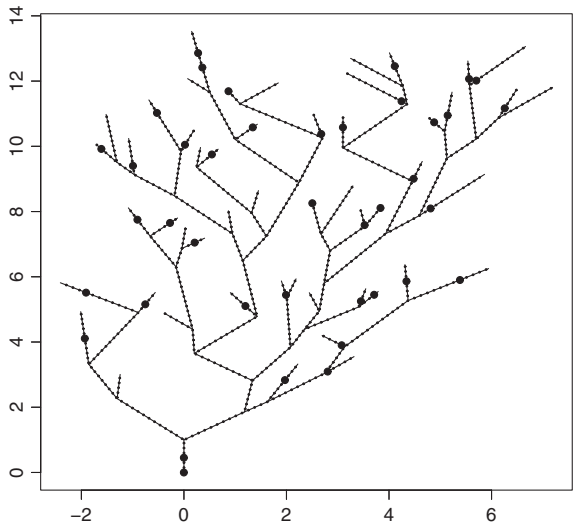

Figure 6. The best scoring design for the prediction with estimated covariance parameters (EK) inference scenario for a tail-up + tail-down spatial process, dendritic-shaped networks, effective range $= 18$, equal segment weighting for the tail-up component, and constant mean structure. This design, labeled as H1 in Table 2, contains two samples within a size-2 cluster at the outlet segment, two samples within a size-2 cluster in a headwater segment, and 36 unclustered samples from individual headwater segments

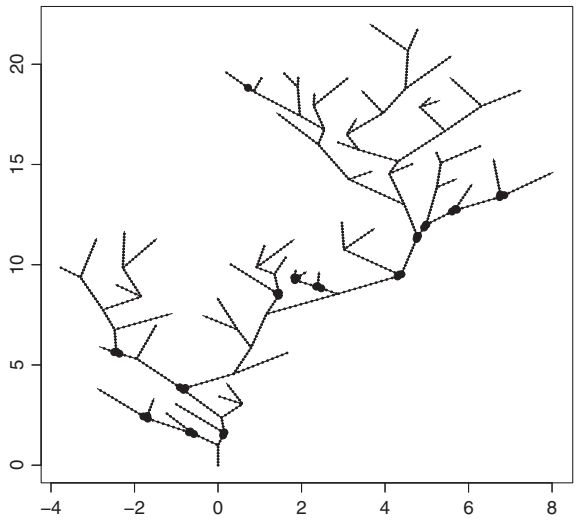

Figure 7. The best scoring design for the estimation of fixed effects and covariance parameters (ED) inference scenario for a tail-up spatial process, dendritic-shaped networks, effective range = 36, Shreve segment weighting, and spatial trend mean structure. This design, labeled as C3 in Table 2, contains 39 samples within size-3 clusters at confluences, and a single sample from a headwater segment

4.3. Covariance parameter estimation

For the tail-up spatial processes, designs consisting predominantly of clusters had better CP-criteria scores than non-clustered designs, regardless of mean function, strength of spatial autocorrelation, network shape, or segment weighting. The best scoring design, U1, focused sample clusters on flow-unconnected segments and ranked as the best design in most of the 1000 simulations from each network shape. For the tail-down process, the criteria scores among all designs were more homogeneous, and there were no apparent effects of mean function, strength of spatial autocorrelation, or network shape. The distribution of GRTS design scores were comparable with those containing clusters at the outlet and headwater segments. The scores for designs based on confluence clusters were markedly worse, a contrast to the toy results. For the tail-up + tail-down process, there was little effect of spatial autocorrelation strength or mean function. In addition, differing methods of segment weighting had little effect on the ranking of best designs. The best designs were similar to the tail-up spatial process, where the best designs were comprised predominantly of flow-unconnected clusters (U1, H3, H5), although for the trellis-shaped networks, the C3 design scored equivalently.

Environmetrics 2014; **25**: 306–323 Published 2014. This article has been contributed to by US Government employees and their work is in the public domain in the USA. wileyonlinelibrary.com/journal/environmetrics

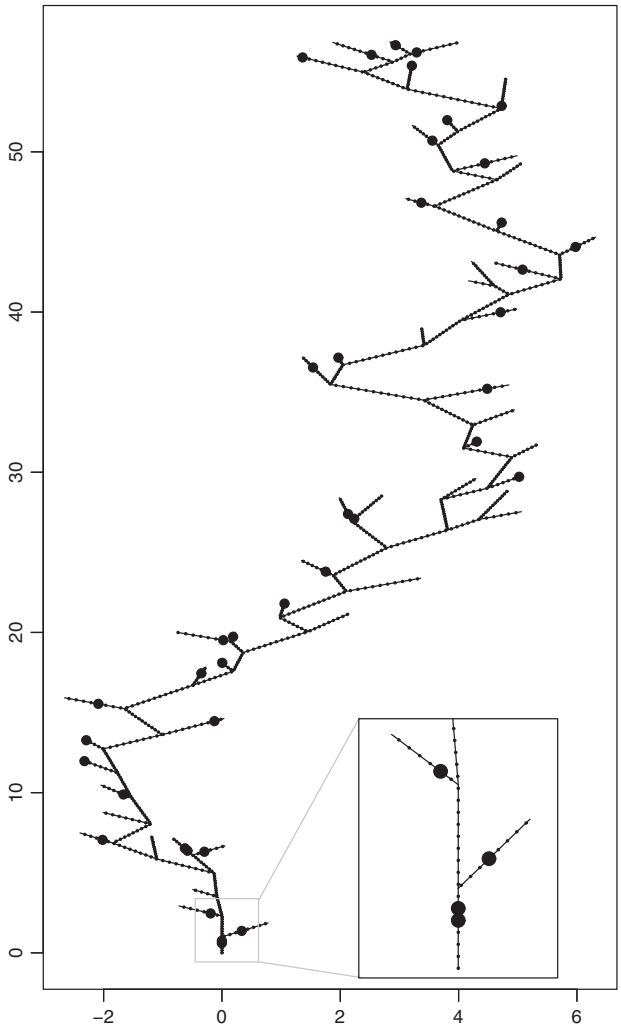

Figure 8. The best scoring design for the estimation of fixed effects and covariance parameters (ED) inference scenario for a tail-up spatial process, trellis-shaped networks, effective range = 18, Shreve segment weighting, and spatial trend mean structure. This design, labeled as H1 in Table 2, contains two samples within a size-2 cluster at the outlet segment, two samples within a size-2 cluster in a headwater segment, and 36 unclustered samples from individual headwater segments. The inset figure highlights the size-2 cluster in the outlet segment, and 2 of the unclustered samples from individual headwater segments

4.4. Prediction with known covariance parameters

For the tail-up spatial process with constant mean, the confluence-based cluster designs had the best K-criteria scores, followed by H1. Additionally, for trellis networks with Shreve-based segment weights, the H1 design had better criteria scores than the majority of the confluence-based designs. For the trend mean structure, results were more homogenous with modestly better scores for the H1 design. Results were similar for both strengths of spatial autocorrelation, except a less homogenization of criteria scores for the trend mean structure in trellis-shaped networks. The results for the tail-down and tail-up + tail-down processes were identical. Regardless of mean function, spatial autocorrelation strength, or network shape, the designs scoring best were the G1 and H1 designs.

4.5. Estimation of fixed effects with known covariance parameters

There was no effect of spatial autocorrelation strength among all spatial processes, and little effect of network shape or mean function on D-criteria scores. For the tail-up process with constant mean, the confluence-based designs were clearly best, and the H1 design again scored competitively. The latter scored better on average for the trellis networks with Shreve-based segment weighting, although the best scores among all designs contained confluence clusters. For the trend mean structure, the H1, H2, C1, and C3 designs were generally best, but C3 was not included in the group of best designs for trellis networks under Shreve-order weighting. For the tail-down process, criteria scores among designs were generally more homogeneous, with H2 joining H1 as designs with modestly better scores. The H1 and H2 designs had the best criteria scores for a constant mean and the tail-up + tail-down process. For the spatial trend mean structure, all designs comprised almost entirely of samples within clusters were among the worst scoring designs, and there was little separation in the criteria scores from the H1, GRTS, C1, and even SRS designs.

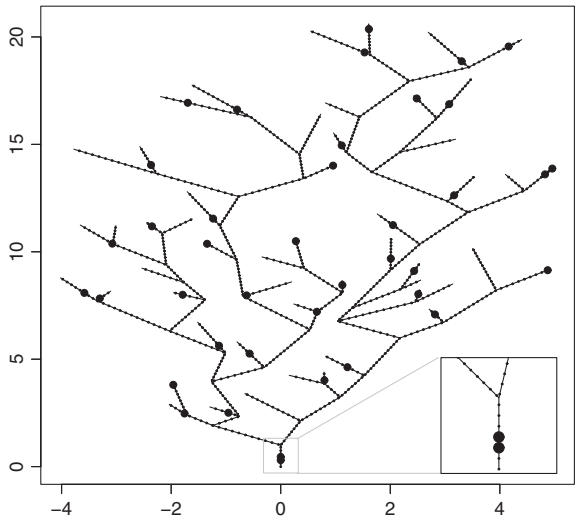

Figure 9. The best scoring design for the estimation of fixed effects and covariance parameters (ED) inference scenario for a tail-up + tail-down spatial process, dendritic-shaped networks, effective range = 36, equal segment weighting for the tail-up component, and constant mean structure. This design, labeled as H1 in Table 2, contains two samples within a size-2 cluster at the outlet segment, two samples within a size-2 cluster in a headwater segment, and 36 unclustered samples from individual headwater segments. The inset figure highlights the size-2 cluster in the outlet segment

5. DISCUSSION

The results from the simulated networks often agreed with the their toy-network counterparts, although an exception is discussed in the following paragraph. For instance, when only covariance parameters were estimated in a tail-up process, clusters of samples from flow-unconnected segments were clearly best. When making predictions with known covariance parameters, the spatially balanced GRTS samples were frequently among the best designs. For fixed effects estimation with known covariance parameters, the designs containing confluence-based clusters were best for the tail-up spatial process, while the GRTS design scored well for the tail-down process. Finally, for both prediction and fixed effects estimation with estimated covariance parameters (common scenarios for scientists and managers), clusters and single samples from the outlet and headwater segments were consistent aspects of the better scoring designs.

Despite the geostatistical evidence for similarities among best designs for estimation and prediction, we were initially surprised that confluence-based clusters were favored for prediction with a constant mean function and tail-up spatial process for the simulated networks. In retrospect, this design does contain many of the characteristics of the toy result. In a tail-up process, a confluence-based cluster can be viewed as two size-2 clusters rather than a size-3 cluster, noting that the two samples on the upstream segments at the confluence are flow-unconnected. Further, these two clusters characterize the disjunct conditions that occur upstream of the confluences if segment weightings are unequal. In our simulated networks, confluences were placed at relatively regular distances along the stream network, and these characteristics match what was observed in the toy network results.

We did expect less homogeneous results among sampling designs for the spatial-trend mean structure in tail-down and tail-up + tail-down processes and more pronounced differences based on autocorrelation strength. One reason for more homogeneous results among the designs, in the context of the spatial trend, may be that candidate designs did not allocate enough samples to the downstream portion of the network. While the suite of designs containing headwater samples and clusters captured one end of the trend leverage, the size-2 outlet cluster was probably insufficient to capture the other. The lack of strong differences due to spatial autocorrelation strength might also be attributed to the values of α_u and α_d we selected. The "weaker" value of 18 units may have been too large, particularly for the dendritic-shaped networks. In the toy networks, the larger α_u and α_d values led to samples at the most extreme locations in the network (i.e., the most downstream location of the outlet segment and most upstream locations of the headwater segments).

In addition to more sampling weight in the downstream portion of the network, future work could consider costs associated with sampling various locations of stream networks, in particular, headwater segments. Headwater segments often lie in the most remote and least accessible sections of stream networks, and difficulties in reaching headwater segments could play a role in sample site choice. Additionally, investigations regarding the inclusion of nugget effects could be considered. Zimmerman (2006) explored the effects of nugget inclusion and amount on design optimality in the geostatistical setting, and although effects were observed for estimation of only covariance parameters, there were not substantive effects on the criteria scores for prediction (with either known or estimated covariance parameters). Given those findings, and the general similarity of results for covariance parameter estimation paired with prediction or fixed effects estimation, we are confident in our results for the two most probable inference scenarios extended to cases where nugget effects are considered. Finally, future work could also focus on an aspect of spatial covariance processes that we assumed known: the form of the covariance function. We opted to use exponential forms of the covariance functions, but this characteristic is generally not known *a priori* (Nowak *et al.*, 2010).

It is worth noting that our results are based on a binary branching structure. This assumption is required for the spatial stream-network models and is also a realistic assumption for many stream ecosystems. There are admittedly situations where the binary assumption does not apply. For example, an optimal design for a binary branching stream may not be as suitable in large alluvial floodplains, where groundwater

and braided channel networks would have a strong influence on flow connectivity (Stanford and Ward, 1993). In other cases where braiding and groundwater influences are less pronounced, it may make sense to focus modeling and prediction on the main channels. Nevertheless, there may be difference in the optimal designs for stream ecosystems that do not exhibit a binary branching structure.

Our results indicated that, generally, the network shapes did not have substantive impacts on the characteristics of the best sampling designs. Although we only considered two shapes for our simulated networks, the dendritic and trellis forms represent extremes of naturally occurring stream-network shapes. However, one interesting shape-based sampling design difference arose for the tail-up spatial process under the constant mean case of the ED inference scenario. By ranking, the confluence-based cluster designs were clearly preferred for the dendritic shapes, while the H1 design was best for the trellis case. While being mindful of network shape might be fruitful in design planning in this case, we also note that the criteria scores indicated that each (either a confluence cluster design or the H1 design) scored nearly as well as the best for each shape in which they did not rank best.

Our results suggest several key themes for sampling design planning on stream networks. First, there is not a demonstrable difference in best designs for fixed effects estimation and prediction; a phenomenon also observed by Müller and Stehlk (2010) and Zimmerman (2006), who noted the importance of obtaining well-estimated fixed effects parameters to improve prediction. If inference regards estimation of an overall mean for a tail-up spatial process, then a design with confluence-based clusters is best, and particulary so in the dendritic-shaped networks. In all other cases considered, a design that includes a small proportion of clustered samples from the outlet and headwater segments, along with samples from areas of high leverage appears the most prudent choice. Further, given the overall performance of the H1 design, it does indeed seem the most robust when little is known about the fixed effects or covariance structure prior to collecting samples.

Acknowledgements

This study was conducted as part of the Spatial Statistics for Streams Working Group supported by the National Center for Ecological Analysis and Synthesis, a center funded by the National Science Foundation (Grant #EF-0553768), the University of California, Santa Barbara and the State of California. Note that any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. The authors thank three anonymous reviewers whose comments led to considerable improvements in the article.

REFERENCES

Al-Chokhachy R, Roper BB, Archer EK, Miller S. 2011. Quantifying the extent of and factors associated with the temporal variability of physical stream habitat in headwater streams in the interior Columbia River Basin. *Transactions of the American Fisheries Society* **140**(2):399–414.

Benda L. 2008. Confluence environments at the scale of river networks. In *River Confluences, Tributaries and the Fluvial Network*, Rice SP, Roy AG, Rhoads BL (eds). John Wiley & Sons: Chichester, UK; 271–300.

Benda L, Poff NL, Miller D, Dunne T, Reeves G, Pess G, Pollock M. 2004. The network dynamics hypothesis: How channel networks structure riverine habitats. *BioScience* **54**(5):413–427.

Cressie NAC. 1993. *Statistics for Spatial Data* Revised, Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons: New York, NY.

Cressie N, Frey J, Harch B, Smith M. 2006. Spatial prediction on a river network. *Journal of Agricultural, Biological, and Environmental Statistics* **11**(2): 127–150.

Cressie N, Wikle CK. 2011. *Statistics for Spatio-Temporal Data*, Wiley Series in Probability and Statistics. John Wiley and Sons: Hoboken, NJ.

Dent CL, Grimm NB. 1999. Spatial heterogeneity of stream water nutrient concentrations over successional time. *Ecology* **80**(7):2283–2298.

Fagan WF. 2002. Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology* **83**(12):3243–3249.

Ganio LM, Torgersen CE, Gresswell RE. 2005. A geostatistical approach for describing spatial pattern in stream networks. *Frontiers in Ecology and the Environment* **3**(3):138–144.

Garreta V, Monestiez P, Ver Hoef JM. 2010. Spatial modelling and prediction on river networks: up model, down model or hybrid? *Environmetrics* **21**(5): 439–456.

Gascuel-Odoux C, Boivin P. 1994. Variability of variograms and spatial estimates due to soil sampling: a case study. *Geoderma* **62**(1):165–182.

Gilmour SG, Trinca LA. 2012. Optimum design of experiments for statistical inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **61**(3):345–401.

Harville DA, Jeske DR. 1992. Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association* **87**(419):724–731.

Herlihy AT, Larsen DP, Paulsen SG, Urquhart NS, Rosenbaum BJ. 2000. Designing a spatially balanced, randomized site selection process for regional stream surveys: the EMAP Mid-Atlantic pilot study. *Environmental Monitoring and Assessment* **63**(1):95–113.

Irvine KM, Gitelman AI, Hoeting JA. 2007. Spatial designs and properties of spatial correlation: effects on covariance estimation. *Journal of Agricultural, Biological, and Environmental Statistics* **12**(4):450–469.

Kackar RN, Harville DA. 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* **79**(388):853–862.

Kiefer J. 1958. On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *The Annals of Mathematical Statistics* **29**(3): 675–699.

Kincaid T, Olsen T, Stevens D, Platt C, White D, Remington R. 2008: spsurvey: Spatial survey design and analysis. R package version 2.0

Labonne J, Ravigné V, Parisi B, Gaucherel C. 2008. Linking dendritic network structures to population demogenetics: the downside of connectivity. *Oikos* **117**(10):1479–1490.

Le ND, Zidek JV. 2006. *Statistical Analysis of Environmental Space-Time Processes*, Springer Series in Statistics. Springer: New York, NY.

Mahjouri N, Kerachian R. 2011. Revising river water quality monitoring networks using discrete entropy theory: the Jajrood River experience. *Environmental Monitoring and Assessment* **175**(1-4):291–302.

Müller WG. 2005. A comparison of spatial design methods for correlated observations. *Environmetrics* **16**(5):495–505.

Müller WG, Stehlk M. 2010. Compound optimal spatial designs. *Environmetrics* **21**(3-4):354–364.

Nowak W, de Barros FPJ, Rubin Y. 2010. Bayesian geostatistical design: Task-driven optimal site investigation when the geostatistical model is uncertain. *Water Resources Research* **46**(3):W03535.

Peterson EE, Ver Hoef JM, Isaak DJ, Falke JA, Fortin MJ, Jordan CE, McNyset K, Monestiez P, Ruesch AS, Sengupta A, Som N, Steel EA, Theobald DM, Torgersen CE, Wenger SJ. 2013. Modelling dendritic ecological networks in space: an integrated network perspective. *Ecology Letters* **16**(5):707–719.
R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0.
Schabenberger O, Gotway CA. 2005. *Statistical Methods for Spatial Data Analysis*, Texts in Statistical Science Series. Chapman and Hall/CRC: Boca Raton, FL.
Shreve RL. 1967. Infinite topologically random channel networks. *Journal of Geology* **75**(2):178–186.
Stanford JA, Ward JV. 1993. An ecosystem perspective of alluvial rivers: connectivity and the hyporheic corridor. *Journal of the North American Benthological Society* **12**(1):48–60.
Stevens DL, Jr. 2006. Spatial properties of design-based versus model-based approaches to environmental sampling. In *Proceedings of 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Caetano M, Painho M (eds). Instituto Geografico Portugues: Lisbon, Portugal; 119–125.
Stevens DL, Jr., Jensen SF. 2007. Sample design, execution, and analysis for wetland assessment. *Wetlands* **27**(3):515–523.
Stevens D, Jr., Olsen AR. 2004. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* **99**(465):262–278.
Strobl RO, Robillard PD, Shannon RD, Day RL, McDonnell AJ. 2006. A water quality monitoring network design methodology for the selection of critical sampling points: part I. *Environmental Monitoring and Assessment* **112**(1-3):137–158.
Ver Hoef JM. 2002. Sampling and geostatistics for spatial data. *Ecoscience* **9**(2):152–161.
Ver Hoef JM, Peterson E, Theobald D. 2006. Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics* **13**(4): 449–464.
Ver Hoef JM, Peterson EE. 2010. A moving average approach for spatial statistical models of stream networks. *Journal of the American Statistical Association* **105**(489):6–18.
Ver Hoef JM, Peterson EE, Clifford D, Shah R. 2014. SSN: An R package for spatial statistical modeling on stream networks. *Journal of Statistical Software* **56**:1–45.
Xia G, Miranda ML, Gelfand AE. 2006. Approximately optimal spatial design approaches for environmental health data. *Environmetrics* **17**(4):363–385.
Zhu Z, Stein ML. 2005. Spatial sampling design for parameter estimation of the covariance function. *Journal of Statistical Planning and Inference* **134**(2):583–603.
Zhu Z, Stein ML. 2006. Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics* **11**(1):24–44.
Zhu Z, Zhang H. 2006. Spatial sampling design under the infill asymptotic framework. *Environmetrics* **17**(4):323–337.
Zimmerman DL. 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics* **17**(6): 635–652.
Zimmerman DL, Cressie N. 1992. Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Annals of the Institute of Statistical Mathematics* **44**(1):27–43.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

