Ecology, 91(3), 2010, pp. 644–651

A mixed-model moving-average approach to geostatistical modeling in stream networks

Erin E. Peterson[1,3] and Jay M. Ver Hoef[2]

[1]*CSIRO Division of Mathematics, Informatics, and Statistics, 120 Meiers Road, Indooroopilly, Queensland 4068 Australia*
[2]*National Marine Mammal Lab, Alaska Fisheries Science Center, NOAA National Marine Fisheries Service, 7600 Sand Point Way NE, Seattle, Washington 98115-6349 USA*

Abstract. Spatial autocorrelation is an intrinsic characteristic in freshwater stream environments where nested watersheds and flow connectivity may produce patterns that are not captured by Euclidean distance. Yet, many common autocovariance functions used in geostatistical models are statistically invalid when Euclidean distance is replaced with hydrologic distance. We use simple worked examples to illustrate a recently developed moving-average approach used to construct two types of valid autocovariance models that are based on hydrologic distances. These models were designed to represent the spatial configuration, longitudinal connectivity, discharge, and flow direction in a stream network. They also exhibit a different covariance structure than Euclidean models and represent a true difference in the way that spatial relationships are represented. Nevertheless, the multi-scale complexities of stream environments may not be fully captured using a model based on one covariance structure. We advocate using a variance component approach, which allows a mixture of autocovariance models (Euclidean and stream models) to be incorporated into a single geostatistical model. As an example, we fit and compare "mixed models," based on multiple covariance structures, for a biological indicator. The mixed model proves to be a flexible approach because many sources of information can be incorporated into a single model.

Key words: *geostatistics; hydrologic distance; moving average; scale; spatial autocorrelation; streams.*

Introduction

Tobler's first law of geography states that "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970). In the field of geostatistics, this phenomenon is referred to as spatial autocorrelation or spatial autocovariance, which quantitatively represents the degree of statistical dependency between random variables using spatial relationships (Cressie 1993). Geostatistical models are somewhat similar to the conventional linear statistical model; they have a deterministic mean function, but the assumption of independence is relaxed and spatial autocorrelation is permitted in the random errors. For example, in a universal kriging model the deterministic mean is assumed to vary spatially and is modeled as a linear function of known explanatory variables (in contrast to ordinary kriging where the mean is unknown, but constant). Local deviations from the mean are then modeled using the spatial autocorrelation between nearby sites. Thus, geostatistical models are typically able to model additional variability in the response variable and make more accurate predictions when the data are spatially autocorrelated. For detailed information about geostatistical methods, please see Cressie (1993) or Chiles and Delfiner (1999).

Spatial autocorrelation is particularly relevant in freshwater stream environments where nested watersheds and flow connectivity may produce patterns that are not captured by Euclidean distance (Dent and Grimm 1999, Torgersen and Close 2004, Ganio et al. 2005, Monestiez et al. 2005, Peterson et al. 2006, Ver Hoef et al. 2006; Ver Hoef and Peterson, *in press*). Thus, aquatic ecologists may have been hesitant to use traditional geostatistical methods, which depend on Euclidean distance, because they did not make sense from an ecological standpoint. Covariance matrices based on Euclidean distance do not represent the spatial configuration, longitudinal connectivity, discharge, or flow direction in a stream network. In addition to being ecologically deficient, many common autocovariance functions are not generally valid when Euclidean distance is simply replaced with a hydrologic distance measure (Ver Hoef et al. 2006). A generally valid autocovariance function is guaranteed to produce a covariance matrix that is symmetric and positive-definite, with all nonnegative diagonal elements, regardless of the configuration of the stream segments or sample sites. If these conditions are not met, it may result in negative prediction variances, which violates the assumptions of geostatistical modeling. These issues

Manuscript received 12 September 2008; revised 21 August 2009; accepted 15 October 2009. Corresponding Editor: J. A. Jones.
[3] E-mail: erin.peterson@csiro.au

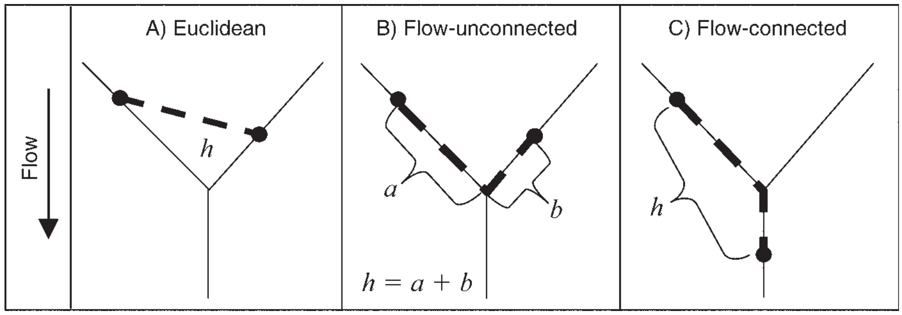

FIG. 1. The distance, h or $h = a + b$, between locations (black circles) is represented by a dashed line. (A) Euclidean distance is used to represent Euclidean relationships. Hydrologic distance can be used to represent both (B) flow-unconnected and (C) flow-connected relationships in a stream network. Water must flow from one site to another in order for the pair to be considered flow connected (C). In contrast, flow-unconnected sites do not share flow but do reside on the same stream network (B).

made it necessary to develop new geostatistical methodologies for stream networks, which permit valid covariances to be generated based on a variety of hydrologic relationships (Cressie et al. 2006, Ver Hoef et al. 2006; Ver Hoef and Peterson, *in press*). Our goal is to introduce these new autocovariance models to ecologists.

There are currently two types of distance measures that can be used to produce valid covariance matrices for geostatistical modeling in stream networks (when used with the appropriate autocovariance function): Euclidean distance and hydrologic distance. Euclidean distance is the straight-line distance between two locations and all locations within a study area have the potential to be spatially correlated when it is used (Fig. 1A). A hydrologic distance is simply the distance between two locations when measurement is restricted to the stream network. In contrast to Euclidean distance, locations within a study area do not automatically have the potential to be spatially correlated when a hydrologic distance is used. Instead, rules based on network connectivity and flow direction can be used to prevent spatial autocorrelation between locations and thus represent different hydrologic relationships. For example, a "flow-connected" relationship requires that water flow from one location to another for two sites to be correlated (Fig. 1C). When this condition is not met, sites have a "flow-unconnected" relationship (Fig. 1B) and these two sites can be made spatially independent. Likewise, whole networks (i.e., stream segments that share a common stream outlet anywhere downstream) can be made dependent.

Freshwater stream environments are typically considered open systems (Townsend 1996) with complex processes and interactions occurring between and within multiple aquatic and terrestrial scales. As a result, multiple patterns of spatial autocorrelation may be present in freshwater ecosystems (Peterson et al. 2006; Ver Hoef and Peterson, *in press*). The strength of each pattern may also vary at different spatial scales since the influence of environmental characteristics has been shown to vary with scale (Sandin and Johnson 2004). Consequently, we believe that a geostatistical model based on a mixture of covariances (i.e., multiple spatial relationships) may better fit the data than a model based on a single covariance structure.

Geostatistical models for stream network data are relatively new and may be unfamiliar to aquatic scientists. Here we review the current state of geostatistical modeling techniques for stream networks. A model-based biological indicator collected by the Ecosystem Health Monitoring Program (Bunn et al., *in press*) is used as a case study to demonstrate the approach. Simple worked examples and more details are given in the on-line appendices.

Stream network models

The stream network models of Ver Hoef et al. (2006) and Cressie et al. (2006) are based on moving-average (MA) constructions. MA constructions are flexible and can be used to create a large number of autocovariance functions (Barry and Ver Hoef 1996). They are developed by creating random variables as the integration of a MA function over a white noise random process. For our purposes, the key point is that spatial autocorrelation occurs when there is overlap between the MA function of one random variable and that of another, which we explain in greater detail for two classes of models.

Tail-up models

Models that are based on hydrologic distance and only allow autocorrelation for flow-connected relationships are referred to as "tail-up" (TU) models (Ver Hoef et al. 2006; Ver Hoef and Peterson, *in press*) because the tail of the MA function points in the upstream direction. There are a large number of TU MA functions that one could use such as the exponential, spherical, linear-with-sill, or mariah models (Ver Hoef et al. 2006), and each has a unique shape, which determines a unique autocorrelation function. Autocorrelation occurs when MA functions overlap among sites, with greater autocorrelation resulting from greater overlap. For example, in Fig. 2A, B, the shape of the TU MA

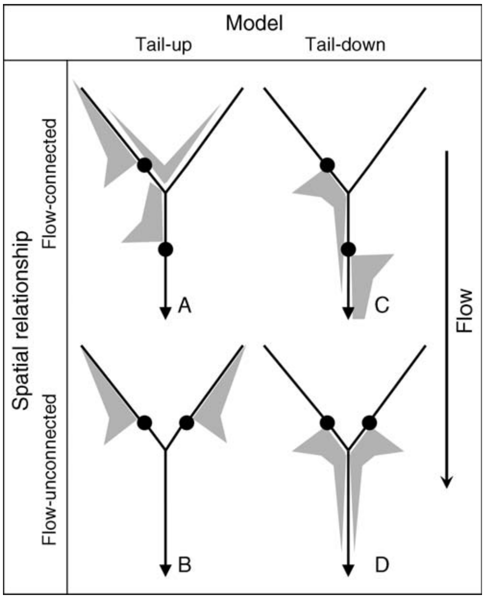

FIG. 2. The moving-average functions for the (A, B) tail-up and (C, D) tail-down models in both (A, C) flow-connected and (B, D) flow-unconnected cases. The moving-average functions are shown in gray with the width representing the strength of the influence for each potential neighboring location. Spatial autocorrelation occurs between locations when the moving-average functions overlap (A, C, and D).

function (shown in gray) dictates the relative influence of upstream values when the random variable for a specific location is generated. Values at short upstream hydrologic distances have a larger influence because there is a greater amount of overlap in the MA functions and this influence tends to decrease with distance upstream.

To better understand how the TU MA function is applied to the unique conditions of a stream network, imagine applying the function to a small network, moving upstream segment by segment. When two locations are flow-unconnected the tails of the MA functions do not overlap (Fig. 2B) and the locations do not have the potential to be spatially correlated. When two locations are flow-connected, the MA functions may overlap and one location has the potential to be spatially correlated with another location (Fig. 2A). When the MA function reaches a confluence in the network, segment weights (called segment PIs in Appendix A) are used to proportionally (i.e., they must sum to 1) allocate, or split, the function between upstream segments (Fig. 2A). The MA function could simply be split evenly between the two (or more) upstream branches, but this would not accurately represent differences in influence related to factors such as discharge. Instead, segment weights can be used to ensure that locations residing on segments that have the strongest influence on conditions at a downstream location are given greater weight in the model (Ver Hoef et al. 2006). Thus, autocorrelation in the TU models depends on flow-connected hydrologic distance. However, it also depends on the number of confluences found in the path between flow-connected sites and the weightings assigned to each of the tributaries. This feature in particular makes stream network models different than classical geostatistical models based on Euclidean distance. It also makes TU models particularly useful for modeling organisms or materials that move passively downstream, such as waterborne chemicals.

The segment weights can be based on any ecologically relevant feature, such as discharge, which is thought to represent relative influence in a stream network. However, discharge data are rarely available for every stream segment throughout a region. As an alternative, watershed area is sometimes used as a surrogate for discharge (Peterson et al. 2007). It is intuitive to think about a site's influence on downstream conditions in terms of discharge or watershed area; stream segments that contribute the most discharge or area to a downstream location are likely to have a strong influence on the conditions found there. However, spatial weights can be based on any measure as long as some simple rules are followed during their construction (see Appendix A for details on how to construct valid segment weights). The segment weights may also be based on measures that represent the sum of the upstream measures (i.e., a segment does not contribute anything to itself), such as Shreve's stream order (Shreve 1966), as used by Cressie et al. (2006). Using segment weights that are normalized so that they sum to one ensures that all random variables have a constant variance (Ver Hoef et al. 2006) if desired, which is typical in geostatistics. As an aside, the MA construction could also allow for non-stationary variances, but those models will not be explored here.

The construction of a TU covariance matrix is based on the hydrologic distance *and* a spatial weights matrix (developed from the segment weights, as illustrated in Appendix A) between flow-connected locations. Furthermore, all flow-unconnected locations are uncorrelated. Additional details and a simple worked example are provided in Appendix A to more clearly illustrate the construction of a TU covariance matrix.

Tail-down models

Tail-down (TD) models allow autocorrelation between *both* flow-connected and flow-unconnected pairs of sites in a stream network (Ver Hoef and Peterson, *in press*). The MA function for a TD model is defined so that it is only non-zero downstream of a location. In other words, the tail of the MA function points in the downstream direction (Fig. 2C, D). Spatial autocorrelation is modeled somewhat differently in a flow-connected vs. flow-unconnected situation due to the way the overlap occurs in the MA functions (Fig. 2C, D). Notice also that the input data requirements are unique for each case. The total hydrologic distance, h (Fig. 1C), is used for flow-connected pairs, but the hydrologic distances, a and b, from each site to a common confluence are used for flow-unconnected pairs (Fig. 1B). As before, more overlap in the MA function implies more autocorrelation. Also, segment weights are not used to model flow-connected relationships since the MA function points downstream (Fig. 2C) and there is no need to split the function to maintain constant variances. Additional details and a simple worked example showing the TD construction of a covariance matrix are provided in Appendix A.

Although the TD model allows spatial autocorrelation between both flow-connected and flow-unconnected pairs, the relative strength of spatial autocorrelation for each type is restricted (Ver Hoef and Peterson, *in press*). For example, consider the situation where there are two pairs of locations, one pair is flow-connected and the other flow-unconnected, and the distance between the two pairs is equal, $a + b = h$. In this case, the strength of spatial autocorrelation is generally equal or greater for flow-unconnected pairs (Ver Hoef and Peterson, *in press*) in the TD models. In fact, none of the current models are able to generate a TD model with significantly stronger spatial autocorrelation between flow-connected pairs (Ver Hoef and Peterson, *in press*) than flow-unconnected pairs for an equal hydrologic distance. These restrictions on spatial autocorrelation in the TD model may make sense for fish populations that have the tendency to invade upstream reaches, such as nonnative brook trout (*Salvelinus fontinalis*; Peterson and Fausch 2003). Yet, there are other situations where it would be useful to generate a model with stronger spatial autocorrelation between flow-connected pairs than flow-unconnected pairs. Peterson and Fausch (2003) also studied the movement characteristics of native cutthroat trout (*Oncorhynchus clarki*) and found that they moved downstream much more often than upstream. Here, a model with stronger spatial autocorrelation between flow-connected locations, which also allows for some spatial autocorrelation between flow-unconnected locations, might best fit the data. Yet, neither the TU or TD model can be used to represent this particular covariance structure. Therefore, another modeling solution is required. Here we turn to the mixed model, which is a very general approach, to address these issues.

Mixed models

The mixed model is closely related to the basic linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where the matrix $\mathbf{X}$ contains measured explanatory variables and the parameter vector $\boldsymbol{\beta}$ establishes the relationship of the explanatory variables to the response variable, contained in the vector $\mathbf{y}$. The random errors are contained in the vector $\boldsymbol{\varepsilon}$, and the general formulation is $\mathrm{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma}$ is a matrix. The mixed-model is simply a variance component approach, which allows the error term to be expanded into several random effects ($\mathbf{z}$.):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\sigma}_{\mathrm{EUC}}\mathbf{z}_{\mathrm{EUC}} + \boldsymbol{\sigma}_{\mathrm{TD}}\mathbf{z}_{\mathrm{TD}} + \boldsymbol{\sigma}_{\mathrm{TU}}\mathbf{z}_{\mathrm{TU}} + \boldsymbol{\sigma}_{\mathrm{NUG}}\mathbf{z}_{\mathrm{NUG}} \tag{2}$$

where $\mathrm{cor}(\mathbf{z}_{\mathrm{EUC}}) = \mathbf{R}_{\mathrm{EUC}}$, $\mathrm{cor}(\mathbf{z}_{\mathrm{TD}}) = \mathbf{R}_{\mathrm{TD}}$, $\mathrm{cor}(\mathbf{z}_{\mathrm{TU}}) = \mathbf{R}_{\mathrm{TU}}$ are matrices of autocorrelation values for the Euclidean (EUC), TD, and TU models, $\mathrm{cor}(\mathbf{z}_{\mathrm{NUG}}) = \mathbf{I}$ where NUG is the nugget effect, $\mathbf{I}$ is the identity matrix, and σ^2_{EUC}, σ^2_{TD}, σ^2_{TU}, and σ^2_{NUG} are the respective variance components. The mixed-model construction implies that covariance matrices based on different types of models, such as the EUC, TU, and TD are combined to form a valid covariance mixture:

$$\Sigma = \sigma^2_{\mathrm{EUC}}\mathbf{R}_{\mathrm{EUC}} + \sigma^2_{\mathrm{TD}}\mathbf{R}_{\mathrm{TD}} + \sigma^2_{\mathrm{TU}}\mathbf{C}_{\mathrm{TU}} \odot \mathbf{W}_{\mathrm{TU}} + \sigma^2_{\mathrm{NUG}}\mathbf{I} \tag{3}$$

where we have further decomposed $\mathbf{R}_{\mathrm{TU}} = \mathbf{C}_{\mathrm{TU}} \odot \mathbf{W}_{\mathrm{TU}}$ into the Hadamard product of the flow-connected autocorrelations $\mathbf{C}_{\mathrm{TU}}$ (unweighted) and the spatial weights matrix $\mathbf{W}_{\mathrm{TU}}$.

The variance component model is attractive for several reasons. First, it solves the problem mentioned in the previous section; namely that the combination of the TU covariance matrix and TD covariance matrix allows for the possibility of more autocorrelation among flow-connected pairs of sites, with somewhat less autocorrelation among flow-unconnected pairs of sites. Secondly, the multiple range parameters can capture patterns at multiple scales. Generally, large scale patterns are the most obvious and explanatory variables are incorporated to help explain them. Spatial patterns of intermediate scale, which have not been measured with explanatory variables, are captured by the range parameters for the EUC, TU, and TD models and the relative strength of each model is given by its variance component. The spatial weights are used to capture the influence of branching in the network, flow direction, and discharge. Finally, some spatial variation occurs at a scale finer than the closest measurements; these are modeled as independent error, which is represented by the nugget effect. We now turn to an example to make these concepts clearer.

Example

The Ecosystem Health Monitoring Program (EHMP) has been collecting indicators of biotic structure and ecosystem function throughout South East Queensland (SEQ), Australia (Fig. 3A) since 2002 (Bunn et al., *in press*). The program aims are to evaluate the condition and trend in ecological health of freshwater environments and to guide investments in catchment protection and rehabilitation. Metrics based on freshwater fish assemblages are commonly used as indicators of ecological health because they are thought to provide a holistic approach to assessment across broad spatial and temporal scales (Harris 1995). In this example, we used a model-based biological indicator, the proportion of native fish species expected (PONSE), which is simply the ratio of observed to expected native freshwater fish species richness (Kennard et al. 2006). The expected species composition data were generated using a referential model and represent the native species that are expected to be present in a physically similar, but undisturbed stream (Kennard et al. 2006). The observed species composition data were collected at 86 survey sites (Fig. 3B) in the spring of 2005. Hereafter, we will refer to these as the "observed sites." PONSE scores at the observed sites ranged from 0.09 to 1, with a mean of 0.76 and a median of 0.83. We also nonrandomly selected 137 "prediction sites" where PONSE was not generated. Additional information about the study area, the sampling methods, and the predictive model used to generate the PONSE scores is provided in Appendix B.

We generated the spatial data necessary for geostatistical modeling in a geographic information system (GIS). These included seven explanatory variables representing watershed-scale land use, land cover, and topographic characteristics for each observed and prediction site, as well as the hydrologic distances and spatial weights, which were based on watershed area. The EHMP classified streams based on elevation, mean annual rainfall, stream order, and stream gradient to create four EHMP regions (Bunn et al., *in press*), which we also included as a site-scale explanatory variable. Additional information about the GIS methodology and explanatory variables can be found in Appendix B.

We used a two-step model selection procedure to compare models. First we fixed the covariance structure and focused on selecting the explanatory variables using the Akaike information criterion (AIC; Akaike 1974). During the second phase of model selection we focused on selecting the most appropriate covariance structure. We fixed the selected explanatory variables, and then compared every linear combination of TU, TD, and EUC covariance structures, where four different autocovariance functions were tested for each model type. This resulted in a total of 124 models, each with a different covariance structure. In addition, we fit a classical linear model assuming independence to compare to models that use spatial autocorrelation. Once the final model was identified, universal kriging (Cressie 1993) was used to make predictions at the 137 prediction sites. Please see Appendix B for details on the model selection procedure.

Our results show that a model based on a mixture of covariances produced more precise PONSE predictions than models based on a single covariance structure (Table 1). When more than one covariance structure was incorporated, mixture models that included the TD model appeared to outperform other mixture types. In addition, all of the geostatistical models outperformed the classical linear model. The lowest RMSPE value was produced by the exponential TU/linear-with-sill TD mixture model. Hereafter this will be referred to as the "final model." The final model only contained one explanatory variable, mean slope in the watershed, which was positively correlated with PONSE. This statistical relationship may represent a physical relationship between PONSE and an anthropogenic disturbance gradient such as land use, water quality, channel or riparian condition, or in-stream habitat (Kennard et al. 2006), which is correlated to slope. For example, watersheds with steeper slopes might be expected to have less cleared or cropped land and, as a result higher PONSE scores. More details on the fitted explanatory variables and diagnostics are given in Appendix B.

We examined the percent of the variance explained by each of the covariance components to provide more information about the covariance structure of the models. In the final model, the TU, TD, and NUG components explained 22.43%, 64.32%, and 13.25% of the variance, respectively. Although the full covariance mixture (TU/TD/EUC) was not the best model in this example based on the RMSPE (Table 1), the loss in predictive ability was only 0.34% when it was used instead of the final model. One of the advantages of a mixed model is its flexibility; a covariance mixture can be used to represent the whole range of covariance structures used in the mixture (i.e., single EUC, TU, or TD or any combination of the three). Therefore, we recommend fitting a full covariance mixture; this allows the data to determine which variance components have the strongest influence, rather than making an implicit assumption about the spatial structure in the data by using a single covariance structure.

The predictions and prediction standard errors produced by the final model (Fig. 3C–F) exhibit some of the common characteristics of kriging predictions. Predictions and their standard errors vary depending on the estimated regression coefficients and distances to observed data sites. If the explanatory variables at the prediction site are not well represented in the observed data set a large standard error will be assigned to the prediction. The predictions change gradually along stream segments (Fig. 3C–F) and the prediction standard errors tend to be smaller near observed data and increase as a function of distance (Fig. 3C).

The predictions and prediction standard errors shown in Fig. 3 also demonstrate some characteristics that are unique to geostatistical models for stream networks. For example, the TU model allows discontinuities in predictions at confluences, which enables flow-unconnected tributaries to receive markedly disparate PONSE predictions (Fig. 3D). The effect of the spatial weights on prediction uncertainty is apparent if upstream segments have not been sampled (Fig. 3E). In this situation, uncertainty is relatively high upstream of a confluence because various combinations of the two PONSE scores could be contributing to the downstream observed score. Two sites located on the main stem are strongly correlated (based on a combination of the tail-up and tail-down models) and uncertainty in the

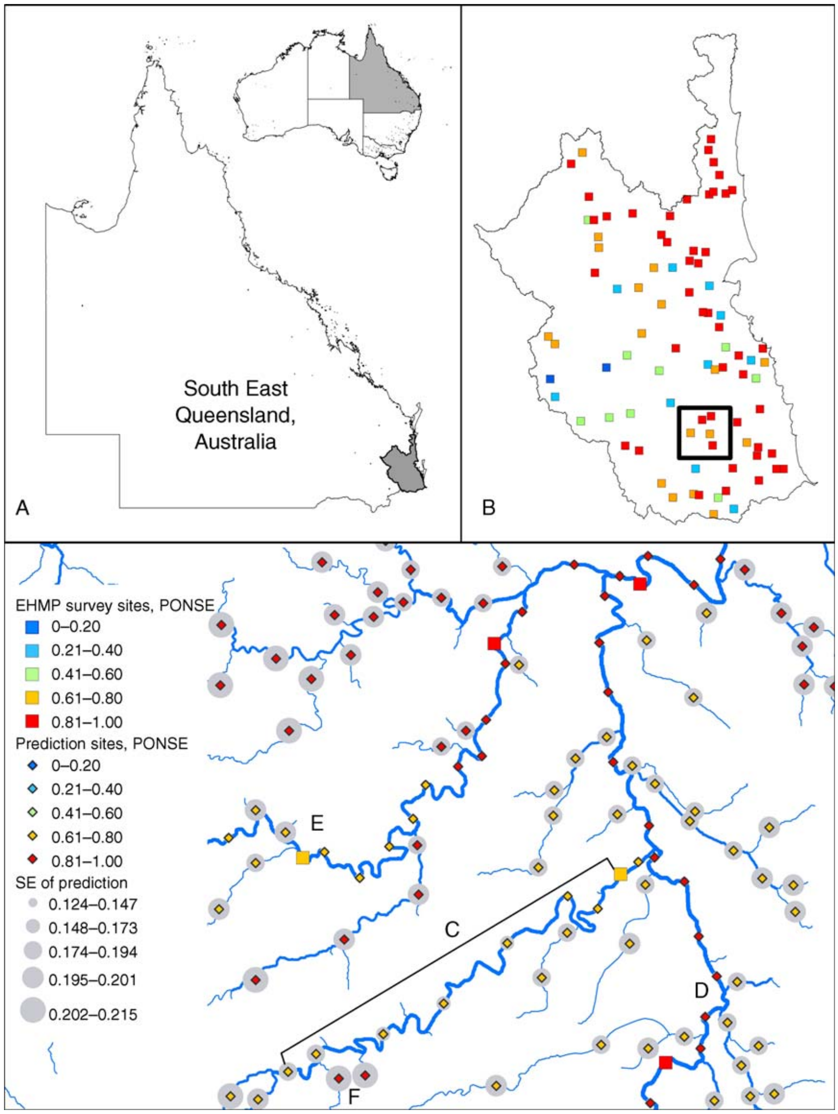

FIG. 3. (A, B) The Ecosystem Health Monitoring Program collects a suite of ecological indicators at survey sites located throughout South East Queensland, Australia. Predictions and prediction variances for the proportion of native fish species expected (PONSE) were produced using the exponential tail-up/linear-with-sill tail-down model (C–F). Observed sites are represented by larger squares, and prediction sites are represented by smaller diamonds. Prediction variances are shown in gray and are sized in proportion to their value (0.124–0.215). Thus, predictions with a large shaded gray area have less precision. Blue line segments symbolize stream segments, and the width of the line is proportional to the watershed area.

TABLE 1. A comparison of mixture models.

Mixture	Model 1	Model 2	Model 3	RMSPE
Nonspatial				0.2573
TU	linear-sill			0.2428
TD	linear-sill			0.2112
EUC	spherical			0.2283
TU/TD	exponential	linear-sill		0.2088
TU/EUC	linear-sill	Gaussian		0.2190
TD/EUC	linear-sill	spherical		0.2103
TU/TD/EUC	mariah	linear-sill	exponential	0.2094

Notes: Models shown represent the best model fit for each mixture type based on the root mean square prediction error (RMSPE). Models are tail-up (TU), tail-down (TD), and Euclidean (EUC).

downstream predictions is low (Fig. 3E). However, when the spatial weights are based on watershed area, and one upstream segment dominates a side branch, uncertainty in the upstream direction may also be low (Fig. 3C). In contrast, prediction uncertainty in small upstream tributaries is relatively large (based on the tail-up model) since the spatial correlation between the observed and prediction site is weak (Fig. 3F).

Discussion and Conclusions

Spatial autocorrelation is clearly a natural phenomenon given the open nature of stream ecosystems (Townsend 1996) and the complexity of process interactions occurring within and between the stream and the terrestrial environment. Observable patterns of spatial autocorrelation are likely caused by multiple spatially dependent factors (Wiens 2002). As a result, conceptualization and modeling in stream ecosystems requires tools that are able to account for dynamic multi-scale patterns ranging from the reach to the network scale (Townsend 1996). Yet many models do not account for these natural interdependencies. Ignoring spatial autocorrelation makes it possible to use traditional statistical methods, which rest on the assumption of independent random errors. Although convenient, our opinion is that it is better to develop new statistical methods that represent the unique ecological conditions found in the environment.

Traditional geostatistical methods account for spatial correlation in the error term, but they may not fully capture spatial autocorrelation structures in stream environments. Euclidean covariance functions, such as the spherical, cubic, exponential, and Gaussian, are all strikingly similar (Chiles and Delfiner 1999). As a result, two sites that are spatially correlated using one function are likely to be spatially correlated using another function. In contrast, the stream models have a markedly different autocovariance structure than the Euclidean models and represent a true difference in the way that spatial relationships are represented. The splitting of the covariance function along a branching network is also unique to stream models. Previously, autocorrelation has been restricted to a linear feature or to two-dimensional space, which cannot be used to capture hydrologic patterns of spatial autocorrelation in a branching network. The stream models given here provide an innovative statistical alternative because they were specifically developed to represent the spatial configuration, longitudinal connectivity, discharge, and flow direction in a stream network.

The mixed model is an extremely flexible approach because many sources of information can be incorporated into a single model. The explanatory variables are used to account for influential factors that can be measured. However, it is common for other influential variables to be left unmeasured due to a lack of resources or an incomplete understanding about the stream process. In a mixed model, autocorrelated errors can be modeled at multiple scales using a variety of distance measures. This produces a rich and complex covariance structure, that when combined with the explanatory variables, can be used to account for the effects of both measured and unmeasured variables at multiple scales. Given the multi-scale complexities of stream ecosystems, we expect these models to better represent the spatial complexity and interdependencies in a streams data sets.

The flexibility of the mixture model makes the method suitable for modeling a variety of variables collected within or near a stream network. Although we used a Gaussian response in the example, the autocovariance functions described here can also be used to produce covariance matrices for kriging Poisson or binomial variables, such as fish counts or the presence or absence of species. It may also be possible to model variables that are present in riparian areas rather than streams, but are expected to exhibit both Euclidean and hydrologic patterns of spatial autocorrelation. This might include riparian plant species that employ waterborne dispersal strategies or animal species that migrate along stream corridors. Finally, we chose to use watershed area to calculate the spatial weights because it made sense given the response variable, but any measurement, such as water velocity, stream width, or discharge, could be used as long as it meets the statistical requirements set out in Appendix A. In principle, multiple weighting schemes could be compared as part of the model selection criteria.

The mixed model is also useful from a management perspective since predictions with estimates of uncertainty may be generated throughout a stream network (Fig. 3). This ability provides a way to move from disjunct stream management, which is traditionally based on site or reach-scale samples, to a more continuous approach that yields location specific predictions and accounts for the network as a whole (Fausch et al. 2002). For example, the predicted PONSE values in two small tributaries were relatively high compared to those found in the main stem (Fig. 3F), which may indicate that those locations have the potential to act as natural refugia for native fish residing in marginally suitable habitats. The potential importance of these tributaries could go unnoticed without the ability to evaluate the network as a whole. This quality is particularly useful because it helps to ensure that management actions are targeted or scaled appropriately (Lake et al. 2007). Including estimates of uncertainty also enables users to gauge the reliability of the predictions and to target future sampling efforts in areas with large amounts of uncertainty or a greater potential for ecological impairment.

Geostatistical modeling in stream networks has the potential to be a powerful statistical tool for freshwater stream research and management. It can be used to capture and quantify spatial patterns at multiple scales, which may provide additional information about ecosystem structure and function (Levin 1992); a key step in developing new ecological theories. Our goal has been to make this methodology accessible to ecologists so that the models can be implemented, modified, and improved to derive additional information from streams data sets. We believe that when geostatistical models for stream networks are used in conjunction with sound ecological knowledge the result will be a more ecologically representative model that may be used to broaden our understanding of stream ecosystems.

Acknowledgments

We thank the SEQ Healthy Waterways Partnership for sharing the data set and Beth Gardner, Mark Kennard, Petra Kuhnert, and an anonymous reviewer for their constructive comments on the manuscript. This project received financial support from the National Marine Fisheries Service of NOAA, the CSIRO Division of Mathematical and Information Sciences, and the Australian Water for a Healthy Country Flagship.

Literature Cited

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6):716–722.

Barry, R. P., and J. M. Ver Hoef. 1996. Blackbox kriging: spatial prediction without specifying variogram models. Journal of Agricultural, Biological, and Environmental Statistics 1:297–322.

Bunn, S., E. Abal, M. Smith, S. Choy, C. Fellows, B. Harch, M. Kennard, and F. Sheldon. *In press*. Integration of science and monitoring of river ecosystem health to guide investments in catchment protection and rehabilitation. Freshwater Biology.

Chiles, J., and P. Delfiner. 1999. Geostatistics: modeling spatial uncertainty. John Wiley and Sons, New York, New York, USA.

Cressie, N. 1993. Statistics for spatial data. Revised edition. John Wiley and Sons, New York, New York, USA.

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

Dent, C. L., and N. B. Grimm. 1999. Spatial heterogeneity of stream water nutrient concentrations over successional time. Ecology 80:2283–2298.

Fausch, K. D., C. E. Torgersen, C. V. Baxter, and H. W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream reaches. BioScience 52: 483–498.

Ganio, L. M., C. E. Torgersen, and R. E. Gresswell. 2005. A geostatistical approach for describing spatial pattern in stream networks. Frontiers in Ecology and the Environment 3:138–144.

Harris, J. H. 1995. The use of fish in ecological assessments. Australian Journal of Ecology 20:65–80.

Kennard, M. J., B. D. Harch, B. J. Pusey, and A. H. Arthington. 2006. Accurately defining the reference condition for summary biotic metrics: a comparison of four approaches. Hydrobiologia 572:151–70.

Lake, P. S., N. Bond, and P. Reich. 2007. Linking ecological theory with stream restoration. Freshwater Biology 52:597–615.

Levin, S. A. 1992. The problem of pattern and scale in ecology. Ecology 73:1943–1967.

Monestiez, P., J.-S. Bailly, P. Lagacherie, and M. Voltz. 2005. Geostatistical modelling of spatial processes on directed trees: application to fluvisol extent. Geoderma 128:179–191.

Peterson, D. P., and K. D. Fausch. 2003. Upstream movement by nonnative brook trout (*Salvenlinus fontinalis*) promotes invasion of native cutthroat trout (*Oncorhynchus clarki*) habitat. Canadian Journal of Fisheries and Aquatic Sciences 60:1502–1516.

Peterson, E. E., A. A. Merton, D. M. Theobald, and N. S. Urquhart. 2006. Patterns of spatial autocorrelation in stream water chemistry. Environmental Monitoring and Assessment 121:569–594.

Peterson, E. E., D. M. Theobald, and J. M. Ver Hoef. 2007. Geostatistical modelling on stream networks: developing valid covariance matrices based on hydrologic distance and stream flow. Freshwater Biology 52:267–279.

Sandin, L., and R. K. Johnson. 2004. Local, landscape and regional factors structuring benthic macroinvertebrate assemblages in Swedish streams. Landscape Ecology 19:501–514.

Shreve, R. L. 1966. Statistical law of stream numbers. Journal of Geology 74:17–37.

Tobler, W. R. 1970. Computer movie simulating urban growth in Detroit region. Economic Geography 46:234–240.

Torgersen, C. E., and D. A. Close. 2004. Influence of habitat heterogeneity on the distribution of larval Pacific lamprey (*Lampetra tridentata*) at two spatial scales. Freshwater Biology 49:614–630.

Townsend, C. R. 1996. Concepts in river ecology: pattern and process in the catchment hierarchy. Archiv für Hydrobiologie 113(Large Rivers Supplement):3–21.

Ver Hoef, J. M., and E. E. Peterson. *In press*. A moving average approach for spatial statistical models of stream networks. Journal of the American Statistical Association.

Ver Hoef, J. M., E. E. Peterson, and D. M. Theobald. 2006. Some new spatial statistical models for stream networks. Environmental and Ecological Statistics 13:449–464.

Wiens, J. A. 2002. Riverine landscape: taking landscape ecology into the water. Freshwater Biology 47:501–515.

APPENDIX A

Formulas and examples for the construction of tail-up and tail-down covariance matrices (*Ecological Archives* E091-048-A1).

APPENDIX B

Additional details about the collection and analysis of the EHMP PONSE data set (*Ecological Archives* E091-048-A2).