A geostatistical state-space (?) model of abundance for stream networks

Daniel J. Hocking

James Thorson

Ben Letcher

Kasper Kristensen?

Disclaimer: This draft manuscript is distributed solely for purposes of scientific peer review. Its content is deliberative and predecisional, so it must not be disclosed or released by reviewers. Because the manuscript has not yet been approved for publication by the US Geological Survey (USGS), it does not represent any official finding or policy.

**Abstract**

**Introduction**

Ecologists are concerned with understanding the abundance and distribution of organisms in space and time and the conditions that influence these patterns. Surveys are frequently employed to study these patterns. However, most statistical methods used in ecology have not explicitly accounted for spatial correlation in the data beyond deterministic covariate parameters (e.g. ANOVA, GLM, linear and generalized linear mixed models). Therefore to use these analyses, researchers must design their studies to ensure that sample points are not overly close in time or space such that unaccounted for autocorrelation could occur. It is difficult to know *a priori* how close is too close. Any residual autocorrelation violates regression model assumptions and leads to biased results and potentially incorrect inference regarding population distributions and environmental relationships. Additionally, information about the spatial and temporal patterns provides potentially interesting ecological insights that would not be gained if the data were collected in a way to avoid autocorrelation (give example). For these reasons, a large field of spatial statistics has been developed and applied to ecological problems (refs: Ross et al 2012, Conn et al 2015, Gelfand example, Cressie example).

Streams are likely to have significant correlation in time and space because of the hydrologic connections creating movements and gradients of chemical and physical properties. For example, (find good example). Similarly, organisms living in streams are likely to respond to these underlying conditions and their movements are often restricted to the dendritic network creating spatial correlation in the abundance and distribution of stream organisms. Spatial models that use Euclidean distance are likely to perform poorly in stream networks because streams in close proximity can be separated by ridges and be completely unconnected or have large hydrologic distances.

A variety of statistical models have been developed to account for spatial correlations in dendritic networks. These include, but are not limited to, deriving valid covariance relationships for linear models (ref: Peterson et al 2007) and linear mixed models with moving averages that account for hydrologic distance and flow (Ver Hoef et al. 2006). Some models also include “tail-up”, “tail-down” or “two-tail” correlations to account for directional autocorrelation (ref: Ver Hoef and Peterson 2010, Peterson and Ver Hoef 2010). Additionally, block Kriging has been used for spatial interpolation (ref: Ver Hoef et al 2006).

While these models provide improved inference for many types of data, there are limitations with the current approaches. Current models account for spatial correlations but do not allow for changing spatial correlations over time as with spatiotemporal models (ref: Peterson et al. 2013 Ecology Letters). A second limitation is the inability to distinguish between process and observation error to account for imperfect detection (ref: Peterson et al. 2013 Ecology Letters). When performing count surveys of organisms, the probability of detecting each individual in the population is often less than one with little to no chance of double counting individuals. This results in a problem of inference regarding the populations and environmental effects on the population, particularly when the probability of detection is variable in time and space. To address this issue, a variety of hierarchical models have been developed separating information regarding abundance and detection (refs: royle, kery, dail and madsen). However, these models generally to not account for spatial correlation among sites. Those that do generally use random group effects assuming clustered sites to be more similar to eachother than to other clusters (refs: Hocking et al.). This coarse grouping does not allow for autocorrelation as a function of distance. For example, if sampling is done in a series of transects, all sites within a transect are treated the same (refs: e.g. Peterman et al., Hocking et al. ) even though it is likely that adjacent sites are more correlated than distant sites at the opposite ends of the transect. A final limitation of current spatial stream models is the computational challenges with analyzing large networks due to estimating large covariance structures (ref: Peterson et al. 2013 Ecology Letters).

We describe a novel and generalizable hierarchical model that accounts for spatial, temporal, and spatiotemporal autocorrelation while accounting for imperfect detection. It also addresses unexplained random variation in abundance (log-normal overdispersion).

We include R and TMB code that allows for rapid, efficient estimation of these models.

The model is assessed with simulations

1. Proof it works

Spatial model varying theta and sigma. I think it makes sense to compare these to a non-spatial model (sorry a sort of 3rd axis). Do you think we need to vary sigma since it affects the spatial correlation but as a constant and not in relation to distance? I don't think we need to test temporal or spatiotemporal components here.

2. Power analysis

spatiotemporal model varying the number of years and sites with data

3. Performance on axis

I'm unsure if this is necessary for this paper but I could vary the detection rate as you suggest holding everything else constant in a spatial model. This would be relevant for other fish species, other taxa (stream salamanders), and \*maybe\* YOY vs adults.

We also applied the model to trout data

* commonly collected by state agencies throughout the US
* Declining and threatened by climate and landuse change, overfishing, exotic species
* Likely to be spatially correlated

**Methods**

**Results**

Semivariograms and Torgegrams

* Of modeled abundance
* Of residuals

**Discussion**

**Acknowledgments**

**Literature Cited**