Climate-Aquatics Blog #27: Part 1, Spatial Statistical Models for Stream Networks: Context & Conceptual Foundations

Having the right tools is a good thing…

Hi Everyone,
This time out we're embarking on a bit of a diversion from our regularly scheduled programming as we start a 3 part mini-series on spatial statistical models for stream networks. It's out of the planned sequence since these blogs will ultimately be filed under the "Climate-Aquatics Cool Stuff Module" but I talked to the blog's publisher and he agreed that their content was significant enough that it warranted an exception & because recent developments are about to make these models much more accessible for broad application to streams anywhere, so without further adieu…

For anyone that's ever tried to build something & get by with poor quality tools or tools not specifically designed for the job at hand, you know the frustration. Something can usually be made to work but it's often a painful process and the thing built could have been much better with the right tools. Well, I'd argue that those of us working with streams and rivers and their critters have long lacked important tools for analyzing and understanding these systems at one of the spatial scales that is fundamentally important for planning and prioritizing conservation efforts. That scale is the network scale—spanning 100's – 1000's of stream kilometers and 10's – 100's of individual streams. It's this scale at which multiple stakeholders across a river basin coordinate and make choices about how & where resources are spent to protect and restore habitat or conserve populations. From a biological perspective, this scale is also relevant because it's one at which metapopulation dynamics and the exchange of individuals between populations occur to lend populations some resilience and facilitate the persistence of species within landscapes. That being the case, the network scale is where "smart maps," informed by lots of data to show the status and trends of important aquatic resources, would be powerful tools for facilitating coordination efforts among resource professionals and different agencies to make sure we do the smartest things in the smartest places.

The good news is that there are literally armies of us running around the woods measuring all manner of stream things at discrete points on those networks and 100's – 1000's of such measurements now exist in many river basins. Those data are the fundamental elements that could be stitched together to build the necessary "smart maps" but doing so also presents a few issues that have been difficult to address until recently. Those issues involve such things as accurate geo-referencing of sample locations (now easily overcome with inexpensive GPS units), computational horsepower for dealing with large databases (more & cheaper horses all the time), and having standardized protocols to obtain unbiased measurements of commonly collected stream and biological attributes (this one's still evolving but rapidly improving with development of inexpensive sensors, genetic assay techniques, and bioassessment protocols. And in practice, having bias in some measurements may be acceptable if it better informs parts of a network-scale "smart map" that would otherwise be devoid of information).

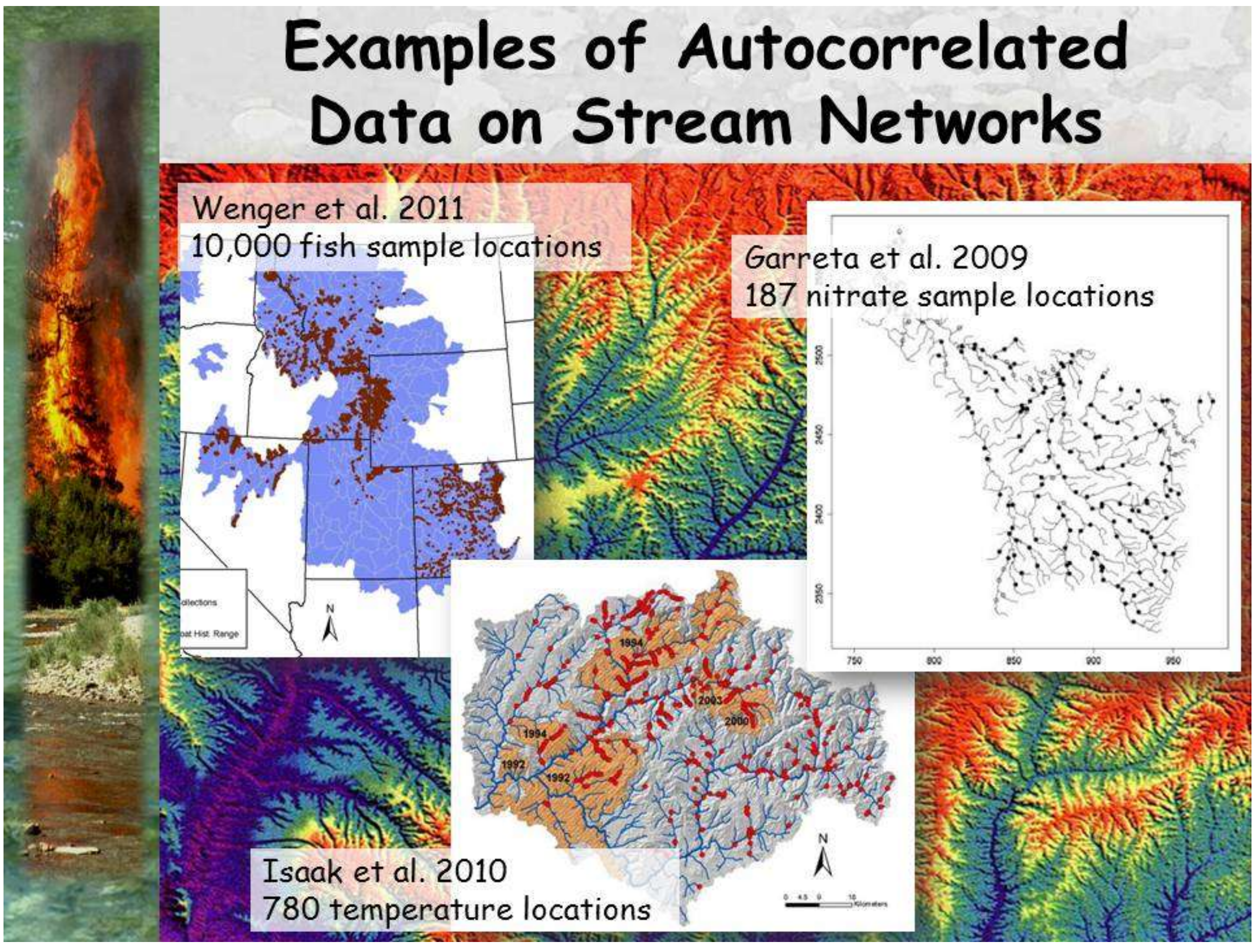
The fourth major issue, and it's a biggie, is having analytical techniques that can properly deal with those 100's – 1000's of nonrandom, clustered measurements that our uncoordinated army is generating (graphic 1). Those nonrandom site locations, especially in areas with high data densities, create a problem called spatial autocorrelation wherein there's a certain amount of spatial dependence in the samples and redundancy in the information they provide. This redundancy violates one of the basic assumptions inherent to many statistical methods, that samples are independent of one another. Analyses that ignore the way sites are located in space and potential spatial dependence among them run the risk of providing biased parameter estimates (and poor maps if used to predict) when applied to these sorts of databases. To deal with the issue, one could simply discard some of the data until the redundancy was reduced to an acceptable level, but whose data do you throw away, how much, and where?

Much better to use it all, if possible, since someone went to the time and expense of collecting the data. Plus, each sample location contains some valuable information if the data can be properly weighted to account for the degree of spatial autocorrelation with nearby sites. Spatial statistical analyses were developed, in part, to address this issue, and can be used to derive unbiased parameter estimates and valid inferences from clustered databases. These analyses have long been widely available for, and applied to, terrestrial systems but they simply haven't existed, other than a few basic descriptive measures, for data on stream networks. I know because I searched high & low for them to no avail during my Ph.D. 15 years ago as I struggled to analyze my own autocorrelated datasets. Although many (myself included) have tried work-arounds that apply spatial terrestrial analyses to stream data, we're back to that wrong tool for the job thing. It works but it's ultimately fitting a square peg in a round hole because the way that information moves through space in a stream network differs fundamentally from a 2-D terrestrial system.

Key to resolving the issue was that somebody had to develop a covariance structure that accounted for the unique topology of stream networks and the spatial configuration of samples measured on those networks. For example, streams flow in one direction, are linear features, have flow connected and unconnected sites, and tributary confluences where big changes occur over small distances. All these factors have consequences for how spatial patterns in networks manifest and how nearby sites influence one another. These network characteristics had to be embedded in the DNA of a new type of statistical model for streams if accurate renditions of spatial patterns in data measured on networks were to be made (graphics 2 and 3). I'm happy to say that that atom has finally been cracked and so this week & in a few subsequent blogs we're highlighting work by two pioneers in the field of stream statistics, Jay Ver Hoef and Erin Peterson. This duo has published steadily on the topic over the last 6 years and have developed a rich literature and solid statistical foundation for their spatial stream network models (attached are a couple recent examples, a more complete bibliography is provided in graphic 4). I can't do the work proper justice here since most of the math is over my head but I encourage you to spend some time with it and for those with especially strong quantitative skills and statistical bents, you'll be in hog heaven.

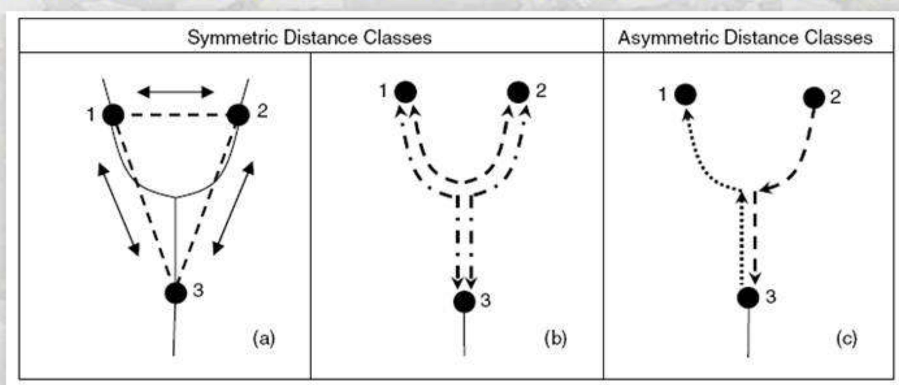
For the rest of us, suffice it to say that a really good tool now exists for analyzing many types of data measured on stream networks. Moreover, the Ver Hoef and Peterson models make it possible, even desirable, to develop integrated, interagency databases because huge amounts of new and accurate information can often be extracted from existing data at relatively small costs and over the span of a few years. Thus, the “smart maps” we need at network scales to describe historical and future stream temperature patterns could be developed from databases like NorWeST (blog #25) or others to inform species conservation efforts (graphic 5). And although these maps won’t be perfect, they’ll be significant improvements over what’s currently available, and in the climate change game, we don’t have the luxury of time, let alone the budgets, to design & execute the perfect monitoring strategy and then wait decades for the answers. In many cases, also, the spatial stream models perform significantly better than their traditional counterparts in terms of raw predictive power and are enabling new types of analyses and inference about streams that simply weren’t possible before. Next time out, we’ll explore a bit of this in more detail but it’s truly the start of a new era when fundamentally better spatial information about streams and their biotas is becoming a real possibility. That’s good because it’s still going to take a lot of work to generate that information and we’re going to need it as we strive to conserve aquatic biodiversity through this transitional century.

Until next time, best regards.
Dan



Spatial Statistical Models for Stream Networks

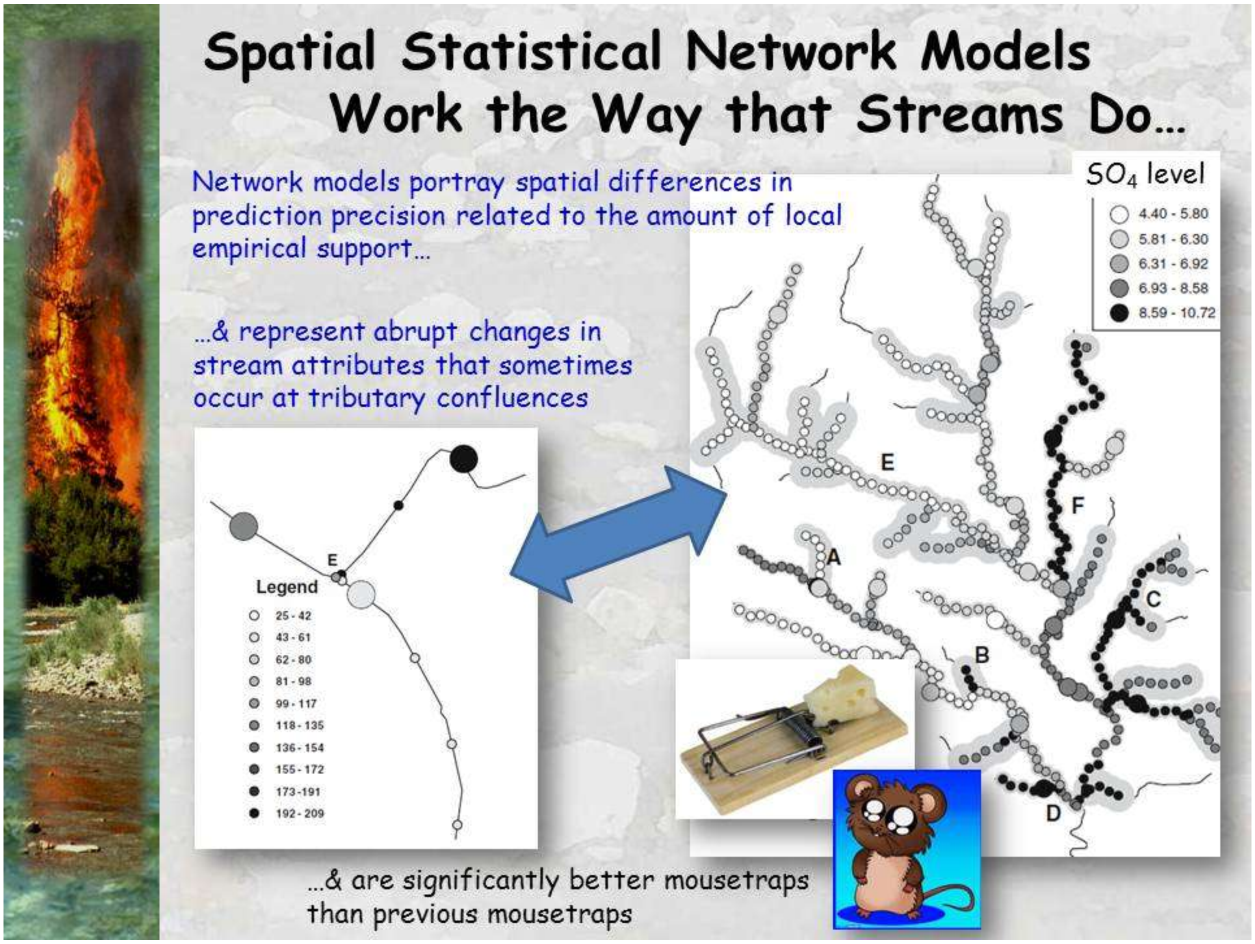
Incorporate accurate representations of how stream information flows into valid covariance structures



Advantages:

- Flexible & valid covariance structures that accommodate network topology & autocorrelation among samples
- Accurate parameter estimation from correlated databases, improved predictive ability relative to non-spatial models, and a suite of new types of estimation

Ver Hoef et al. 2006; Peterson & Ver Hoef 2010; Ver Hoef & Peterson 2010



Spatial Statistical Network Models
Work the Way that Streams Do…
Network models portray spatial differences in prediction precision related to the amount of local empirical support…
…& represent abrupt changes in stream attributes that sometimes occur at tributary confluences
Legend
25 - 42
43 - 61
62 - 80
81 - 98
99 - 117
118 - 135
136 - 154
155 - 172
173 - 191
192 - 209
E
SO_4 level
4.40 - 5.80
5.81 - 6.30
6.31 - 6.92
6.93 - 8.58
8.59 - 10.72
A
B
C
D
E
F
…& are significantly better mousetraps than previous mousetraps

Stream Network Models – Key References

Peterson, E.E., J.M. Ver Hoef. In Preparation. STARS: An ArcGIS toolset used to calculate the spatial data needed to fit spatial statistical models to stream network data. *Journal of Statistical Software* x:xxx-xxx.

Peterson, E.E., J.M. Ver Hoef. 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology* 91:644-651.

Ver Hoef, J.M., E.E. Peterson, D. Clifford, and R. Shah. In Preparation. SSN: An R package for spatial statistical modeling on stream networks. *Journal of Statistical Software* x:xxx-xxx.

Ver Hoef, J.M., and E.E. Peterson. 2010. A moving average approach for spatial statistical models of stream networks. *J American Statatistical Association* 105:6-18.

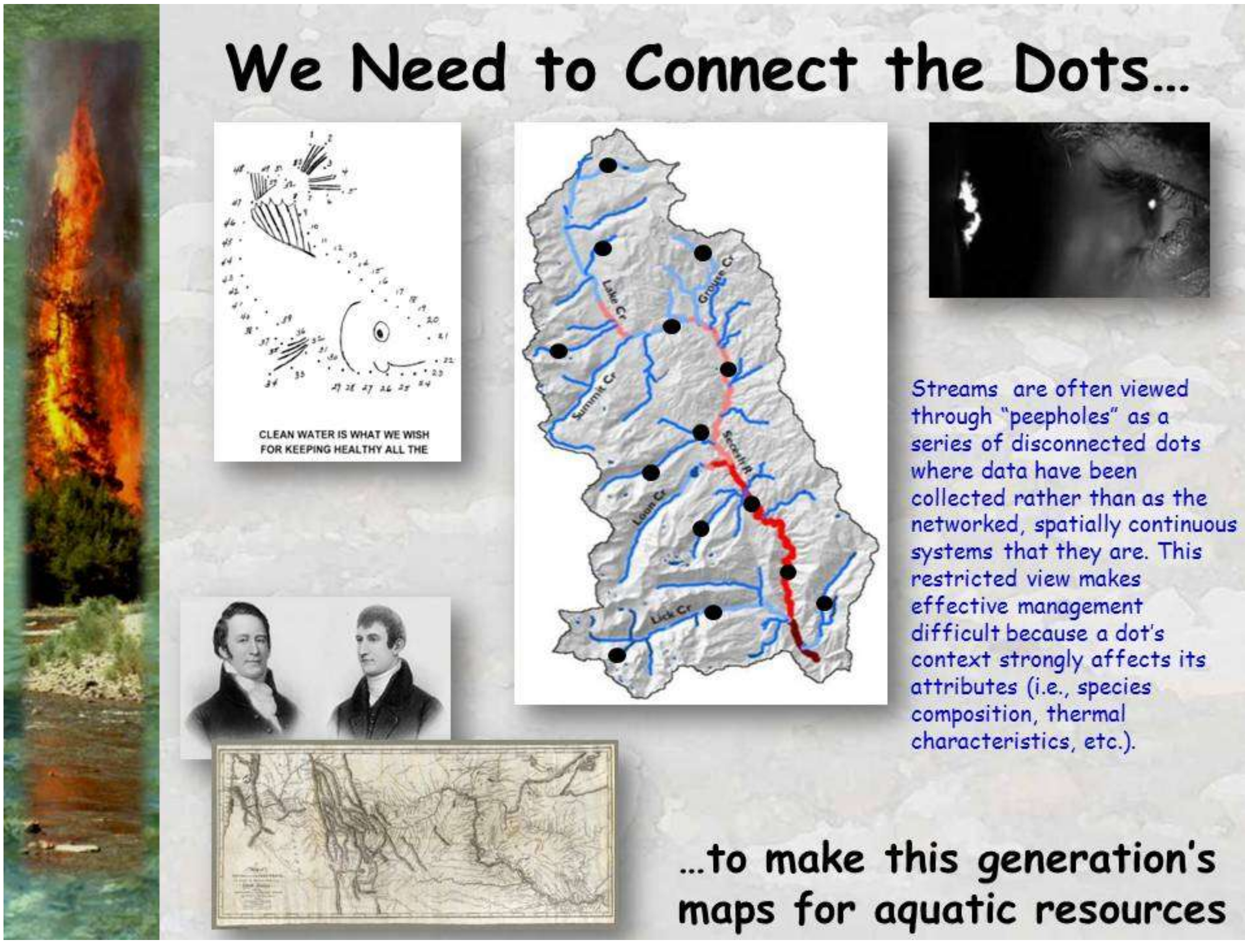
Ver Hoef, J.M., E.E. Peterson, and D.M. Theobald. 2006. Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics* 13:449–464.

Garreta, V, Monestiez P, and Ver Hoef JM. 2010. Spatial modeling and prediction on river networks: up model, down model or hybrid? *Environmentrics* 21:439-456.

Cressie N, Frey J, Harch B, and Smith M. 2006. Spatial Prediction on a River Network. *J Agricultural, Biological, and Environmental Statistics*. 11:127–150.

Website for Freeware Tools & R stats package…

- Coming Soon…"SSN and STARS"



Welcome to the Climate-Aquatics Blog. For those new to the blog, previous posts with embedded graphics can be seen by clicking on the hyperlinks at the bottom or by navigating to the blog archive webpage on our Forest Service site at: (http://www.fs.fed.us/rm/boise/AWAE/projects/stream_temp/stream_temperature_climate_aquatics_blog.html). To discuss these topics with other interested parties, a Google discussion group has also been established and instructions for joining the group are also on the webpage. The intent of the Climate-Aquatics Blog and associated discussion group is to provide a means for the 4,161 (& growing) field biologists, hydrologists, anglers, students, managers, and researchers currently on this mailing list across North America, Europe, and Asia to more broadly and rapidly discuss topical issues associated with aquatic ecosystems and climate change.

Messages periodically posted to the blog will highlight new peer-reviewed research and science tools that may be useful in addressing this global phenomenon. Admittedly, many of the ideas for postings have their roots in studies I and my colleagues have been a part of in the Rocky Mountain region, but attempts will be made to present topics & tools in ways that highlight their broader, global relevance. Moreover, I acknowledge that the studies, tools, and techniques highlighted in future missives are by no means the only, or perhaps even the best, science products in existence on particular topics, so the hope is that this discussion group engages others doing, or interested in, similar work and that healthy debates & information exchanges will occur to facilitate the rapid dissemination of knowledge among those most concerned about climate change and its effects on aquatic ecosystems.

If you know of others interested in climate change and aquatic ecosystems, please forward this message and their names can be added to the mailing list for notification regarding additional science products on this topic. If you do not want to be contacted regarding future such notifications, please reply to that effect and you will be removed from this mailing list.

Previous Posts

Future topics…

Climate-Aquatics Biology Module
Climate-Aquatics Management Module
Climate-Aquatics Cool Stuff Module