Climate-Aquatics Blog #28: Part 2, Spatial Statistical Models for Stream Networks: Applications and Inference

Using the right tools is even better…

Hi Everyone,
Last time out it was argued that Ver Hoef and Peterson's spatial statistical network models are a fundamentally better tool for analyzing many types of stream attributes, particularly when the locations of samples are characterized by non-randomness and spatial clustering as will often be the case with aggregated databases. This time we're highlighting some of spatial model applications to show the sorts of improved information they may provide about the attributes of stream networks. I thought the easiest way to do this is simply by stepping through an example because the map graphics will convey a lot more information more efficiently than I can write about it. Before starting, however, let me re-emphasize the fact that there's an untapped goldmine of data out there to learn from if/when it's organized into functional databases. There are thousands upon thousands of stream sites that have been sampled to determine the occurrence and abundance of species (graphic 1), there are rapidly growing databases of genetic attributes for these species (graphic 2), there are thousands of sites where regulatory agencies monitor water quality attributes (graphic 3), and of course, 10's of thousands of sites with stream temperature measurements (graphic 4, blog #25). Each of those individual samples is ultimately just a local representation of much broader spatial patterns when viewed at the stream or river network scale. The Ver Hoef & Peterson models simply allow us to describe these patterns more accurately, and sometimes in ways that were previously impossible.

So in the example, we'll use a temperature database compiled from several state and federal agencies across a 7,000 km^2 mountain river basin in central Idaho (graphic 5). In this basin, there were almost 800 summers of data available across a stream network of 2,500 kilometers, so autocorrelation & spatial redundancy among some of these measurements was a strong possibility. These data were fit with 2 models; a traditional, non-spatial multiple regression model (graphic 6, upper panel) & the spatial statistical stream regression model (lower panel). The same set of predictor covariates was used in each model, but notice that we get different parameter estimates describing the relationships to stream temperature in each model. That's because the non-spatial model estimates were biased by the autocorrelation in the database. Moreover, this bias has consequences when we use the models to make predictions. Predictions from the non-spatial model deviate systematically from the 1:1 line; in this case under predicting temperatures by a few degrees in warm streams and over predicting in cold streams. That bias is largely eliminated by the spatial models, which also have the advantage of considerably greater predictive power & precision (R^2 improves from 0.68 to 0.93; RMSE decreases from 1.54 °C to 0.74 °C).

As a bit of an aside, I've now been involved in projects to fit the spatial stream models to 3 different temperature databases that were composites from multiple agencies & some interesting patterns are beginning to emerge when making comparisons between spatial and non-spatial regression estimates. If, for example, we look at the parameter estimate for elevation across those 3 datasets (graphic 7), we see a lot of variability in the answers that the non-spatial models provide (-0.0036 to -0.0064 °C/meter) and more consistency from the spatial models (-0.0034 to -0.0045 °C/meter). Thus, a meta-estimate for this parameter averaged across the 3 datasets would have a standard error that is more than 50% smaller using the spatial models than the non-spatial models and an overall mean that is also less biased (graphic 7, bottom panel). It again highlights some of the dangers associated with autocorrelation if it's not properly accounted for. In this case the ***apparent*** variation in the relationship between stream temperature & elevation would have been much greater than the reality & we'd have been misled to some extent by biased model results.

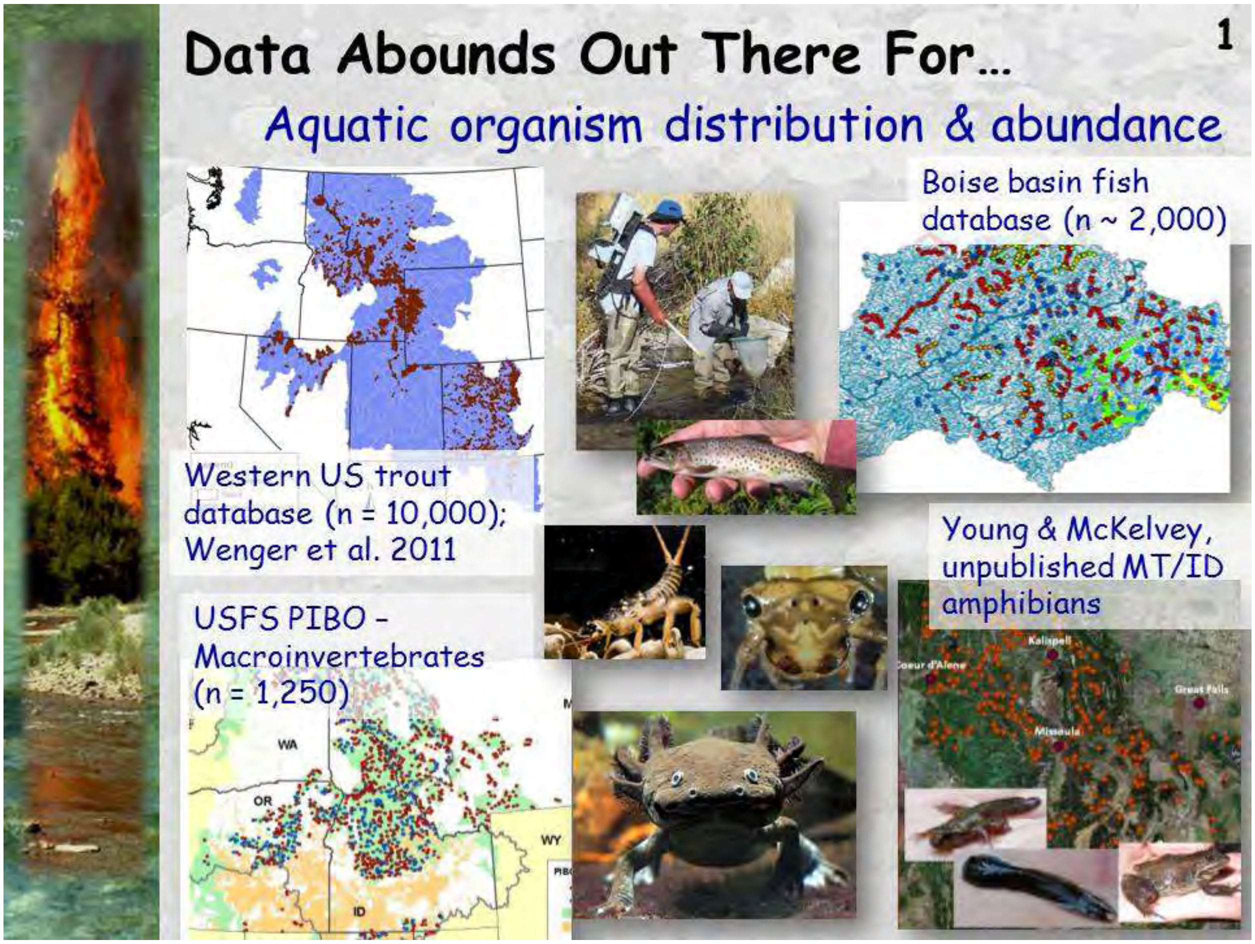
So in the spatial stream models now, we have a flexible analytical structure for accurately describing patterns in many datasets collected on networks & that's a really powerful scientific tool. If this tool is coupled with good ecological theory and insightful, a priori hypotheses, we'll be able to describe new relationships and test or refine many old hypotheses to increase the rigor of our science (graphic 8). That, in turn, will fundamentally improve what we know about streams & should also improve our ability to manage & conserve them. The attached paper by McIntire & Fajardo, "Beyond description: the active and effective way to infer processes from spatial patterns" is a great one for discussing the potential interplay between spatial patterns, hypothesis formulation, and inference regarding underlying processes.

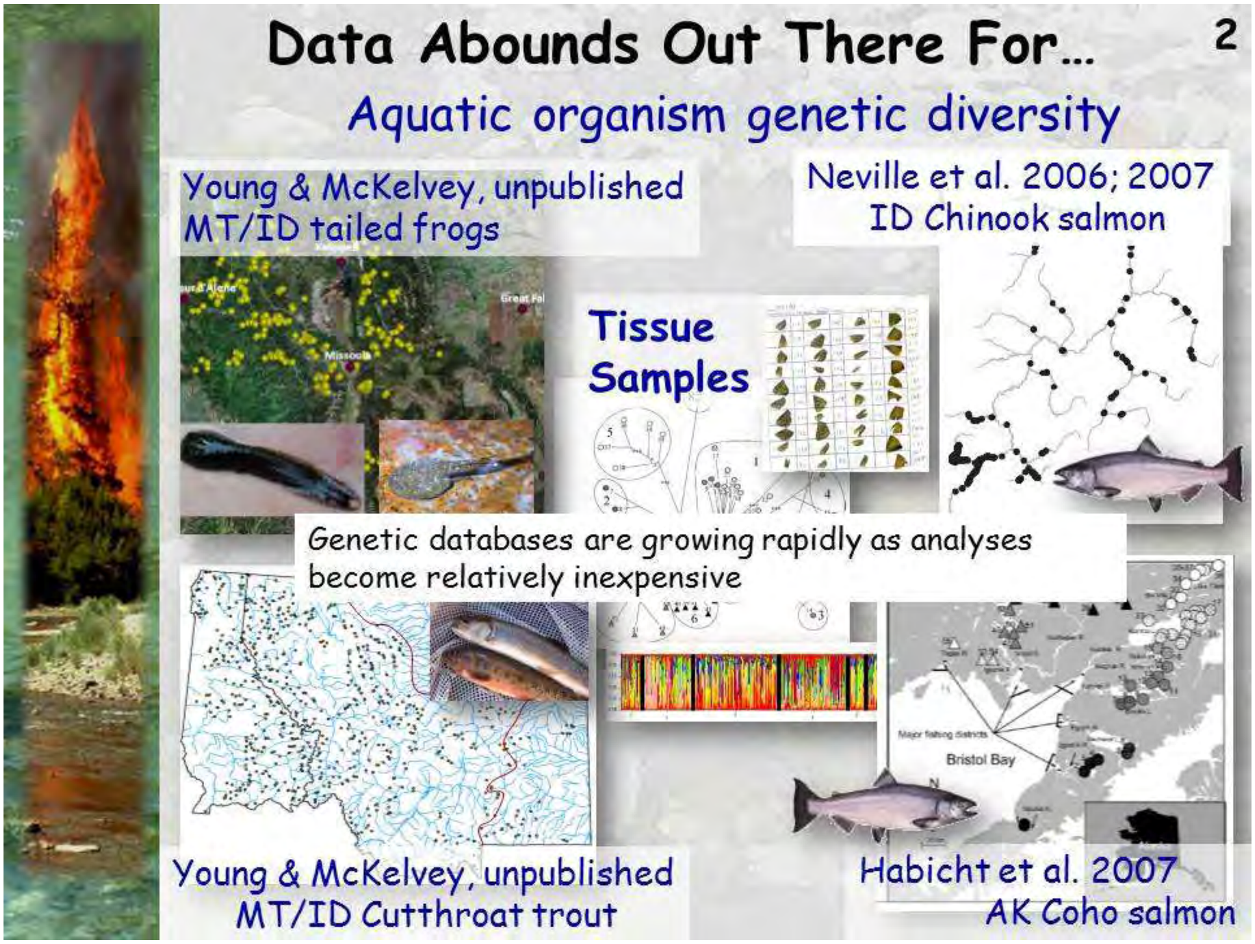
Once we've accounted for the spatial autocorrelation in our temperature dataset & have an accurate model, it can be used for many purposes that include: 1) making predictions at unsampled locations to develop those "smart maps" we need for prioritizing conservation efforts across river networks (graphic 9; blog #26), 2) quantifying the effects of climate change on stream temperatures (graphic 10; blog #7), and 3) translating stream temperature increases to species-specific maps of thermal habitat (graphic 11; blog #7). Those are the standard temperature model applications that may often be useful but the spatial models also provide a suite of new applications that will be interesting to explore in future years. These include: 1) designing efficient temperature monitoring strategies using information regarding autocorrelation distance to ensure that monitoring sites are not redundant (graphic 12); 2) developing spatially explicit maps of uncertainty in temperature predictions that could also aid in monitoring strategies or be used in decision support tools (graphic 13); and 3) block-kriging estimates of stream temperature parameters within subsections of a river network that are of particular interest (graphic 14). And remember, although this example is based on a stream temperature dataset, these same basic analyses & inferences are possible for many of the attributes we commonly sample on streams because the Ver Hoef and Peterson models are generalizable to the standard set of Gaussian, Poissan, and binomial response variable types (graphic 15). For more on additional applications of the spatial stream models, graphic 16 contains a short bibliography.

For all the benefits the spatial models provide, there are no free lunches in life and so here are the downsides. First, there are more parameters to estimate in these models because of the complex stream covariance structure (blog #27), which means we need more data, and a good general rule of thumb regarding a minimum sample size is probably around 100 sites. There also needs to be some spatial clustering among those sites and autocorrelation in the dataset if the spatial models are going to provide performance enhancements relative to non-spatial models. Second, the spatial models are not for the quantitatively faint of heart. They require relatively advanced GIS skills to develop the spatial data that describe stream network topology and the spatial relationships among samples taken on those networks, a working knowledge the R statistical program, and some graduate level training in statistics is always handy for fitting sensible models and interpreting the results. It will often be the case, therefore, that using the spatial models requires small teams of people with complimentary skillsets. Third, fitting the spatial models in the past required special R code and GIS tools that have not been widely available and aren't going to appear any time soon in commercial statistical programs like SAS or SyStat. This hurdle is close to being removed, however, as Erin Peterson and Jay Ver Hoef are putting the finishing touches on a set of freeware GIS tools, an R statistical package, example datasets, and extensive tutorials that will be distributed through a new website (more on that later…).

So in some regards the spatial models may be less convenient than many traditional analyses but there are big payoffs, including the ability to: 1) use data aggregated across multiple agencies without worries about spatial autocorrelation, 2) extract massive amounts of new information, and more accurate information, from existing databases, and 3) map information back to real-world coordinates so that it's format is accessible to those making on-the-ground decisions and choices about where to prioritize conservation efforts. In many ways, the spatial models have the potential to bring people together as we work to manage and conserve aquatic resources this century. And so even as budgets shrink & pressures on natural resources continue to grow, there's a real possibility that not only will we be able to do more with less, but we may be able to do much more.

Until next time, best regards,
Dan


1
Data Abounds Out There For…
Aquatic organism distribution & abundance
Western US trout database (n = 10,000); Wenger et al. 2011
Boise basin fish database (n ~ 2,000)
Young & McKelvey, unpublished MT/ID amphibians
USFS PIBO – Macroinvertebrates (n = 1,250)
WA
OR
ID
WY
Kalispell
Coeur d'Alene
Missoula
Great Falls


Data Abounds Out There For…
2
Aquatic organism genetic diversity
Young & McKelvey, unpublished
MT/ID tailed frogs
Neville et al. 2006; 2007
ID Chinook salmon
Tissue
Samples
Genetic databases are growing rapidly as analyses
become relatively inexpensive
Bristol Bay
Young & McKelvey, unpublished
MT/ID Cutthroat trout
Habicht et al. 2007
AK Coho salmon

Data Abounds Out There For…

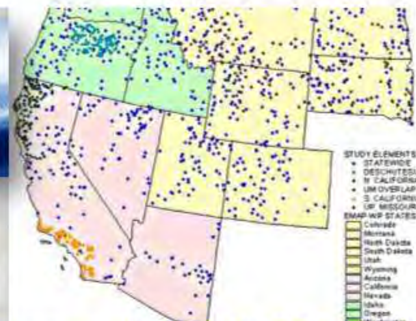
Water Quality/Chemistry Information
(Nitrates, alkalinity, ph, DOC, conductivity, etc.)

Peterson et al. 2006

Gardner & McGlynn 2009

The Clean Water Act and other regulatory standards mean lots of water quality sampling is routinely done.

USGS, unpublished



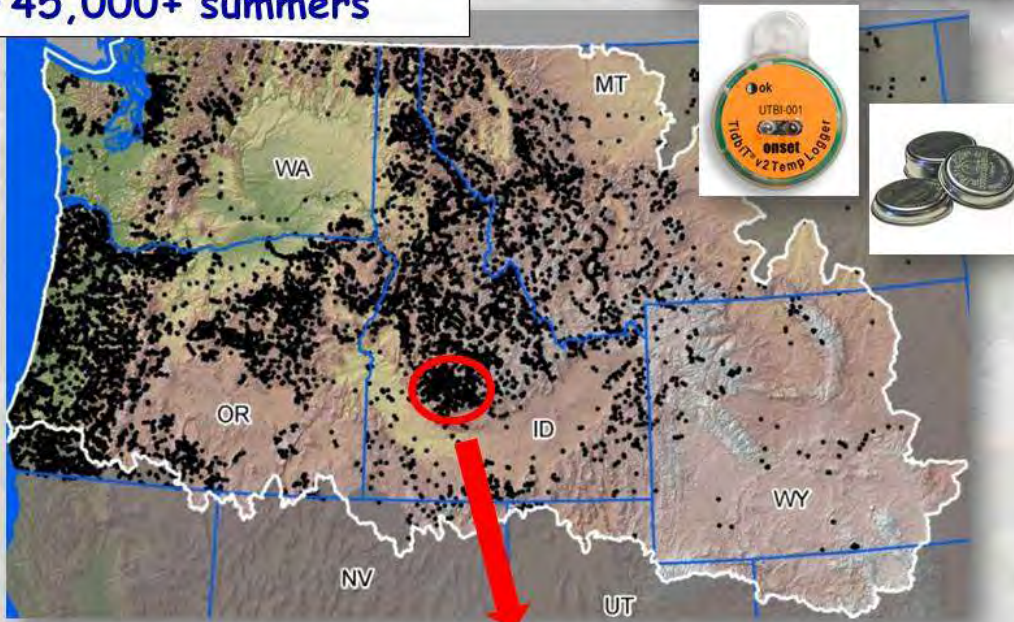
Pont et al. 2009. EPA EMAP

Data Abounds Out There For...

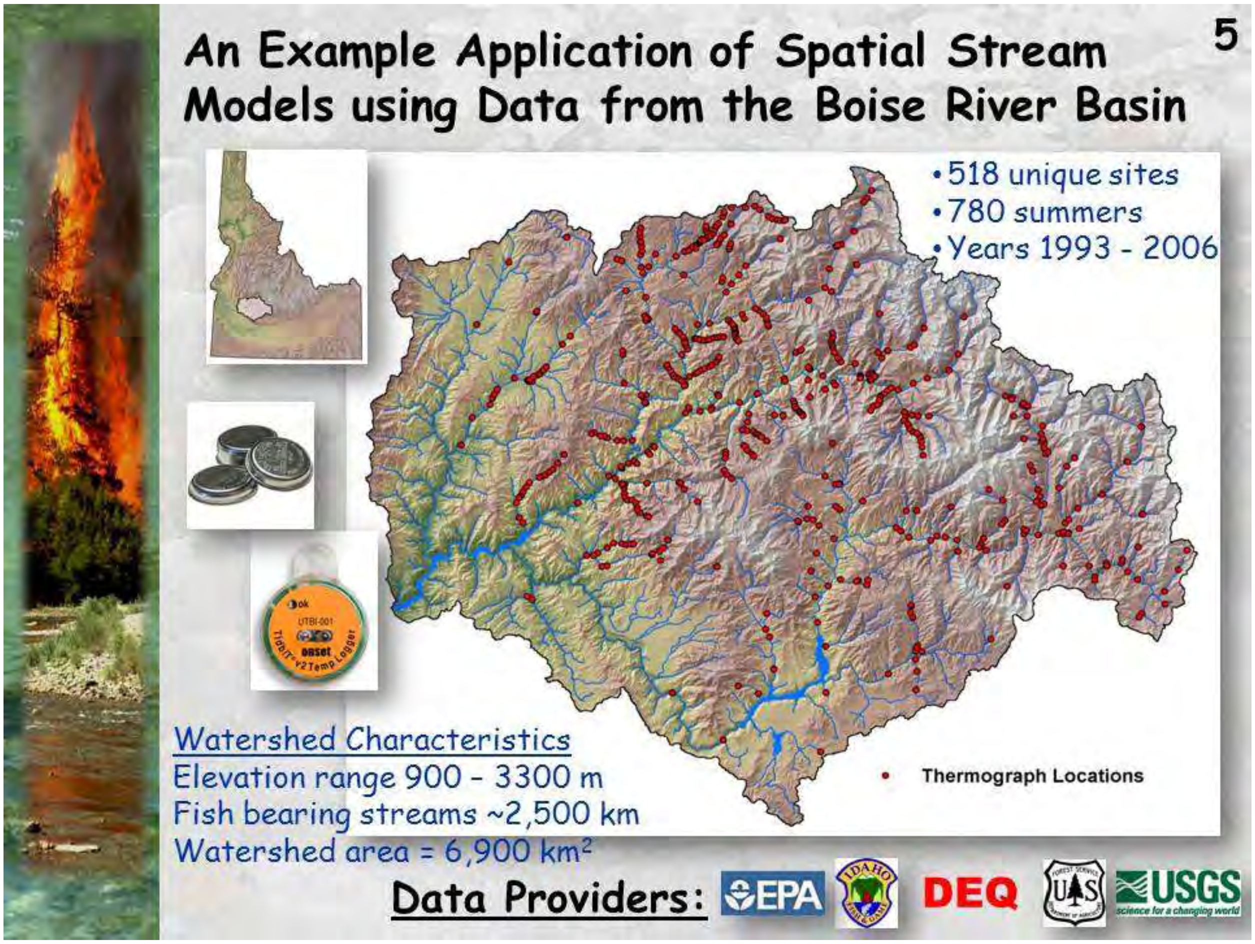
Stream Temperatures

NorWeST Stream Temp

- **15,000+ unique sites**
- **45,000+ summers**



Let's look here in more detail...

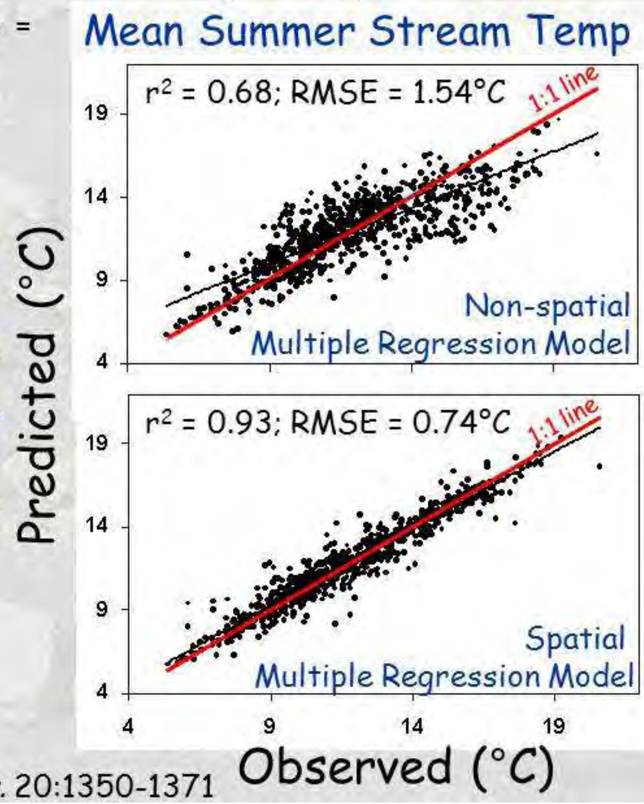

5
An Example Application of Spatial Stream Models using Data from the Boise River Basin
• 518 unique sites
• 780 summers
• Years 1993 - 2006
Thermograph Locations
Watershed Characteristics
Elevation range 900 - 3300 m
Fish bearing streams ~2,500 km
Watershed area = 6,900 km²
Data Providers: EPA DEQ USGS

Comparison of temperature model results from traditional, non-spatial regression model & new spatial statistical stream models fit to the Boise River database (n = 780). **6**

Non-spatial Stream Temp =
- 0.0064*Elevation (m)
+ 0.0104*Radiation
+ 0.39*AirTemp (°C)
- 0.17*Flow (m^3/s)

Notice the changes in parameter estimates. Estimates in the non-spatial model were biased because the autocorrelation among temperature samples was not accounted for. The spatial models also have more predictive power because they model any spatial structure that remains in residuals after accounting for the effects of the predictor variables.

Spatial Stream Temp =
- 0.0045*Elevation (m)
+ 0.0085*Radiation
+ 0.48*AirTemp (°C)
- 0.11*Flow (m^3/s)



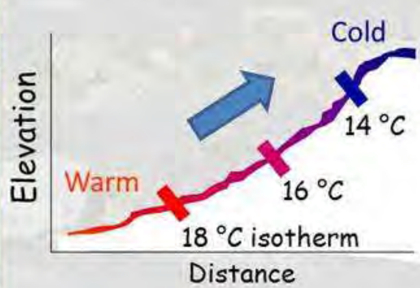
Isaak et al. 2010. *Ecol. Apps.* 20:1350-1371

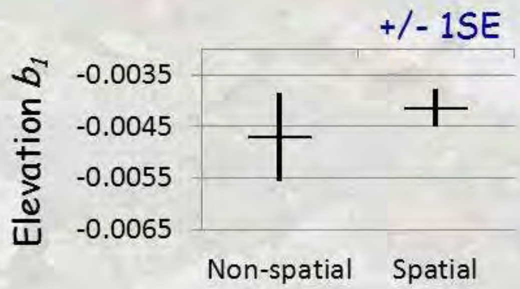
Elevation parameter estimates from 3 different temperature databases fit with non-spatial & spatial regression models. Notice the greater imprecision and bias of the non-spatial model estimates when comparisons are made across areas. Accounting for autocorrelation in these databases reveals a more consistent elevation-stream temperature relationship than non-spatial models would have suggested.

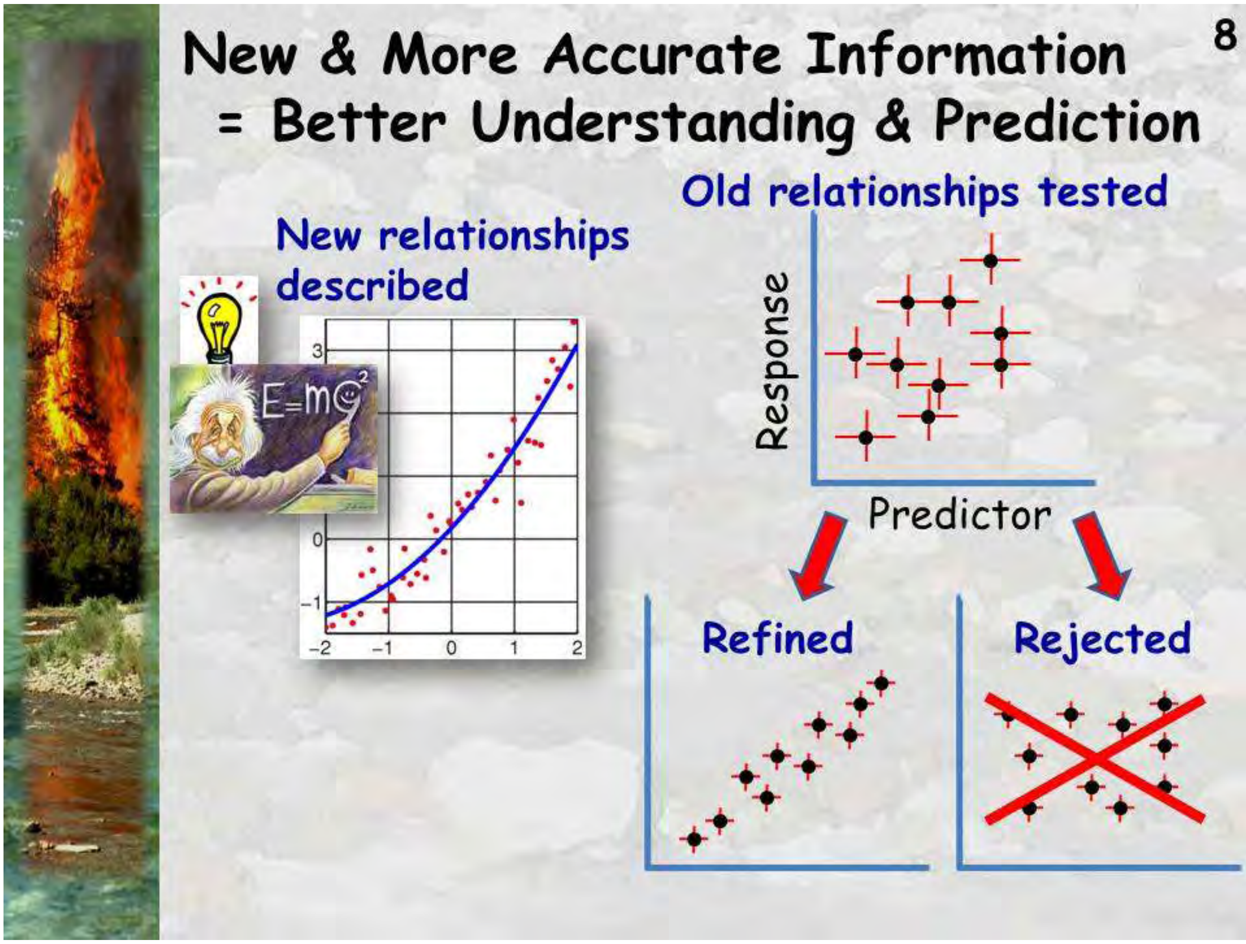
7

Temp Model	Non-spatial	Spatial
Boise basin	-0.0064	-0.0045
Payette NF	-0.0036	-0.0034
Lower Snake	-0.0041	-0.0045

Elevation Parameters (°C / m)



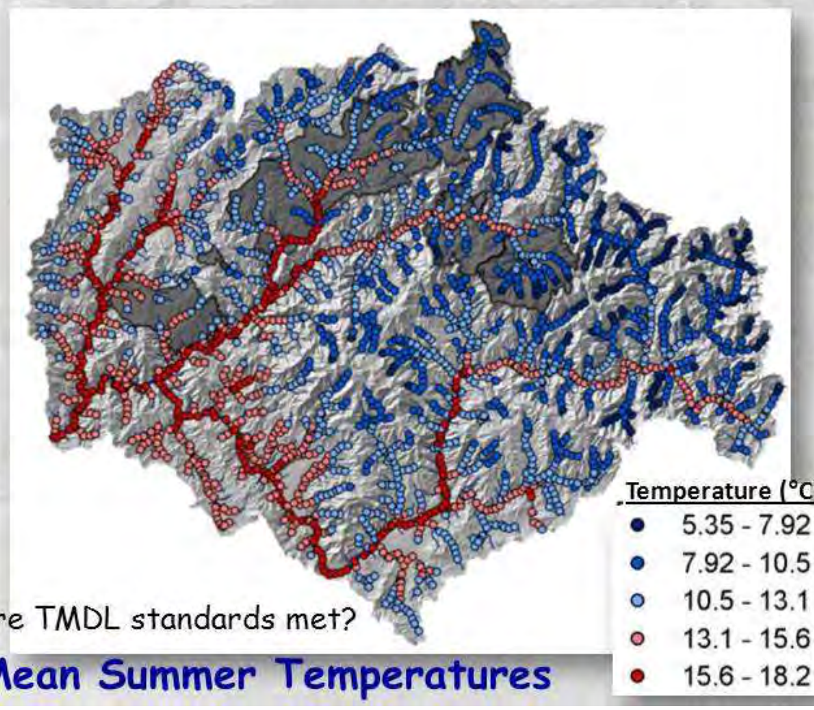




New & More Accurate Information
= Better Understanding & Prediction
8
New relationships described
E=mc²
3
0
-1
-2
-1
0
1
2
Old relationships tested
Response
Predictor
Refined
Rejected

Application: Mapping River Network Temps

Once the spatial model is fit to the temperature database, it can be used to make predictions & interpolate an unbiased, spatially continuous map of temperatures throughout the river network. Maps like this might be useful for regulatory agencies to see when/where water quality standards are met.

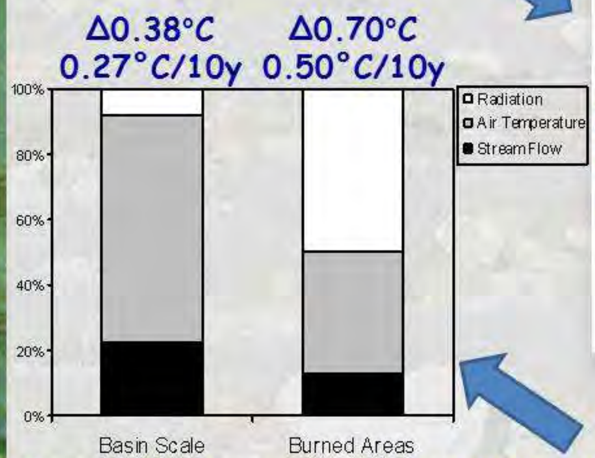


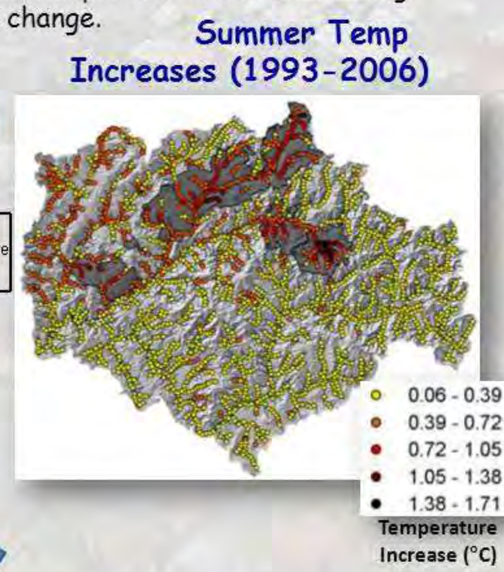
Where are TMDL standards met?

2006 Mean Summer Temperatures

Application: Mapping Climate Change Effects

Differences between temperature maps representing different time periods (past or future) can be used to quantify the effects of climate change on thermal conditions throughout a river network. The map below shows the increase in mean summer temperatures from 1993-2006 based on long-term trends (i.e., 30 - 50 years) of increasing air temperatures and decreasing stream discharge associated with climate change.



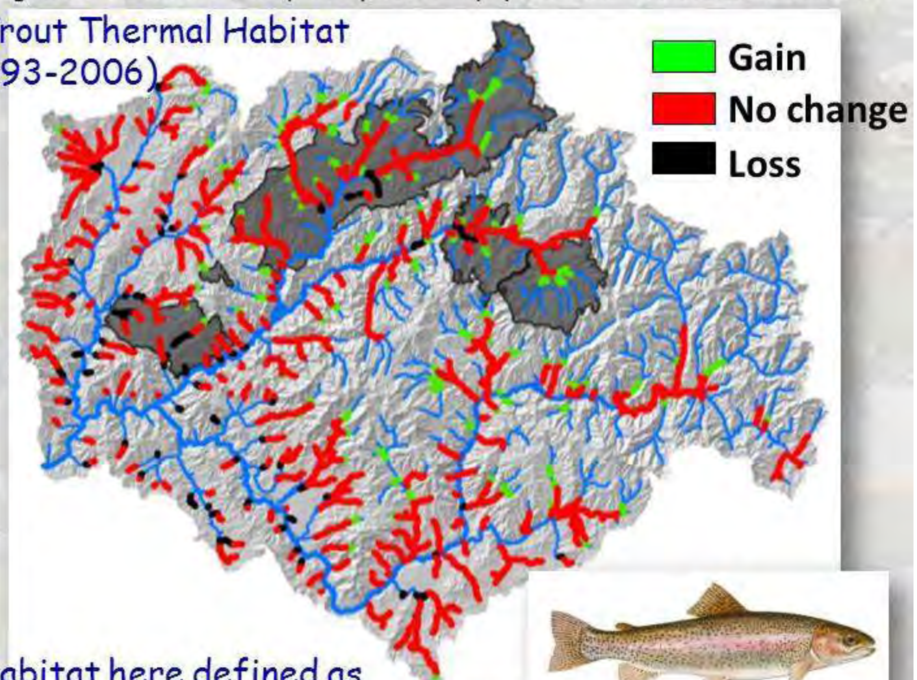


Temperature increases can be summarized over the entire network or subdomains within the network to calculate the amount of warming, the rate of warming, and assess the relative importance of factors associated with warming.

Application: Mapping Thermal Habitat Effects

Once accurate maps of stream temperature climate scenarios are developed, mapping effects on thermal habitat is easily done by applying species-specific thermal criteria. Maps like this might be useful to natural resource agencies for understanding climate vulnerability of species & populations within river basins.

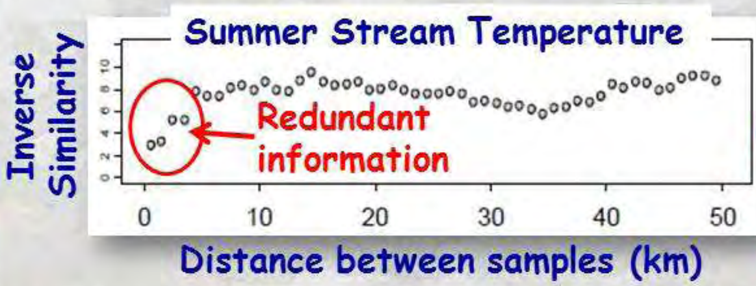
Rainbow Trout Thermal Habitat Shifts (1993-2006)

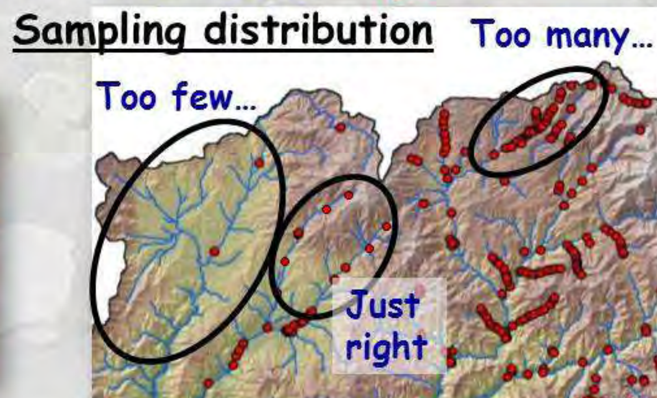


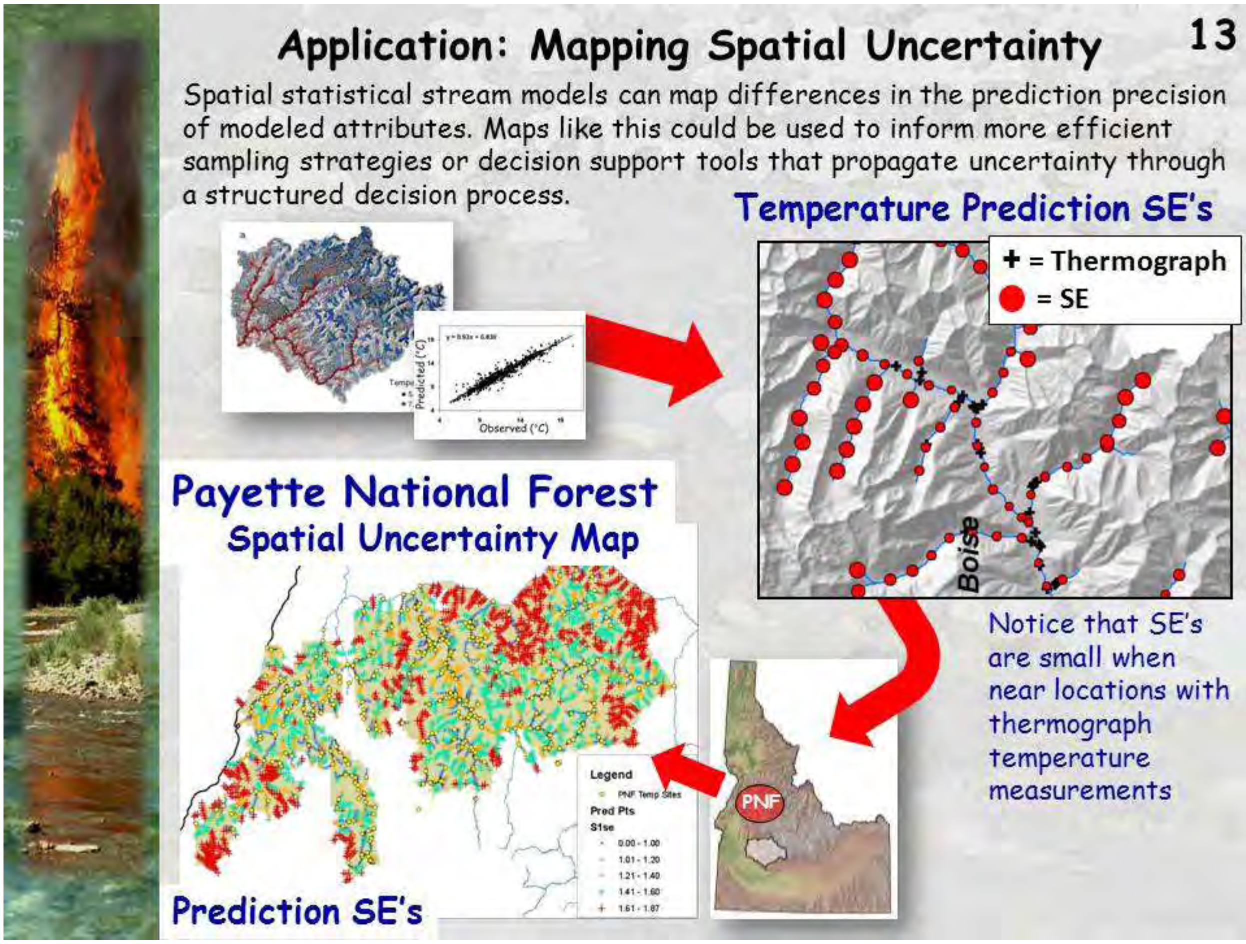
Thermal habitat here defined as streams w/mean temperatures from 11.0-14.0°C

Application: Efficient Monitoring Designs

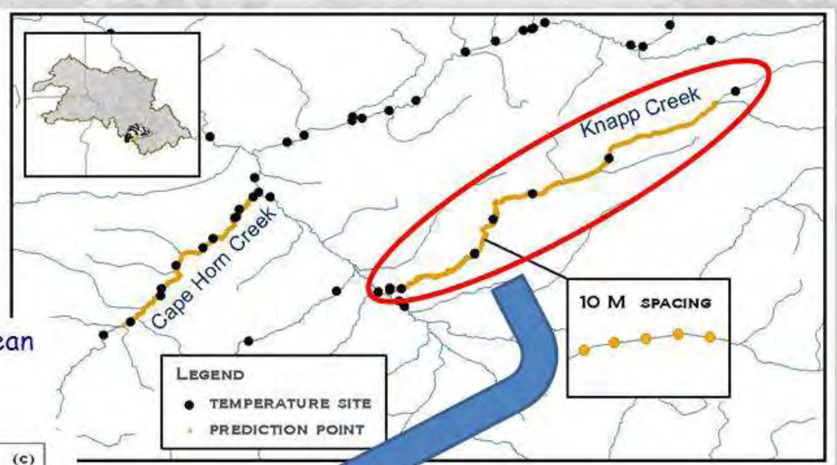
The spatial statistical models also describe the distances along the network over which samples are autocorrelated & provide redundant information. This information could be used to design optimal sampling strategies that maximize information gains relative to costs & will be valuable when designing new data collection efforts.



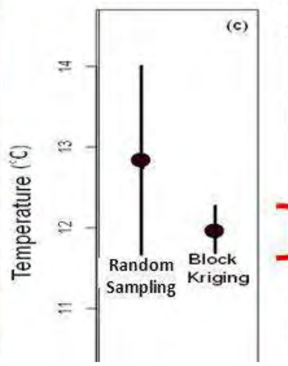



Application: Mapping Spatial Uncertainty
13
Spatial statistical stream models can map differences in the prediction precision of modeled attributes. Maps like this could be used to inform more efficient sampling strategies or decision support tools that propagate uncertainty through a structured decision process.
Temperature Prediction SE's
+ = Thermograph
= SE
Boise
Predicted (°C)
Observed (°C)
Payette National Forest
Spatial Uncertainty Map
Notice that SE's are small when near locations with thermograph temperature measurements
PNF
Legend
PNF Temp Sites
Pred Pts
S1se
0.00 - 1.00
1.01 - 1.20
1.21 - 1.40
1.41 - 1.60
1.61 - 1.87
Prediction SE's

Application: Block Kriging for Accurate Stream or Reach Scale Estimates

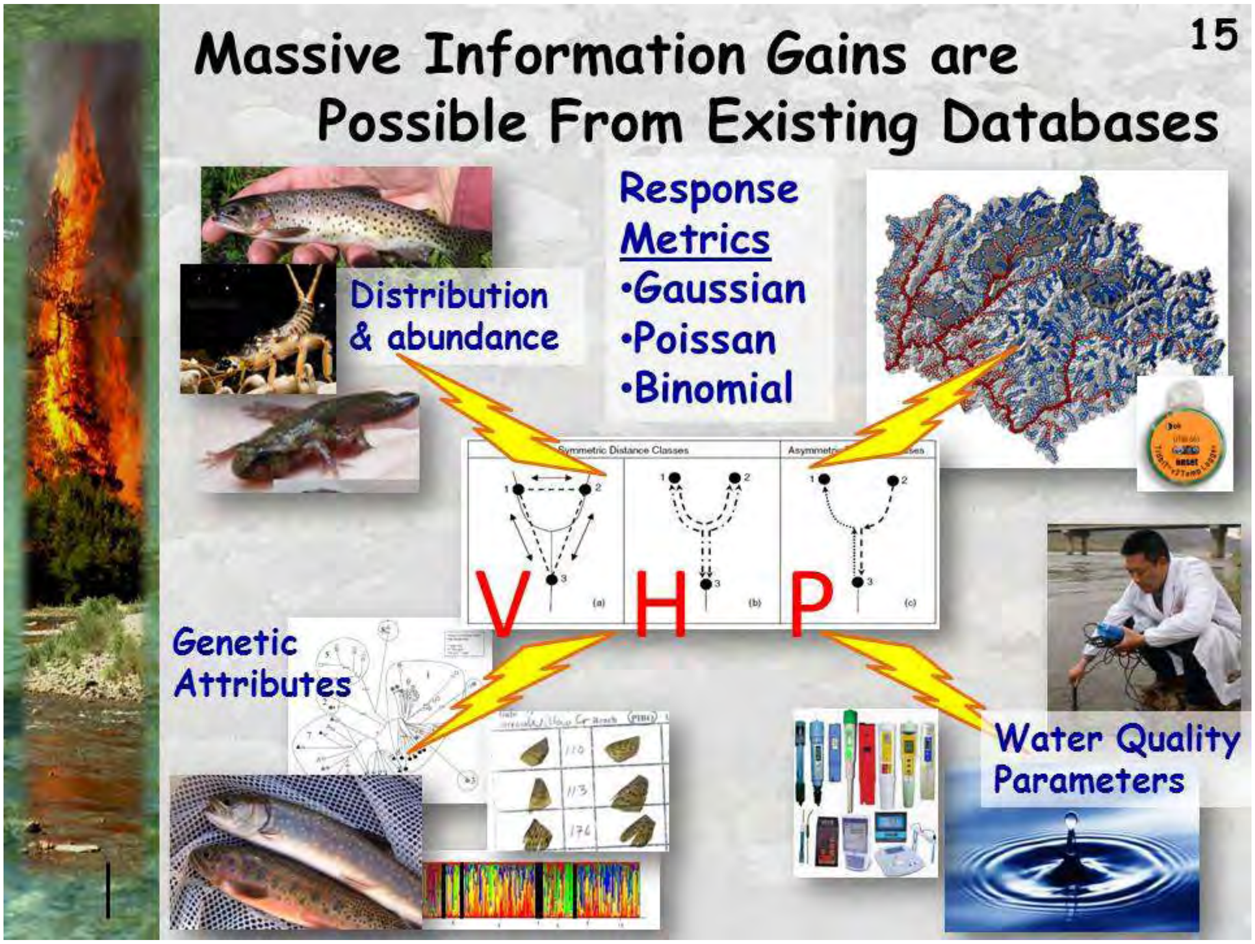


Knapp Creek Mean Temperature Estimates



Estimates of the mean and variance for subsections of a river network can be made that leverage information from a larger set of observations to provide more accurate & precise answers.

Are TMDL standards met in this reach?


15
Massive Information Gains are Possible From Existing Databases
Response Metrics
•Gaussian
•Poissan
•Binomial
Distribution & abundance
Genetic Attributes
Water Quality Parameters
Symmetric Distance Classes
V
H
P
(a)
(b)
(c)

Stream Network Models - Applications

Gardner K, McGlynn B. 2009. Seasonality in spatial variability and influence of land use/land cover and watershed characteristics on stream water nitrate concentrations in a developing watershed in the Rocky Mountain West. *Water Resources Research* 45, DOI: 10.1029/2008WR007029.

Isaak DJ, Luce CH, Rieman BE, Nagel DE, Peterson EE, Horan DL, Parkes S, Chandler GL. 2010. Effects of climate change and recent wildfires on stream temperature and thermal habitat for two salmonids in a mountain river network. *Ecological Applications* 20:1350-1371.

Money E, Carter G, and Serre M. 2009. Modern space/time geostatistics using river distances: data integration of turbidity and *E. coli* measurements to assess fecal contamination along the Raritan River in New Jersey. *Environ. Science and Technology* 43:3736–3742.

Money E, Carter G, and Serre M. 2009. Using river distances in the space/time estimation of dissolved oxygen along two impaired river networks in New Jersey. *Water Research* 43:1948–1958.

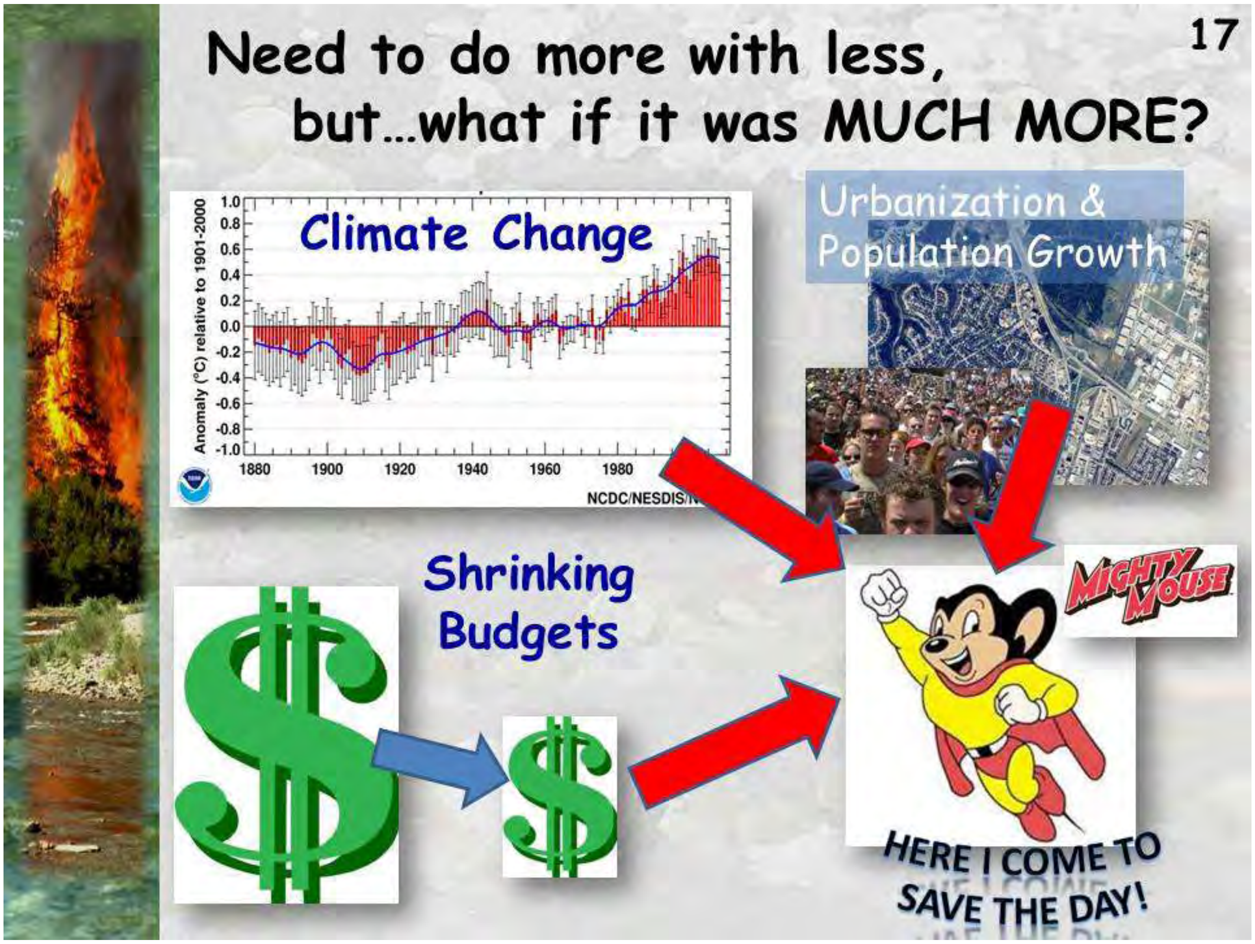
Peterson, EE, Merton AA, Theobald DM, and Urquhart NS. 2006. Patterns of spatial autocorrelation in stream water chemistry. *Environmental Monitoring and Assessment* 121:569–594.

Peterson, EE, and Urquhart NS. 2006. Predicting water quality impaired stream segments using landscape-scale data and a regional geostatistical model: a case study in Maryland. *Environmental Monitoring and Assessment* 121:615–638.

Ruesch AS, Torgersen CE, Lawler JJ, Olden JD, Peterson EE, Volk CJ, and Lawrence DJ. In press. Projected climate-induced habitat loss for salmonids based on a network model of stream temperature. *Conservation Biology*.

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 4,175 (& 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

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

Future topics…
Climate-Aquatics Biology Module
Climate-Aquatics Management Module