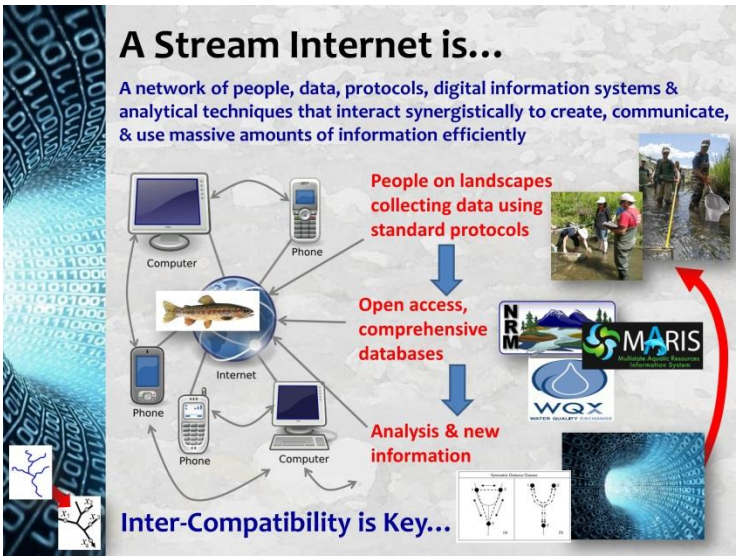
Climate-Aquatics Blog #74: The National Stream Internet Project



"High-quality information" is the currency of the realm in the climate game we're in. The better the information, the easier it is to make effective, well informed, defensible investment decisions that maximize conservation bang for the buck. As described in last week's blog (#73), we are literally sitting on mountains of data that could be mined to extract high-quality information at very low cost compared to what it cost to collect the data originally, or what it would cost to collect new data. But mining data requires shifting one's mentality from "collect more…" to "use what's here…" It also requires hunkering down to learn new technical skills & create the teams that are required to build & tap those databases. It might not be as fun in some ways as mucking around in the field, but nobody can argue with the value and cost effectiveness. Moreover, once we're up-to-snuff on bringing our dark-data legacy datasets into the light to create fully functional databases, we've set ourselves up to be far more efficient in the future anytime we collect new data.

But no one's sanity can withstand years of organizing old data just for the purpose of organizing data. Going through the process is much more endurable (not that it's ever enjoyable) with light at the end of the tunnel in the form of tangible end-products, analyses, and new types of valuable information. To help enable that, we're today launching the National Stream Internet Project (website: http://www.fs.fed.us/rm/boise/AWAE/projects/NationalStreamInternet.html). The NSI provides compatibility among key national databases and statistical software for data on stream networks to provide a nationally consistent analytical infrastructure. The NSI works with any type of data commonly collected for streams (e.g., species distributions/abundance/genetic attributes, habitat surveys, and water quality data) in previous years by many agencies, individuals, and organizations for a variety of purposes. All data are good data in the NSI universe, as long as they are accurately measured and georeferenced onto the stream network. The analytical engine under the NSI hood is powered by the spatial-stream network (SSN) models, which accommodate spatial autocorrelation and redundancy among data locations (blogs #27, #28, and #29) so the lack of a grand sampling design and/or coordination among agencies during data collecting activities doesn't matter. In fact, the more redundancy among data sites the

better because that only increases the predictive power & quality of inference from the SSN models. The sophisticated spatial analyses that can now be done routinely anywhere in the U.S. based on NSI technologies provide the sexy "killer apps" that will unlock a wealth of new information to help incentivize all that slogging through legacy datasets.

As with the real Internet, the stream internet is possible only through the synergy that emerges from the confluence of multiple components. Technical aspects of that synergy are described in the NSI user-guide (hyperlinked here), but the components warrant some additional discussion as provided below.

NSI Component:

1) The National Hydrography Dataset, Version 2 (NHD; graphic 1) provides a consistent digital geodatabase that represents all 5.5 million kilometers of streams & rivers in the contiguous U.S. (Canada & Europe have approximate equivalents). Each reach in that network has a unique spatial identifier and the network itself has many useful characteristics as described in the NHD user-guide (hyperlinked here). Having a nationally consistent data model is a really big & useful deal, so kudos to those on the NHD development team that EPA & USGS have run jointly for much of the past 20 years. There are two basic flavors to the NHD network, the medium-resolution version (1:100,000-scale) and the high-resolution version (1:24,000-scale). For now, the NSI is designed to work with the medium-resolution network because that's the one that has 100s of reach & watershed descriptors already calculated, and those serve invaluable roles as predictor variables in models or to enable spatial queries of the stream network. The NHD team is working towards someday having a comparable set of reach descriptors for the high-resolution version of the network, but that dataset is still several years in the future from broad availability.

2) The "Plus" part of what is referred to as NHD-Plus are those reach & watershed descriptors alluded to above. In last week's BIG aquatic DATA blog (#73), we highlighted two of the larger reach descriptor databases (i.e., EPA's new StreamCat & a similar precursor developed by the National Fish Habitat Partnership; graphic 2), but there are others as summarized on this subpage of the NSI website (hyperlinked here). For many common types of stream analyses, therefore, there is no need to reinvent the wheel & create custom sets of descriptors/predictors. It's more a matter of understanding the full array of what's now available, then logically selecting the best subset for project purposes (& from personal experience on many projects I've participated in the last 5 years, I can attest that this is a huge time saver!).

3) Mountains of existing water quality or biological survey data (graphic 3). To be usable within the NSI context, those data locations simply have to be referenced to the specific NHD reaches where data were collected. That done, any of the 100s of existing reach/watershed descriptors can be linked to the data to create a dataset for SSN analysis (or analysis by many statistical procedures for that matter). Even if a new analysis isn't the goal, it's still often useful to organize existing data & link them to reach descriptors so that summaries can be made for reporting purposes or to describe the characteristics of locations with & without data. It then becomes possible to design sampling strategies that efficiently supplement

existing datasets as described by Som and colleagues in detail here or as postulated previously (blogs #8 and #39).

4) Statistical theory & software for running the SSN models (graphic 4). This software is distributed for free through the SSN/STARS website (hyperlinked here) and was described extensively in previous blogs (#27, #28, and #29). Among the many reasons for liking SSN models, they've been shown to outperform traditional statistical techniques applied to stream data for many applications, work particularly well with data aggregated from multiple sources that contain clustered sampling locations, and enable predictions at unsampled locations to create accurate status maps for full river networks.

5) NSI hydrography network (graphic 5). SSN models can be used with any digital hydrography network but these require reconditioning to create specific topological relationships prior to conducting a SSN analysis. That process may be time consuming and requires considerable GIS skills, so to expedite SSN analysis, the medium resolution NHD network has been reconditioned for the entirety of the contiguous U.S. to create the NSI network. ArcGIS shapefiles of the NSI network are available for download from this subpage of the NSI website (hyperlinked here); as are shapefiles of prediction points that can be used to make predictions at unsampled locations and map the results of SSN analyses. The NSI network is fully back-compatible with the original NHD-Plus layer so users can tap the large databases of reach/watershed descriptors for use as predictor variables in spatial analyses.

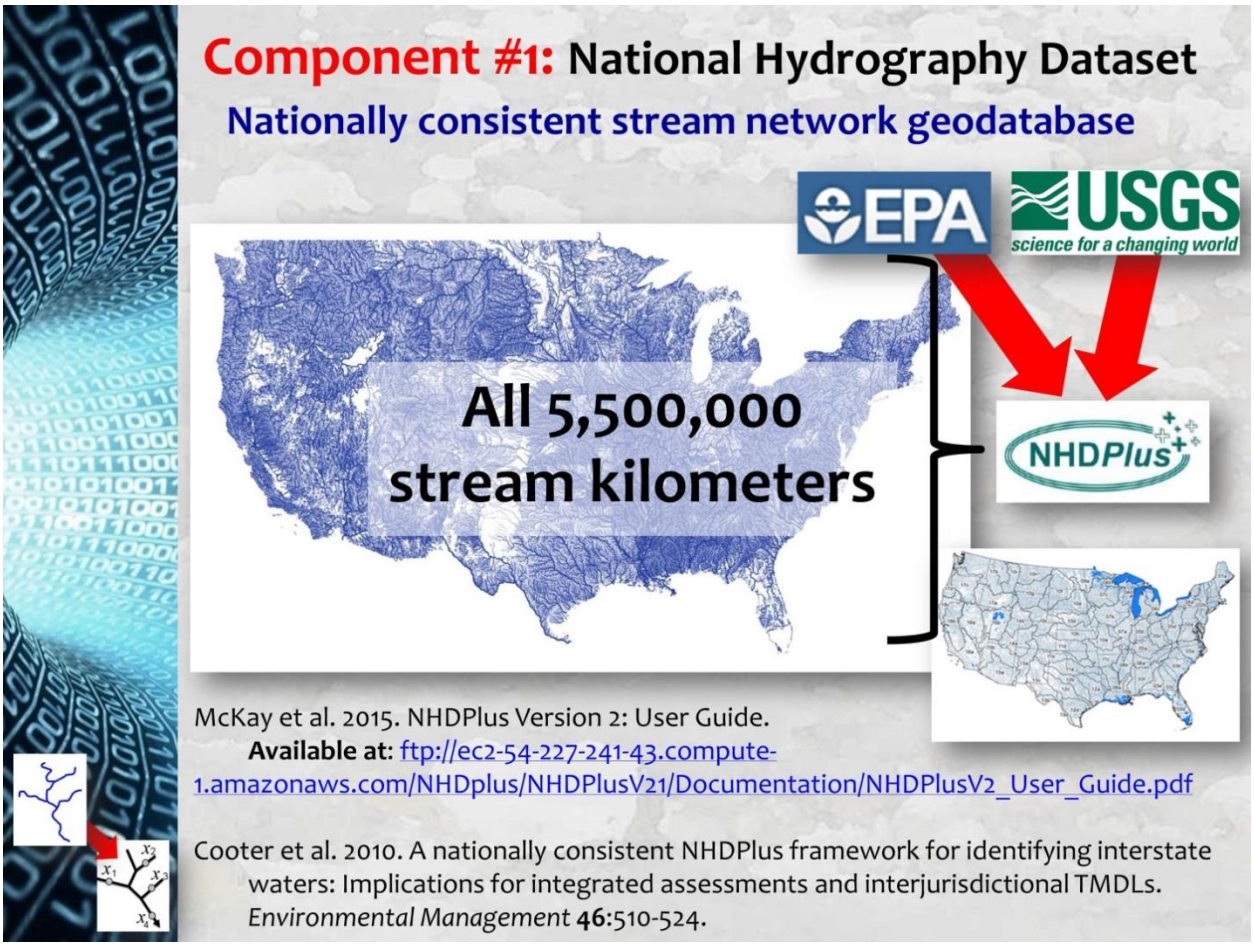
6) Killer apps (graphic 6). These are the many new & powerful types of analyses & inference that NSI technologies now make possible for any type of stream data anywhere in the country. Parameter estimation & spatially-explicit predictions, power analyses, simulations of data on networks, block-kriging, data mining—you name it, you can now do it better than was previously possible for stream data. Isaak & colleagues (hyperlinked here) provide an overview of the many possibilities using example datasets and questions common to aquatic research and management.

7) YOU. The real Internet wouldn't exist without a set of standard protocols, a community of users, content creation, transmission, communication, and evolution. So it is with the NSI, which will remain only a bunch of digital bits & unrealized possibilities unless people use it. So whether you're the one collecting data, building databases, analyzing data, or making decisions, you have to buy into the system and use it to make it work. If you're collecting the data, that means doing so with a standard protocol, georeferencing the collection sites, and archiving everything properly in a database so that it's accessible & usable. If you're the one building databases, that means developing a consistent, well documented, efficient & transparent system so that people know how to feed you data & that they develop the expectation of doing so whenever they collect it. If you're the one analyzing data, it means investing time to continually learn & apply new types of analyses, talking/listening to many people inside/outside of your group to continually think about/refine the types of analyses you do to create useful information for decision makers. And if you're the one making decisions, you ultimately set the tone for how your team or agency values data, collects & manages data, and creates information. It all starts at the top.

Time will tell whether the potential that now exists is realized, but that's why we call it the NSI "Project." It has an indefinite future completion date that depends ultimately on whether, and how widely, this set of resources is adopted and used by the aquatics community. Initial signs are promising, as there are already a large number of NHD applications across the country (hyperlinked here), a diversity of SSN applications (hyperlinked here), and the uptake and use of these resources seems to be accelerating. But the NSI concept, like the website itself, is really just a nexus or hub that helps us link together complementary resources and people so that information creation can be accelerated when we most need it to be. If it provides value, then the NSI won't ever be finished because it will simply begin to co-evolve with the user community and proliferate into a diversity of new analyses, bigger & better databases, more efficient monitoring schemes, and useful information types. And that will benefit our understanding & appreciation of aquatic ecosystems, and our ability to conserve and manage them in the face of not only climate change, but all the challenges they face.

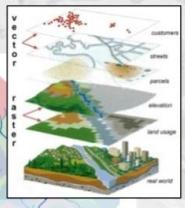
Until one last time next time, best regards. Dan

Tweeting at Dan Isaak@DanIsaak



Component #2: The "PLUS" part of NHDPlus (Stream Reach Descriptors)

Unique reach IDs

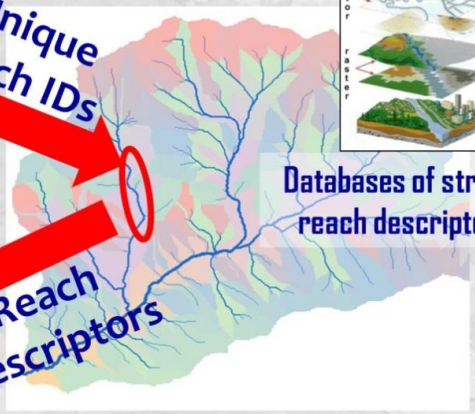


Databases of stream reach descriptors

- Elevation
- Slope
- %Landuse
- Precipitation

100's more…

Reach descriptors

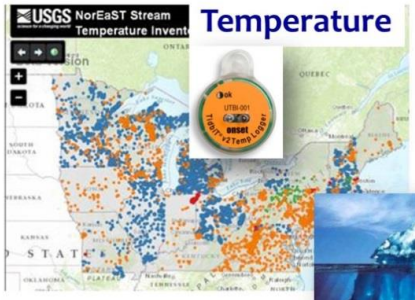


Wang et al. 2011. A hierarchical spatial framework and database for the national river fish habitat condition assessment. *Fisheries* **36**: 436-449.
Available at: https://www.researchgate.net/profile/Lizhu_Wang2

Hill et al. In Press. The stream-catchment (StreamCat) dataset: A database of watershed metrics for the conterminous USA. *The Journal of the American Water Resources Association*.
Available at: http://www2.epa.gov/national-aquatic-resource-surveys/streamcat

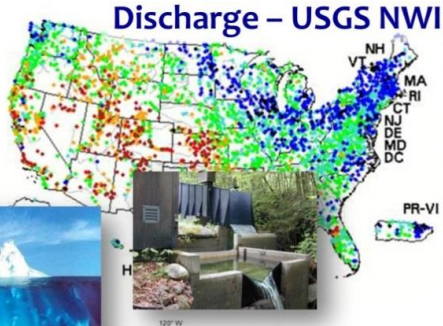
Component #3: Mountains of Data Exist for Information Creation

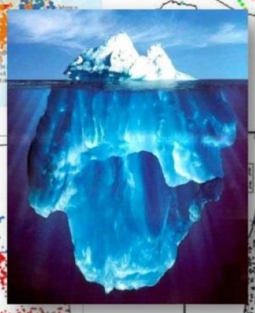
Databases of stream measurements



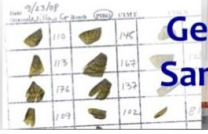
Temperature

Discharge – USGS NWIS

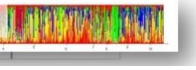


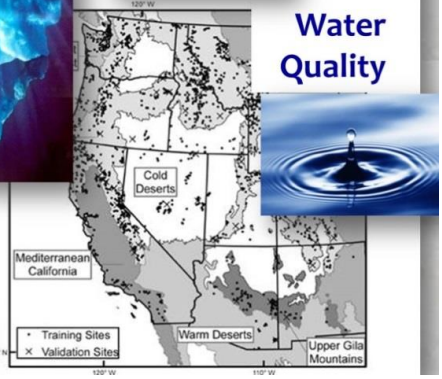


Species distributions



Genetic Samples





Water Quality

Component #4: Spatial-statistical network model theory & software

Ver Hoef et al. 2006. Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics* **13**:449–464.

Ver Hoef & Peterson. 2010. A moving average approach to spatial statistical models of stream networks. *The Journal of the American Statistical Association* **489**: 6-18.

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

Peterson et al. 2013. Modelling dendritic ecological networks in space: An integrated network perspective. *Ecology Letters* **16**:707-719.

Peterson & Ver Hoef. 2014. STARS: An ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. *Journal of Statistical Software* **56(2)**:1-17.

Ver Hoef et al. 2014. SSN: An R package for spatial statistical modeling on stream networks. *Journal of Statistical Software* **56(3)**:1-45.

Component #5: NSI Network for SSN analysis is Derived from the NHD Network

SSN analysis requires specific network topology

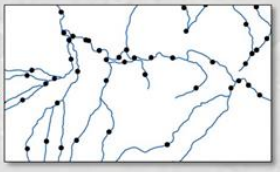
NSI Network Created by Addressing…

- Uninitialized flow
- Braids and diverging flow
- Converging flow
- Complex confluences
- Outlets and sinks

Back compatibility with NHD-Plus is preserved

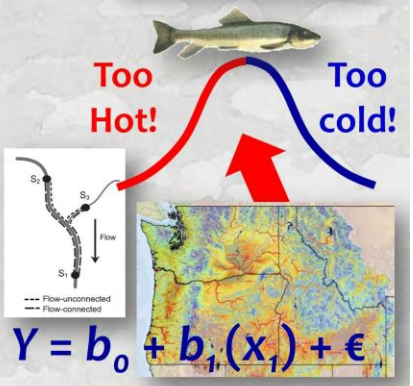
Bonus!

- Prediction points created for NSI network



Component #6: SSN Model "Killer Apps"

- Parameter estimation & prediction
- Status & trend assessments
- Efficient monitoring designs
- Block-kriging for reference site comparisons & fish population estimates
- Mining of BIG DATA databases
  - Climate scenarios
  - Temperature criteria
  - Species distribution models

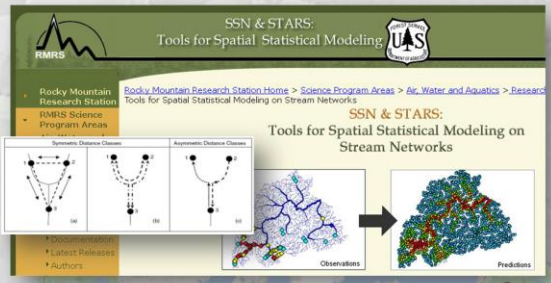


$$Y = b_0 + b_1(x_1) + \epsilon$$

Isaak et al. 2014. Applications of spatial statistical network models to stream data. *Wiley Interdisciplinary Reviews - Water* 1:277-294.

Component #7: Growing User Community

SSN/STARS Website



>28,500 website visits in first 3.5 years

Free, high-quality software

>1,000 software downloads

Locations of visits to SSN/STARS website in last month

The National Stream Internet Project (Website)

LANDSCAPE CONSERVATION COOPERATIVES

NOAA

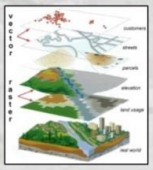
ACEMS

USGS

NSI Resources

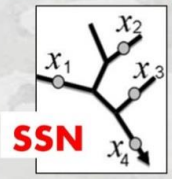
Workshop & presentations

NSI hydrography network



Databases of stream reach descriptors

Databases of stream measurements



Spatial stream-network models

Ideas → Data → Analysis → Information

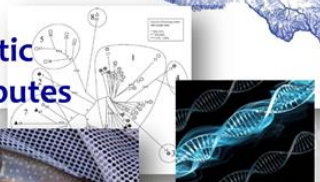
Develop a Database & Do It!

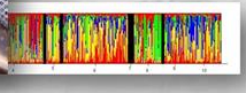
Distribution & abundance

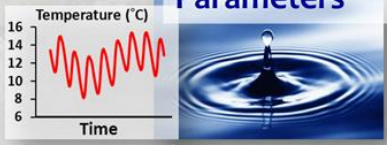
Habitat condition

Water Quality Parameters

Genetic Attributes

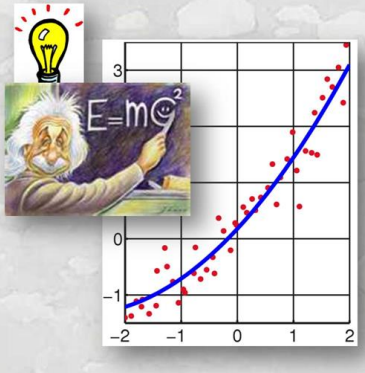




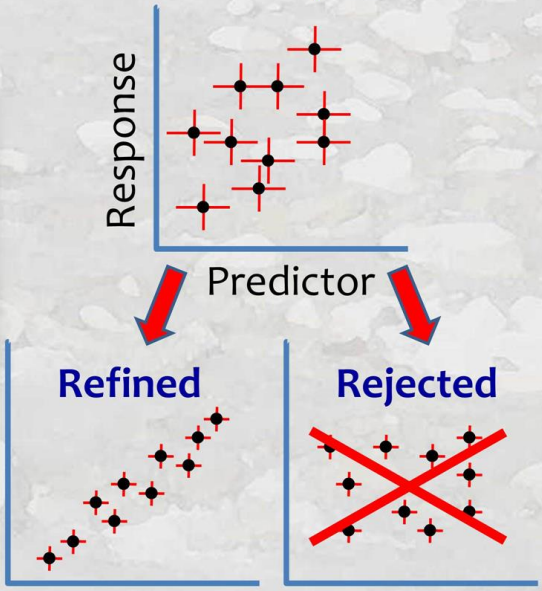


A New Era of Better Prediction & Understanding for Stream Things…

New relationships described

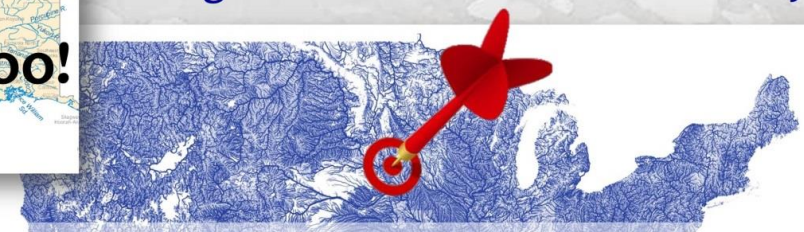


Old relationships tested

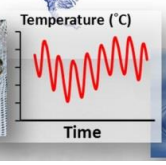


Vision: TSI (Total Stream Information) through the NSI

High-resolution information for all stream things in all reaches across the country



5,500,000 stream kilometers



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 here: (http://www.fs.fed.us/rm/boise/AWAE/projects/stream_temp/stream_temperature_climate_aquatics_blog.html). The intent of the Climate-Aquatics Blog is to provide a means for the ~9,000 field biologists, hydrologists, anglers, students, managers, and researchers currently on this mailing list across North America, South 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 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 my colleagues & I have been conducting in the Rocky Mountain region, but attempts will be made to present topics & tools in ways that highlight their broader, global relevance. I acknowledge that the studies, tools, and techniques highlighted in these 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 occur to facilitate the rapid dissemination of knowledge among those concerned about climate change and its effects on aquatic ecosystems.

If you know others interested in climate change and aquatic ecosystems, please forward this message to them. If you do not want to be contacted again in the future, please reply to that effect and you will be de-blogged.

Previous Blogs…

Climate-Aquatics Overviews

Blog #1: Climate-aquatics workshop science presentations available online
Blog #2: A new climate-aquatics synthesis report
Blog #52: Review & key knowable unknowns

Climate-Aquatics Thermal Module

Blog #3: Underwater epoxy technique for full-year stream temperature monitoring
Blog #4: A GoogleMap tool for interagency coordination of regional stream temperature monitoring
Blog #5: Massive air & stream sensor networks for ecologically relevant climate downscaling
Blog #6: Thoughts on monitoring air temperatures in complex, forested terrain
Blog #7: Downscaling climate change effects on river network temperatures using SSN models
Blog #8: Thoughts on monitoring designs for temperature sensor networks across river and stream basins
Blog #9: Assessing climate sensitivity of aquatic habitats by direct measurement of stream & air temperatures
Blog #10: Long-term monitoring shows climate change effects on river & stream temperatures
Blog #11: Long-term monitoring shows climate change effects on lake temperatures
Blog #12: Climate trends & climate cycles & weather weirdness
Blog #13: Tools for visualizing local historical climate trends
Blog #14: Leveraging short-term stream temperature records to describe long-term trends
Blog #15: Wildfire & riparian vegetation change as the wildcards in climate warming of streams
Blog #23: New studies describe historic & future rates of warming in northwestern US streams
Blog #24: NoRRTN: An inexpensive regional river temperature monitoring network
Blog #25: NorWeST: A massive regional stream temperature database
Blog #26: Mapping thermal heterogeneity & climate in riverine environments
Blog #40: Crowd-sourcing a BIG DATA regional stream temperature model
Blog #60: New report describes data collection protocols for continuous monitoring of temperature & flow in wadeable streams
Blog #61: Significant new non-American stream temperature climate change studies

Blog #62: More bits about the how, what, when, & where of aquatic thermalscapes
Blog #64: Building real-time river network temperature forecasting systems

Climate-Aquatics Hydrology Module
Blog #16: Shrinking snowpacks across the western US associated with climate change
Blog #17: Advances in stream flow runoff and changing flood risks across the western US
Blog #18: Climate change & observed trends toward lower summer flows in the northwest US
Blog #19: Groundwater mediation of stream flow responses to climate change
Blog #20: GIS tools for mapping flow responses of western U.S. streams to climate change
Blog #21: More discharge data to address more hydroclimate questions
Blog #22: Climate change effects on sediment delivery to stream channels

Climate-Aquatics Cool Stuff Module
Blog #27: Part 1, Spatial statistical models for stream networks: context & conceptual foundations
Blog #28: Part 2, Spatial statistical models for stream networks: applications and inference
Blog #29: Part 3, Spatial statistical models for stream networks: freeware tools for model implementation
Blog #30: Recording and mapping Earth's stream biodiversity from genetic samples of critters
Blog #53: DNA barcoding & fish biodiversity mapping
Blog #71: Harnessing social & digital network technologies to maximize climate effectiveness
Blog #X: Blogging fish science
Blog #72: The eDNA revolution and developing comprehensive aquatic biodiversity archives
Blog #73: BIG DATA as an engine for aquatic information creation
Blog #74: The national stream internet project

Climate-Aquatics Biology Module
Blog #31: Global trends in species shifts caused by climate change
Blog #32: Empirical evidence of fish phenology shifts related to climate change
Blog #33: Part 1, Fish distribution shifts from climate change: Predicted patterns
Blog #34: Part 2, Fish distribution shifts from climate change: Empirical evidence for range contractions
Blog #35: Part 3, Fish distribution shifts from climate change: Empirical evidence for range expansions
Blog #36: The "velocity" of climate change in rivers & streams
Blog #37: Part 1, Monitoring to detect climate effects on fish distributions: Sampling design and length of time
Blog #38: Part 2, Monitoring to detect climate effects on fish distributions: Resurveys of historical stream transects
Blog #39: Part 3, Monitoring to detect climate effects on fish distributions: BIG DATA regional resurveys
Blog #41: Part 1, Mechanisms of change in fish populations: Patterns in common trend monitoring data
Blog #42: BREAKING ALERT! New study confirms broad-scale fish distribution shifts associated with climate change
Blog #56: New studies provide additional evidence for climate-induced fish distribution shifts
Blog #43: Part 2, Mechanisms of change in fish populations: Floods and streambed scour during incubation & emergence
Blog #44: Part 3, Mechanisms of change in fish populations: Lower summer flows & drought effects on growth & survival
Blog #45: Part 4, Mechanisms of change in fish populations: Temperature effects on growth & survival
Blog #46: Part 5, Mechanisms of change in fish populations: Exceedance of thermal thresholds
Blog #47: Part 6, Mechanisms of change in fish populations: Interacting effects of flow and temperature
Blog #48: Part 7, Mechanisms of change in fish populations: Changing food resources
Blog #49: Part 8, Mechanisms of change in fish populations: Non-native species invasions
Blog #50: Part 9, Mechanisms of change in fish populations: Evolutionary responses

Blog #51: Part 10, Mechanisms of change in fish populations: Extinction
Blog #63: Navigating stream thermalscapes to thrive or merely survive
Blog #65: The fish jumble as they stumble along with the shifting thermalscape

Climate-Aquatics Management Module
Blog #54: Part 1, Managing with climate change: Goal setting & decision support tools for climate-smart prioritization
Blog #55: Part 2, Managing with climate change: Streams in channels & fish in streams
Blog #58: Part 3, Managing with climate change: Maintaining & improving riparian vegetation & stream shade
Blog #59: Part 4, Managing with climate change: Keeping water on the landscape for fish (beaverin' up the bottoms)
Blog #66: Part 5, Managing with climate change: Barrier placements to facilitate fish flows across landscapes
Blog #67: Part 6, Managing with climate change: Assisted migration to facilitate fish flows across landscapes
Blog #68: Part 7, Managing with climate change: Identifying & protecting climate refugia as a strategy for long-term species conservation
Blog #57: Managing with climate change: Identifying & protecting climate refuge lakes for coldwater fishes
Blog #69: Part 8, Managing with climate change: Building climate-smart conservation networks (metapopulations + biodiversity + refugia)
Blog #70: Part 9, Managing with climate change: Restoration success stories that improve population resilience to climate change