Spatial Statistical Network (SSN) Models for Data on Stream Networks: Background, Theory, and Applications

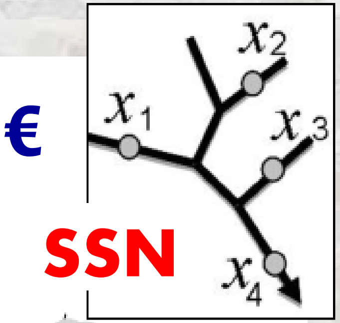
Dan Isaak

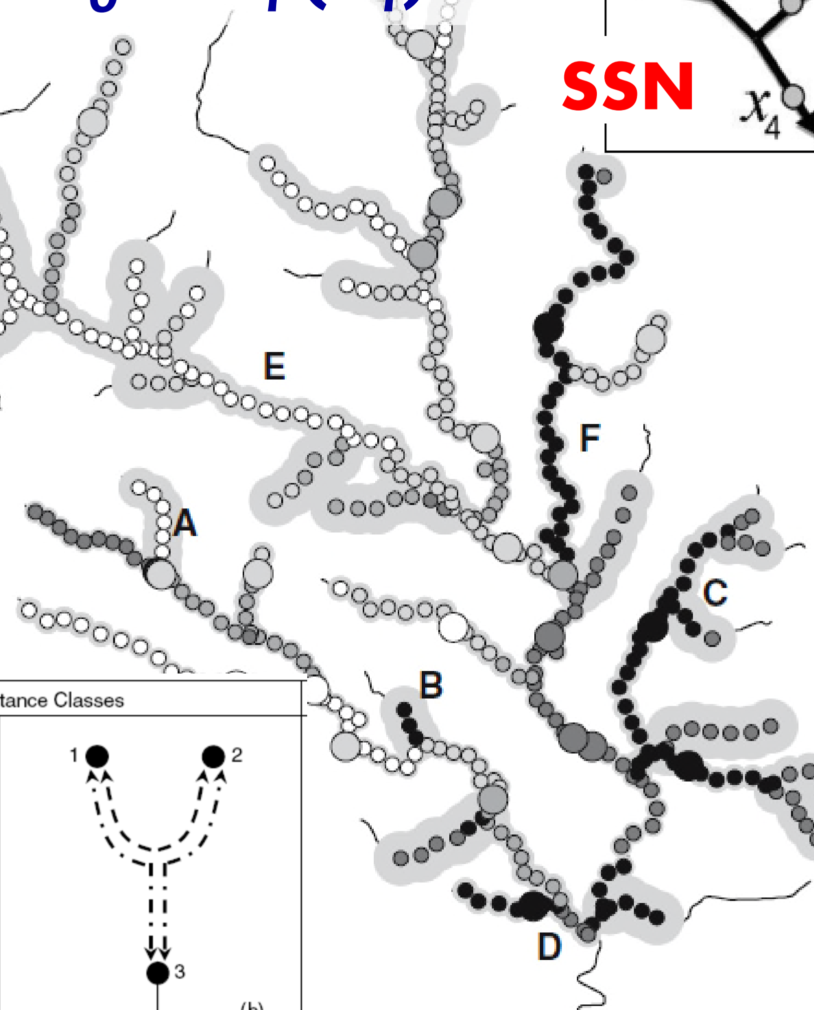
Jay Ver Hoef

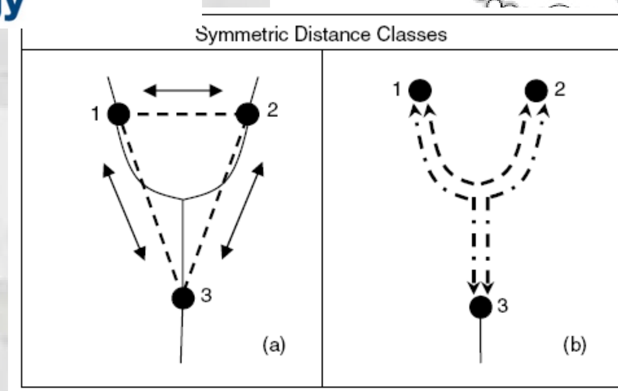
Erin Peterson

Queensland University of Technology

$$Y = b_0 + b_1(x_1) + €$$





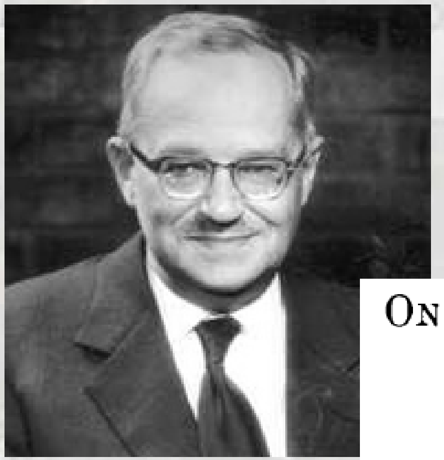


Streams Ignored by Statisticians

THE ARRANGEMENT OF FIELD EXPERIMENTS

R. A. Fisher, Sc.D.,
Rothamsted Experimental Station.

1926



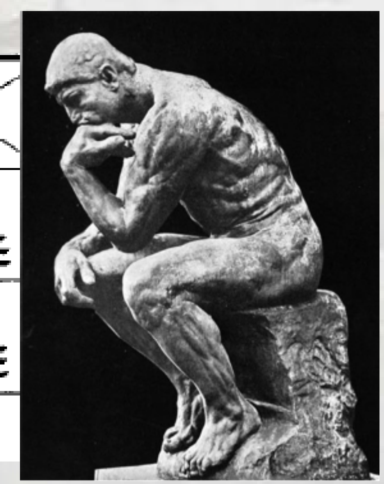
On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection.

By Jerzy Neyman

1934

2M LATE	2 S LATE	1 S EARLY		
	1M LATE	1M EARLY	2 S EARLY	2M LATE
2 S EARLY			2M EARLY	2 S LATE

Fig. 1.—A Complex Experiment with Winter Oats.



Statistical Models for Data on Stream Networks… Finally!

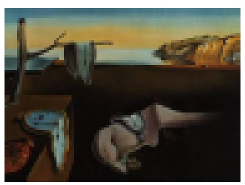
Environ Ecol Stat (2006) 13:449–464
DOI 10.1007/s10651-006-0022-8

ORIGINAL ARTICLE

Spatial statistical models that use flow and stream distance

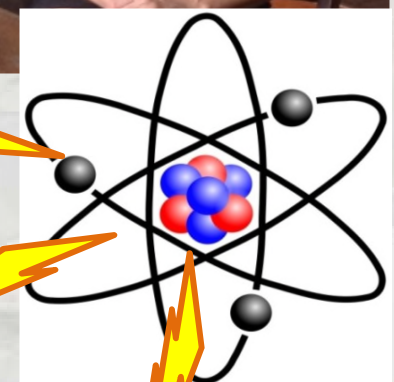
Jay M. Ver Hoef · Erin Peterson · David Theobald

2006



Journal of Statistical Software

January 2014, Volume 56, Issue 3. *http://www.jstatsoft.org/*



SSN: An R Package for Spatial Statistical Modeling on Stream Networks

Jay M. Ver Hoef, NOAA National Marine Mammal Laboratory
Erin E. Peterson, CSIRO, Brisbane
David Clifford, CSIRO, Brisbane
Rohan Shah, CSIRO, Brisbane

STARMAP
Space-Time Aquatic Resources Modeling and Analysis Program

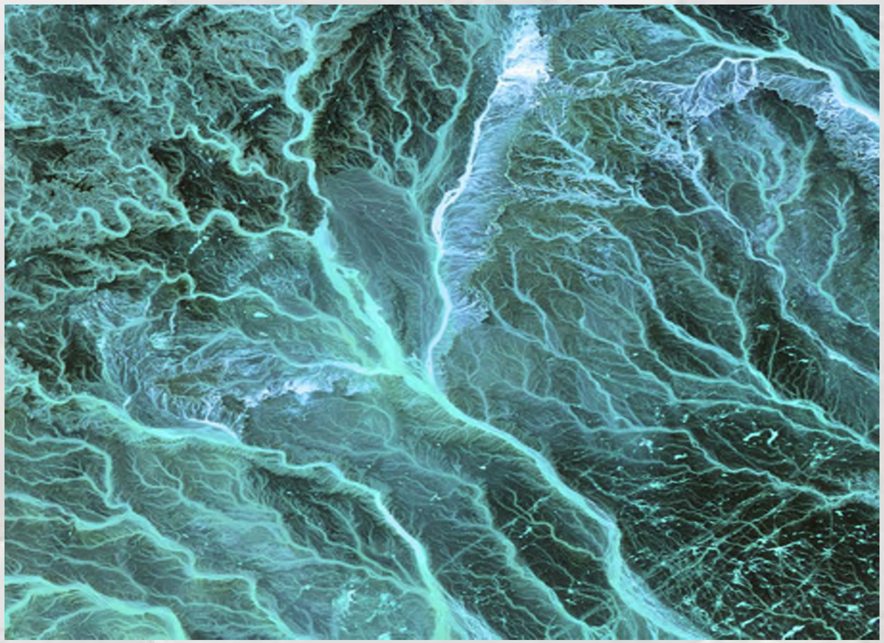
Functional Linkage of Watersheds and Streams (FLoWS)

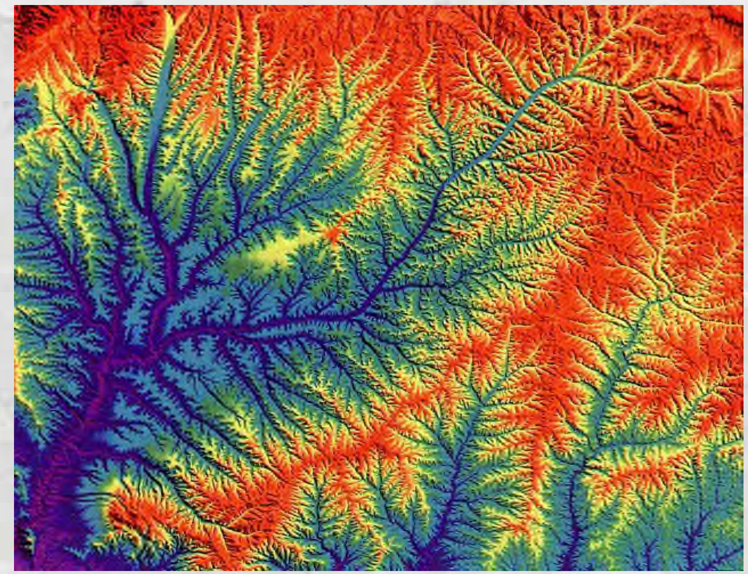
- ArcGIS Geoprocessing Toolbox written in Python v2.5 for ArcGIS v
- Developed by Dr. Dave Theobald and John Norman at Colorado St

EPA

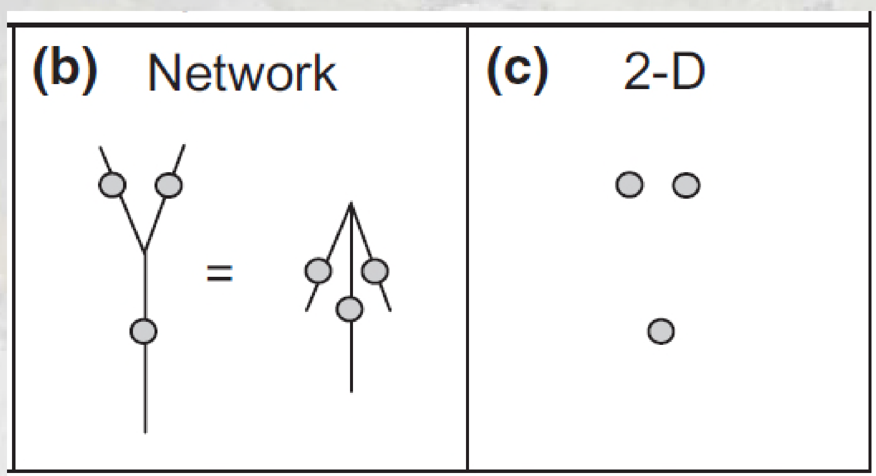
Streams Have Unique Properties

Networked Systems, Directional Flow, & Embedded Within Terrestrial Landscapes



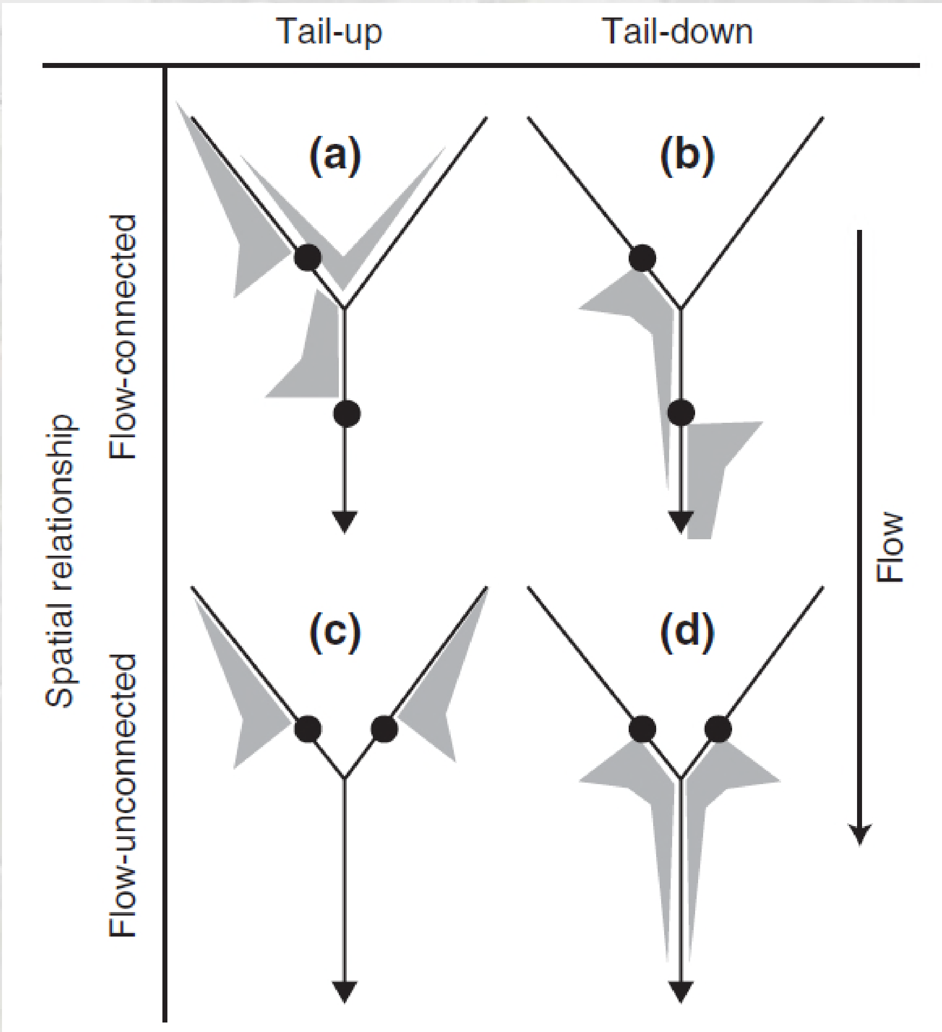


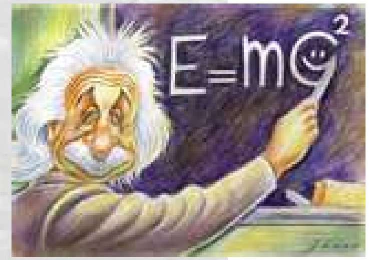
Dual coordinate system required for networks

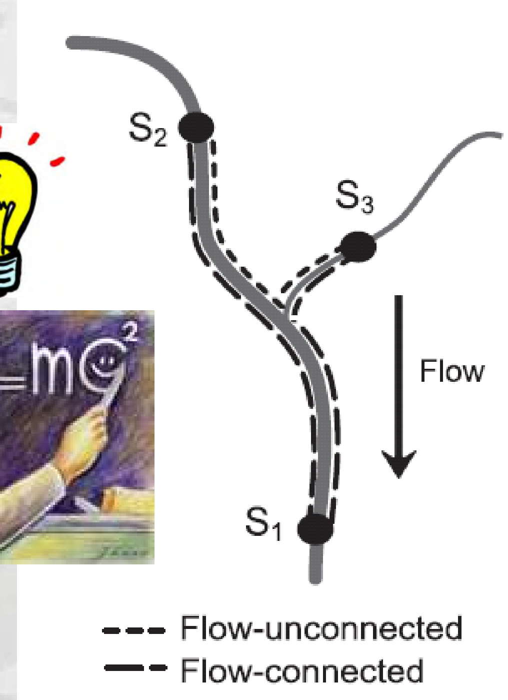


Peterson et al. 2013. *Ecology Letters* **16**:707-719.

Key Innovation is Covariance Structure Based On Network Structure





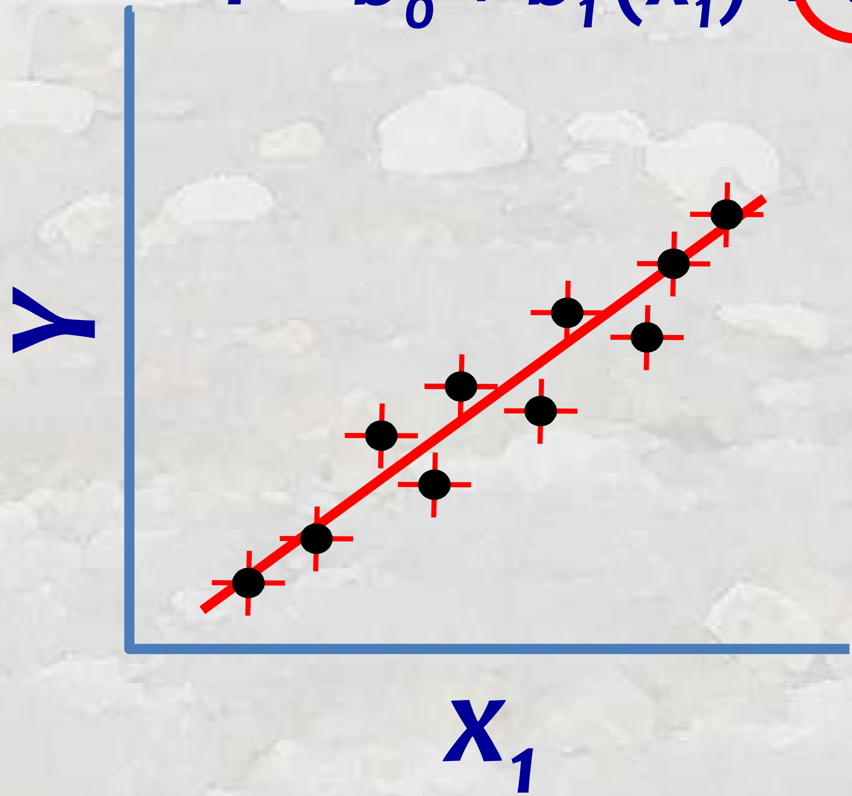


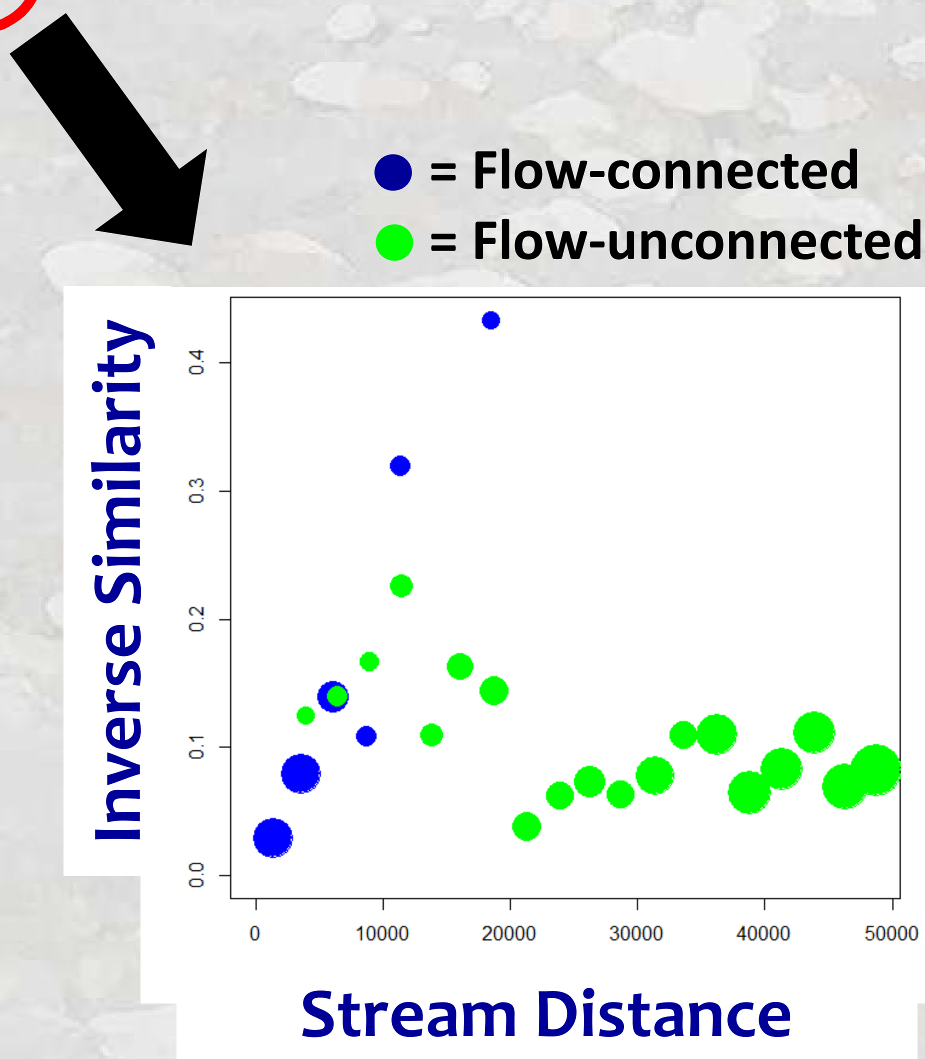
- Models "understand" how information moves among locations
- Models account for spatial autocorrelation among observations

Peterson et al. 2007. *Freshwater Biology* **52**:267-279;
Peterson & Ver Hoef. 2010. *Ecology* **91**:644-651.

Linear Statistical Model w/AutoCovariance Error Structure Based on Network Topology

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



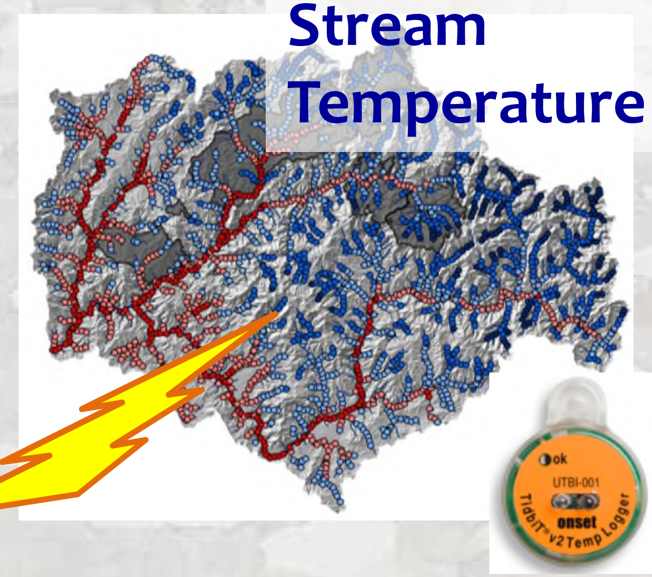


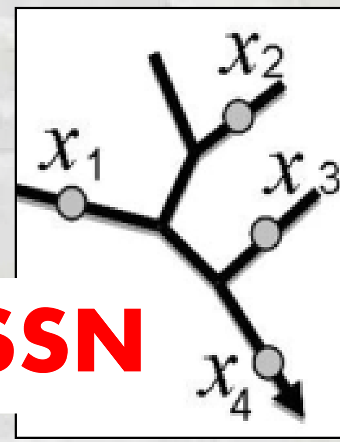
Stream Models are Generalizable…

Distribution & abundance

Response Metrics
- •Gaussian
- •Poissan
- •Binomial

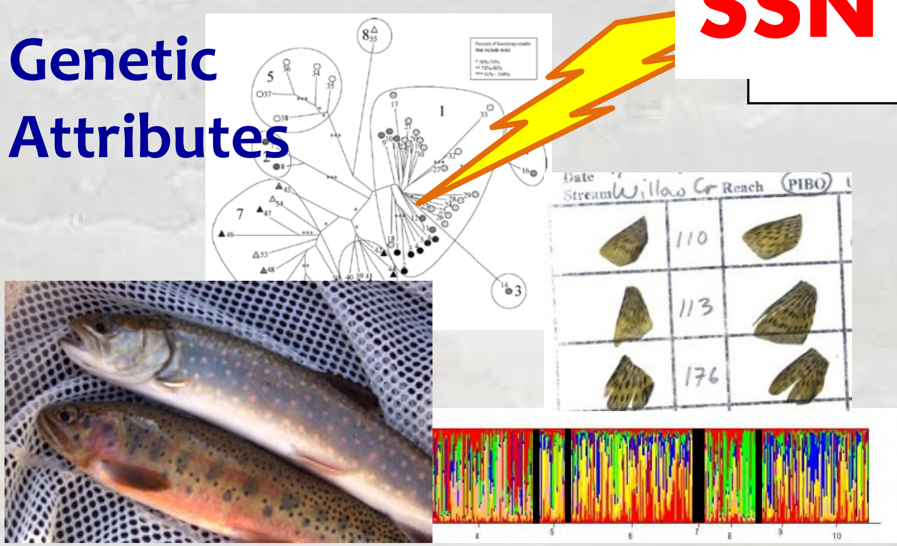
Stream Temperature





SSN

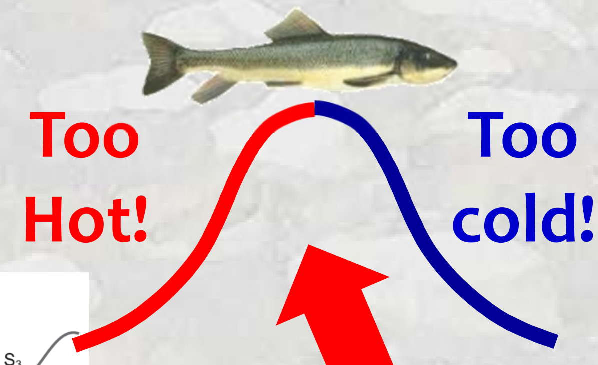
Genetic Attributes

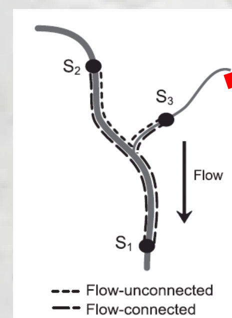


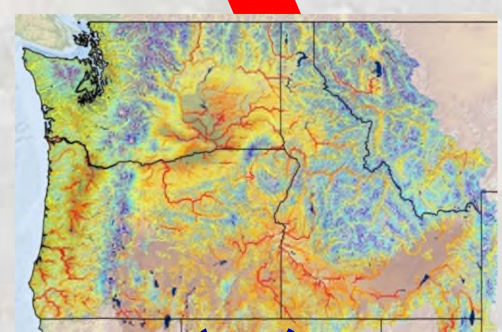
Water Quality Parameters

SSN Model are Versatile…

- Parameter estimation & significance testing
- Predictions at unsampled locations for 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



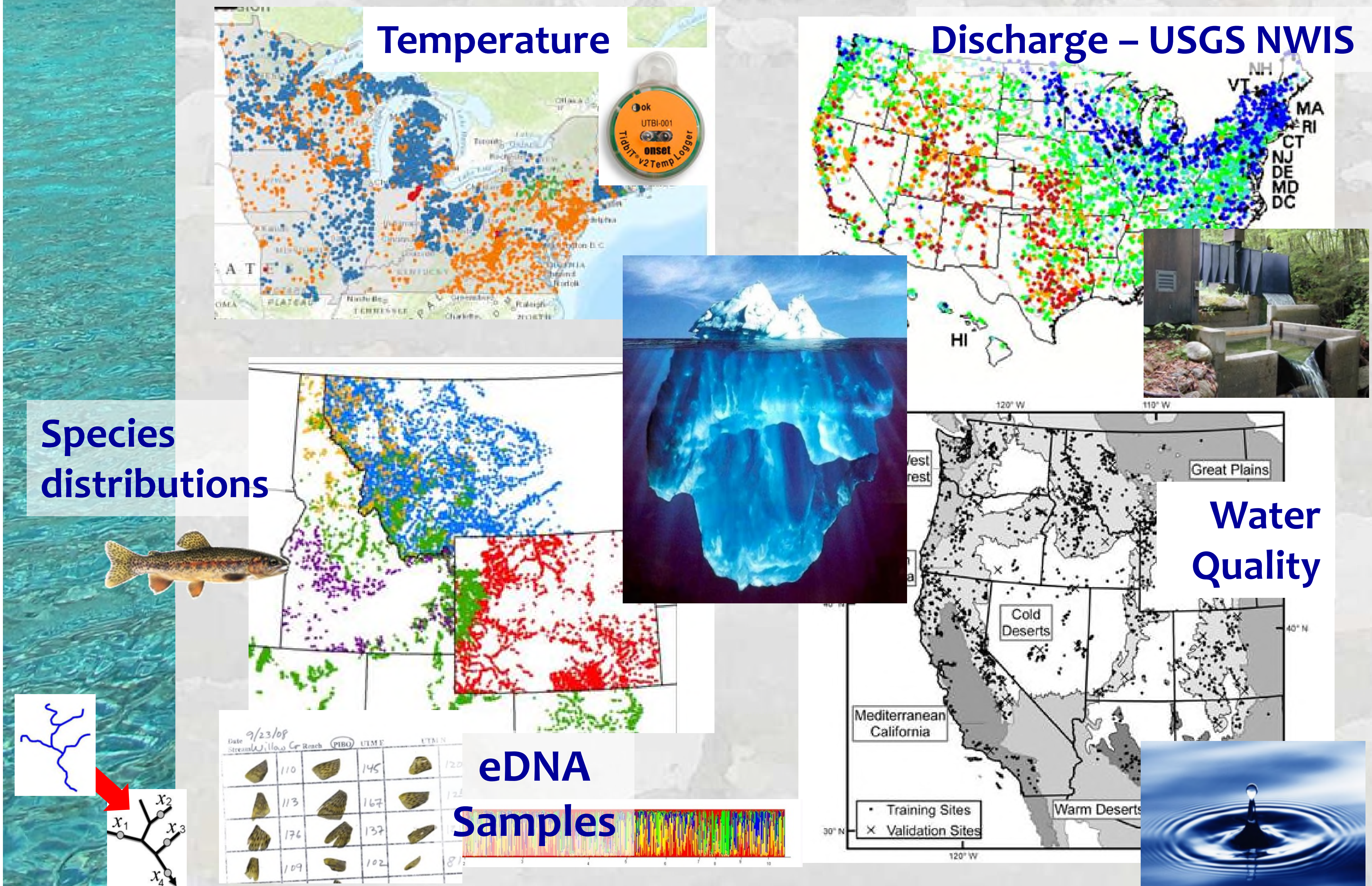




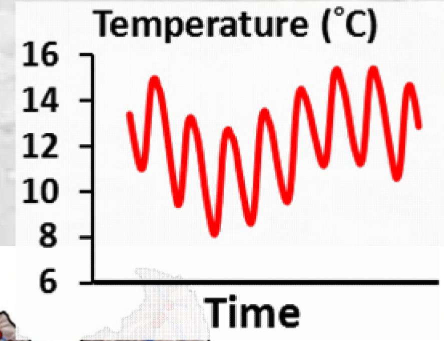
$$Y = b_0 + b_1(x_1) + €$$

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

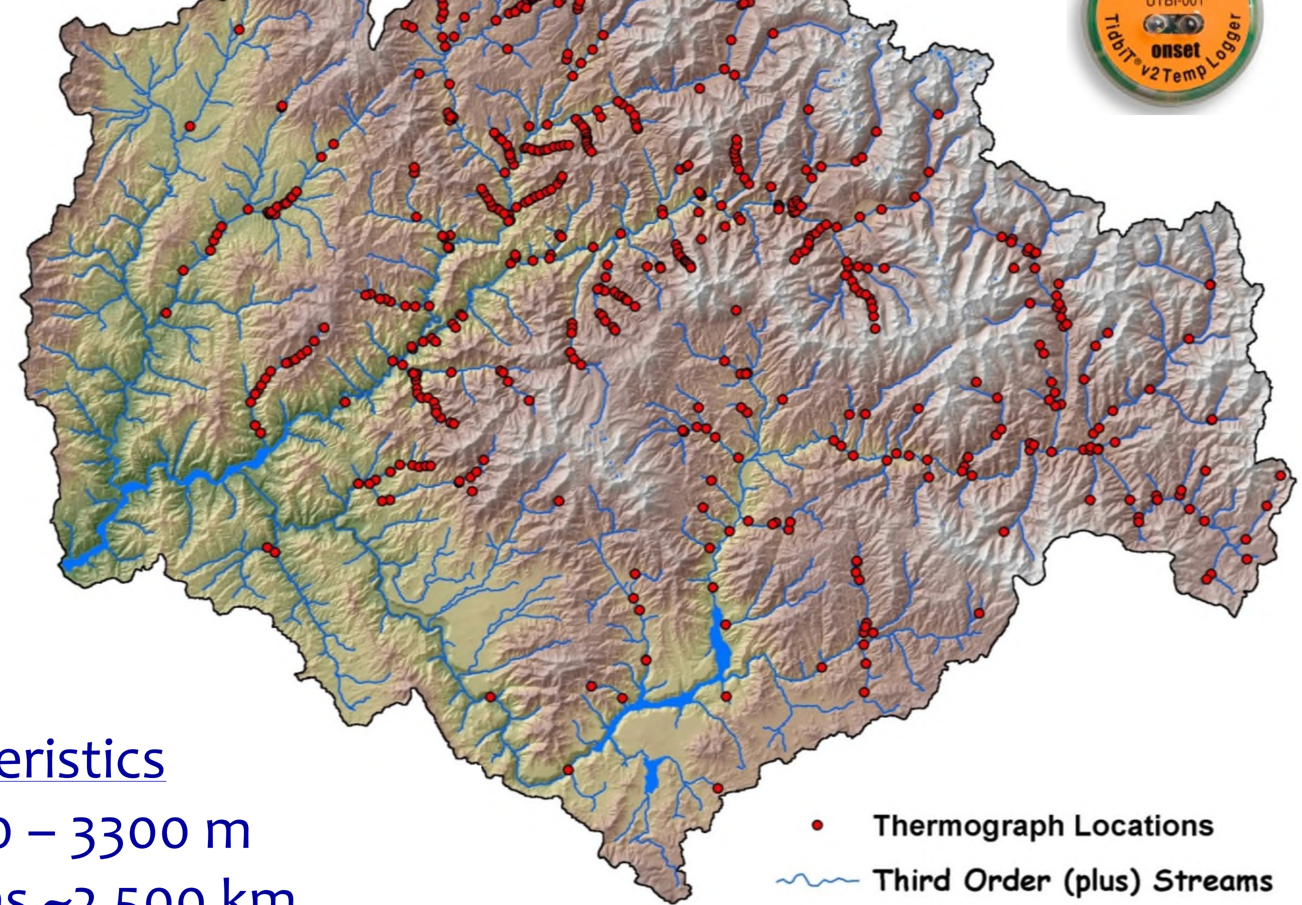
It's Becoming a **BIG DATA** World



Example: a River Network Temperature Model – Boise River Basin



Stream Temperature Database
14 year period (1993 – 2006)
780 observations
518 unique locations



Watershed Characteristics
Elevation range 900 – 3300 m
Fish bearing streams ~2,500 km
Watershed area = 6,900 km²

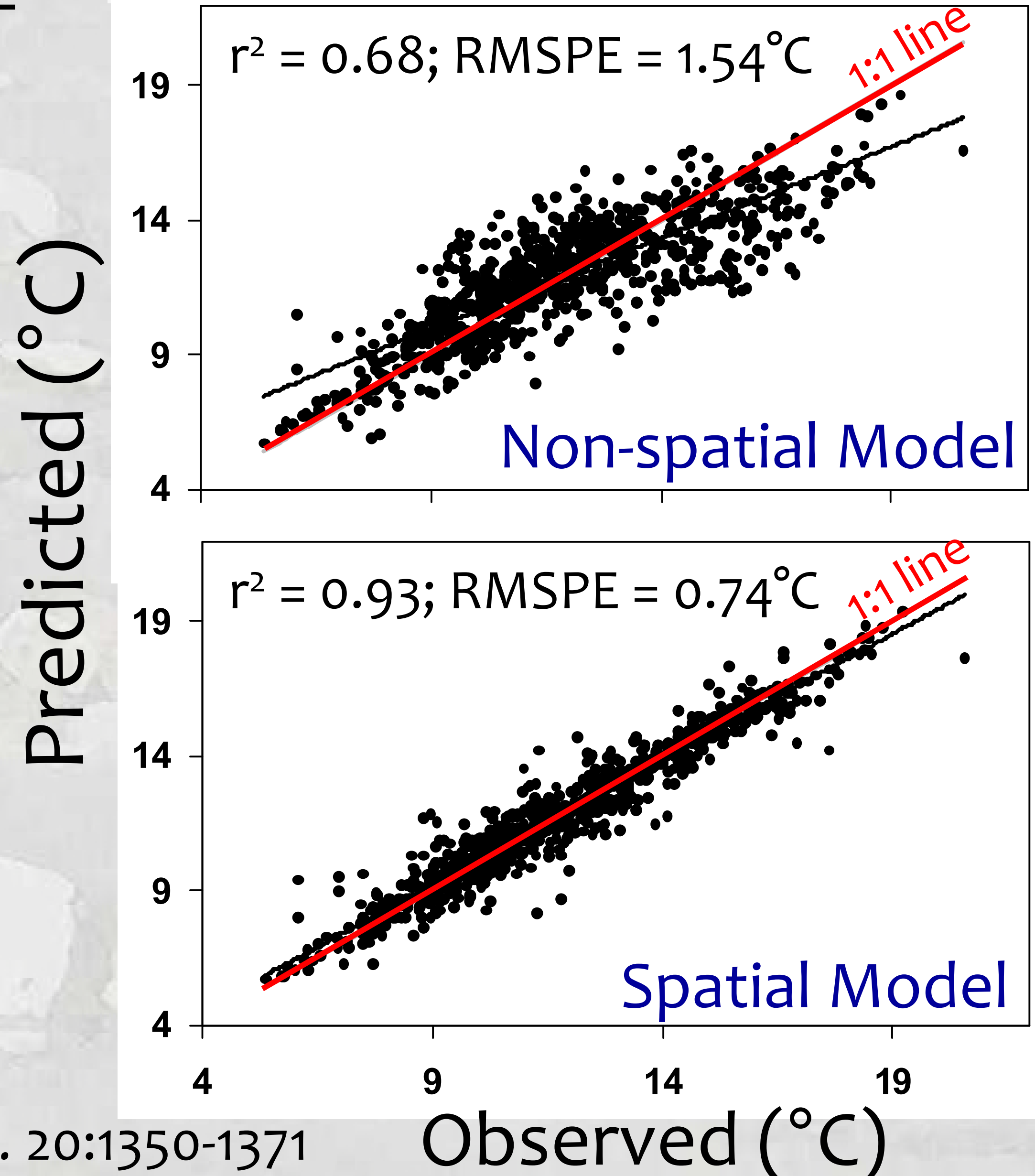
Accurate & Precise Information from a Crowd-Sourced, Interagency Database

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

Autocorrelation affects parameter estimates

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

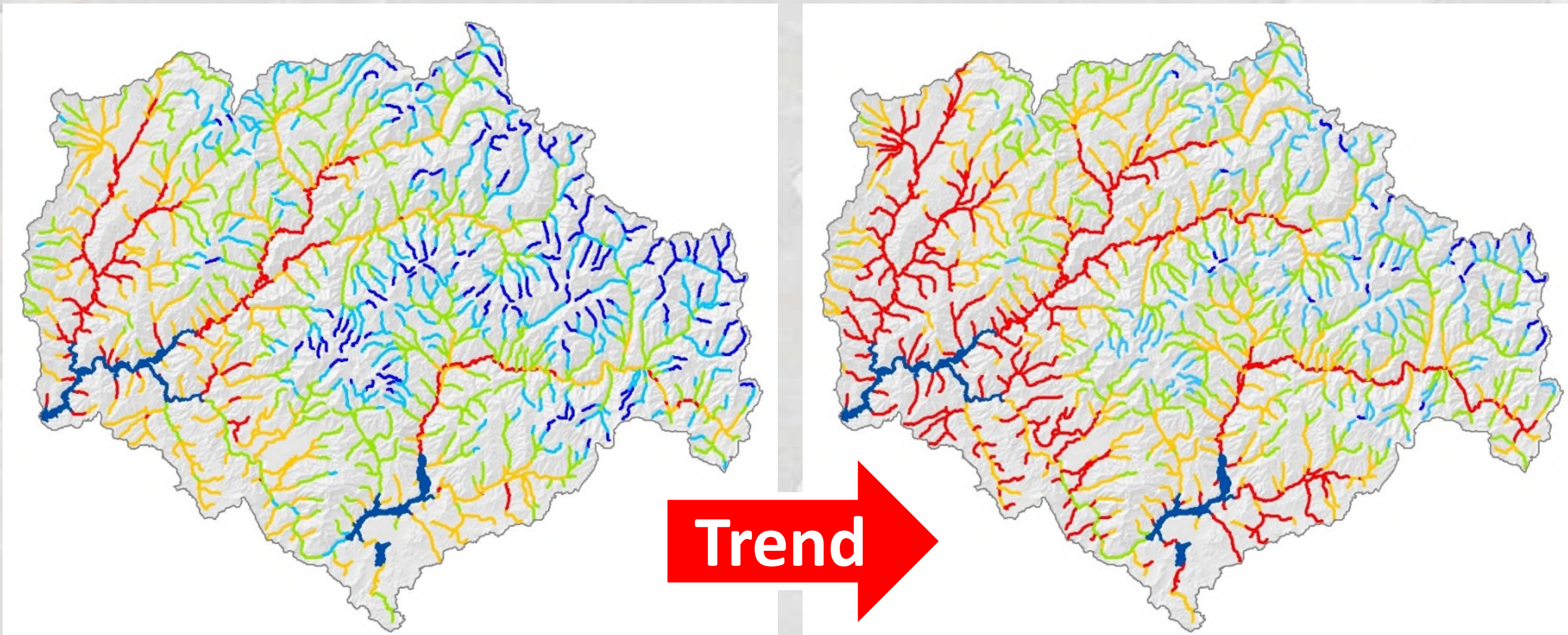
Mean Summer Stream Temp



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

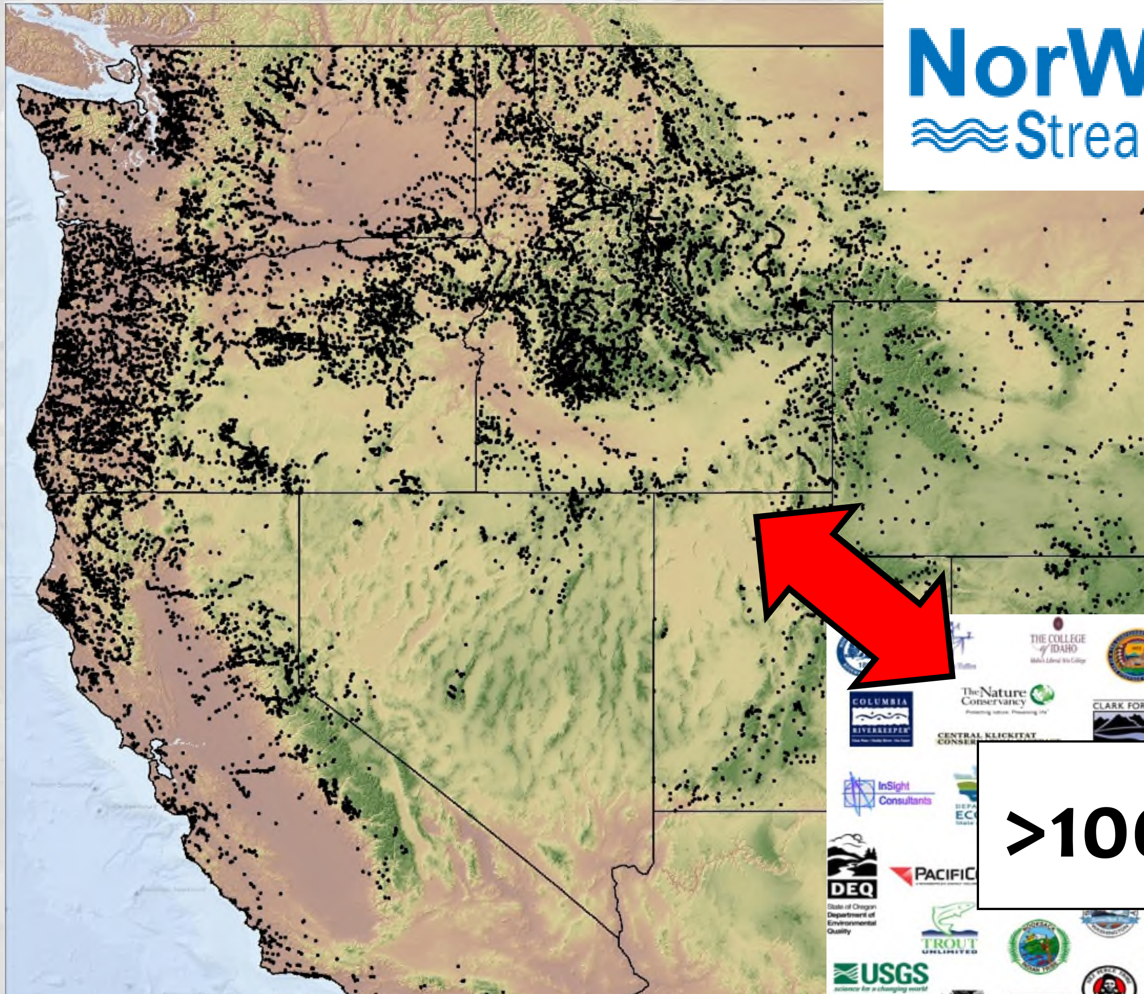
Interpolated Predictions Provide High-Resolution Network *Status* Maps

Time 1 **Time 2**

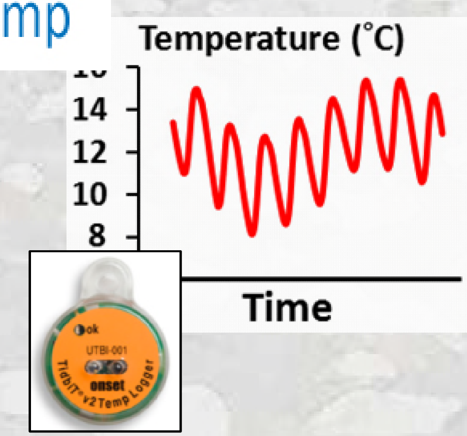


Which sets the stage for *trend* assessments…

SSN Models Scale Easily…



NorWeST Stream Temp

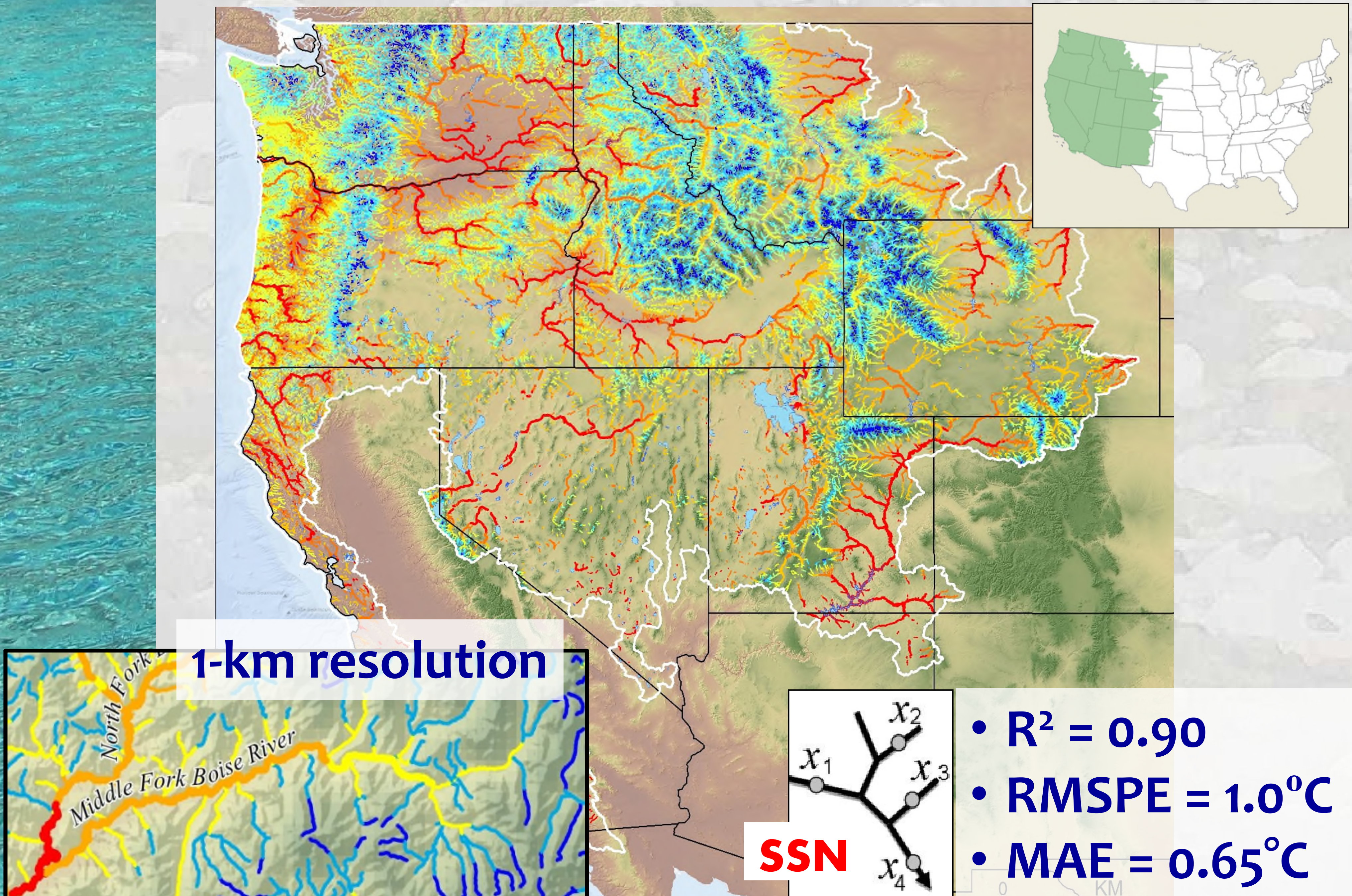


>100 agencies

>200,000,000 hourly records

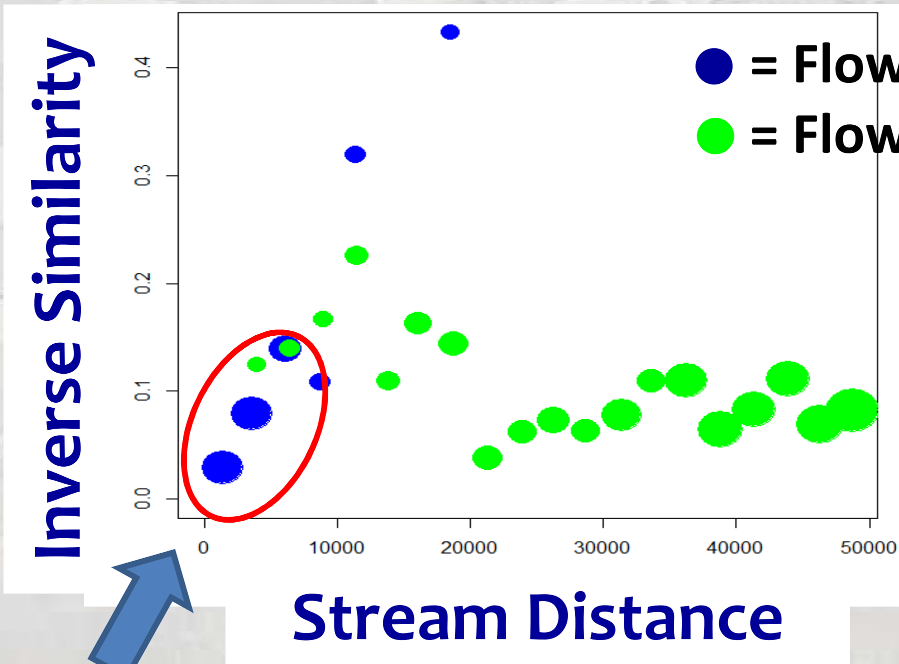
>20,000 unique stream sites

… To Go From Headwaters to Sea



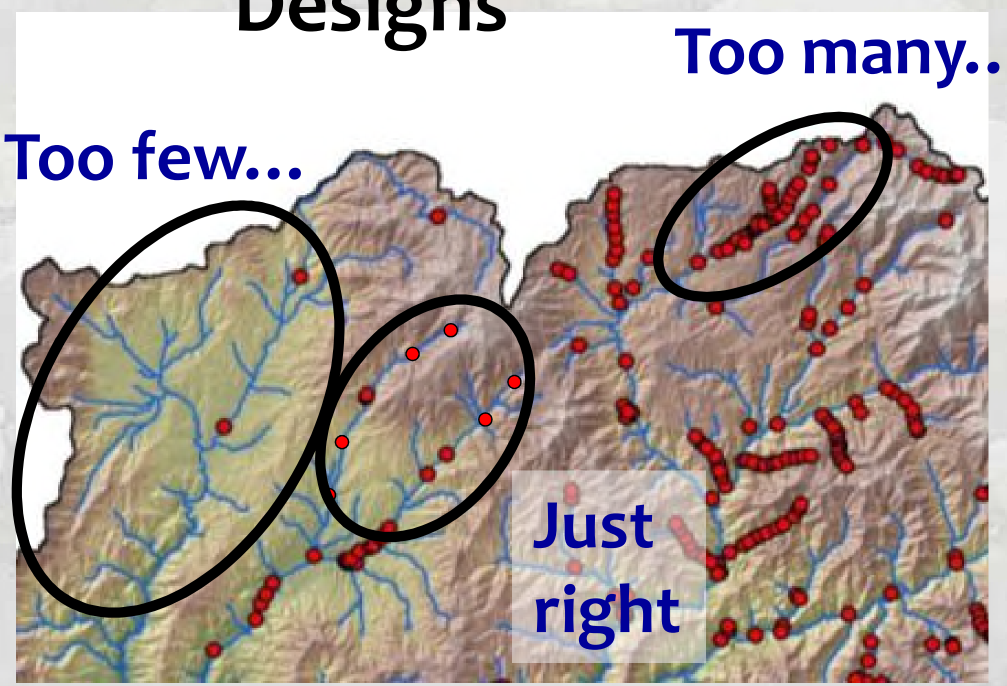
- $R^2 = 0.90$
- RMSPE = 1.0°C
- MAE = 0.65°C

Models Describe Autocorrelation Distances



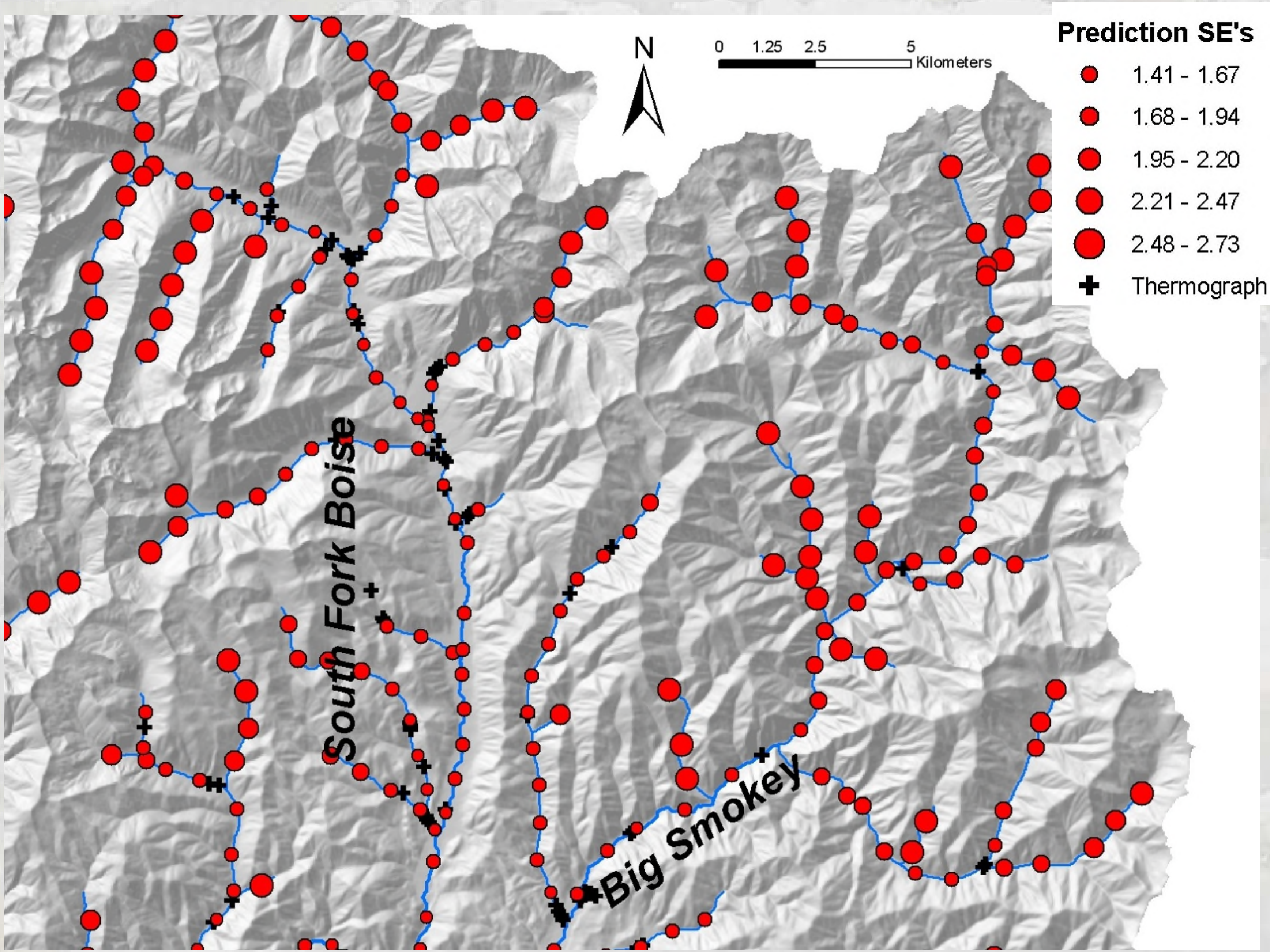
Redundant information

Efficient Monitoring Designs

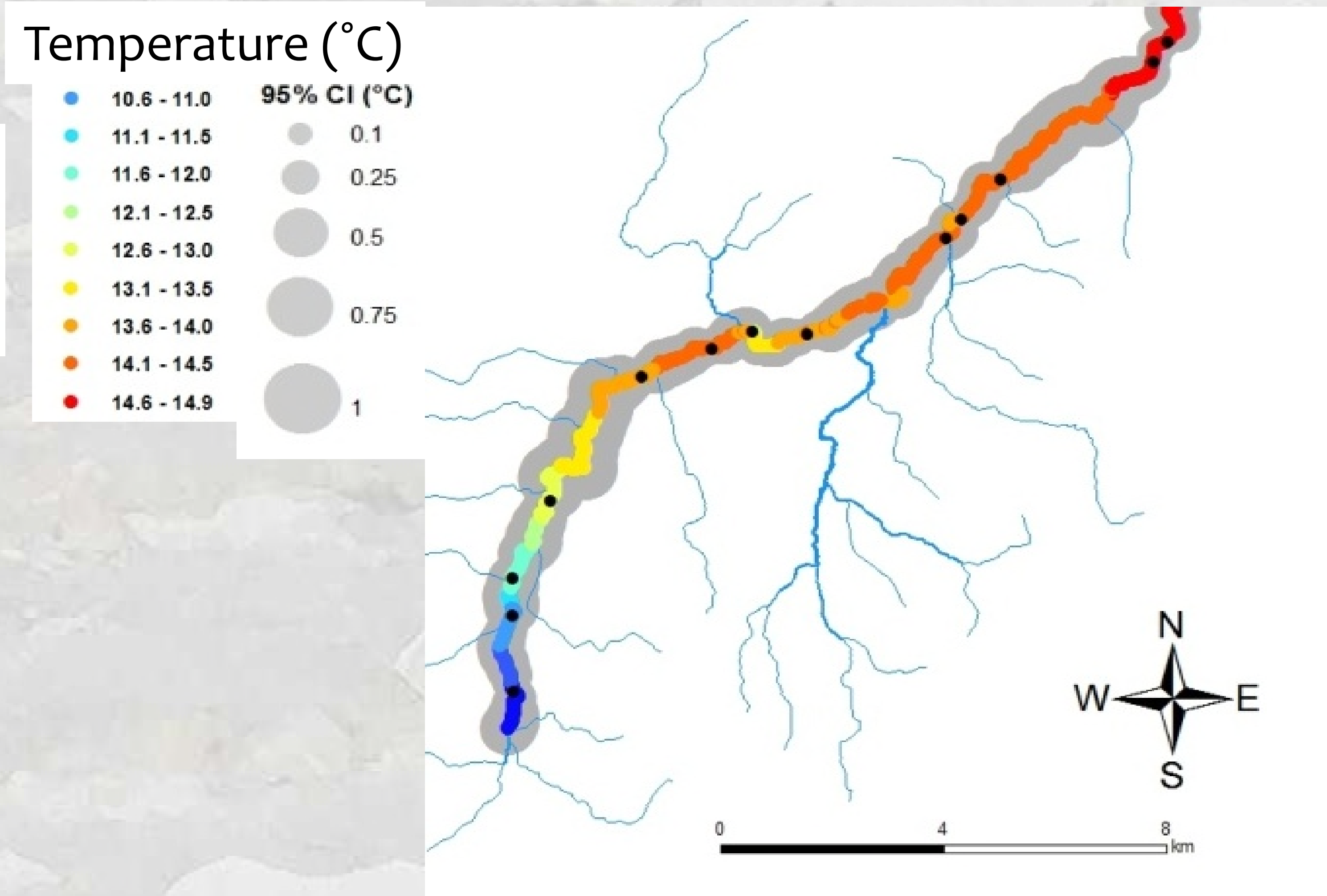


Som et al. 2014. Spatial sampling on streams: principles for inference on aquatic networks. *Environmetrics* **25**: 306-323.

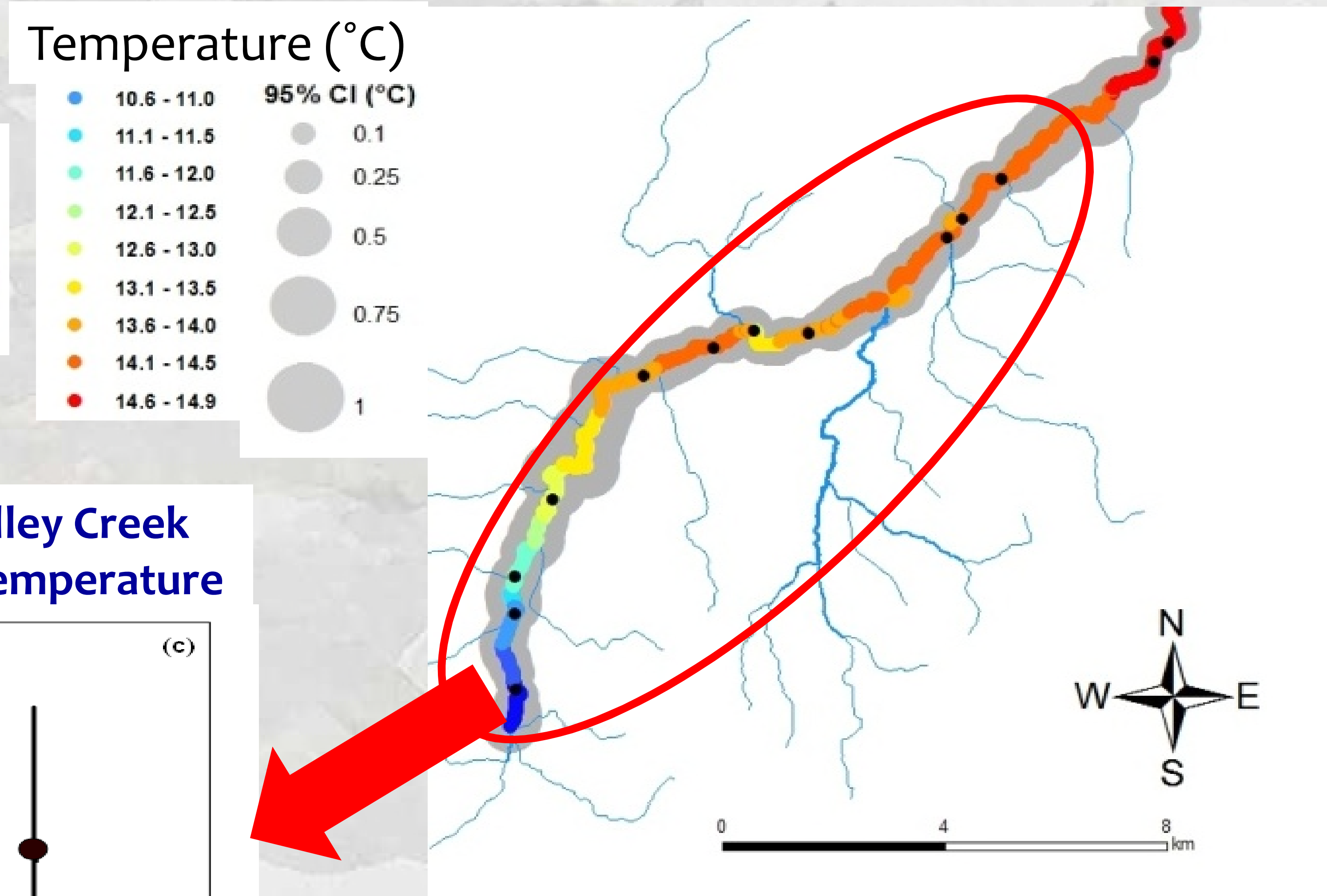
Spatial Variation in Prediction Precision



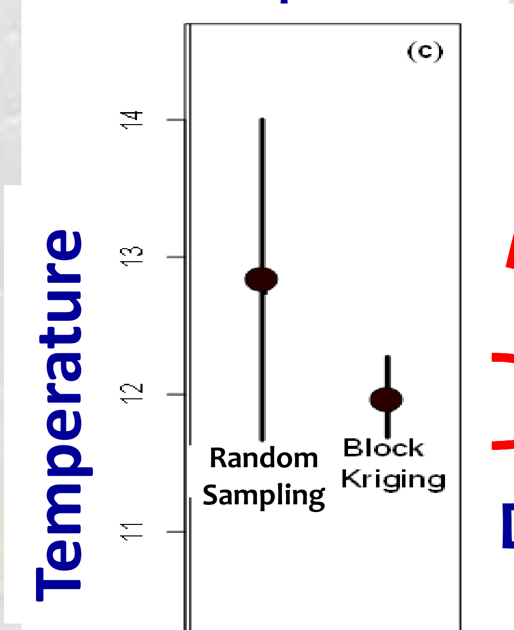
Block-krige Estimates of Mean & Variance at User-Defined Scales



Block-krige Estimates of Mean & Variance at User-Defined Scales



Bear Valley Creek
Mean Temperature



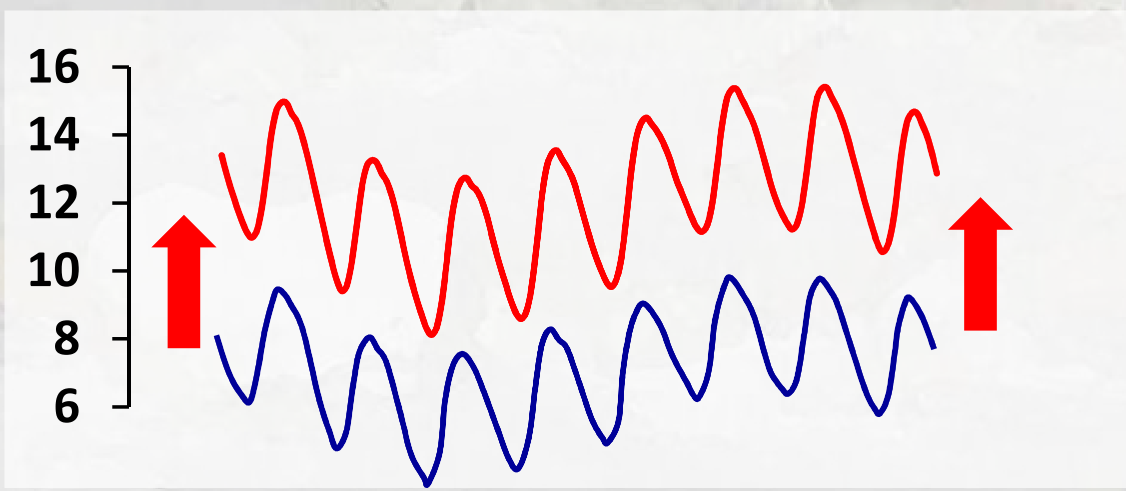
Precise & unbiased estimates

Does this reach meet the TMDL standard?

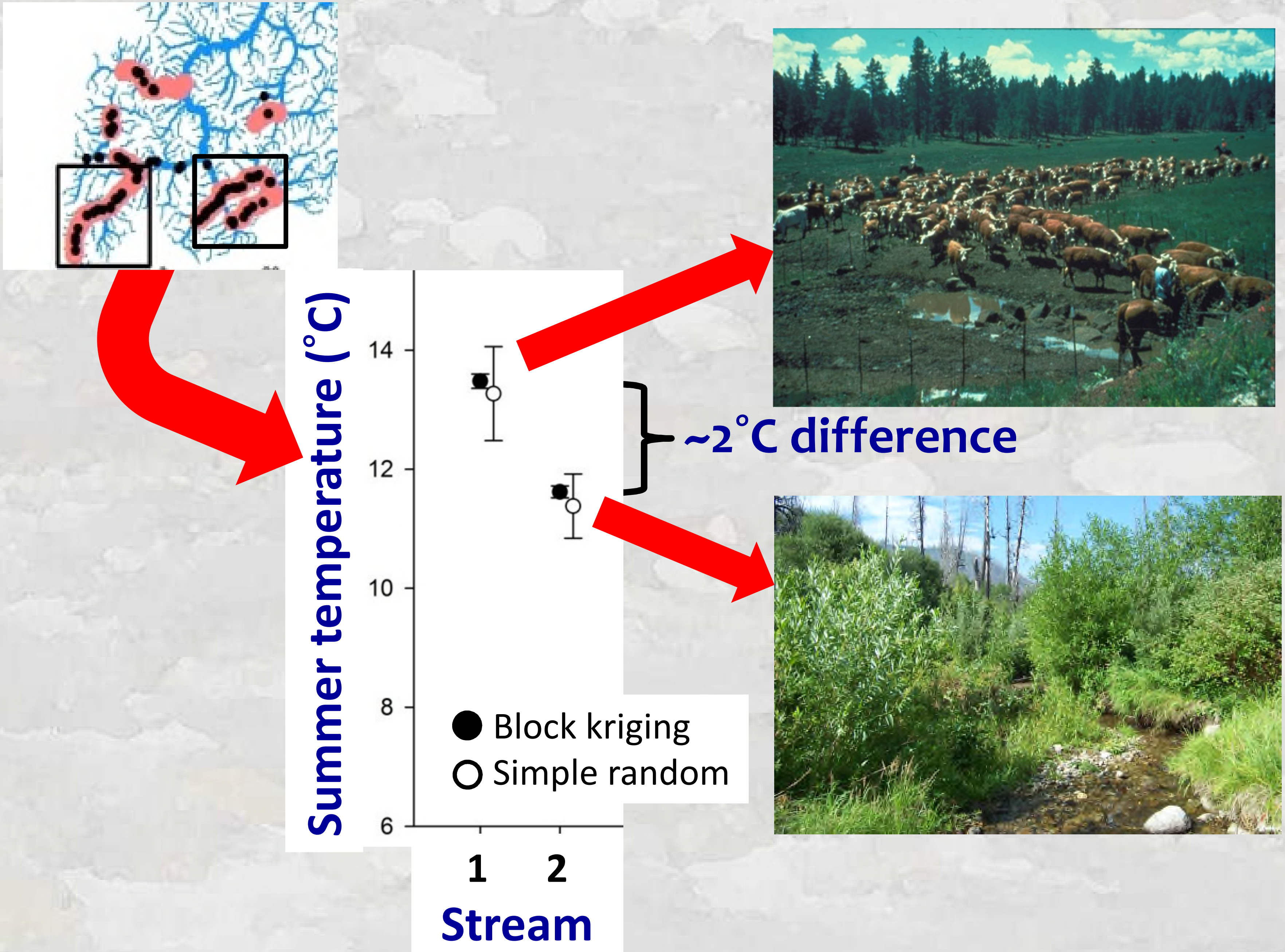
Reference Site Comparison Approach

Pick “degraded” & “healthy” streams to compare

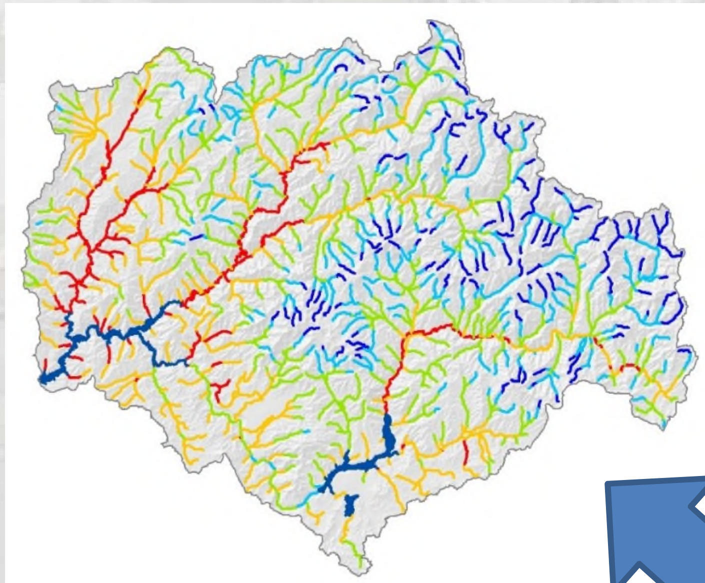
How altered is this stream?

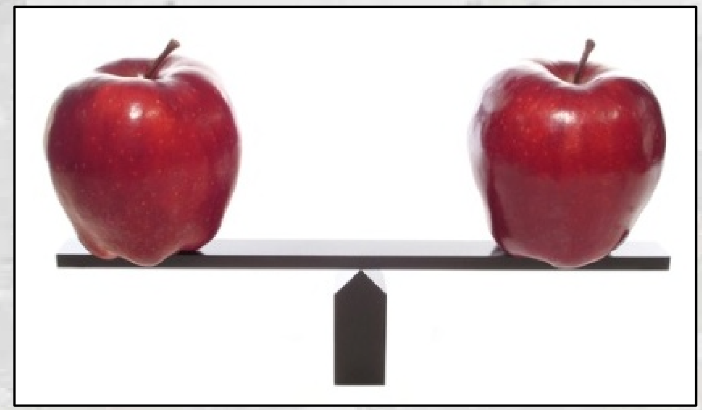


Block-Krige Estimates for Both Streams

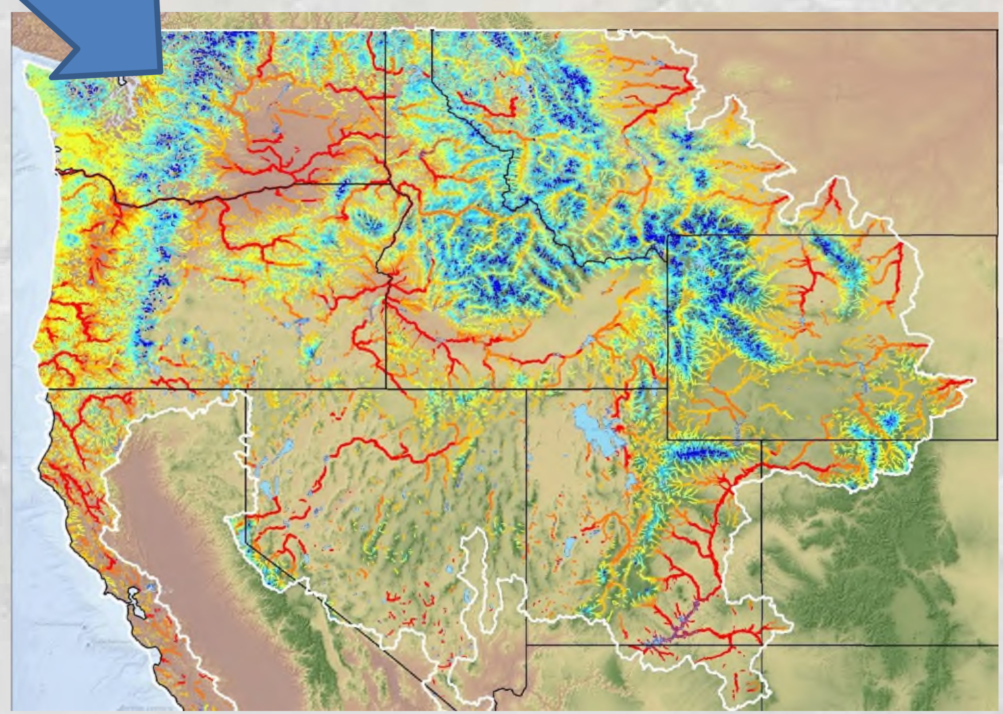


Anywhere Within a River Network





Anywhere within the country (someday)



Block-Kriging & Reference Site Approach Broadly Applicable for Many Water Parameters…

Sediment…

Urban runoff…

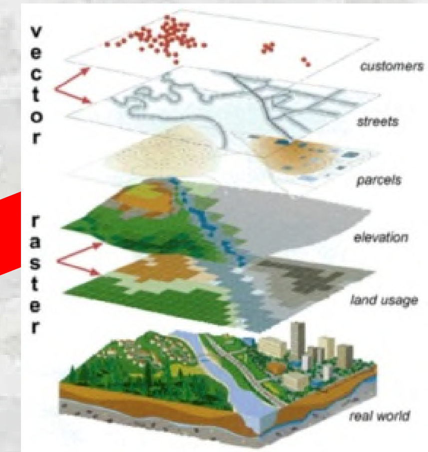
Nutrients…

Mining…

SSN Models Benefit from Team Science Approach

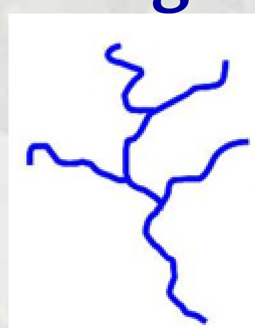
Managers

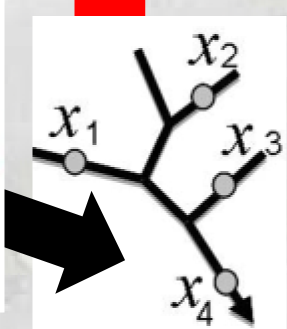
GIS analysts



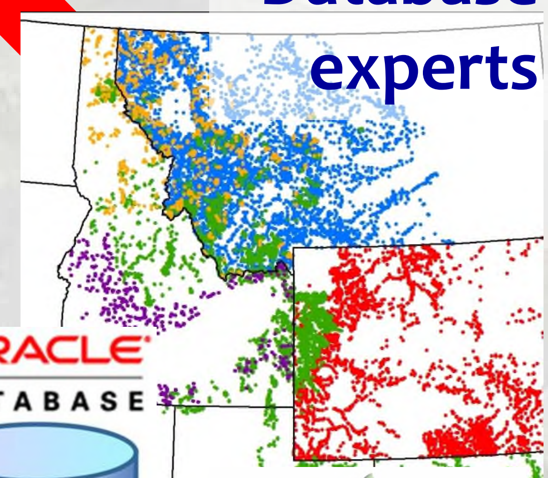
Scientists

Ecological Modelers





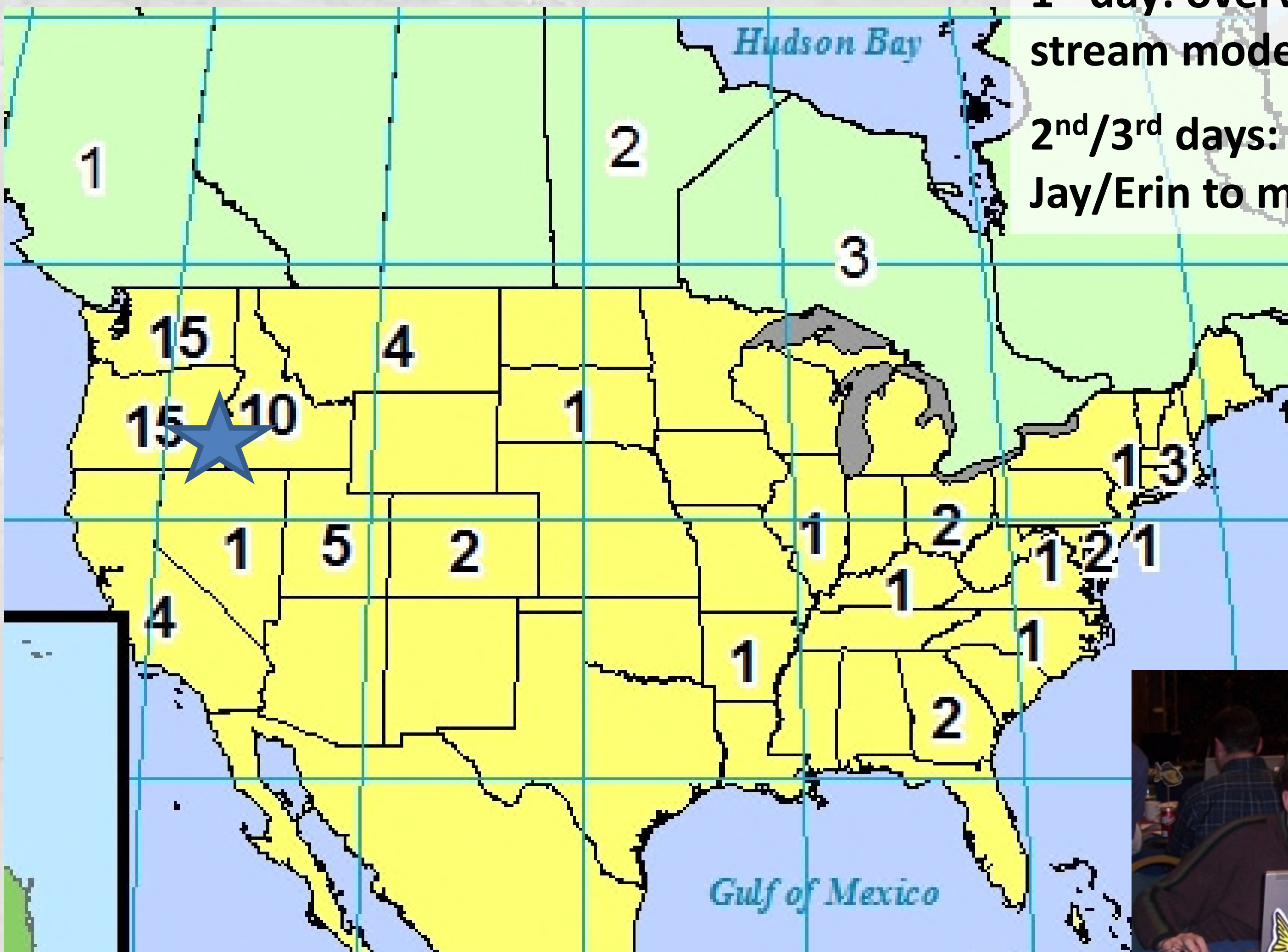
Database experts



Annual Stream Statistics Training Workshop

Building a Grassroots Analysis Army

Boise, Idaho each spring…



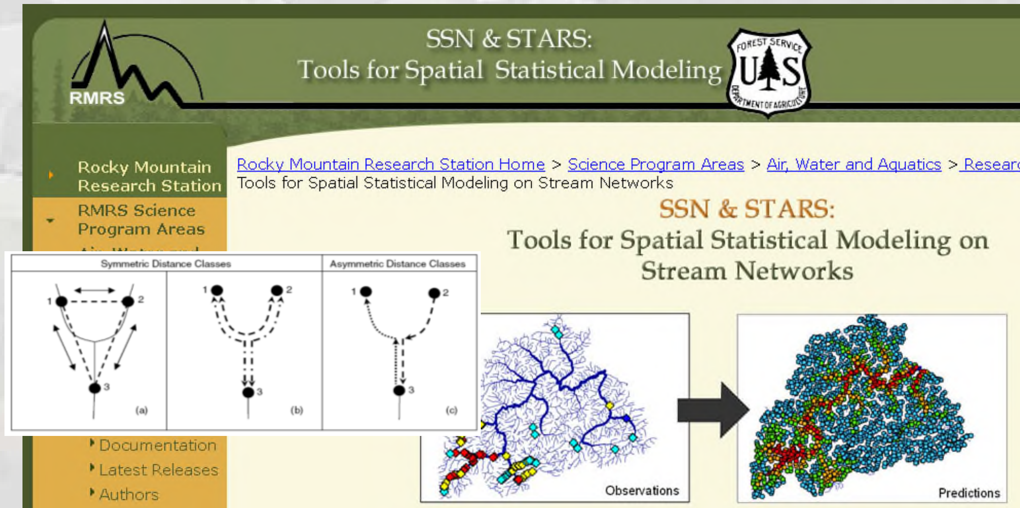
3 day workshop

1st day: overview of spatial stream models (webinar)

2nd/3rd days: work 1-on-1 with Jay/Erin to model your data

Growing User Community…

SSN/STARS Website



>28,500 website visits in first 3.5 years

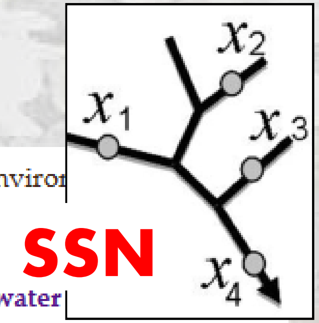
Free, high-quality software

>1,000 software downloads

Locations of visits to SSN/STARS website in last month

Growing Literature & Applications

24 studies & counting…



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

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

Peterson, EE, and NS Urquhart. 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.

Gardner K, and B McGlynn. 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.

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

Ver Hoef J.M. and Peterson E.E. 2010. **A moving average approach to spatial statistical models of stream networks: Rejoinder**. Journal of the American Statistical Association 105: 22-24.

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

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.

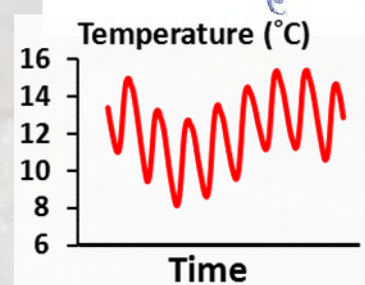
Isaak D., Luce, C., Rieman B., Nagel D., and Peterson E., Horan D., Parkes S., and G. Chandler. 2010. **Effects of climate change and wildfire on salmonid thermal habitats within a mountain stream network**. Ecological Applications 20:1350-1371.

Ruesch, A. S., Torgersen, C. E., Lawler, J. J., Olden, J. D., Peterson, E. E., Volk, C. J., & Lawrence, D. J. 2012. **Projected Climate-Induced Habitat Loss for Salmonids in the John Day River Network, Oregon, USA**. Conservation Biology 26:873-882.

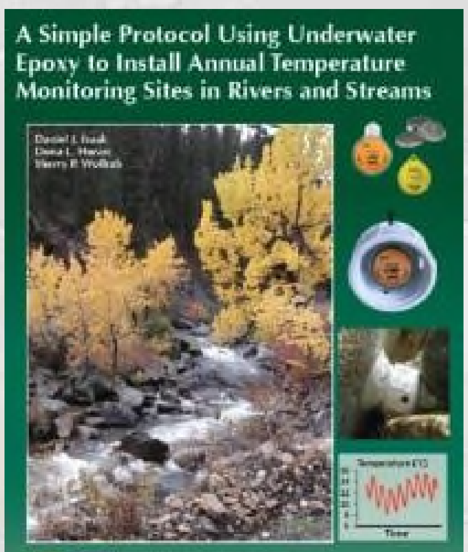
Related Websites Google search on…

1) **SSN/STARS** – software tools, tutorials, and example datasets for spatial statistical network modeling
2) ***NorWeST*** – regional stream temperature database & climate scenarios
3) ***National Stream Internet*** **–** NHD networks reconditioned for application of SSN models, SSN bibliography, stream databases

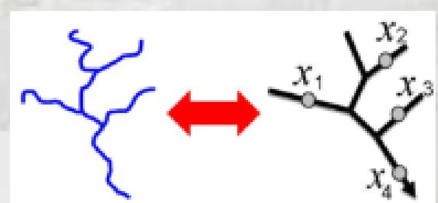
Data…



Publications…



Software…




The End

When & Where Are SSN Models Useful?

Not here

Small datasets (n < 50 sites)

Spatially sparse datasets

NARS – National Aquatic Resource Surveys

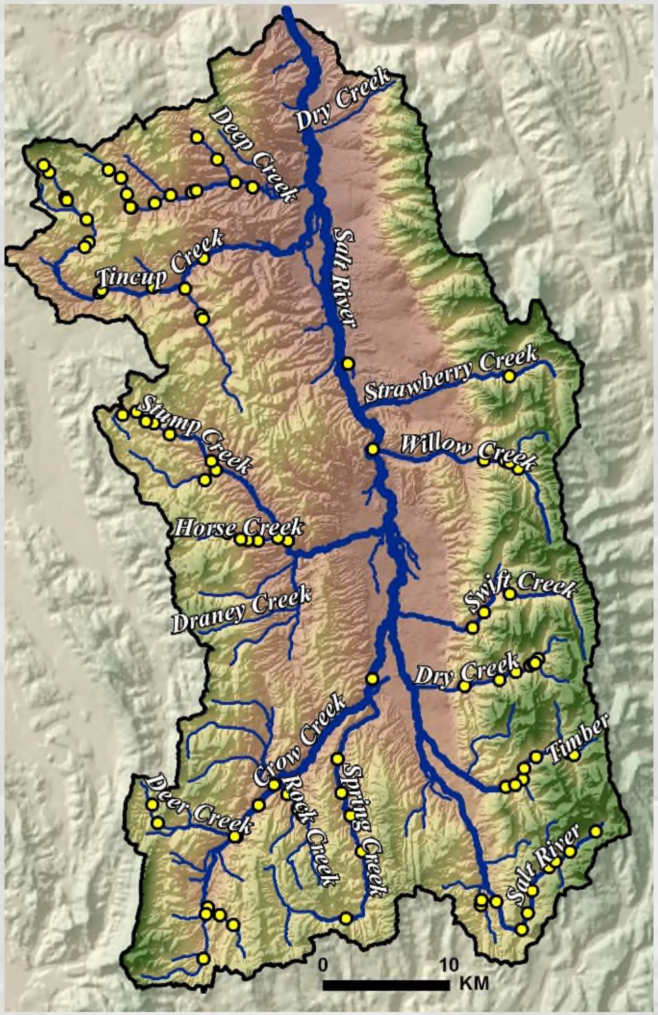
EPA

Network connectivity effects weak or difficult to estimate

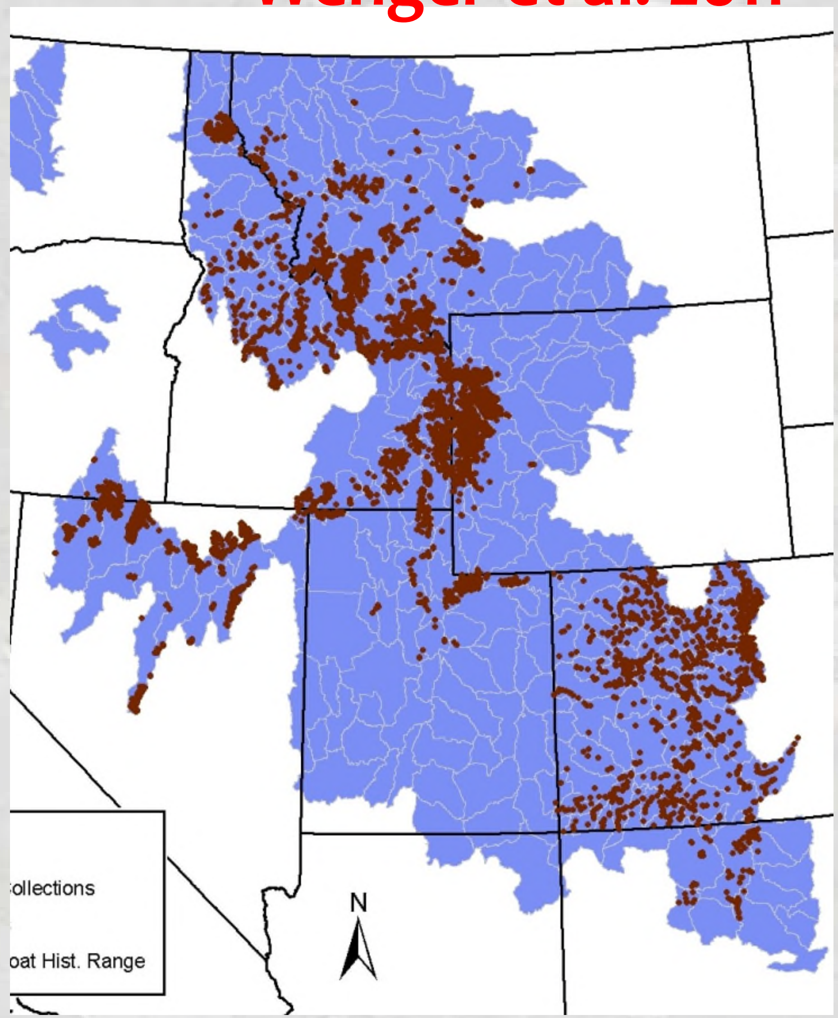
When & Where Are SSN Models Useful?

Large or Spatially Dense Datasets

103 Alkalinity, pH, & conductivity measurements



10,000 fish sample sites
Wenger et al. 2011

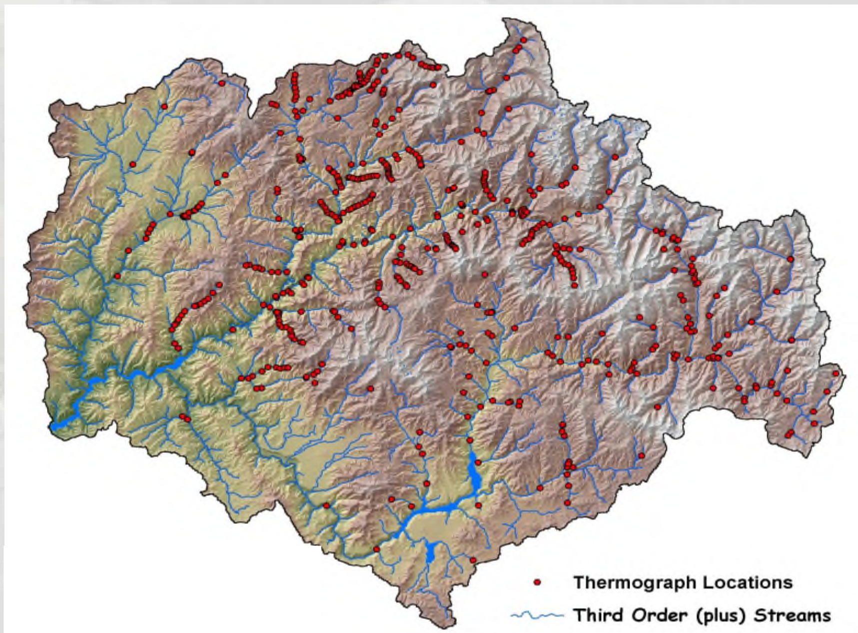


Network connectivity affects patterns among sites

Sample Size & Computational Requirements

Minimum sample size ~ n ≥ 50 / 100

- -more parameters with autocovariance
- -spatial clustering needed



TO site \ FROM Site	A	B	C	D
A	0	28	30	0
B	0	0	15	0
C	0	13	0	0
D	0	0	0	0

Distance matrix

Maximum sample size ~ n < 10,000

- -inversion of n x n matrix

Streams Ignored by Spatial Statisticians

Vegetatio 80: 107–138, 1989.
© 1989 *Kluwer Academic Publishers. Printed in Belgium.*

1989

Spatial pattern and ecological analysis

Pierre Legendre[1] & Marie-Josée Fortin[2]

Integrating of 2D Location Information to Model Structure

Ecology, 74(6), 1993, pp. 1659–1673
© 1993 by the Ecological Society of America

1993

SPATIAL AUTOCORRELATION: TROUBLE OR NEW PARADIGM?[1]

PIERRE LEGENDRE
Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale A, Montréal, Québec, Canada H3C 3J7

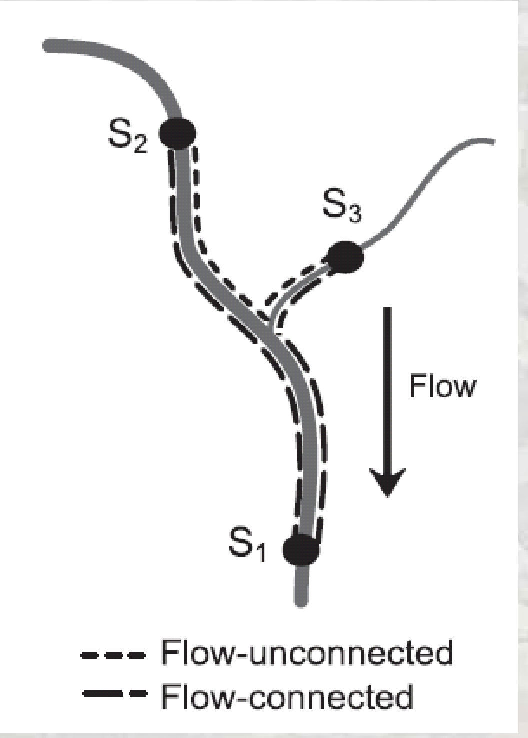
NOEL A.C. CRESSIE

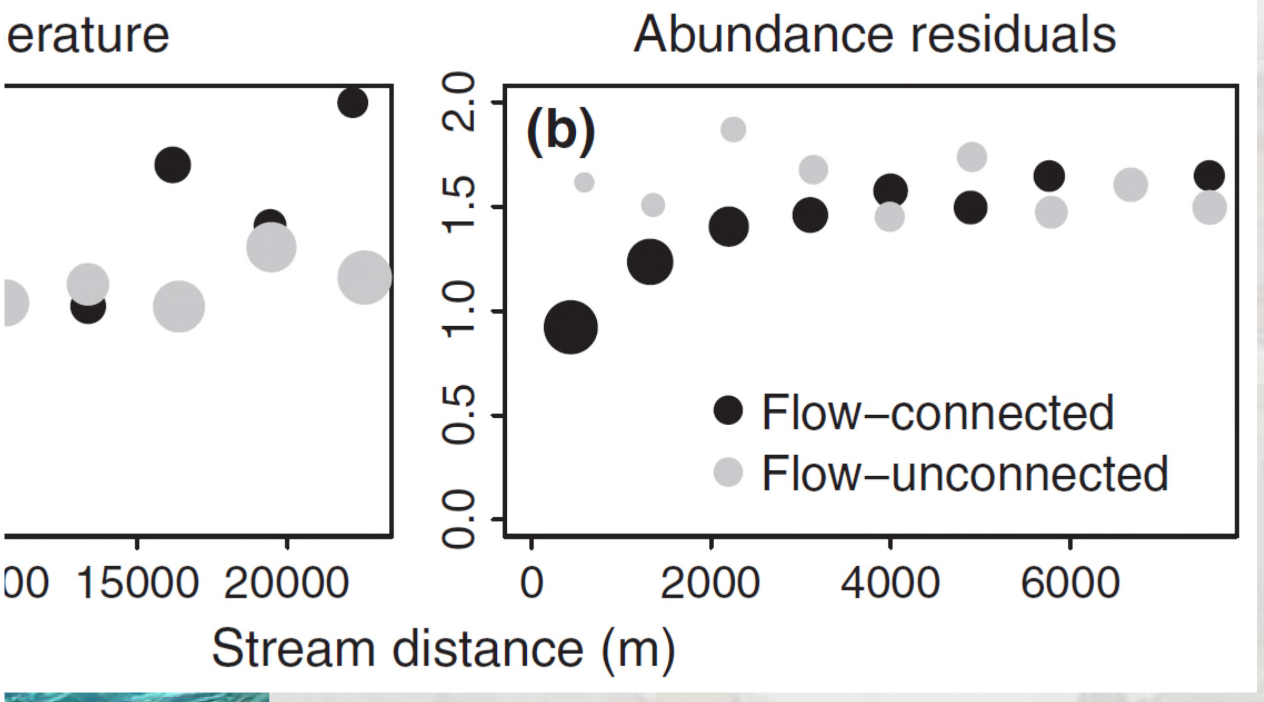
STATISTICS FOR SPATIAL DATA

Revised Edition

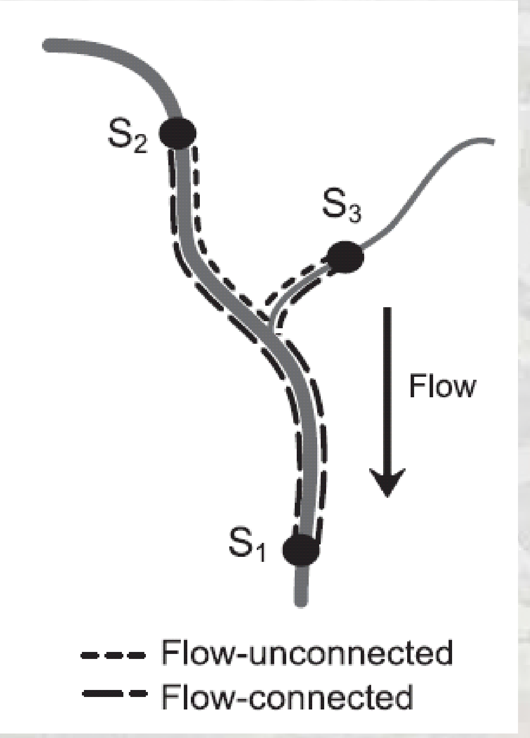
1993

Covariance Among Sites Depends on Flow Connectivity





Covariance Depends on Flow Magnitude



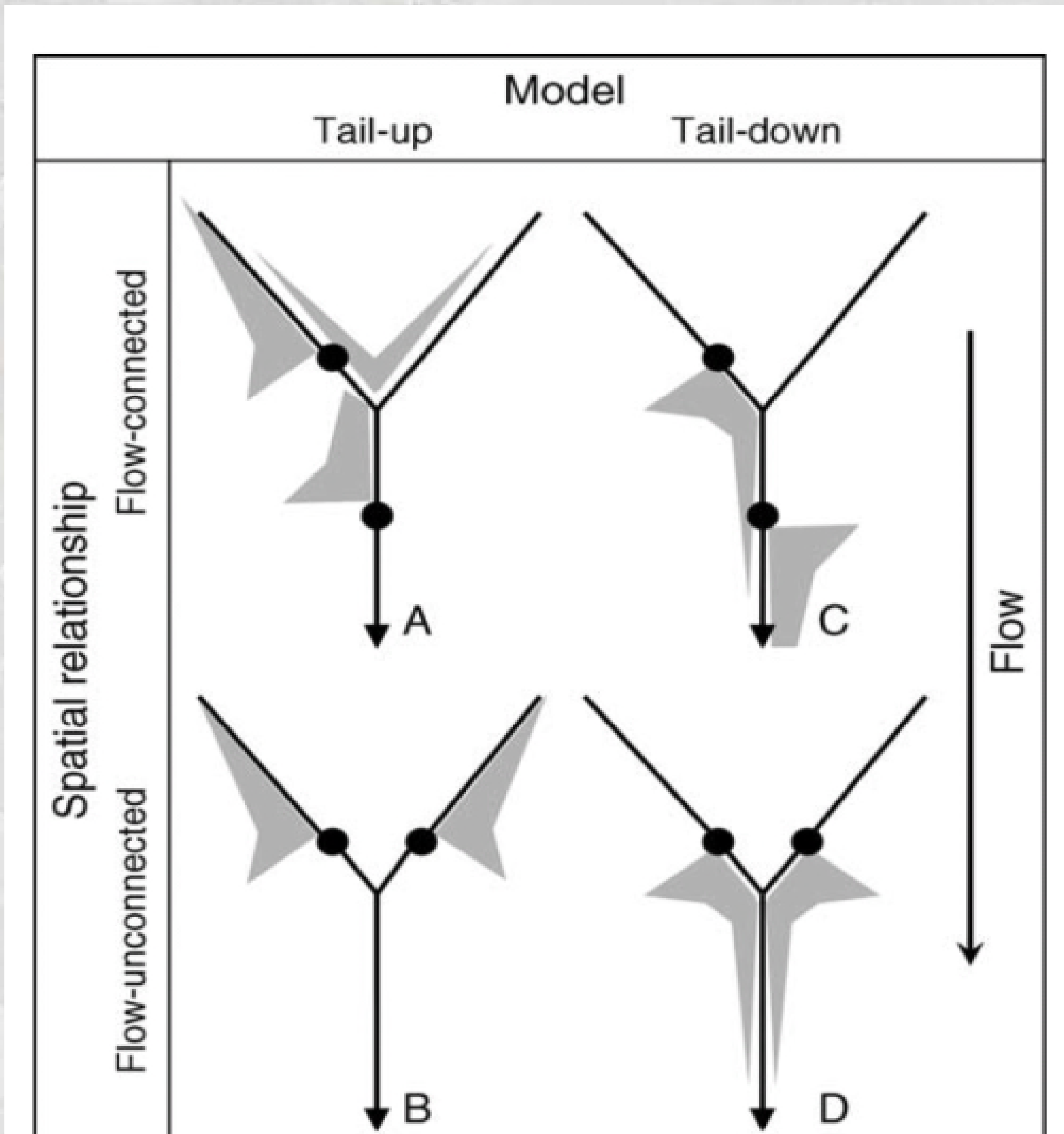
Show equal streams vs

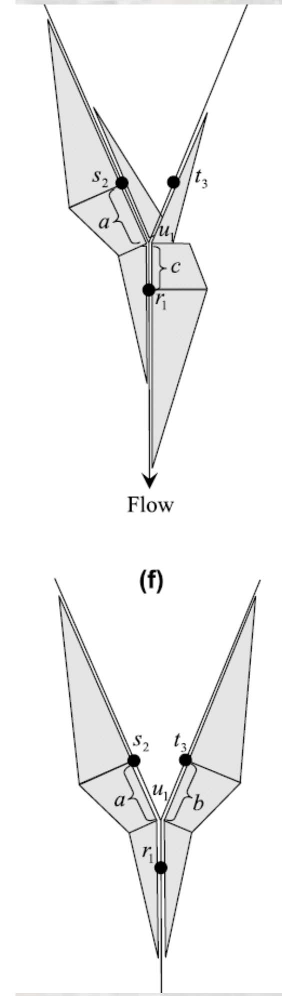
Big little stream confluence

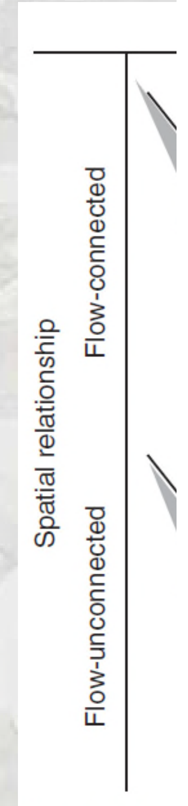
Mixed Model Covariance Structure

& Euclid
Terre:

TU/TD/Euclid





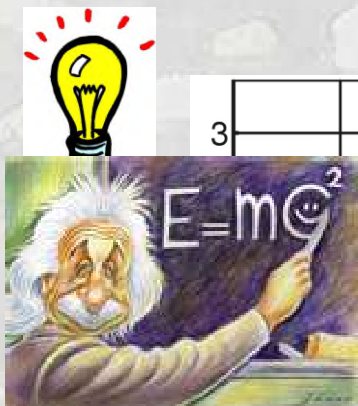


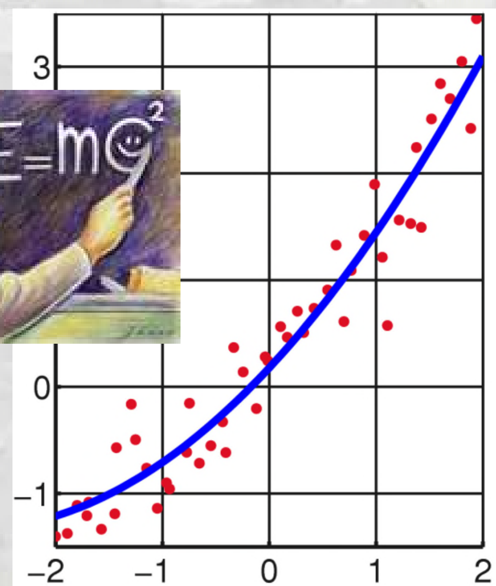
GURE 3 | Fl
patial relations
ving-average fu
tionships are s

cal Association **105:**6-18.

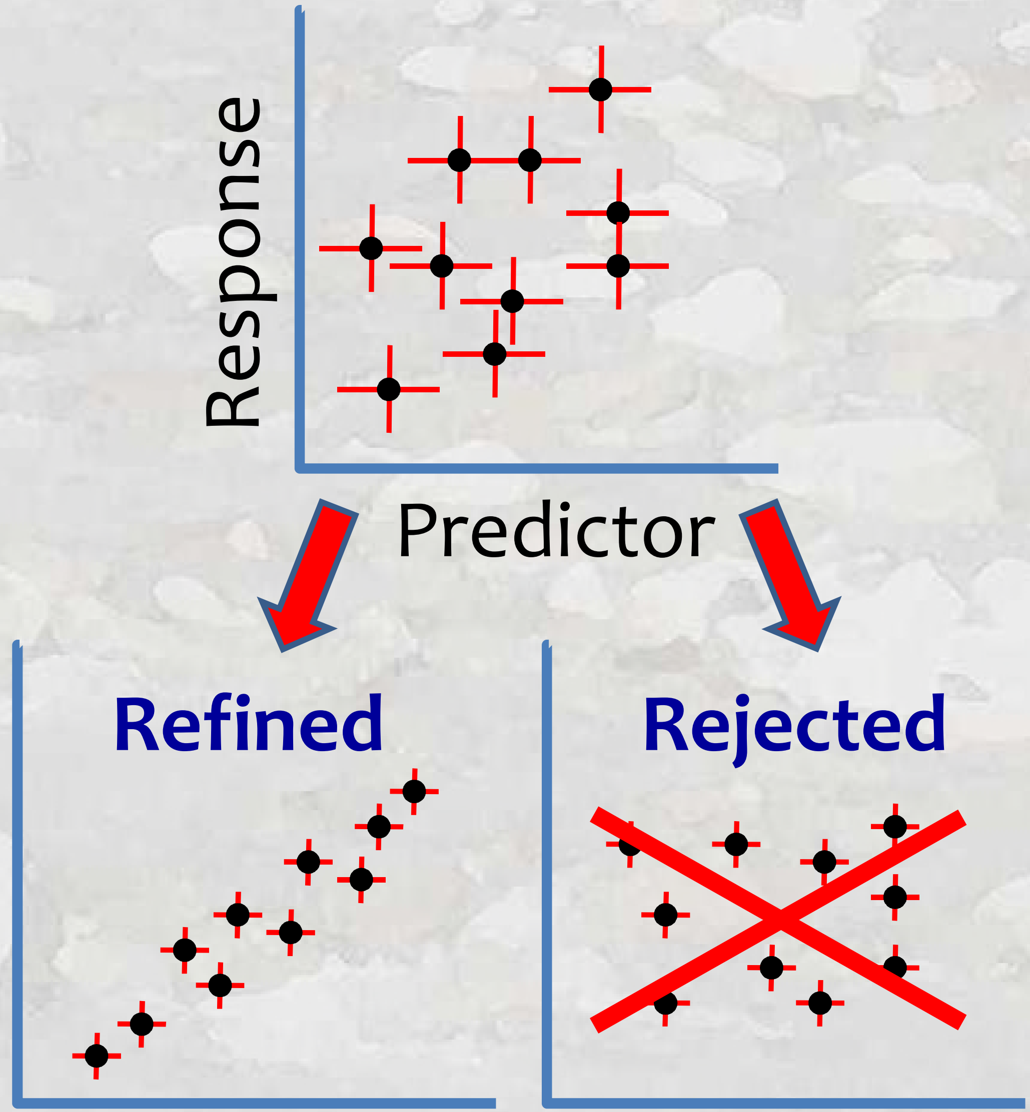
A New Era of Better Prediction & Understanding for Stream Things…

New relationships described

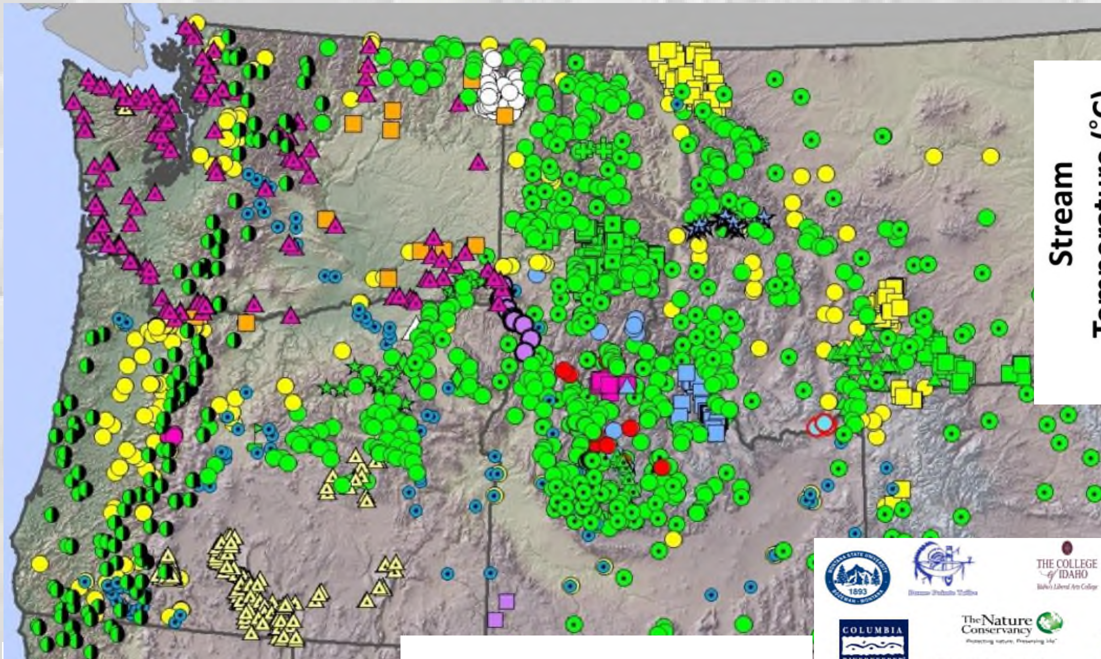


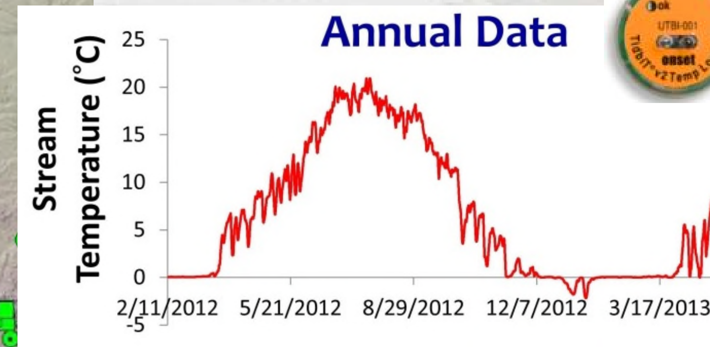


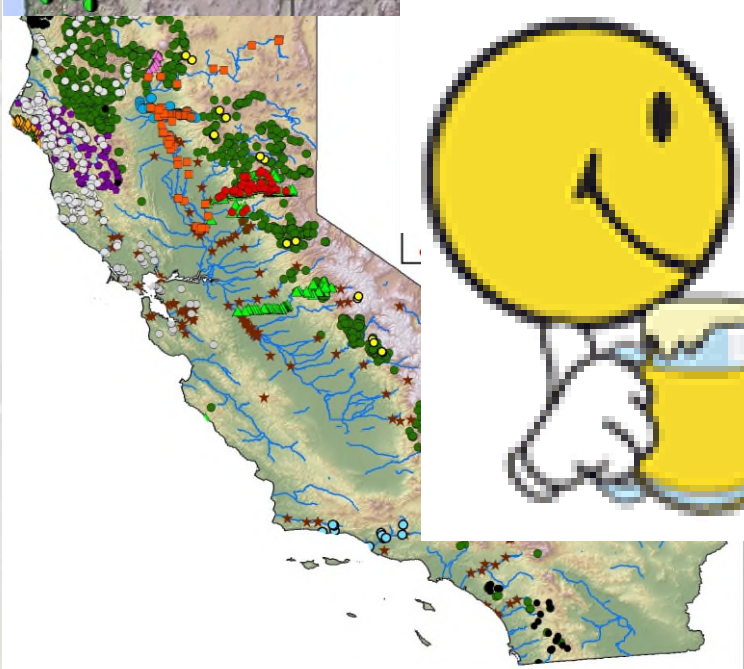
Old relationships tested



Efficient Inter-Agency Coordination

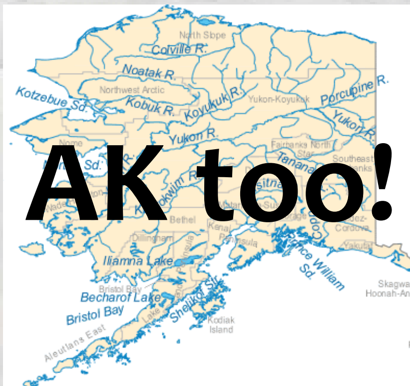




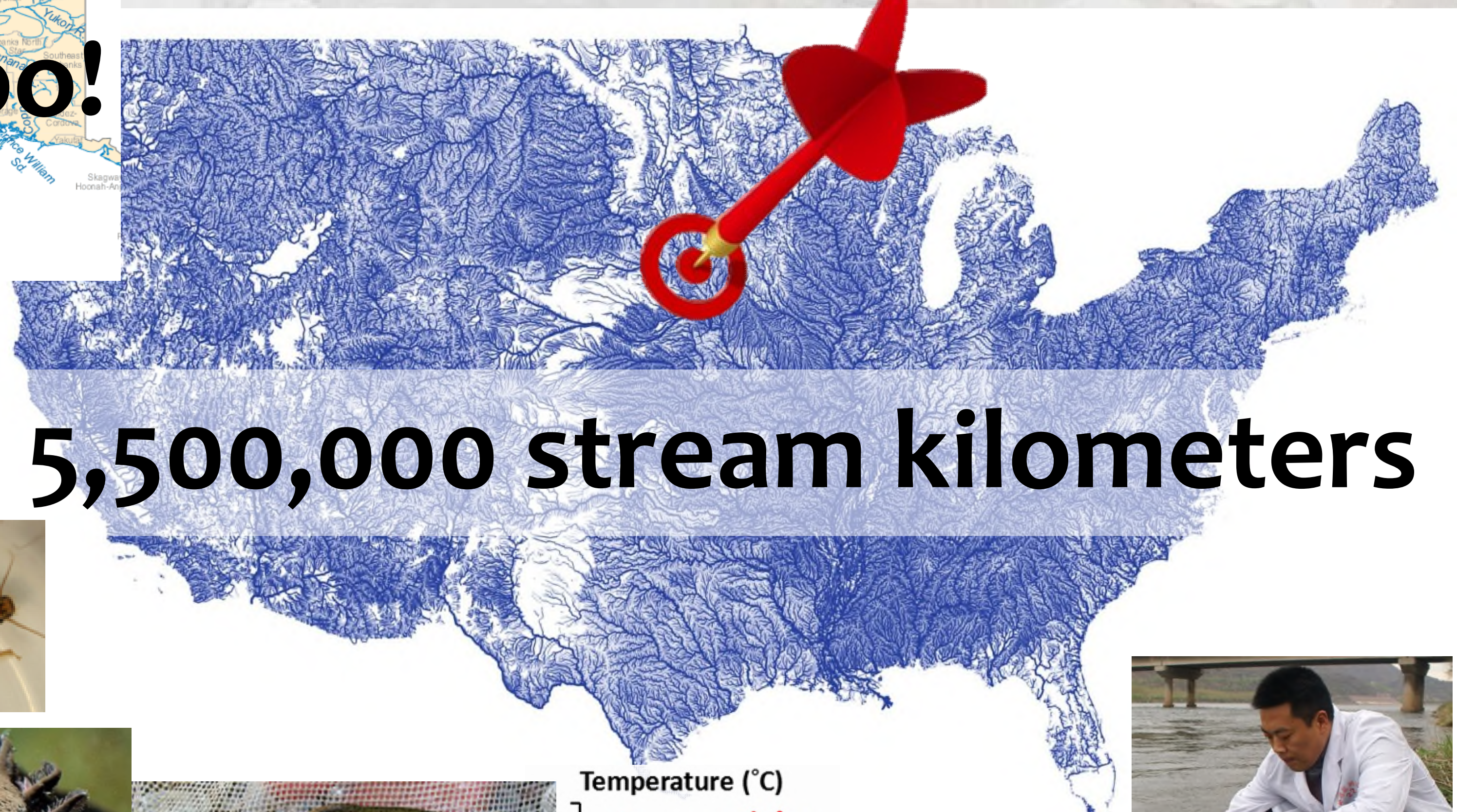


Vision: TSI (Total Stream Information) through the NSI

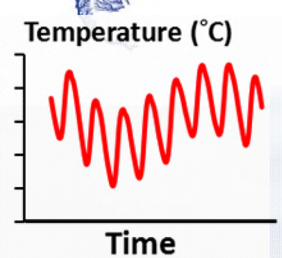
High-resolution information for all stream things in all reaches across the country

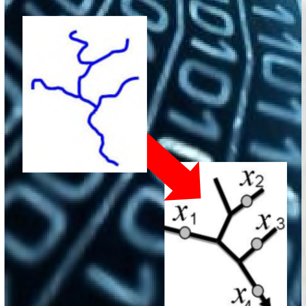


AK too!



5,500,000 stream kilometers





Spatial Statistical Network Models (SSNM) for Data on Stream Networks: Background, Theory, and Applications

Daniel J. Isaak[1], Erin Peterson[2], Jay Ver Hoef[3],

[1]U.S. Forest Service, Rocky Mountain Research Station, Boise, ID 83702; [2]CSIRO Division of Mathematics, Informatics, and Statistics, Dutton Park, Queensland, Australia; [3]NOAA National Marine Mammal Laboratory, Fairbanks, Alaska

Most statistical techniques traditionally applied to data measured on stream networks were developed for terrestrial applications and are not optimized for streams. The recent development of spatial statistical network models (SSNM) with covariance structures based on network topology opens a wealth of possibilities to improve the prediction and understanding of many stream phenomenon. Moreover, because SSNMs account for spatial autocorrelation (i.e., non-independence) among measurements, they can be applied to common datasets (e.g., water chemistry, habitat conditions, biological attributes) aggregated from multiple sources to perform powerful data-mining exercises and generate huge amounts of information at low cost. SSNMs are geostatistical models, so can also be used for kriging and block-kriging to make statistically valid predictions throughout river networks or subdomains within networks to aid in a variety of applications related to status assessments, enumeration, and the design of efficient monitoring strategies. Here, we provide a brief history of SSNM development and example applications. Free software for SSNM analysis is available at the SSN/STARS website (http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml) and the National Stream Internet website hosts topologically corrected digital stream networks to

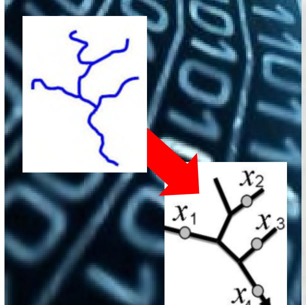
Tools for Information Creation

Better information =

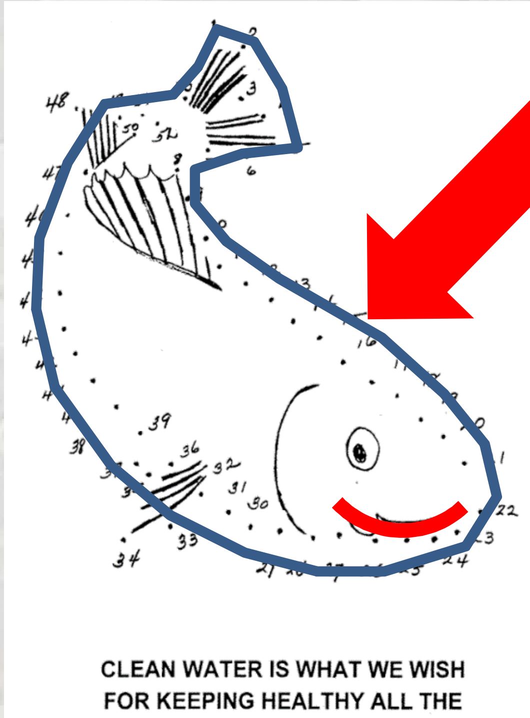
Better Conservation & Management

^

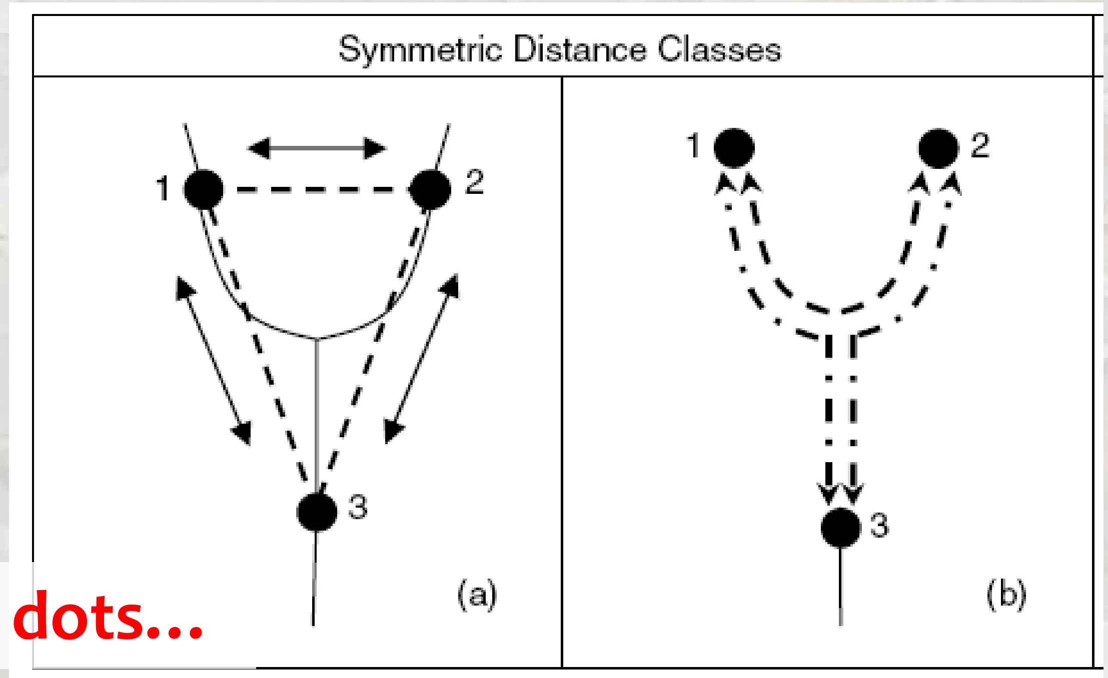
stream



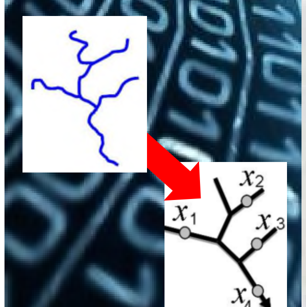
Predictive Models with Covariates



Valid Interpolation on Networks



Let's us connect the dots...



Advantages:

-flexible & valid covariance structures
by accommodating network topology

-weighting by stream size

-improved predictive ability & parameter
estimates relative to non spatial models