Spatial Statistical Models on Stream Networks

Jay Ver Hoef[1], Erin Peterson[2], Dan Isaak[3]

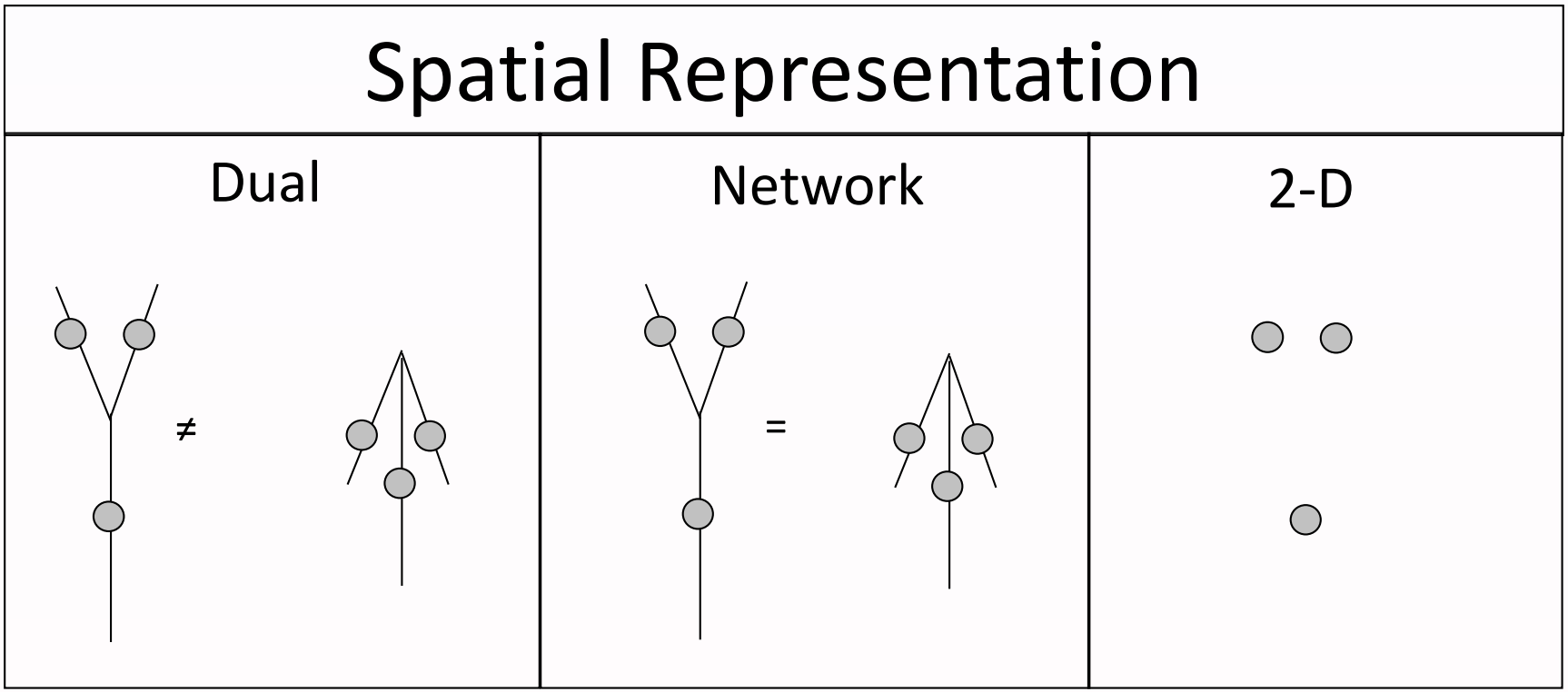
[1]NOAA National Marine Mammal Lab, Seattle, USA
[2]CSIRO Mathematics, Informatics, and Statistics, Brisbane, Australia
[3]USDA Forest Service, Boise, USA

Concepts

NCEAS Working Group

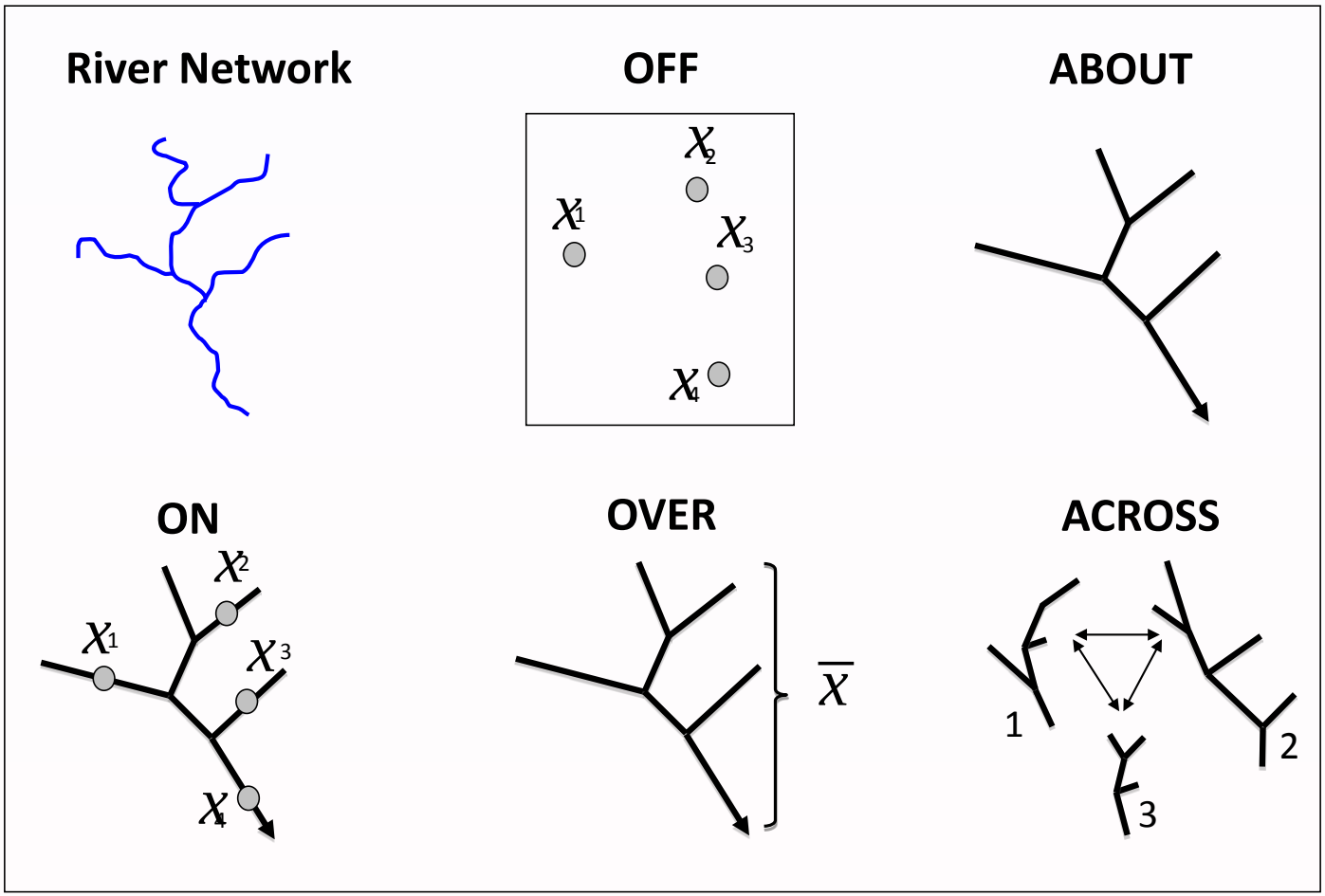
Concepts

Dual Spatial Coordinate System



Concepts

Classification of Analytical Methods



Concepts

Spatial Linear Model

$$\begin{pmatrix} \mathbf{z}_{observed} \\ \mathbf{z}_{unobserved} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$$

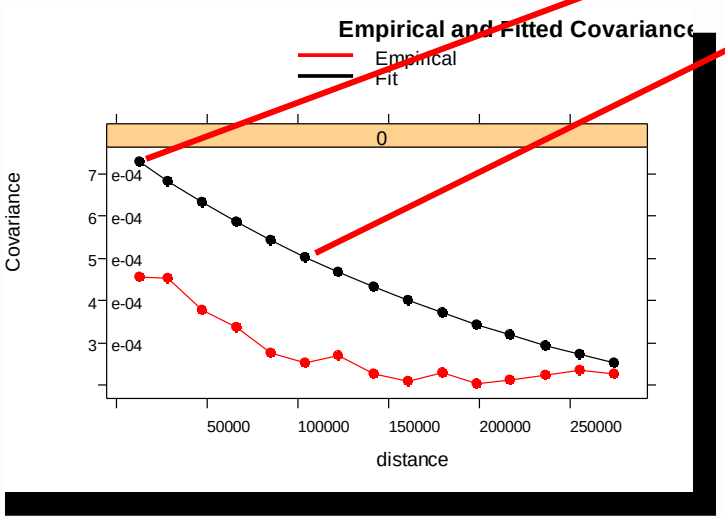
Prediction

Estimation

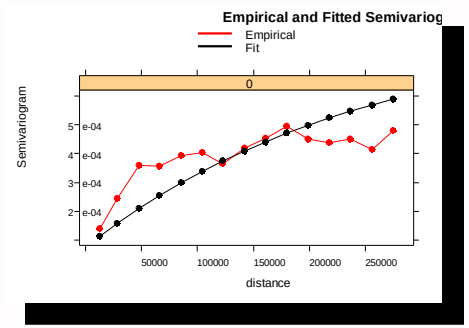
Concepts

Stream Network Covariance

$$\begin{pmatrix} \mathbf{z}_{observed} \\ \mathbf{z}_{unobserved} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$$

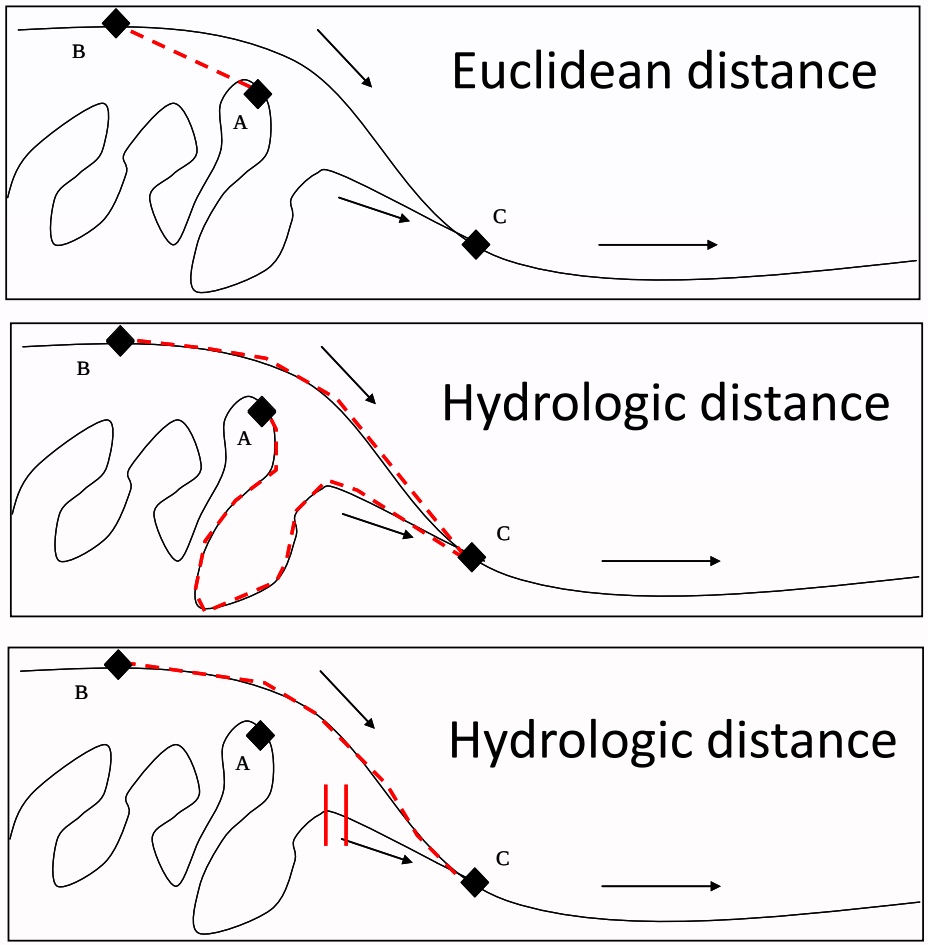


$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{21} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$



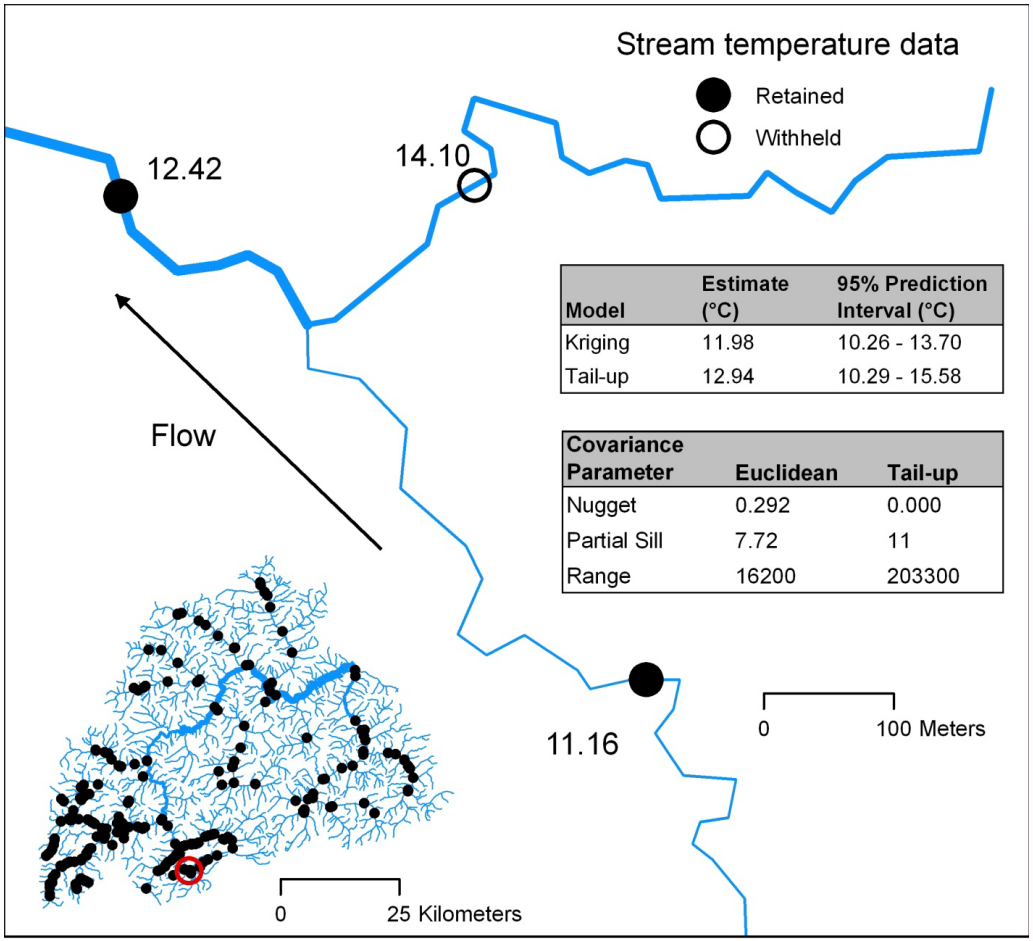
Concepts

Distance and Flow



Why Bother?

Better Predictions

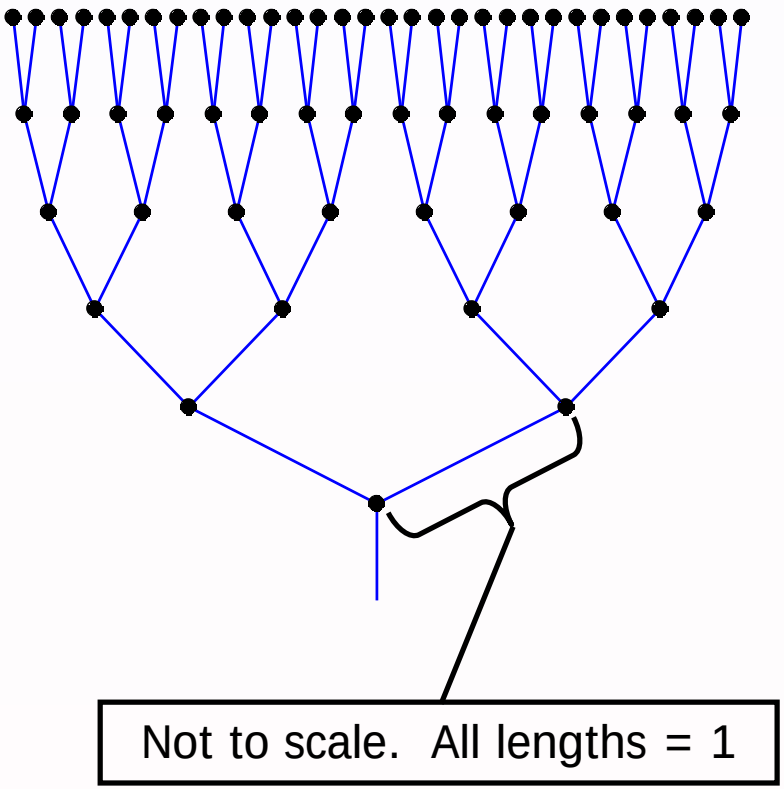


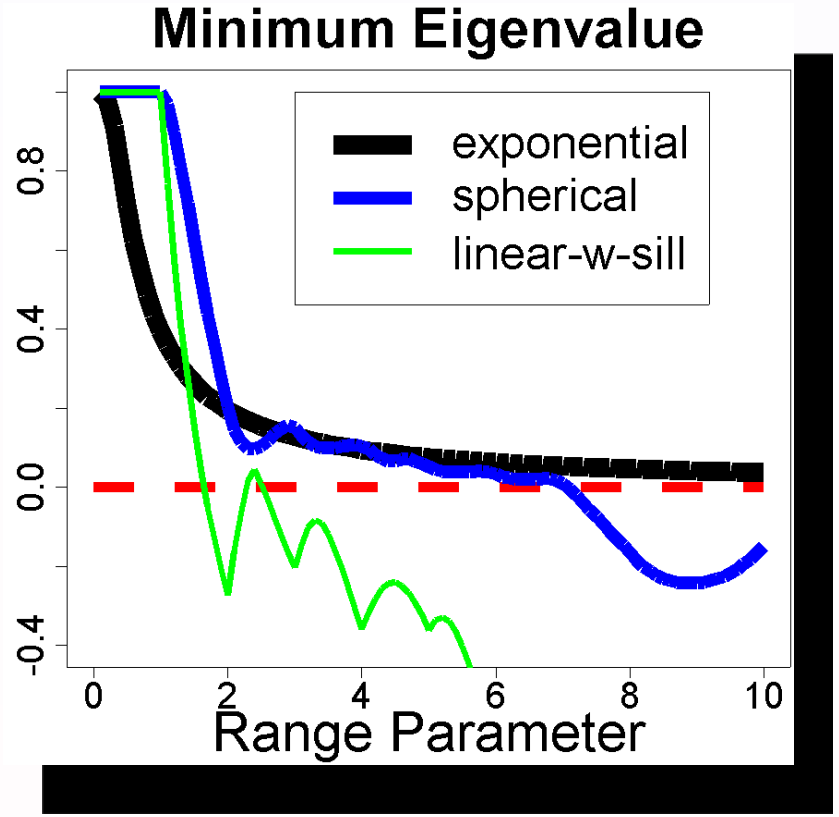
Model	Estimate (°C)	95% Prediction Interval (°C)
Kriging	11.98	10.26 - 13.70
Tail-up	12.94	10.29 - 15.58

Covariance Parameter	Euclidean	Tail-up
Nugget	0.292	0.000
Partial Sill	7.72	11
Range	16200	203300

Problem:

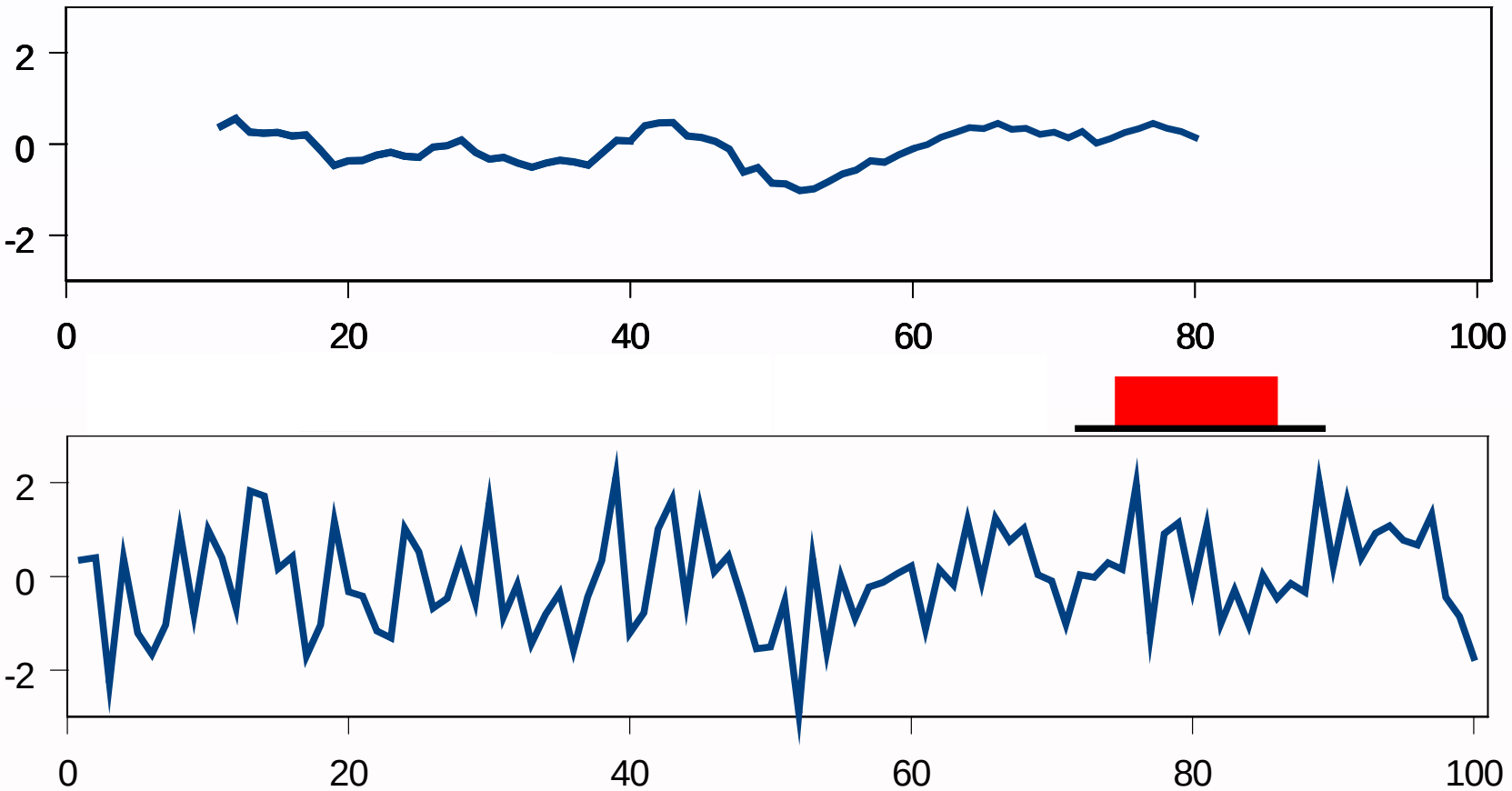
Valid Covariance Matrices Using Stream Distance





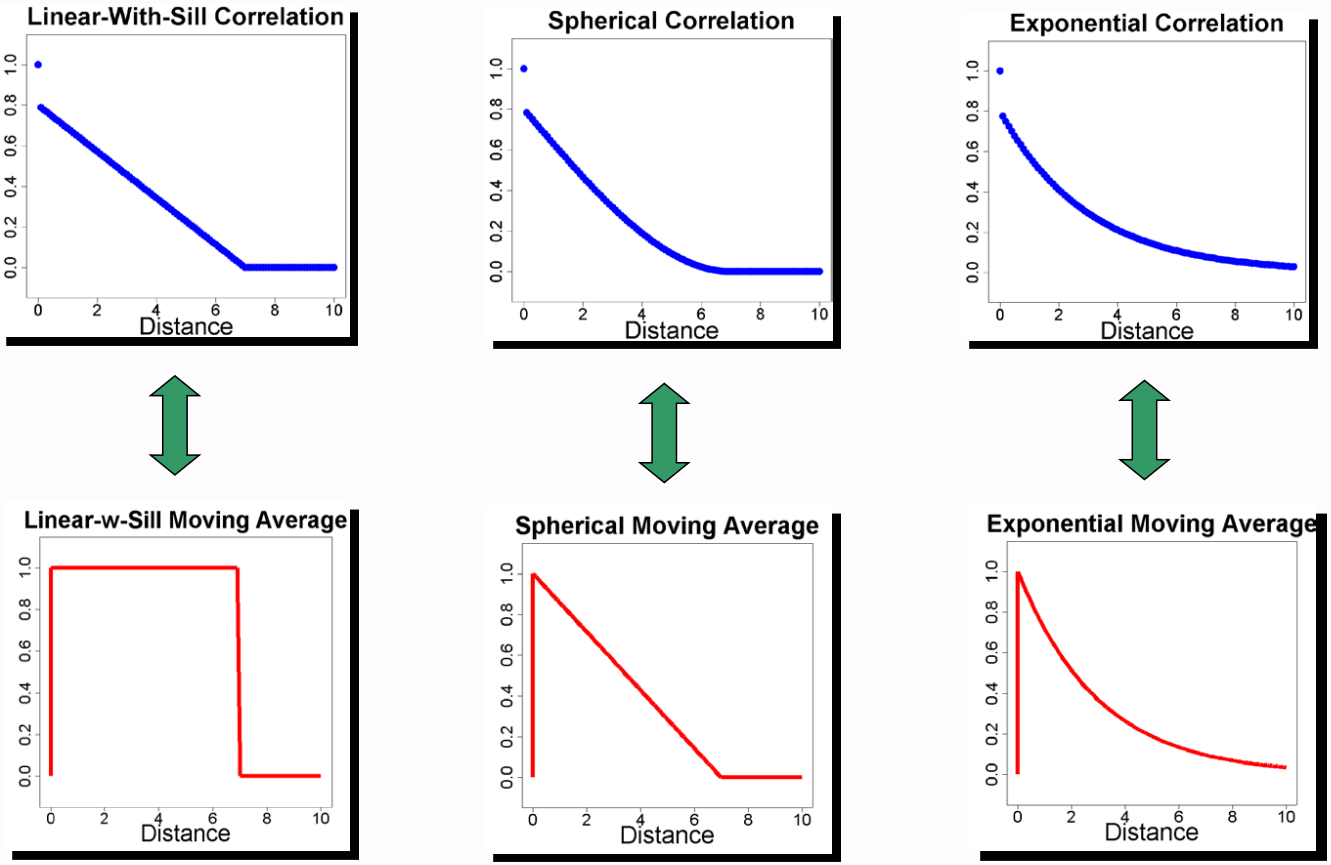
Models

Time Series Moving Average Construction



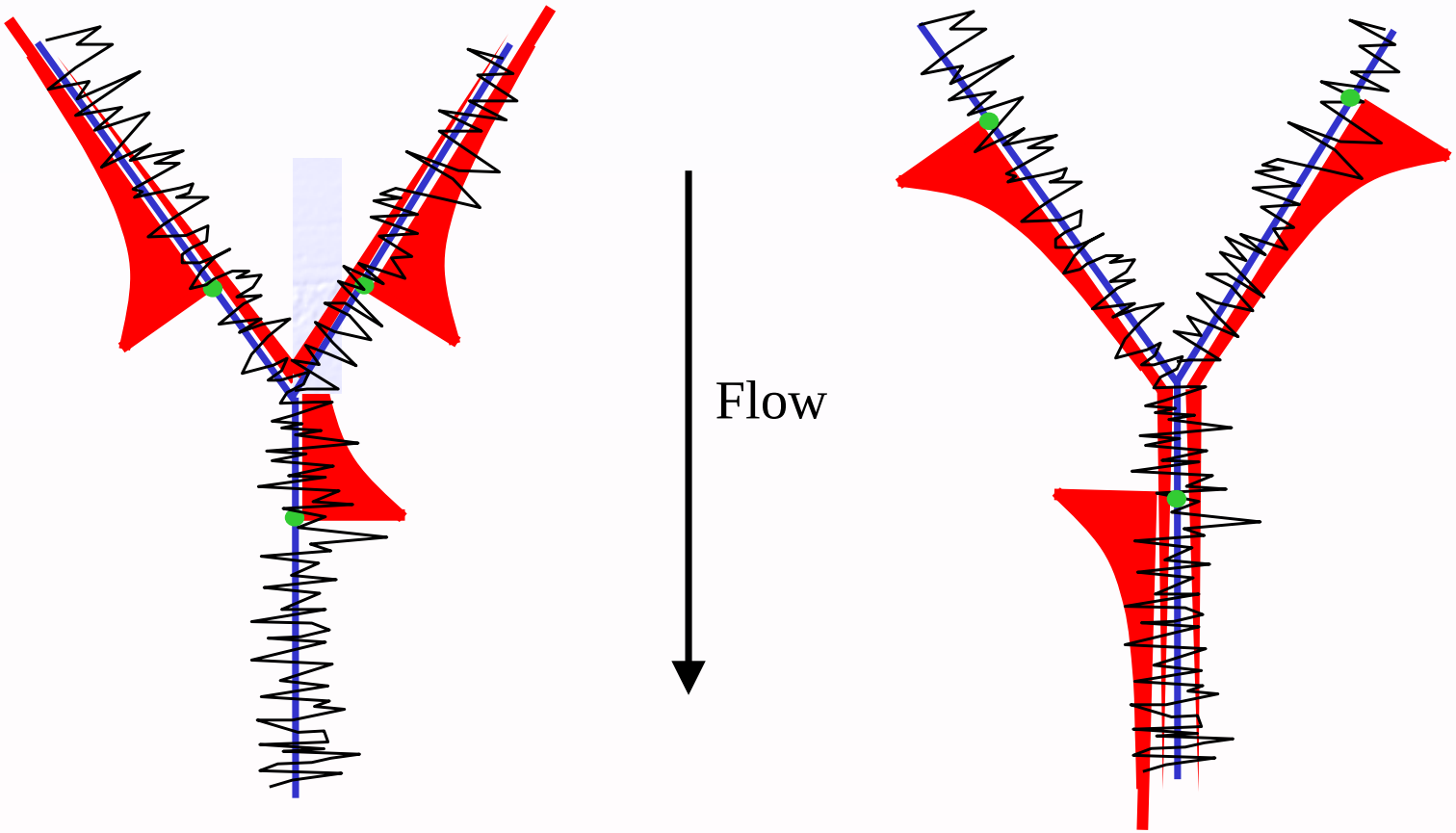
Models

Time Series Moving Average Construction



Models

Stream Network Moving Average Construction



Models

Tail-up Models

$$C_u(r_i, s_j|\boldsymbol{\theta}_u) = \begin{cases} \pi_{i,j} C_t(h|\boldsymbol{\theta}_u) & \text{if } r_i \text{ and } s_j \text{ are flow-connected,} \\ 0 & \text{if } r_i \text{ and } s_j \text{ are flow-unconnected,} \end{cases}$$

- Tail-Up Linear-with-Sill Model

$$C_t(h|\boldsymbol{\theta}_u) = \sigma_u^2 \left(1 - \frac{h}{\alpha_u}\right) I\left(\frac{h}{\alpha_u} \leq 1\right),$$

- Tail-Up Exponential Model

$$C_t(h|\boldsymbol{\theta}_u) = \sigma_u^2 \exp(-3h/\alpha_u),$$

Models

Tail-down Models

- Tail-Down Linear-with-Sill Model, $b \geq a \geq 0$,

$$C_d(a,b,h|\boldsymbol{\theta}_d) = \begin{cases} \sigma_d^2 \left(1 - \frac{h}{\alpha_d}\right) I\left(\frac{h}{\alpha_d} \leq 1\right) & \text{if flow-connected,} \\ \sigma_d^2 \left(1 - \frac{b}{\alpha_d}\right) I\left(\frac{b}{\alpha_d} \leq 1\right) & \text{if flow-unconnected,} \end{cases}$$

- Tail-down Exponential Model,

$$C_d(a,b,h|\boldsymbol{\theta}_d) = \begin{cases} \sigma_d^2 \exp(-3h/\alpha_d) & \text{if flow-connected,} \\ \sigma_d^2 \exp(-3(a+b)/\alpha_d) & \text{if flow-unconnected,} \end{cases}$$

Models

Mixed Models

- Mixed Model (Variance Component Model),

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}_u + \mathbf{z}_d + \mathbf{z}_e + \mathbf{W}_1\boldsymbol{\gamma}_1 + \ldots + \mathbf{W}_p\boldsymbol{\gamma}_p + \boldsymbol{\epsilon},$$

- Tail-down Exponential Model,

$$\text{var}(\mathbf{Y}) = \boldsymbol{\Sigma} = \sigma_u^2\mathbf{R}(\alpha_u) + \sigma_d^2\mathbf{R}(\alpha_d) + \sigma_e^2\mathbf{R}(\alpha_e) +$$

$$\sigma_1^2\mathbf{W}_1\mathbf{W}_1' + \ldots + \sigma_p^2\mathbf{W}_p\mathbf{W}_p' + \sigma_0^2\mathbf{I}.$$

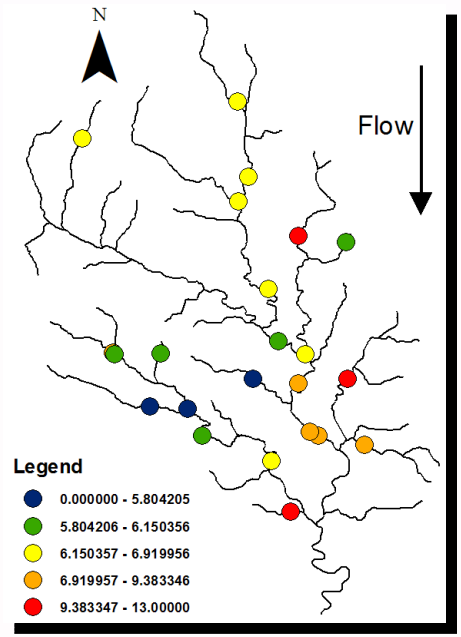
Models

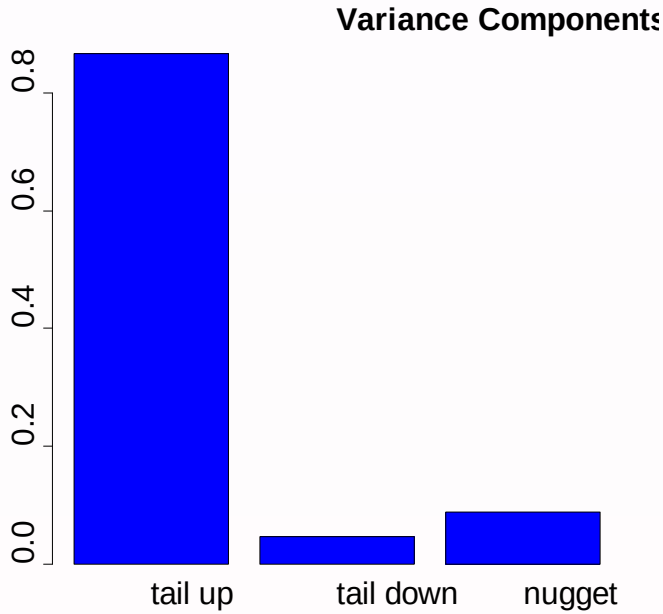
Mixed Models

Estimation

$$\begin{pmatrix} \mathbf{z}_{observed} \\ \mathbf{z}_{unobserved} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$$

SO_4 Data





Models

Estimation

SO4, Tail-up Model Only

effect	estimate	std.err	df	t.value	prob.t
Intercept	9.07	0.984	21	9.22	<0.0001
Distance Upstream	-0.1337	0.0537	21	-2.488	0.0213

SO4, Components Model

effect	estimate	std.err	df	t.value	prob.t
Intercept	9.20	1.399	21	6.58	<0.0001
Distance Upstream	-0.1259	0.0664	21	-1.893	0.0722

Estimation

$$\begin{pmatrix} \mathbf{z}_{observed} \\ \mathbf{z}_{unobserved} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$$

Models

Prediction

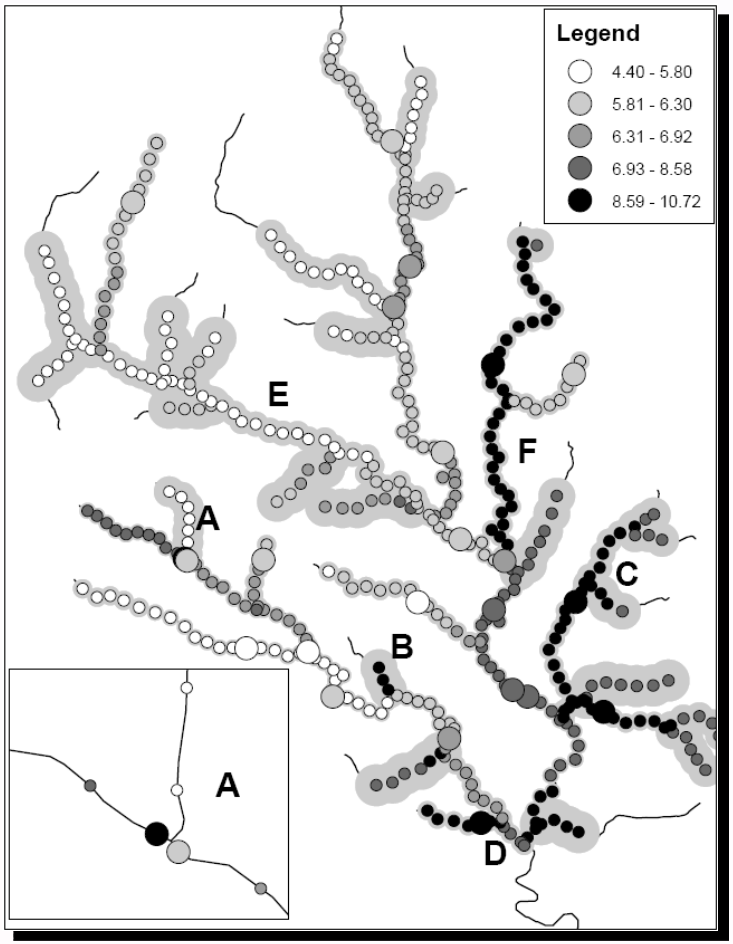
$$\begin{pmatrix} \mathbf{z}_{observed} \\ \mathbf{z}_{unobserved} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$$

Prediction

Block Kriging

E = 5.82 se =1.13

F = 9.67 se = 0.67



Software / Tools - Collaborators

Jay Ver Hoef, Statistician

NOAA National Marine Mammal Laboratory

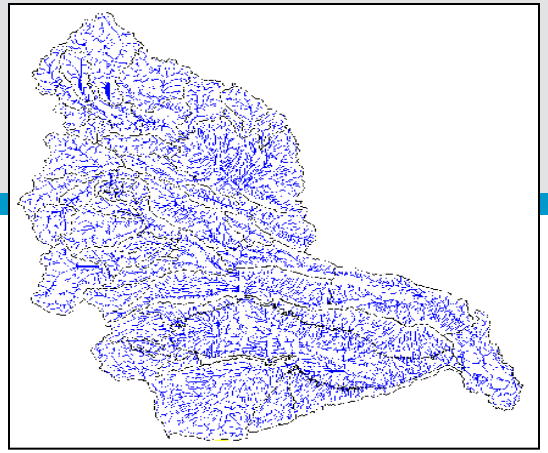
David Clifford, Statistician

CSIRO Mathematics, Informatics & Statistics

Rohan Shah, Mathematician & Programmer

CSIRO Mathematics, Informatics & Statistics

Software / Tools



- Multidisciplinary skills are required
  - Knowledge of aquatic ecology
  - Specialized skills using geographic information systems (GIS)
  - Spatial statistics

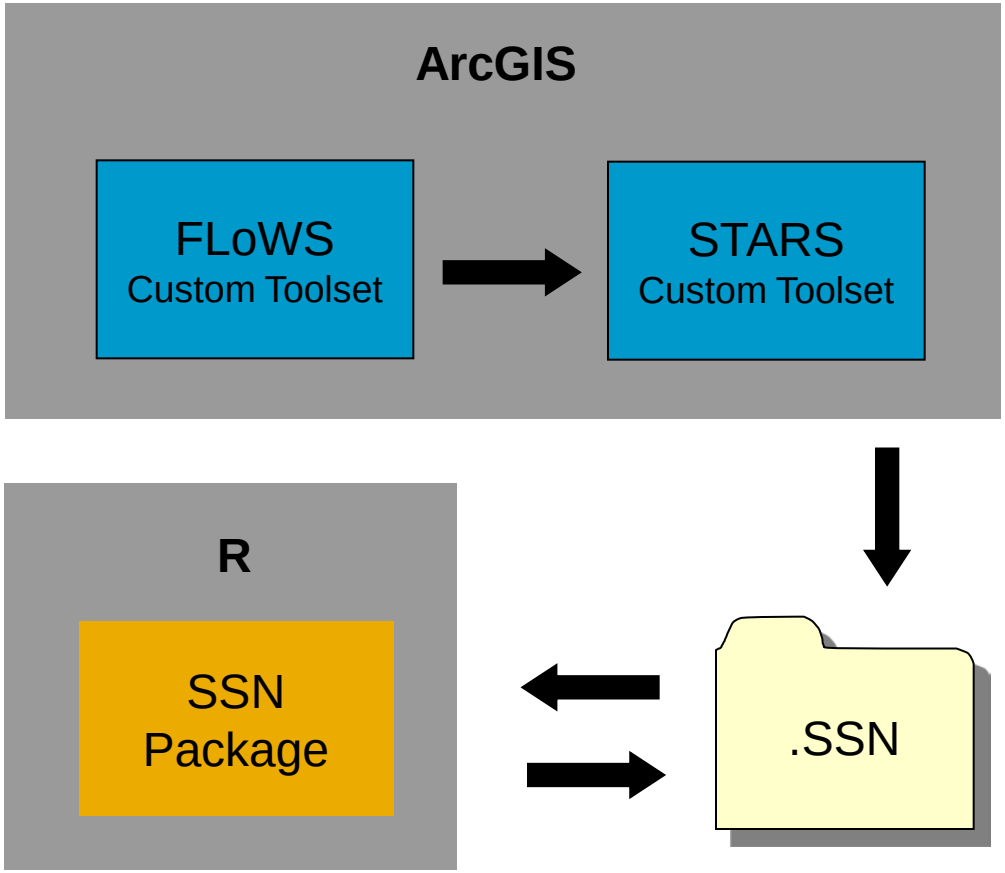
http://ddsgeo.com

- Knowledge transfer & methodological uptake requires software/tools
  - Journal articles are not enough

$$\begin{pmatrix} \mathbf{z}_{observed} \\ \mathbf{z}_{unobserved} \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \mathrm{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$$

Software / Tools

Suite of GIS and Statistical Tools



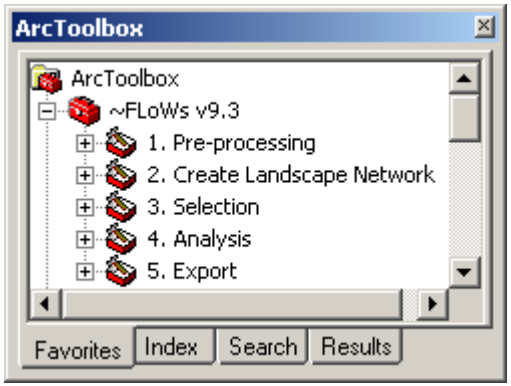
Software / Tools – FLoWS

Functional Linkage of Watersheds and Streams (FLoWS)

- ArcGIS Geoprocessing Toolbox written in Python for ArcGIS v9.3
- Developed by Dave Theobald and John Norman at Colorado State University
- Website: http://www.nrel.colostate.edu/projects/starmap/flows_index.htm

What are the FLoWS v9.3 tools?

- Graph theoretic-based analysis tools
- Functionally link aquatic and terrestrial components of a landscape based on hydrologic processes
- Hydrologic modelling framework

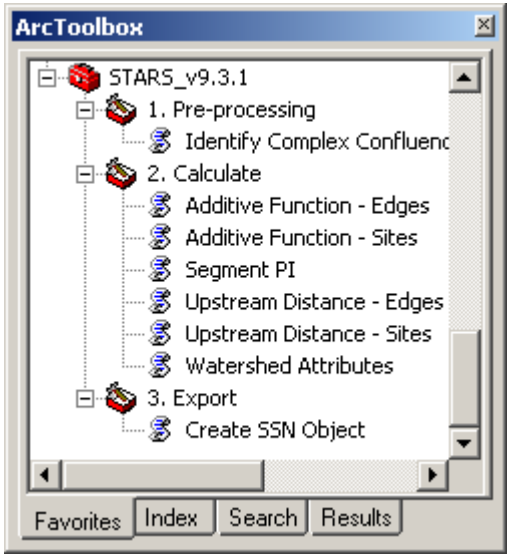


Software / Tools – STARS

Spatial Tools for the Analysis of Streams (STARS)

- ArcGIS Geoprocessing Toolbox written in Python for ArcGIS v9.3.1

Toolsets



- Pre-processing: Identify unique topological relationships that are prohibited in the SpatialStreamNetwork class

- Calculate: Derive spatial data needed to fit a spatial statistical model to streams data:
  - Hydrologic distances
  - Spatial weights
  - Covariates for both observed & unobserved locations

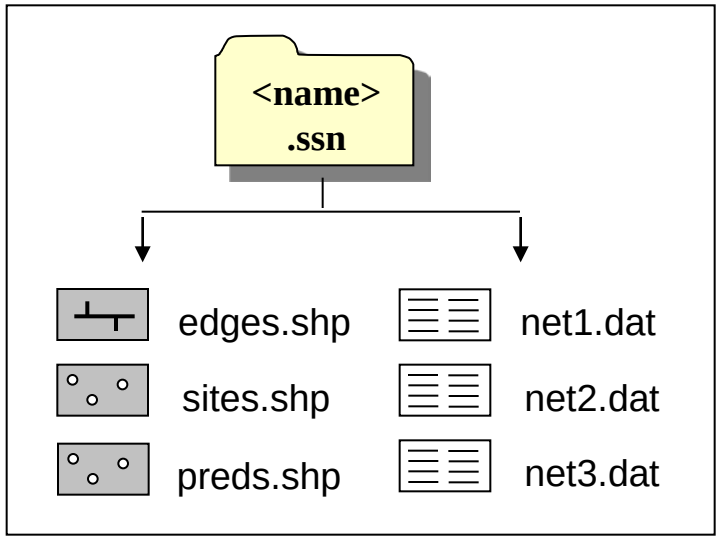
- Export: Export the spatial features, topological relationships, and attribute information to a format that can be easily accessed using R

Software / Tools - STARS

Export – Create .ssn Directory

Stores the spatial features, topological relationships, and attribute information to a format that can be easily accessed using R

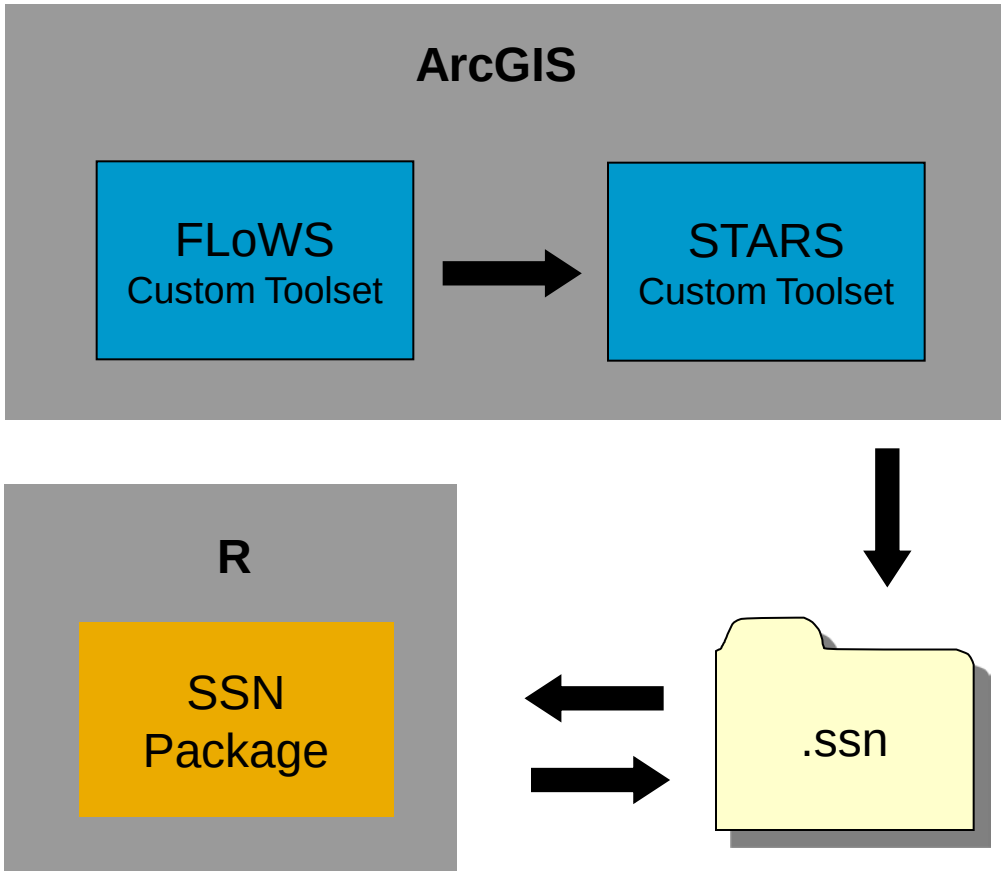
- .SSN directory components:
  - Sites & edges shapefiles
  - Text file containing topological information for each network
  - Prediction site shapefile(s): optional



- Easily transferable
  - If it's in the .ssn directory, you need it

Note, shapefiles will always have at least three files associated with them (.shp, .shx, .dbf) and may have additional files (.sbn, .sbx, .prj, etc.)

Software / Tools



Software / Tools – SSN Package

Spatial Stream Networks (SSN) Package

Purpose

- Read, store, visualize, analyse & export spatial streams data

Our design goals

- Make the package user friendly
  - Easy to install on multiple operating systems
  - Fit models to multiple response types
  - Include helper functions so that you don't have to be an R developer or a statistician to fit a model
  - Reduce opportunities for user error

Software / Tools – SSN Package

SpatialStreamNetwork Class: S4 object

- Formally defined class structure
  - Ensures data is in the “right” format
- Conventions for spatial classes set out by Bivand et al. 2008
  - Co-authored sp package to extend R classes and methods for spatial data
  - 49 other packages depend on sp
- SpatialStreamNetwork class
  - Contains both point and line features: edges, sites, prediction sites
  - Unique coordinate system used to navigate network

Bivand R.S., Pebesma E.J., and V. Gomez-Rubio (2008) Applied Spatial Data Analysis with R, UseR! Series, Springer, 378 p.

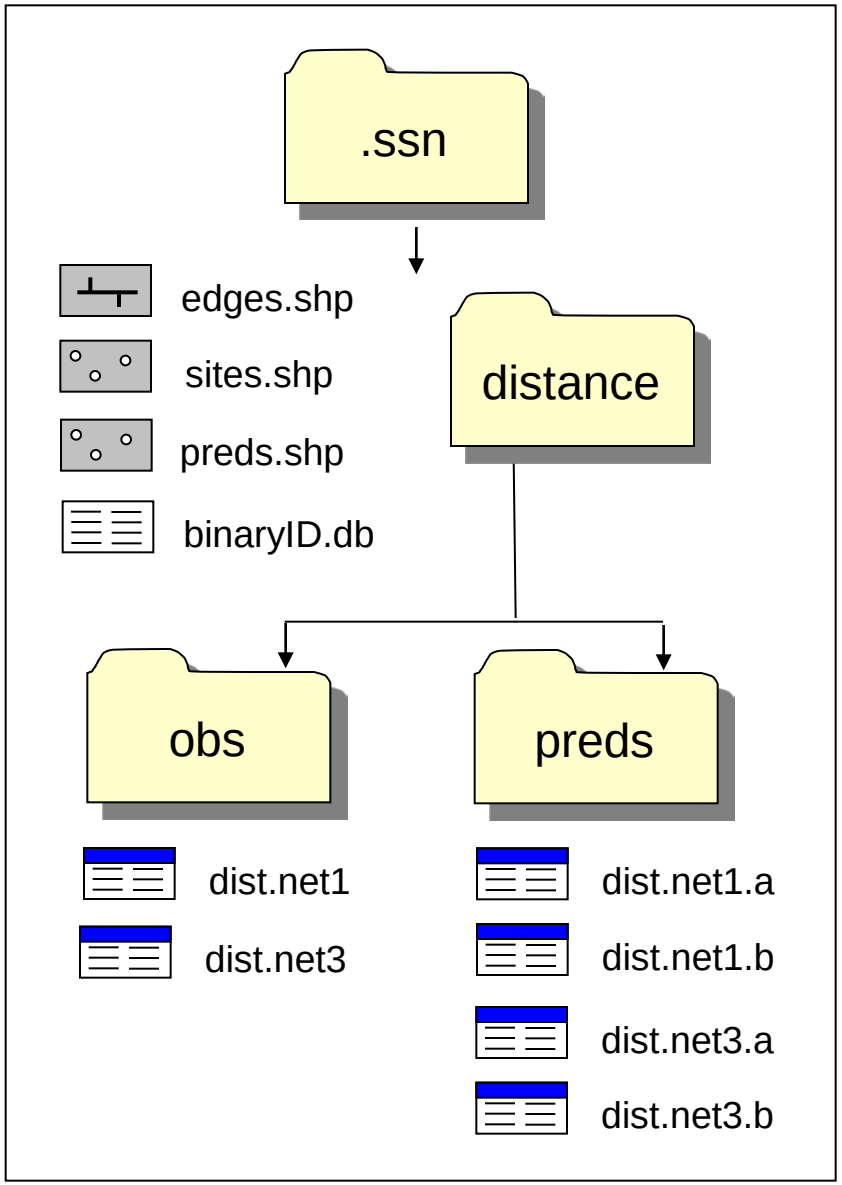
Software / Tools – SSN Package

.ssn directory

- Explicitly linked to SpatialStreamNetwork object
- R reads/writes data to this location

- binaryID.db: SQLite database
  - Stores some topological information for line segments
  - Automatically created when data are imported into R

- Distance matrices
  - Stored as .Rdata files
  - External storage means they only have to be calculated once



Software / Tools – SSN Package

SSN Package Functionality…

- Fit generalised linear models with spatially-autocorrelated errors
  - Normal likelihood methods (including REML)
  - Quasi-likelihood for Poisson and Binomial families
- Supported distributions: Gaussian, binomial, Poisson
  - Repeated measurements are permitted
  - Random effects may be included
- Block-kriging estimates
- Multiple covariance models supported:
  - Tail-up & tail-down: Exponential, Linear-with-Sill, Spherical, Mariah
  - Euclidean: Exponential, Spherical, Linear-with-Sill, Cauchy

Software / Tools – SSN Package

SSN Package Functionality…

- Influence & cross-validation diagnostics
  - Raw residuals, studentized residuals, cross-validation residuals, Leverage values, Cooks D, cross-validation predictions, and standard errors for the cross validation predictions
- Predictions & std errors at unobserved locations
  - Predicted response variable & standard errors for the predictions
- Create SSN objects
  - Generate network(s) with observations & prediction locations
  - Point distribution based on different sample designs: Poisson, hard core, systematic
- Simulate data
  - Gaussian, binomial, Poisson distribution
  - Different covariance structures & fixed effects

Software / Tools – SSN Package

Helper Functions for 4 Classes

Allow users to examine data, produce model diagnostics, visualize data, and export results easily

Model Diagnostics

- AIC, generalized R^2 for covariates, cross-validation statistics, variance components

Visualisation

- Plot: map of observed values, residuals, or predictions and/or standard errors
- Torgegram: flow-connected & flow-unconnected
- QQplots, histograms, boxplot

Others:

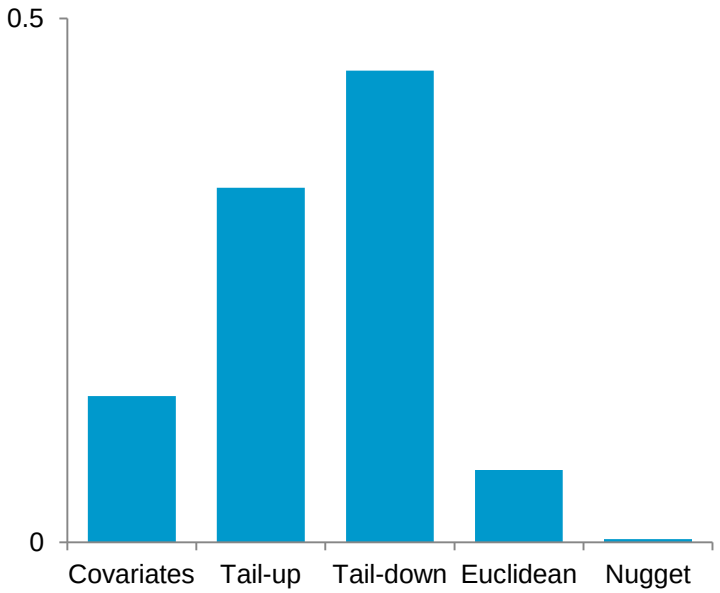
Access, alter, export, subset, summary…

Software / Tools – SSN Package

Model Fitting

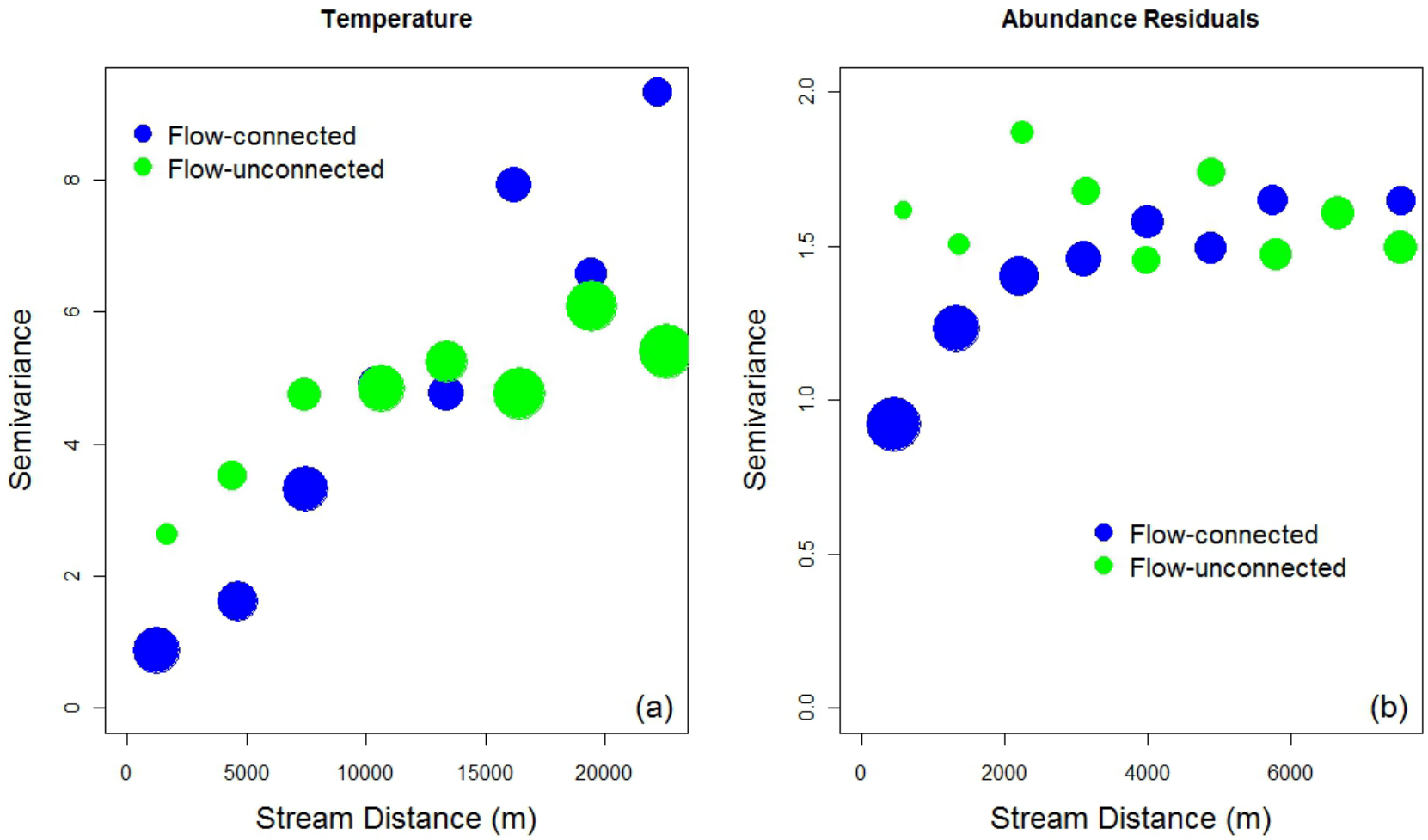
```
> mf04.glmssn0c <- glmssn(Summer_mn ~ ELEV_DEM + SLOPE, mf04c,
+     CorModels = c("Exponential.tailup", "Exponential.taildown",
+     "Exponential.Euclid"),
+     addfunccol = "afvArea")
>
> summary(mf04.glmssn0c)
$fixed.effects.estimates
  FactorLevel      Estimate      std.err   t.value  prob.t
1 (Intercept)  29.348881916  5.540699171  5.296964 0.00000
2    ELEV_DEM  -0.009970748  0.002758303 -3.614812 0.00051
3       SLOPE -12.812193742 12.340454558 -1.038227 0.30207

$covariance.parameter.estimates
      Covariance.Model Parameter     Estimate
1   Exponential.tailup   parsill 1.782467e+00
2   Exponential.tailup     range 7.845687e+05
3 Exponential.taildown   parsill 2.370570e+00
4 Exponential.taildown     range 1.951213e+05
5   Exponential.Euclid   parsill 3.625637e-01
6   Exponential.Euclid     range 1.631959e+05
7               Nugget   parsill 1.536211e-02
```

```
> varcomp(mf04.glmssn0c)
              VarComp  Proportion
1    Covariates (R-sq) 0.139588592
2   Exponential.tailup 0.338483306
3 Exponential.taildown 0.450161531
4   Exponential.Euclid 0.068849368
5               Nugget 0.002917203
```



Software / Tools – SSN Package

Torgegrams



Future Tool Development

Statistical tools

- Continue to improve the SSN package…

Update GIS tools

- Personal geodatabase is not supported in ArcGIS v10
- FLoWS toolset is only available up to ArcGIS v9.3

Tools *must* be more computationally efficient

- Sensor networks = massive datasets
- Model-fitting algorithms can't handle observations

CSIRO Accelerated Computing Projects

- Speed up matrix operations
    - Intel's MKL library (commercial software)
    - Multiple GPU-based calculations
    - Implement MAGMA multi-GPU qr() decomposition function in R

NCEAS Spatial Stats for Streams Working Group

Implicit goal

Make the tools and dataset available so that others can extend/apply these methods

http://www.nceas.ucsb.edu/featured/peterson

Accessing the Tools, Documentation & Data

SSN & STARS:
Tools for Spatial Statistical Modeling

RMRS

US FOREST SERVICE
Research and Development

US Forest Service website (currently under construction):

http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml

Tools:

- FLoWS, STARS, & SSN package

Documentation:

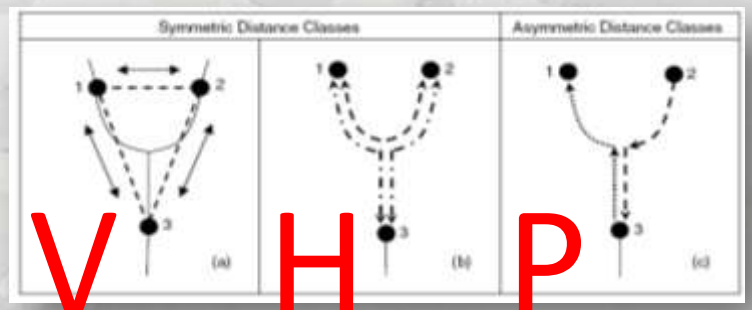
- Tutorials, journal articles, examples

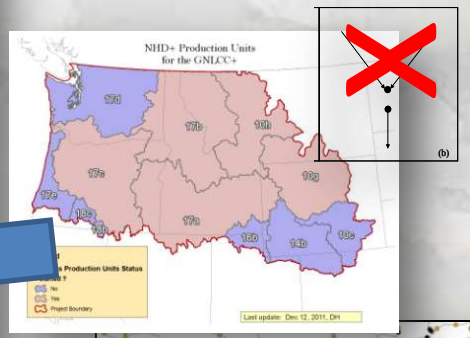
SSN & STARS:
Tools for Spatial Statistical Modeling on Stream Networks

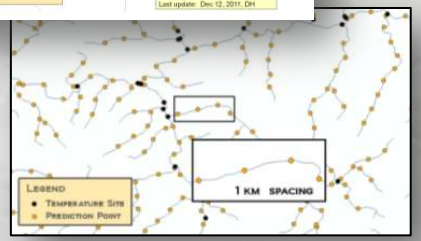
Datasets:

- Full Lower Snake dataset
- Smaller subsets of the Lower Snake
- Space-time, block kriging, macroinvertebrates

Regional Infrastructure for VHP Analysis of Stream Data



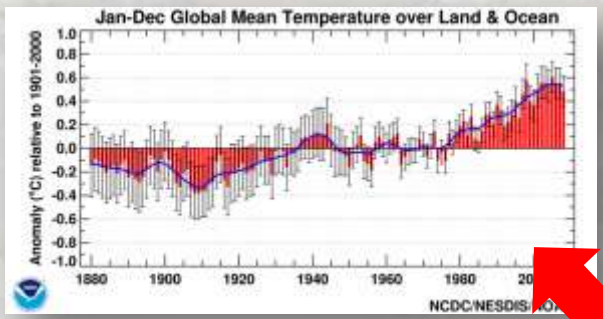


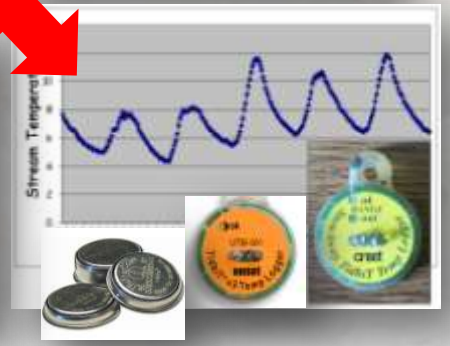


~350,000 stream kilometers

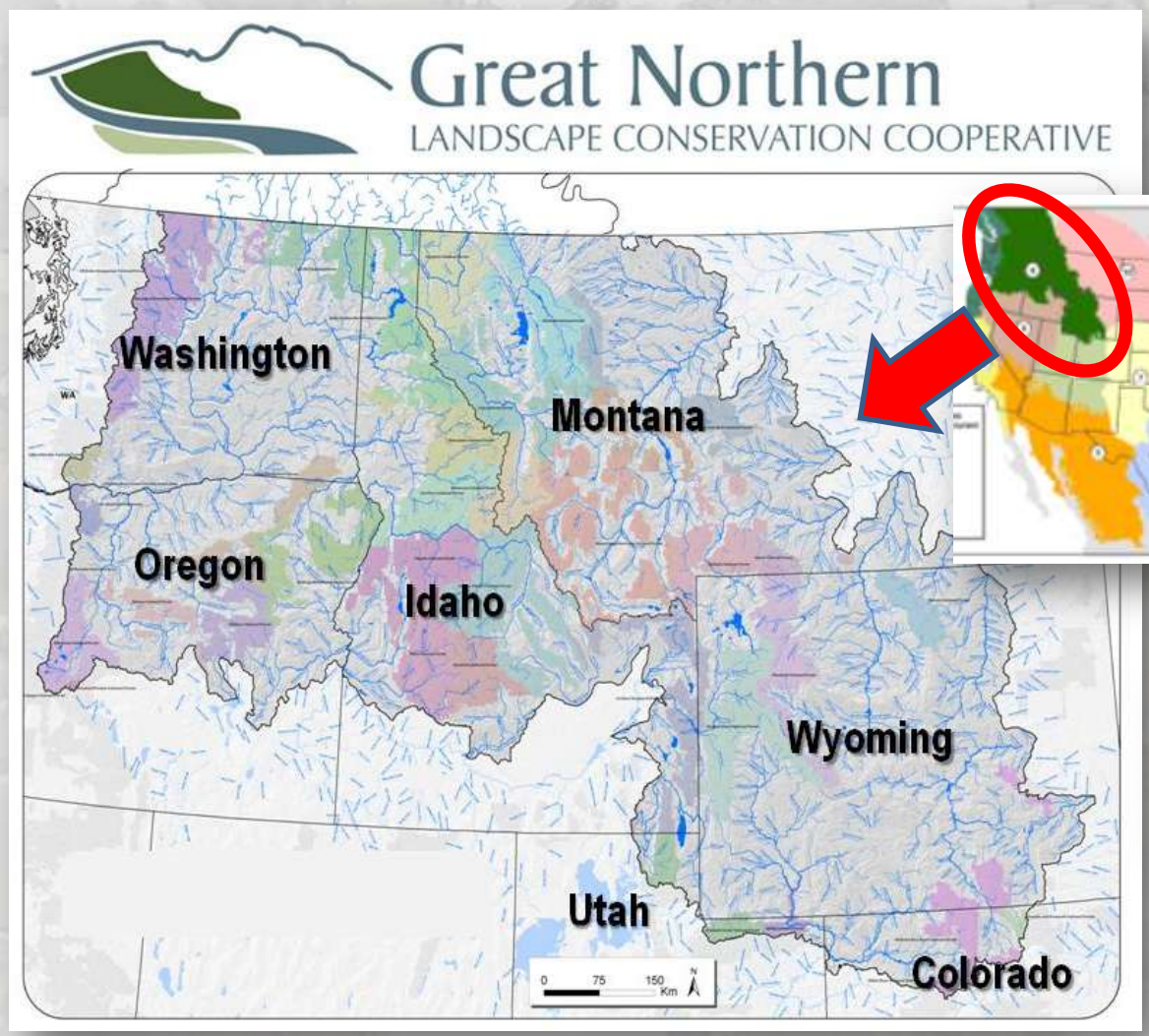
A Regional Stream Temperature Model for Mapping Thermal Habitats & Species Climate Vulnerability Across the GNLCC

Dan Isaak[1], Erin Peterson[2], Seth Wenger[3], Jay Verhoef[4], Charlie Luce[1], Jason Dunham[5], Dave Nagel[1], Jeff Kershner[5], Brett Roper[1], Steve Hostetler[5], Dona Horan[1], Gwynne Chandler[1], Sherry Wollrab[1], Sharon Parkes[1], Dave Hockman[5]



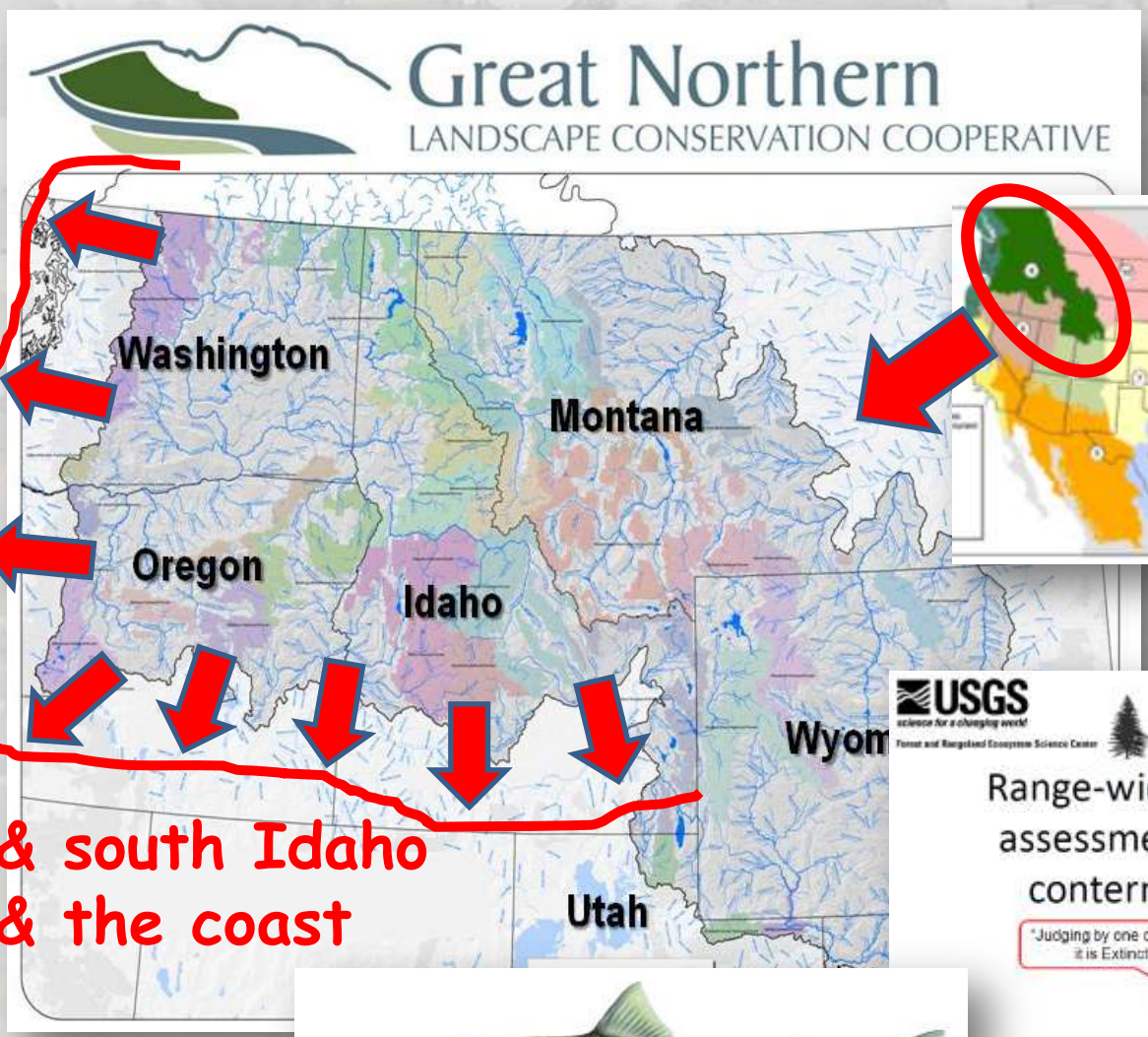


GNLCC Temperature Project

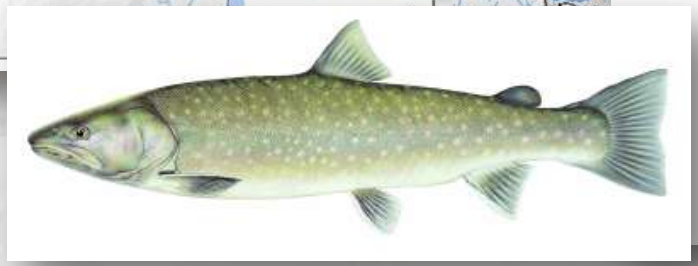


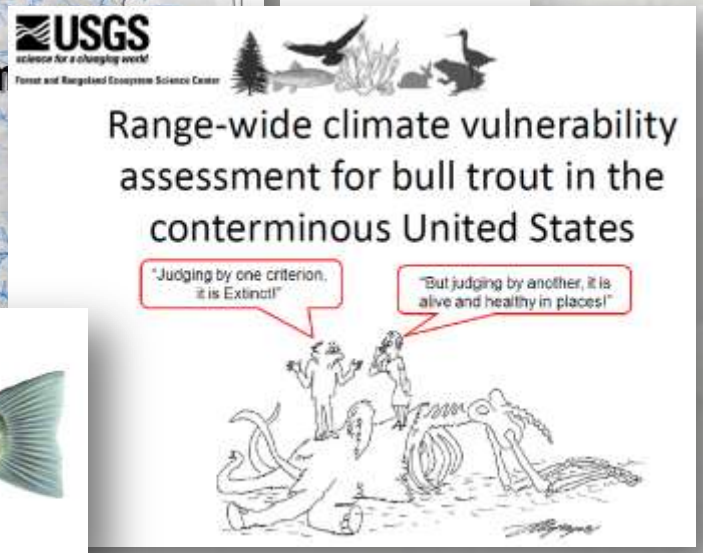
Landscape Conservation Cooperatives

GNLCC plus Temperature Project



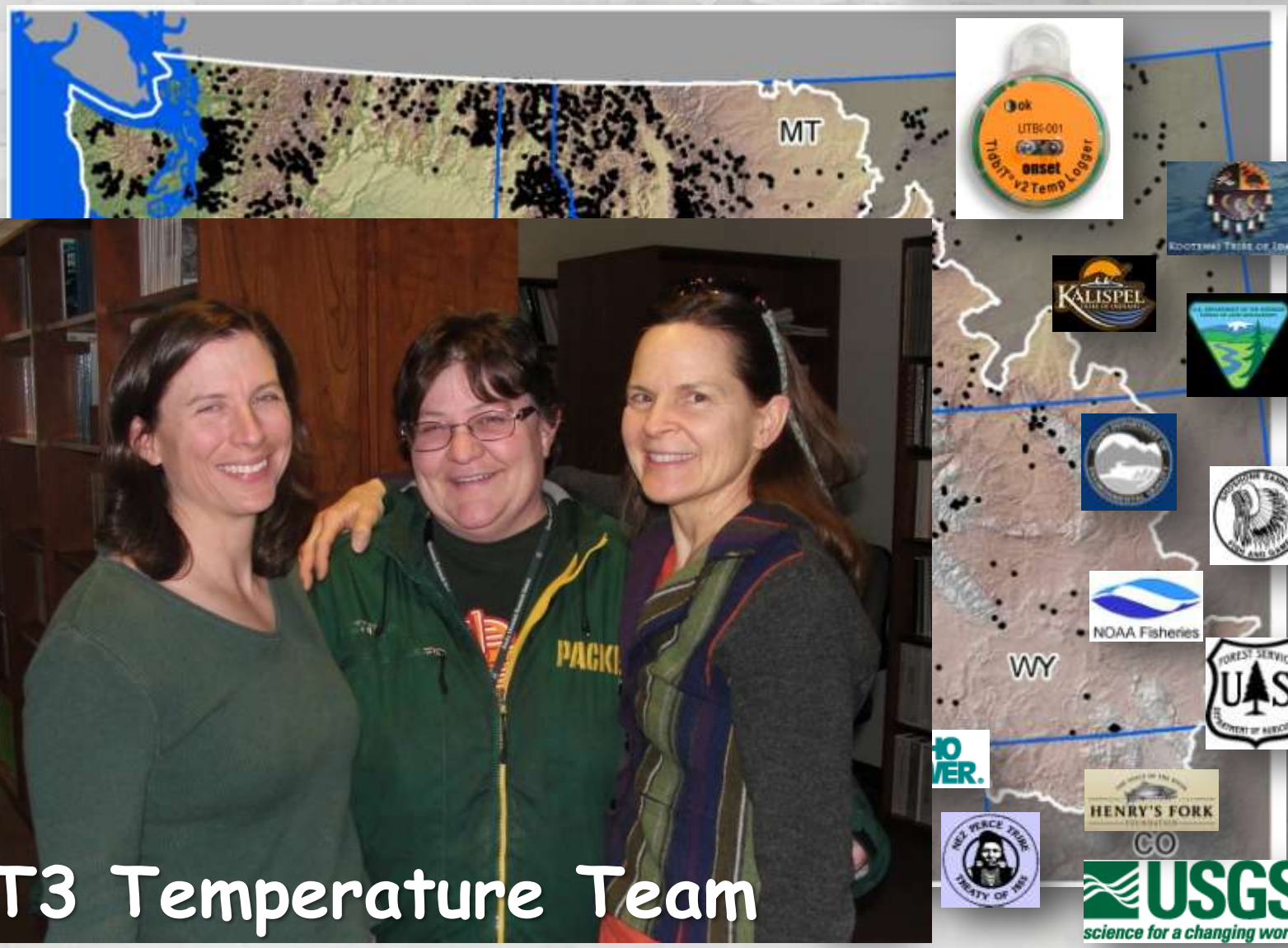
Landscape Conservation Cooperatives





Interagency Database

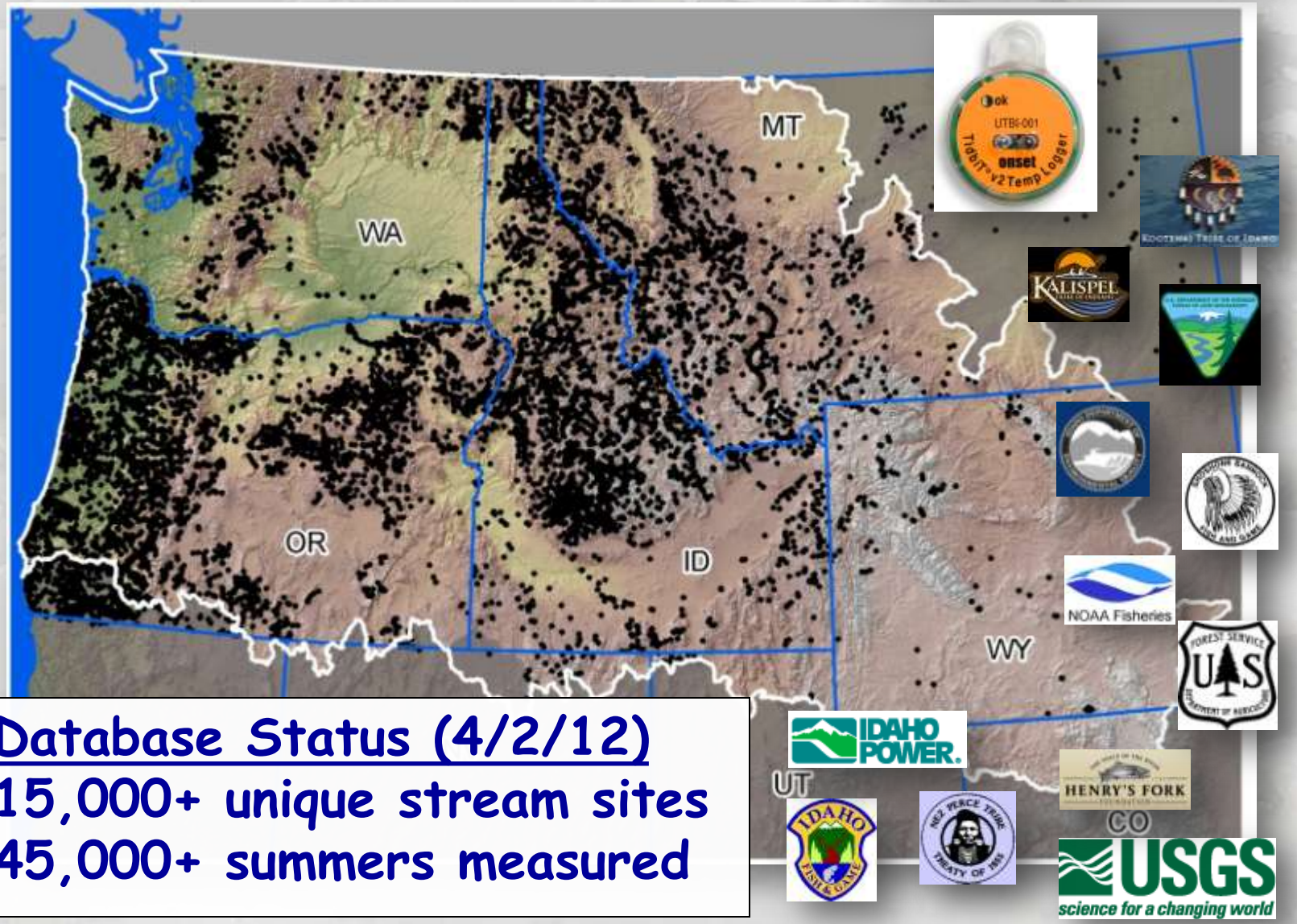
NorWeST Stream Temp



T3 Temperature Team

Interagency Database

NorWeST Stream Temp

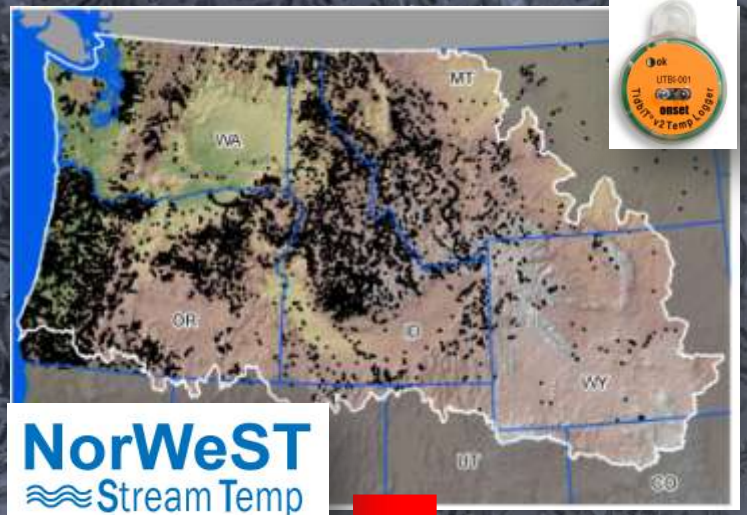


Database Status (4/2/12)

15,000+ unique stream sites

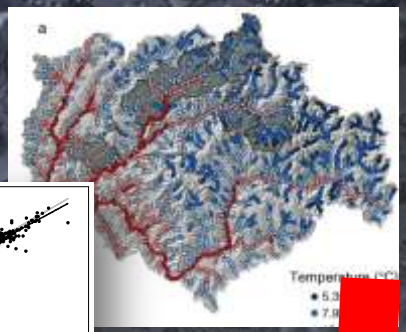
45,000+ summers measured

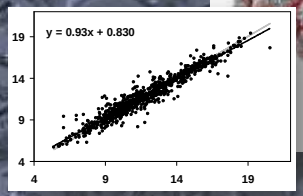
Regional Temperature Model



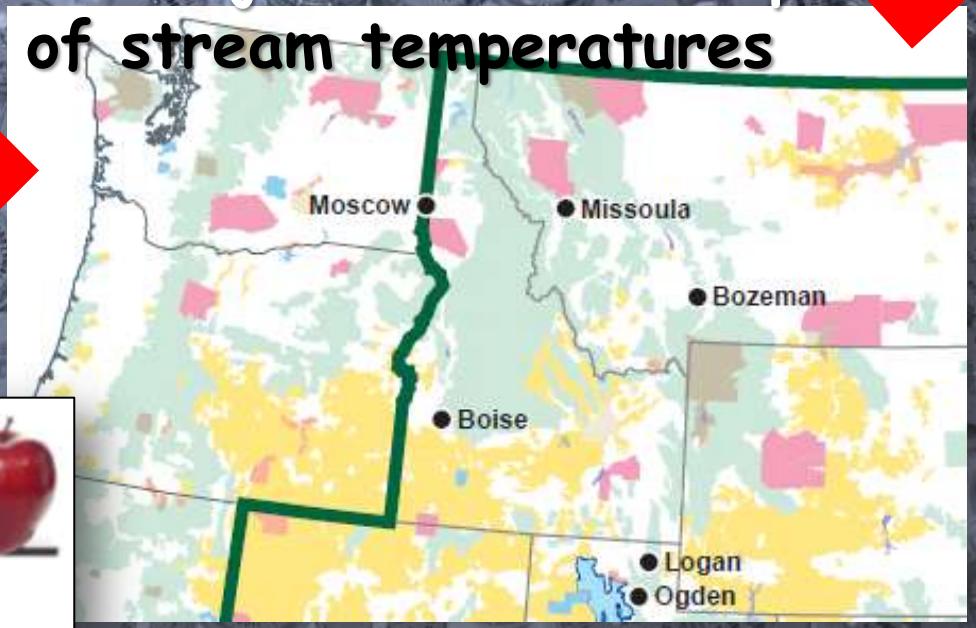
NorWeST Stream Temp

VHP models

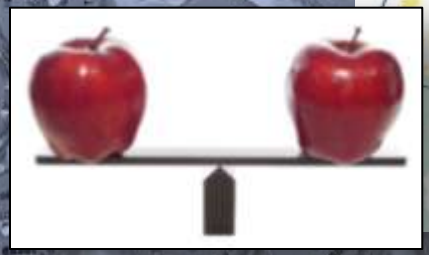




Cross-jurisdictional "maps" of stream temperatures



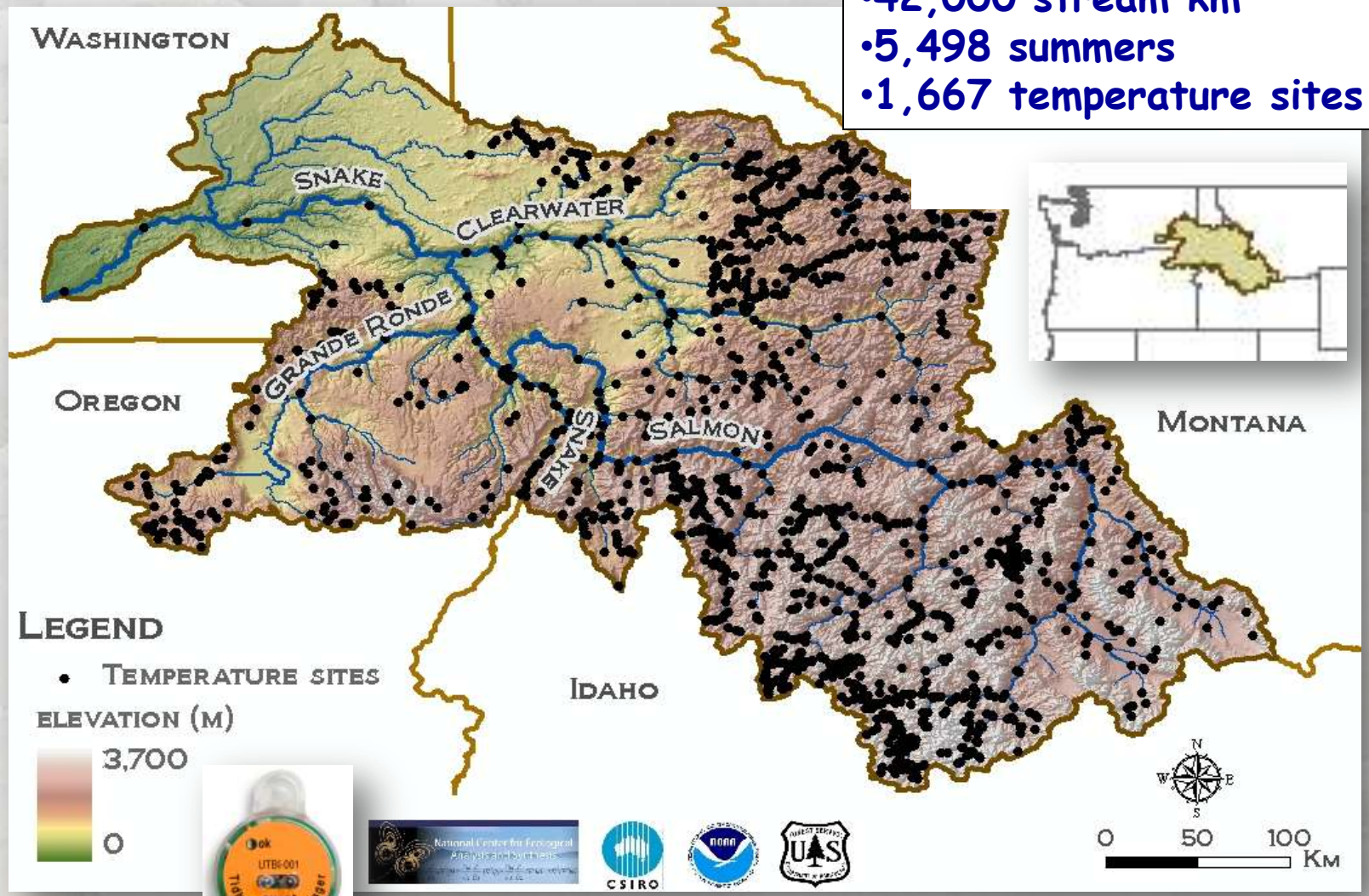
Consistent datum for strategic assessments



Computation Challenges: Model Fitting

NCEAS - Lower Snake Hydrologic Region

- 42,000 stream km
- 5,498 summers
- 1,667 temperature sites

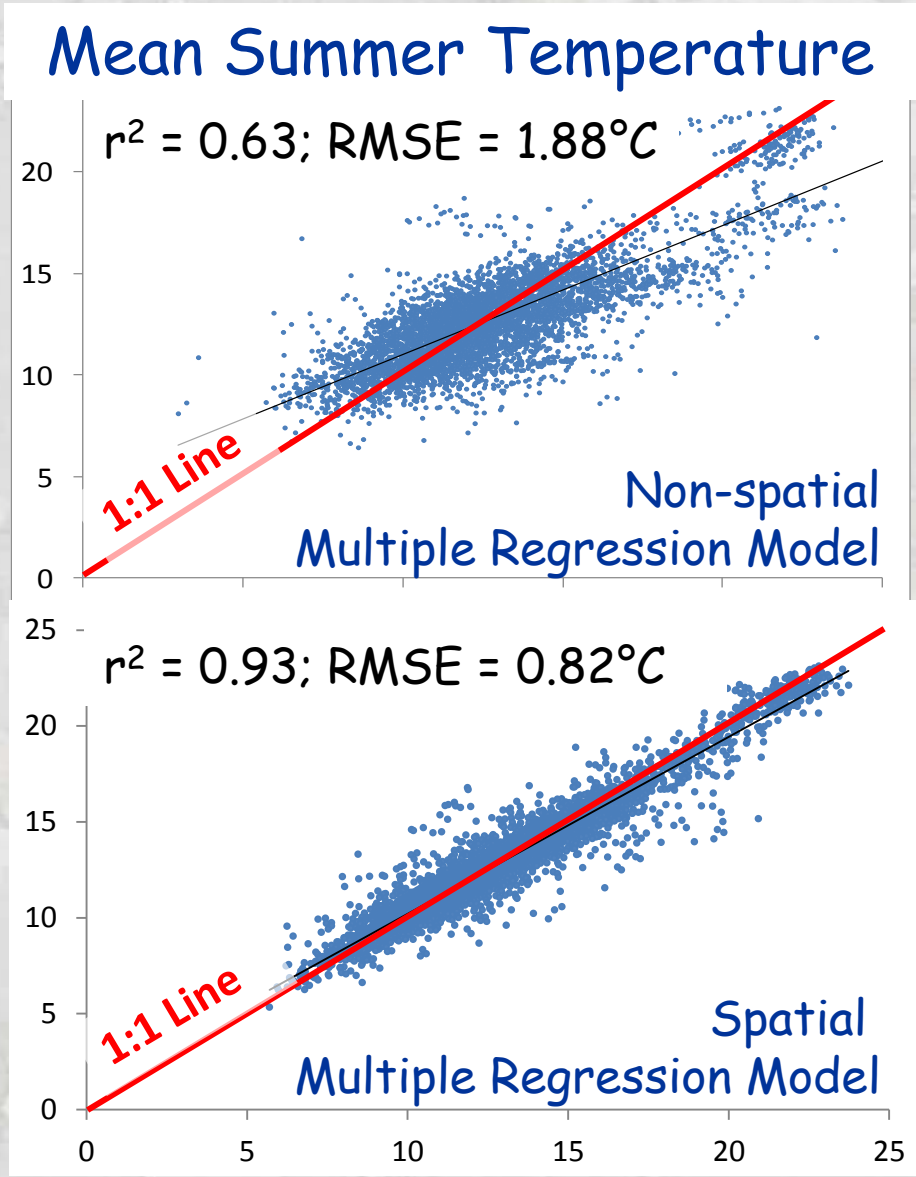


Peterson, Ver Hoef, and Isaak 2010

Lower Snake VHP Temperature Model

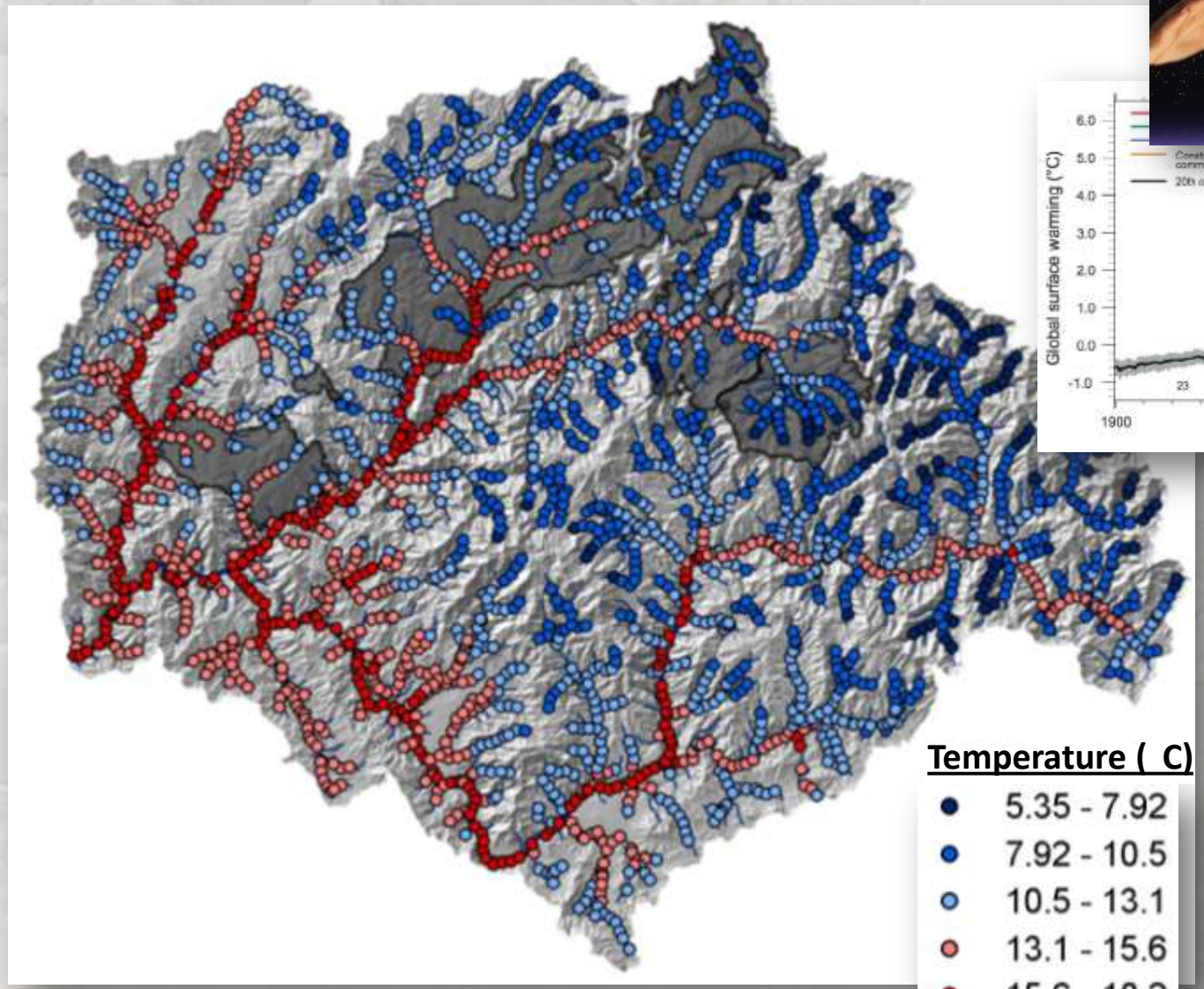
Non-spatial Stream Temp =
- 0.0041*Ele (m)
- 13.9*Slope (%)
+ 0.016*Wat_size (100km²)
-0.0022*Ave_Precip
- 0.041*Flow (m³/s)
+ 0.42*AirMean (C)

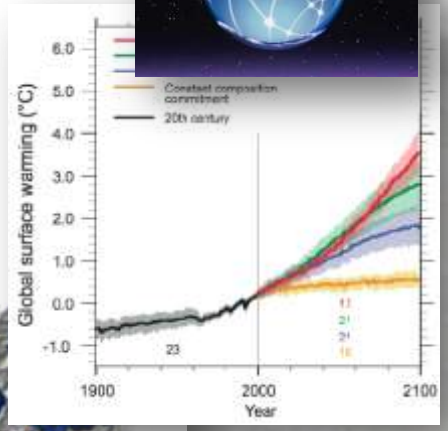
Spatial Stream Temp =
- 0.0045*Ele (m)
- 9.8*Slope (%)
+ 0.012*Wat_size (100km²)
- 0.00061*Ave_Precip
- 0.037*Flow (m³/s)
+ 0.46*AirMean (C)



Kriged Temperature "Maps"

2006 - Summer Mean - Boise River





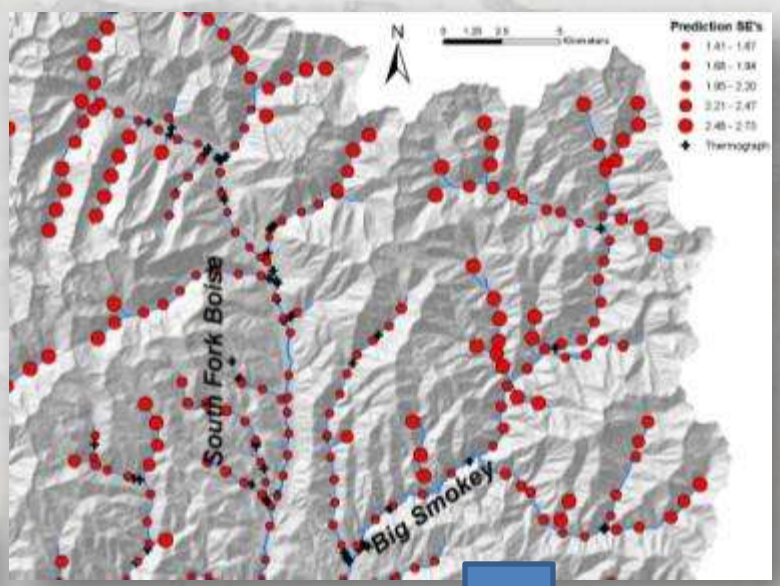
Temperature (C)

- 5.35 - 7.92
- 7.92 - 10.5
- 10.5 - 13.1
- 13.1 - 15.6
- 15.6 - 18.2

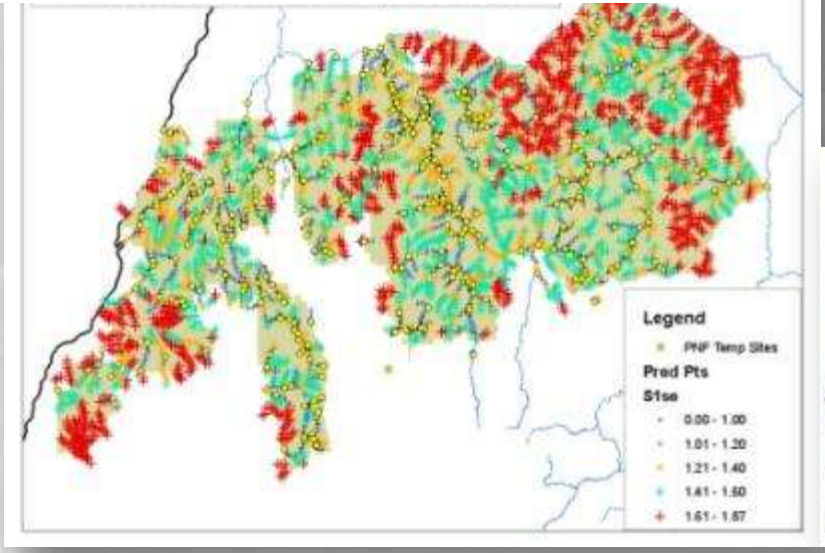
Isaak et al. 2010. *Eco. Apps.* 20:1350-1371

Prediction Precision "Maps"

Temperature Prediction SE's

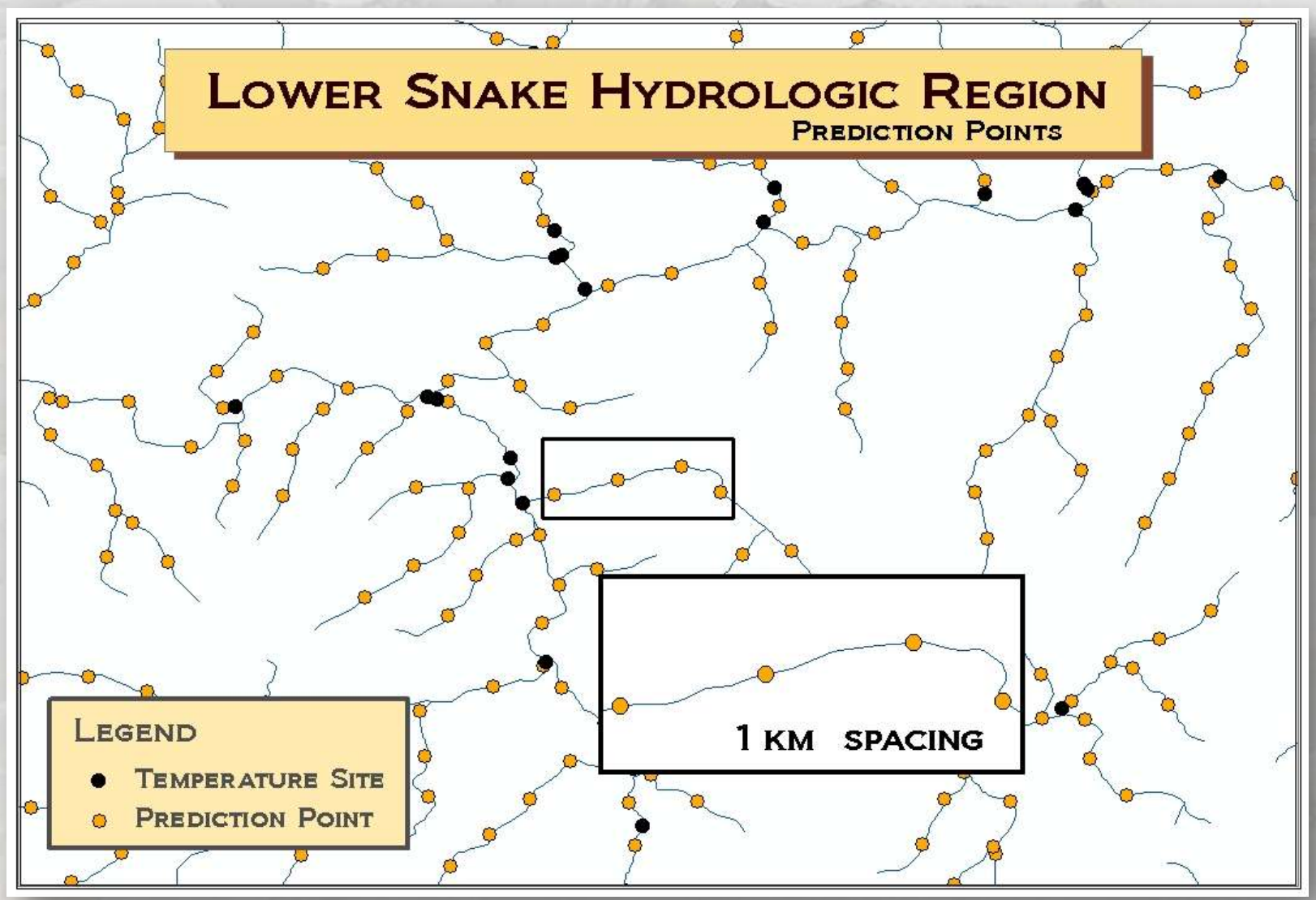


Payette National Forest
Spatial Uncertainty Map



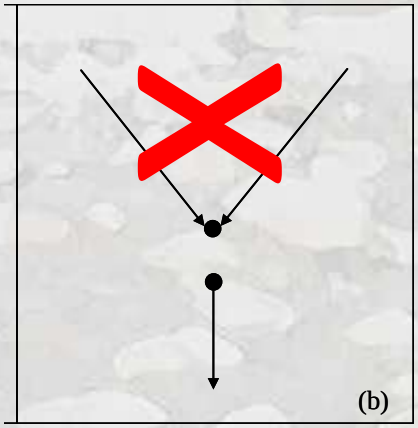
Data Model for Prediction Points

1 km default, but customizable

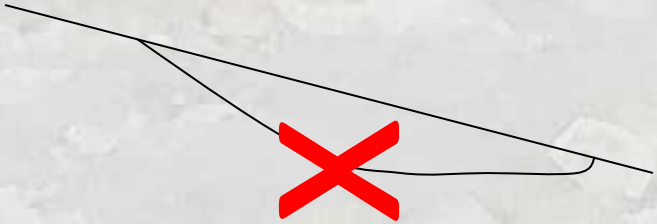


NHD+ Hydrolayer Conditioning

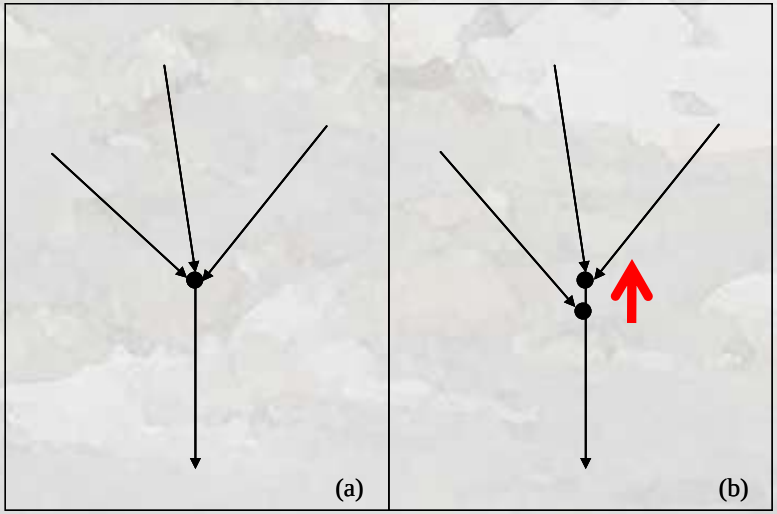
1. **Converging streams** – stream segments upstream that are not connected to the main channel were deleted.



2. **Downstream divergences** - side-channels were deleted.

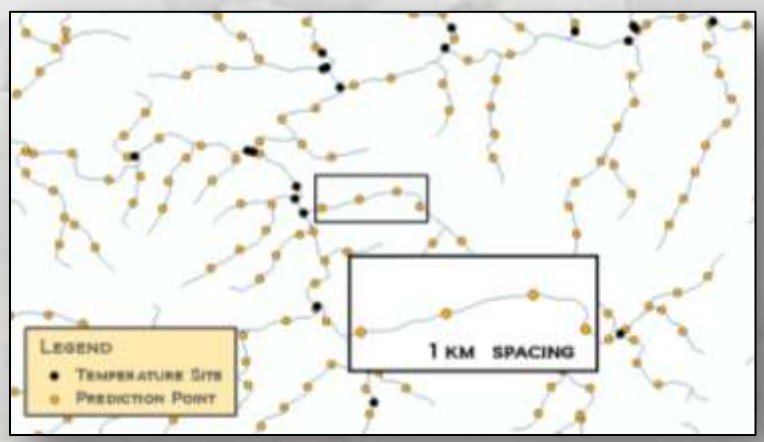


3. **Multi-segment confluences** (i.e., > 2 stream converging) were adjusted

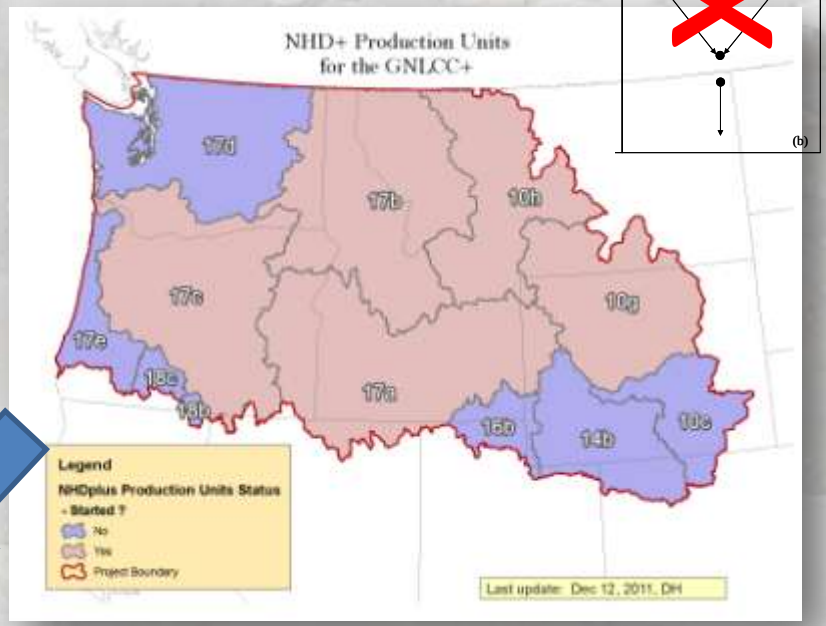


GIS Layers Online

Shapefiles of prediction points

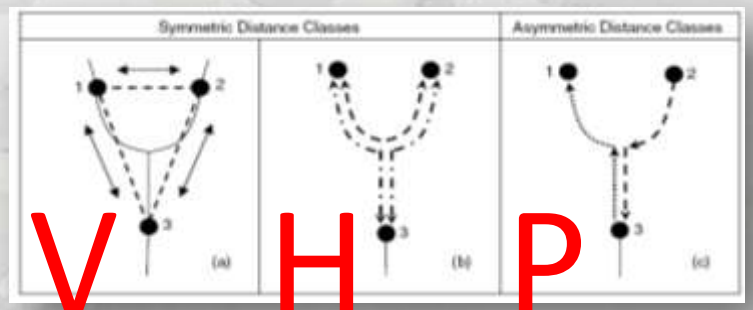


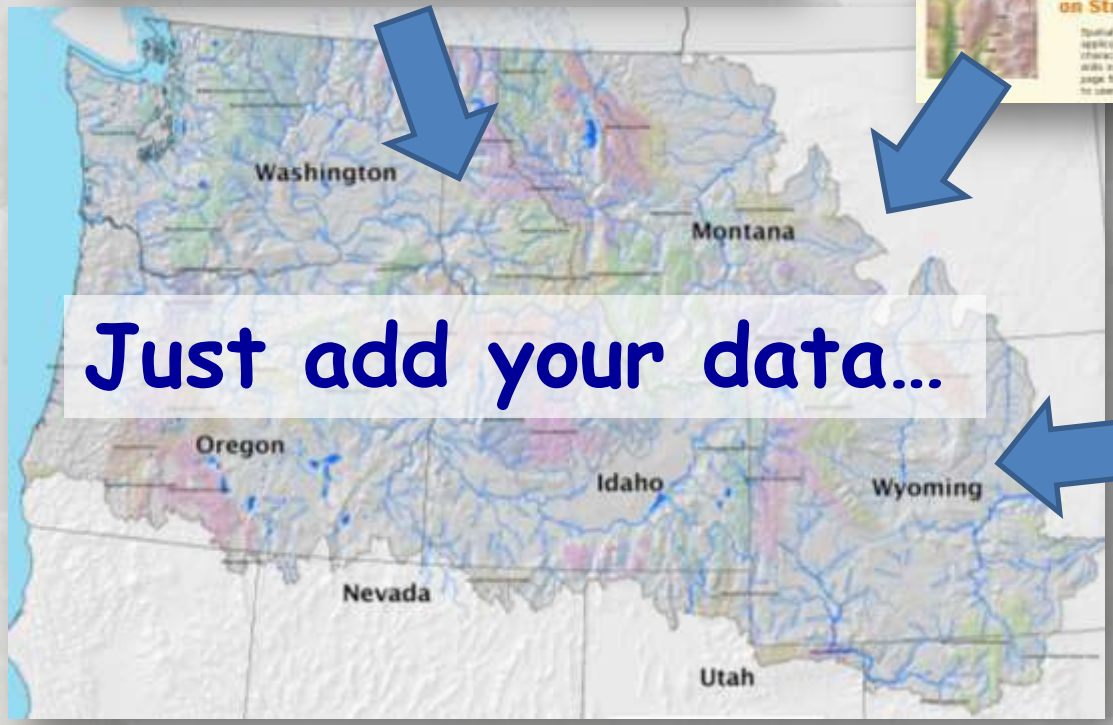
Reconditioned NHD+ Hydrolayers



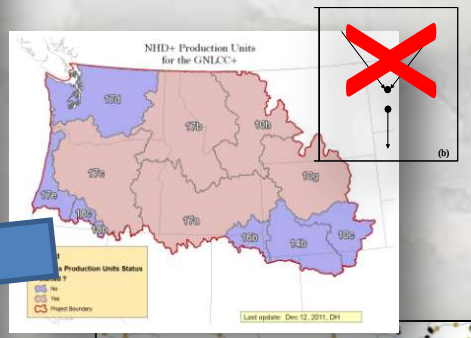
Websites for Distribution

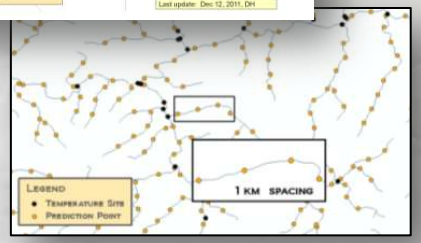
Regional Infrastructure for VHP Analysis of Stream Data



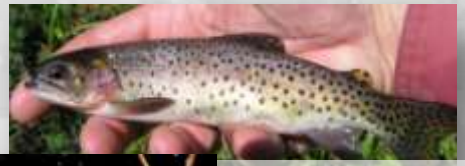


Just add your data…





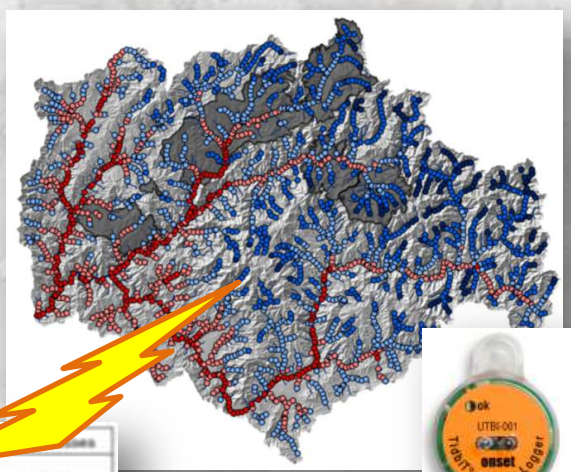
Temperature Data, but also…

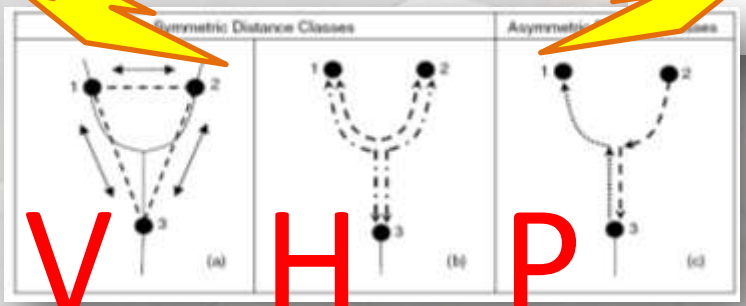


Distribution & abundance

Response Metrics

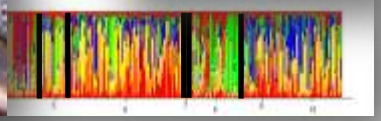
- Gaussian
- Poissan
- Binomial





V H P

Genetic Attributes



Water Quality Parameters


The End