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

Dan Isaak

Jay Ver Hoef

Erin Peterson

Y = b_0 + b_1(x_1) + \epsilon
Streams Ignored by Statisticians

The Arrangement of Field Experiments
R. A. Fisher, Sc.D., Rothamsted Experimental Station.


By Jerzy Neyman

1926

1934

Fig. 1.—A Complex Experiment with Winter Oats.
Statistical Models for Data on Stream Networks... Finally!

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
http://www.jstatsoft.org/

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

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

Functional Linkage of Watersheds and Streams (FLoWS)

- ArcGIS Geoprocessing Toolbox written in Python v2.5 for ArcGIS v10.1
- Developed by Dr. Dave Theobald and John Norman at Colorado State University
Streams Have Unique Properties
Networked Systems, Directional Flow, & Embedded Within Terrestrial Landscapes

Dual coordinate system required for networks

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

\[ Y = b_0 + b_1(x_1) + \epsilon \]

- Black dots: Flow-connected
- Green dots: Flow-unconnected

- Inverse Similarity
- Stream Distance
Stream Models are Generalizable...

Response Metrics
- Gaussian
- Poisson
- Binomial

Distribution & abundance

Genetic Attributes

Stream Temperature

Water Quality Parameters

SSN
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

\[ Y = b_0 + b_1(x_1) + \epsilon \]

It’s Becoming a BIG DATA World

Species distributions

Temperature

Discharge – USGS NWIS

Water Quality

eDNA Samples
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 =
\[-0.0064 \times \text{Elevation (m)} + 0.0104 \times \text{Radiation} + 0.39 \times \text{AirTemp (°C)} - 0.17 \times \text{Flow (m}^3/\text{s})\]

Spatial Stream Temp =
\[-0.0045 \times \text{Elevation (m)} + 0.0085 \times \text{Radiation} + 0.48 \times \text{AirTemp (°C)} - 0.11 \times \text{Flow (m}^3/\text{s})\]

Isaak et al. 2010. Ecol. Apps. 20:1350-1371
Interpolated Predictions Provide High-Resolution Network Status Maps

Which sets the stage for trend assessments...
SSN Models Scale Easily...

>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

1-km resolution
Models Describe Autocorrelation Distances

Efficient Monitoring Designs

Spatial Variation in Prediction Precision
Block-krige Estimates of Mean & Variance at User-Defined Scales

Temperature (°C)

- 10.6 - 11.0
- 11.1 - 11.5
- 11.6 - 12.0
- 12.1 - 12.5
- 12.6 - 13.0
- 13.1 - 13.5
- 13.6 - 14.0
- 14.1 - 14.5
- 14.6 - 14.9

95% CI (°C)

- 0.1
- 0.25
- 0.5
- 0.75
- 1
Block-krige Estimates of Mean & Variance at User-Defined Scales

Temperature (°C)

Precise & unbiased estimates

Bear Valley Creek Mean Temperature

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

~2°C difference

Summer temperature (°C)

- Block kriging
- Simple random

Stream

1 2
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
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...


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

Publications...

Data...

Software...
When & Where Are SSN Models Useful?

- Small datasets (n < 50 sites)
- Spatially sparse datasets
- Network connectivity effects weak or difficult to estimate

NARS – National Aquatic Resource Surveys
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 $\sim n \geq 50 / 100$
- more parameters with autocovariance
- spatial clustering needed

Maximum sample size $\sim n < 10,000$
- inversion of $n \times n$ matrix
Streams Ignored by Spatial Statisticians

Spatial pattern and ecological analysis

Pierre Legendre¹ & Marie-Josée Fortin²

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

The Final Frontier
Covariance Among Sites Depends on Flow Connectivity

![Diagram showing three sites (S1, S2, S3) with flow direction and stream distance graph with flow-connected and flow-unconnected data points.](image)
Covariance Depends on Flow Magnitude

Show equal streams vs Big little stream confluence
Mixed Model Covariance Structure

TU/TD/Euclid
A New Era of Better Prediction & Understanding for Stream Things...

New relationships described

Old relationships tested

Refined

Rejected

Predictor

Response
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. Isaak¹, Erin Peterson², Jay Ver Hoef³,

¹U.S. Forest Service, Rocky Mountain Research Station, Boise, ID 83702; ²CSIRO Division of Mathematics, Informatics, and Statistics, Dutton Park, Queensland, Australia; ³NOAA 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
Advantages:
- flexible & valid covariance structures by accommodating network topology
- weighting by stream size
- improved predictive ability & parameter estimates relative to non spatial models

Valid Interpolation on Networks

Let’s us connect the dots…

Predictive Models with Covariates