BIG DATAbases for Streams

Part 1a: Designed databases
NAWQA

Daren Carlisle, Ecological Studies Coordinator,
National Water-Quality Assessment Program,
US Geological Survey

Stream Internet Workshop
Boise, ID

Website: http://water.usgs.gov/nawqa/
NAWQA Regional Assessments

- **GOAL**: Assess stream health and the potential stressors that affect it
- 90-130 stream sites
- Gradient of anthropogenic stressors
- Characterize chemical & physical stressors for 10 weeks
- Characterize physical & biological conditions during week 10 of chemical sampling
Pacific Northwest Stream-Quality Assessment
Regional Assessments

NRSA Ecoregions
- Northern Appalachians
- Southern Appalachians
- Coastal Plains
- Temperate Plains
- Northern Plains
- Western Mountains
- Southern Plains
- Xeric

Present/Future study areas
Water: Monitoring & Assessment

You are here: Water » Our Waters » Watersheds » Monitoring & Assessment » National Aquatic Resource Surveys

National Aquatic Resource Surveys

National Rivers & Streams Assessment

> Learn About

Richard Mitchell - EPA

Website: http://water.epa.gov/type/watersheds/monitoring/aquaticsurvey_index.cfm
NRSA 2008/2009 Sample Locations
NRSA 2013/2014 Sample Locations
NRSA Between Survey Resample Sites
BugLab & WMC

Scott Miller, Chuck Hawkins – USU/BLM

BugLab Website: http://www.usu.edu/buglab/
WMC Website: http://www.cnr.usu.edu/wmc/
NAMF (National Aquatic Monitoring Framework)

Lotic ecosystem monitoring (started 2013)
- Standardized methods for chemical, physical and biological indicators
- Statistically valid sample designs
- Collaborations with EPA, USFS, States

Nearly 600 pts in first 3 years

Scott Miller (BLM/USU)
Website: http://www.usu.edu/buglab/Projects/CurrentProjects/
WMC/BugLab Invertebrate Spatial Database

1.5 million records for 5,473 taxa from 49,450 samples

15,279 unique stream locations
Additional Alaska samples

Borrowed data
MAPIT website: http://www.cnr.usu.edu/wmc/htm/data

MAPIT - a Mapping Application for Freshwater Invertebrate Taxa
Type in a taxon’s name
Caudatella
Read Me First!

Results 1 - 2000 of 4468

<table>
<thead>
<tr>
<th>ProjectName</th>
<th>State</th>
<th>Sampled</th>
<th>Station</th>
<th>ProjectName</th>
<th>Method</th>
<th>Habitat</th>
<th>Area</th>
<th>LabSplit</th>
<th>Qualitative</th>
<th>SampledDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREMP (422)</td>
<td>CA (1370)</td>
<td>148461</td>
<td>CSNM_11</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>7.81</td>
<td>N</td>
<td>2011/08/29 00:00:00</td>
</tr>
<tr>
<td>CA.R3 (91)</td>
<td>CA (1370)</td>
<td>151068</td>
<td>2013CAEIN1003</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>62.5</td>
<td>N</td>
<td>2013/06/25 00:00:00</td>
</tr>
<tr>
<td>EMAP.WEST (104)</td>
<td>CA (1370)</td>
<td>151090</td>
<td>2013CAIND1008</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>100</td>
<td>N</td>
<td>2013/05/15 00:00:00</td>
</tr>
<tr>
<td>HAWKINS.CHE (159)</td>
<td>CA (1370)</td>
<td>151098</td>
<td>2013CALIN1007</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>100</td>
<td>N</td>
<td>2013/07/03 00:00:00</td>
</tr>
<tr>
<td>NAOWA.REP (41)</td>
<td>CA (1370)</td>
<td>151099</td>
<td>WCAP59-R037</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>100</td>
<td>N</td>
<td>2013/07/03 00:00:00</td>
</tr>
<tr>
<td>PIBO (1088)</td>
<td>CA (1370)</td>
<td>151100</td>
<td>2013CALIN1099</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>100</td>
<td>N</td>
<td>2013/07/03 00:00:00</td>
</tr>
<tr>
<td>RS-USU.EAST (10)</td>
<td>CO (41)</td>
<td>151101</td>
<td>JBR-3000</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>100</td>
<td>N</td>
<td>2013/07/03 00:00:00</td>
</tr>
<tr>
<td>RS-USU.WEST (286)</td>
<td>ID (1324)</td>
<td>151102</td>
<td>JBR-1000</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>100</td>
<td>N</td>
<td>2013/07/03 00:00:00</td>
</tr>
<tr>
<td>USU-STAR-I (38)</td>
<td>ID (1324)</td>
<td>151103</td>
<td>JBR-2000</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>100</td>
<td>N</td>
<td>2013/07/03 00:00:00</td>
</tr>
<tr>
<td>USU-STAR-II (8)</td>
<td>ID (1324)</td>
<td>151104</td>
<td>JBR-3000</td>
<td>USU/BLM-BUGLAB</td>
<td>Surber net</td>
<td>Targeted Riffle</td>
<td>0.74</td>
<td>100</td>
<td>N</td>
<td>2013/07/03 00:00:00</td>
</tr>
<tr>
<td>USU/BLM-BUGLAB (2041)</td>
<td>WA (430)</td>
<td>151105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More...
Columbia Habitat Monitoring Program & Integrated Status and Effectiveness Monitoring Program

Website: https://www.champmonitoring.org/

Chris Jordan - NOAA

Website: http://ISEMP.org/
CHaMP and ISEMP

• CHaMP samples stream habitat conditions at ~500 unique locations in PNW, paired with ISEMP sampling stream fish (salmonids) at 250 CHaMP sample sites annually.

• Goals
  – Developing fish-habitat relationships across all CHaMP sampling sites
  – Developing generalized fish-habitat relationships across all CRB salmonid domain
  – Distributing metric data from sampling sites

• Need help thinking about
  – Distributing derived metrics
  – Distributing network-scale data products
Spatially and Temporally Continuous Predictions of Stream Temperature
PIBO (PacFish/InFish Biological Opinion)
AREMP (Aquatic & Riparian Effectiveness Monitoring Program)

- Rotating panel designs
- 250 sites/year
- 1,500 sites total

- Habitat
- Macroinvertebrates
- Stream temperature

Jeff Kershner, Brett Roper, Stephanie Miller (USFS)
PIBO Website: http://www.fs.fed.us/biology/fishecology/emp/index.html
AREMP Web: http://reo.gov/monitoring/reports/watershed/aremp/Welcome.htm
BIG DATAbases for Streams

Part 1b: Aggregated databases
MARIS (Multistate Aquatic Resources Information System)

The Multistate Aquatic Resources Information System (MARIS) is an online resource that contains over one million fish sampling and water quality records for over one thousand fish species compiled from state fish and wildlife agency and other sampling programs. (Learn More)

Download
Download the entire MARIS database or selected states in various formats.

Browse Map
View a clickable map of every lake and stream surveyed by the MARIS partner agencies.

Search Data
Search the MARIS database using several parameters and view the results.

Andy Loftus (Natural Resources Consultant)
Website: http://www.marisdata.org/
MARIS (Multistate Aquatic Resources Information System)

“contains over one million fish sampling and water quality records for over one thousand fish species compiled from state fish and wildlife agency and other sampling programs.”

Andy Loftus (Natural Resources Consultant)
Website: http://www.marisdata.org/
Western Trout Interagency Database Compilation

~20,000 unique sites

Seth Wenger

(Wenger et al. 2011a; 2011b; In preparation)
Thermal Criteria in Batch Mode...

NorWeST Stream Temperature (S1)

Wenger et al. *In Preparation*. Description of realized thermal niches using massive biological and temperature databases.
Genomics Frontier

eDNA, DNA Barcoding, etc.

Website: http://edna.fisheries.org/

Website: http://www.fishbol.org/
Dana Infante

Website: http://wim.usgs.gov/NorEaST/
NorWeST Stream Temp

>50,000,000 hourly records
>15,000 unique stream sites
>80 resource agencies

Dan Isaak  Website:
http://www.fs.fed.us/rm/boise/AWAE/projects/NorWeST.html
AKOATS (Alaska Online Aquatic Temperature Sites)

Where (do we monitor year round)

Sensor Season
Interval = Continuous
Status = Active
Frequency ≤ 60 min.

★ year round, n=211
open water, n=159
? unknown, n=17

n= 211

Joel Reynolds – FWS, Western Alaska LCC

Website: http://aknhp.uaa.alaska.edu/aquatic-ecology/akoats/#content
<table>
<thead>
<tr>
<th>ALASKA’S WATER</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 3,000,000</td>
</tr>
<tr>
<td>Mapped Lakes and Ponds</td>
</tr>
<tr>
<td>&gt; 847,000</td>
</tr>
<tr>
<td>Length of Mapped Streams and Rivers in miles</td>
</tr>
<tr>
<td>47,000</td>
</tr>
<tr>
<td>Length of Mapped Coastline in miles</td>
</tr>
<tr>
<td>1,980</td>
</tr>
<tr>
<td>Length of the longest river, the Yukon, in miles</td>
</tr>
<tr>
<td>616</td>
</tr>
<tr>
<td>Named Glaciers</td>
</tr>
<tr>
<td>40%</td>
</tr>
<tr>
<td>Share of the Nation’s Surface Water</td>
</tr>
</tbody>
</table>
Lucinda - Lakes?

Dan Magoulick – SE additions?
Update: Open Water Data Initiative

Stream Internet Workshop
April 23, 2015

Al Rea, USGS National Geospatial Program
Ed Clark, NOAA National Weather Service

Subcommittee on Spatial Water Data
OWDI Activities To Date

• **FGDC** Steering Committee (6/26) and **ACWI** (8/19) voted unanimously to revitalize and charge the Subcommittee on Spatial Water Data to scope and design a national Open Water Data Infrastructure

• **AWRA**
  - National Meeting (Nov 3-6, 2014) special track on OWDI
  - National Meeting, Denver (Nov 16-19, 2015) proposed special track

• **Subcommittee on Spatial Water Data (SSWD)**
  - Monthly meetings since August
  - 40+ regular attendees, mailing list of 80+
  - > 30 organizations represented
  - Applying “Lean Startup” principles
  - Three initial use cases
  - Four data work groups
  - Technology/standards work group
**OWDI Use-Case Working Groups**

**Work Group 1:**
National Flood Interoperability Experiment
- Identify flood data including stream-flow observations, forecasts and impacts
- Developing *Hydrofabric* v 0.1 and exploring data conflation
  *Supported by 3 sub-teams*

**Work Group 2:**
Drought Decision Support System
- Identify water resources data including natural flow, reservoir storage and drought impacts
- Explore visualization of drought in Lower Colorado

**Work Group 3:**
Spill Response Tool
- Identify water quality data including potential points sources and impacts
- Exploring requirements for new/additional data (e.g. velocity forecasts and reservoir residence times)
Hydro Event Services

- Hydro Event Data Catalog (HEDC)
  - Open – federated catalog
  - Web-service based
  - Starting with USGS ScienceBase platform for ESRI services, USGS CIDA group will host open (GeoServer) services
  - Intent is to complement, not replace WATERS web services
  - Iterative approach, will seek feedback
Aquatic Surveys Module in Natural Resource Monitor

Corporate database for National Forests & Grasslands

Survey Type Modules
- Stream temperature
- Fish, amphibian
- Reach habitat
- Passage conditions

Callie McConnell, Brian Sanborn

Website: http://www.fs.fed.us/nrm/index.shtml
BIG DATAbases for Streams

Part 2: Environmental predictors
Wang et al. (2011) Environmental predictors

A Hierarchical Spatial Framework and Database for the National River Fish Habitat Condition Assessment

Many databases in one

Dana Infante, Gary Whelan
Phase 1 Indicators

Examples of possible Phase 2 indicators:

- Topographic Wetness Index
- Mean summer temperature/precip
- Functional Process Zone (e.g., valley floor width, channel belt width and sinuosity)
- Pesticide applications rates
- Agricultural fertilizer application rates
- Recent forest loss
Quality Assurance

Landscape layers (LL) (CONUS-wide layers of climate, geology, soils, land cover/land use, and others)

Problem – Most Landscape Layers do not cross international boundaries or have missing values in some locations

Solution – Calculate the % completeness of each Landscape Layer within each watershed

48% complete

Tony Olsen (EPA)
StreamCat: Watershed Variables for Predicting Aquatic Condition and Watershed Integrity

Data Assembly Products

- 30 landscape layers with QA documentation
- 84 initial watershed metrics for:
  - 2.7 million NHD watersheds
  - 1,883 NRSA watersheds
- Data dictionary for 84 watershed metrics
- Python and R code for each step in metric calculation process
- 7 page description of methods

Tony Olsen (EPA)
Not very useful for ecologists

Surficial Geology maps

John Olson/Chuck Hawkins (USU)
John Olson translations

Derived from Integrated Geologic Maps Databases and other sources

www.sciencebase.gov
Use derived geology to model water chemistry

Predicted spatial variation in natural base flow salinity (μS/cm) in Utah streams

John Olson’s nationwide model
Some new spatial analyses
POPULATION GENETICS & SSN

David Cowley, Sabela Lois, Erin Peterson, Dan Isaak et al.
Canadian, 7 streams, N=160
Rio Grande, 26 streams N=757
Pecos, 9 streams, N=254

Georeference Genetic Samples

GIS: spread individual genotypes @ 50 m

Create network outlets @

Format Genetic Data (0, 1, 2)

<table>
<thead>
<tr>
<th>Sample</th>
<th>allele 1</th>
<th>allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (heterozy.)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2 (homozy.)</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

PCA on allele-count data

PC scores for each individual

Separate SSN Analyses on “Important” PCs

Select Weighting & Fixed Effects

Simulation of Genetics on SSNs
Multi-Species Conservation Using SSNs

Sabela Lois & David Cowley
Characterizing spatial dependence on streams
A “typical” empirical semivariogram (in Euclidean space):
A Torgeogram (empirical stream-network semivariogram) consists of separate semivariograms for flow-connected and flow-unconnected sites, and may also account for flow volume:
• The Torgegram has tremendous diagnostic value, and for very large stream datasets may be the only estimate of spatial dependence that is feasible to compute.

• I’m developing formal hypothesis testing methods which use the Torgegram to determine whether the best model for a given stream dataset is tail-up, tail-down, or a hybrid.

• I’m also developing methods for testing for (spatial) stationarity based on comparisons of watershed-specific Torgegrams.

• Jay and I plan to extend the Torgegram to space-time and multivariate data.

Introduction

• **Associate Professor (CSU)**

  Fish, Wildlife, and Conservation Biology
  Statistics

• **Assistant Unit Leader (USGS)**

  Colorado Cooperative Fish and Wildlife Research Unit
Spatial Occupancy Models

Spatial occupancy models for large data sets

Devon S. Thompson,1,2 Paul R. Cook,1,2 Mevin B. Hooten,1,2,4 Justin C. Royle,1,2 and Bruce A. Fink1

1National Marine Mammal Laboratory, SIO, 7600 Sand Point Way NE, Seattle, Washington 98115 USA
2U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, Colorado 80526 USA
3Department of Psych, Carleton University, Ottawa, Ontario, Canada K1S 5B6
4Department of Statistics, Colorado State University, Fort Collins, Colorado 80523 USA

ABSTRACT Since its development, occupancy modeling has become a popular and useful tool for ecologists and conservationists. However, the number of environmental covariates used to model occupancy has been limited by time and cost. In this paper, we present a new method for estimating occupancy using a spatially structured data set. The method uses a Bayesian hierarchical model to estimate the probability of occupancy for each individual in a large data set. The method is compared to other approaches, and it is found that the method provides more accurate estimates of the probability of occupancy than the other approaches. The method is also more efficient than the other approaches, and it is easily implemented in a variety of software packages. The method is suitable for use in a variety of environmental applications, and it is a valuable tool for ecologists and conservationists.
Colorado Plains Fish

Accounting for imperfect detection in Hill numbers for biodiversity studies

Kristen M. Broms*, Mevin B. Hooten 1,2,3 and Ryan M. Fitzpatrick*

1Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523 USA; 2U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Unit, Fort Collins, CO 80523 USA; 3Department of Statistics, Colorado State University, Fort Collins, CO 80523 USA; and 4Aquatic Wildlife Research Group, Colorado Parks and Wildlife, Fort Collins, CO 80523 USA

Summary

1. Hill numbers are early biodiversity metrics by combining several into one expression. For example, species richness, Shannon’s diversity index and the Gini-Simpson index are all of the form used in diversity measure, and they are expressed as Hill numbers. Traditionally, Hill numbers have been calculated from relative abundance data, but this expression has been modified to use presence data as well. We demonstrate an approach for estimating Hill numbers using an occupancy modeling framework that accounts for imperfect detection.

2. We also use the Hill number framework to use occupancy probabilities as exposed to the number of species that have been sampled and to adjust its interpretation from the unhandled species richness. After introducing the occupancy-based Hill numbers, we determine the differences between them and the occupancy-based Hill numbers previously used through simulation study and two applications.

3. In the simulation study and the two examples using real data, the occupancy-based Hill numbers were larger than the maximum-based Hill numbers, although species richness was estimated similarly using both methods.

4. The occupancy-based Hill number estimates are always at their simple component (i.e., as if all entries not

Key-words: Bayesian methods, Gini-Simpson index, incidence matrix, multi-species occupancy models, occupancy, species richness

Introduction

Biodiversity is one of the most important concepts in the study of ecosystems and is commonly measured by species richness, the Gini/Simpson index and Shannon diversity (Lande 1986; Jain 2006; Mac Nally 2007; Gotelli & Chao 2010; Chao et al. 2013). Multispecies assessments of biodiversity are valuable because species richness does not account for presence among species. To adopt the example from Gotelli & Chao (2013), suppose one community has exactly five species. The first community has five species and species 69 in common of the total number of individuals, with all other species each comprising 1/6 of the individuals. In the second community, each species occupies 1/3 of the population. Again, the second community should have more diversity, but the species richness approach will not reflect the similarity. In other words, both the Shannon index and the Gini-Simpson index take relative abundance of each species into account. The Shannon index “quantifies the uncertainty in the species density of a randomly chosen individual in the sample” (Gotelli & Chao 2013). It is also called the Shannon’s diversity index or the Shannon-Wiener index (Unni 1996). The Gini-Simpson index measures the probability that two randomly chosen individuals in the same sample belong to two different species (Gotelli & Chao 2013). Variations of the Gini-Simpson index include the Simpson index, the inverse Simpson index, the general order Banta entropy or the order-entropy quotient (Chao 2013).

Hill numbers are a modification of all these diversity measures in a single expression, providing a calculation (Hill 1973; Chao et al. 2013; Chao & Chao 2015) and a framework for which its additive alpha and beta diversity (Chao 2005). We describe the Hill number framework and its exact relationship to Shannon diversity and the Gini-Simpson index has the “Hill Numbers” section.

Traditionally, biodiversity measurement is functions of the relative abundance of each species in an assemblage, as calculated from the sampling design. However, Hill numbers have also been calculated using presence-absence data (Colwell & Bouldin 1984; Chao et al. 2007). This approach allows us to use presence data, but they can be weaker than others, they may allow for comparisons

Mevin Hooten
Landscape Genetics

Mevin Hooten

Fish, Wildlife, & Conservation Biology and Statistics
Colorado State University
Colorado Cooperative Fish and Wildlife Research Unit
U.S. Geological Survey
NFWF (National Fish & Wildlife Foundation)

Dave Lawrence
Website: http://www.nfwf.org/
NFHP (National Fish Habitat Partnership)

National Fish Habitat Conservation Act Introduced in U.S. Senate

Reservoir FHP featured on Bass Pro Shops Outdoor World Radio

Jeff Boardman, coordinator for the Reservoir Fisheries Habitat Partnership recently did an interview with Rural Radio of Sirius XM's Rural Radio (Channel 80) during Bass Pro

Join the effort!
- become a partner coalition member today

Create an account to receive Fish Habitat News and join the growing Fish Habitat Partner Coalition!

National Fish Habitat Partnership Data System

The National Fish Habitat Partnership (NFHP) Data System supports coordinated efforts of scientific assessment and data exchange among the partners and

Gary Whelan  Website: http://www.fishhabitat.org/
Anything else?