STATISTICAL SURVEY DESIGN AND ANALYSIS FOR AQUATIC RESOURCES

Department of Statistics, Colorado State University
September 20 - 21, 2002

ABSTRACTS

in the order of presentation, and as available
and associated presentation materials

Statistics in EPA's Science to Achieve Results (STAR) Program.

Barbara Levinson, Program Manager
STAR Program, USEPA, Washington, D.C.

This PowerPoint presentation is available as a pdf file at:

epa.ord.ncer.pdf

Sound Science for Monitoring Ecological Condition

Anthony Olsen, EMAP Design,
WED, NHEERL, USEPA, Corvallis, OR

This PowerPoint presentation is available as a pdf file at:

emap.overview.olsen.pdf

Design-based/Model-assisted Survey Methodology for Aquatic Resources: An EPA STAR Cooperative Research Program

Don L. Stevens
Department of Statistics, Oregon State University, Corvallis, OR

This PowerPoint presentation is available as a pdf file at:

osu.overview.pdf

Spatial Interpolation of Environmental Variables using Bayesian Techniques and Ancillary Information

Steven S. Carroll
Department of Statistics, Oregon State University, Corvallis, OR

In this presentation, we review spatial modeling techniques used to assess the ecological condition of streams in the eastern US. Spatial interpolation techniques that incorporate ancillary information in the covariance structure and use Bayesian techniques to account for covariance parameter uncertainty are described. The data used in the analysis was collected in the Mid-Atlantic Integrated Assessment and Mid-Atlantic Highland Assessment programs.

This PowerPoint presentation is available as a pdf file at:

carroll.pdf

ODFW Coho Salmon Trend Analysis Development

Don L. Stevens and Cynthia Cooper
Department of Statistics, Oregon State University, Corvallis, OR

This PowerPoint presentation is available as a pdf file at:

cooper.pdf

Model-Aided Sampling Designs for Salmon Population Status

Jean-Yves Pip Courbois, Department of Statistics, VIGRE
Loveday Conquest, Aquatic/Fisheries Science
Rebecca Buchanan, Quantitative Ecology and Resource Management
University of Washington, Seattle, WA

This PowerPoint presentation is available as a pdf file at:

courbois.pdf

Inferring Species Richness from Samples

Paul Murtaugh
Department of Statistics, Oregon State University, Corvallis, OR

The ultimate goal is to devise ways of estimating species richness from samples obtained in a variable-probability survey, like those used in the EMAP program. As an initial step, I have explored ways of estimating species richness from a single sample, e.g., one obtained by simple random sampling from some community.

Let $\boldsymbol{S}$ be the true number of species in the sampled community, and $\boldsymbol{S}_N$ be the number of species observed in our sample of $\boldsymbol{N}$ individuals. Suppose $\boldsymbol{n}_1$, $\boldsymbol{n}_2$, ..., $\boldsymbol{n}_{SN}$ are the sampled numbers of individuals of species 1, 2, ..., $\boldsymbol{S}_N$, $\boldsymbol{N} = \sum_{i=1}^{N_S} \boldsymbol{n}_i$. Finite samples are likely to miss rare species. Our goal is to find the value of $\boldsymbol{S}$ that is most consistent with the observed $\boldsymbol{S}_N$.

Any inference along these lines requires some assumption about the relative frequencies of different species in the community. Some authors have used Bayesian approaches, starting with noninformative prior distributions for the true number of species $\boldsymbol{S}_N$. and the relative frequencies of those species in the community (e.g., see Solow 1994). These priors are combined with data on the number of species in a sample(s) to arrive at a posterior distribution of $\boldsymbol{S}$.

I am exploring a frequentist approach, in which a distribution of relative frequencies is assumed and an interval of plausible values of $\boldsymbol{S}$ is inferred from the sample. This avoids the need to specify a prior distribution for $\boldsymbol{S}$, but, obviously, it still requires a guess about the nature of the distribution of relative frequencies.

The “broken-stick model" introduced by MacArthur (1957) is a simple model of the relative abundance of species in a community in which there are not strong interactions among species. The relative abundances of $\boldsymbol{S}$ species π_1 ,..., π_S are obtained by randomly throwing down $\boldsymbol{S}_1$ points on the interval [0,1]. The lengths of the resulting $\boldsymbol{S}$ segments give the species' relative abundances. Pielou (1969, p. 214) gives a readable account of the statistical properties of the segment lengths and their order statistics.

If we assume that the $\boldsymbol{S}_N$ species occurring in the sample are the $\boldsymbol{S}_N$ most abundant of the $\boldsymbol{S}$ species in the community, then we can compare the observed relative abundances of the $\boldsymbol{S}_N$ observed species to what would be predicted by the broken-stick model. The value of $\boldsymbol{S}$ yielding the closest agreement between observed and expected frequencies is chosen as our best guess of the true number of species in the community.

I have developed an algorithm, implemented in S+, to estimate $\boldsymbol{S}$ from $\boldsymbol{S}_N$. For a particular number of species $\boldsymbol{S}$, the program does a chi-square goodness-of-fit test comparing the relative abundances of the $\boldsymbol{S}_N$ species observed in the sample, and the zero abundances of the remaining $\boldsymbol{S}$-$\boldsymbol{S}_N$ species, to the abundances expected under the broken-stick model. Increasing values of $\boldsymbol{S}$, starting from $\boldsymbol{S}_N$, are tested, until the $\boldsymbol{P}$-value from the goodness-of-fit test drops below 0.05;
the $\boldsymbol{S}$ at which that happens is the upper limit of a 95% confidence interval for the true species richness. The value of $\boldsymbol{S}$ yielding the **maximum** $\boldsymbol{P}$-value is chosen as the point estimate of the species richness.

Applied to simulated samples from communities with known values of $\boldsymbol{S}$ and relative frequencies following the broken-stick distribution, this method gives unbiased estimates of $\boldsymbol{S}$ (as well it should!). The associated confidence intervals are often conservative (i.e., have greater than the nominal coverage), perhaps because the expected numbers of individuals in particular species can be quite small, especially when $\boldsymbol{S}$ is large.

Next steps include evaluating the performance of the method on communities not following the broken-stick distribution; comparing the new method to existing methods of estimating species richness; developing ways of extracting additional information about species richness from the relationship between $\boldsymbol{S}_N$ and $\boldsymbol{N}$ in multiple samples from the same community; and, ultimately, modifying the approach to account for the variable-probability sampling done in EMAP-like surveys.

References

MacArthur, R.H 1957. On the relative abundance of bird species. Proceedings of the National Academy of Sciences Washington. **43**: 293-295.

Pielou, E.C. 1969. **An introduction to mathematical ecology**. Wiley, New York.

Solow, A.R. 1994. On the Bayesian estimation of the number of species in a community. Ecology **75**: 2139-2142

This presentation is available as a pdf file at:

murtaugh.pdf

Causal Modeling of Macro-Invertebrate Data

Alix I. Gitelman
Department of Statistics, Oregon State University, Corvallis, OR

Causal models such as Bayesian belief networks (BBN) and structure equation models (SEM) are seeing increased use in ecological modeling (e.g., Malaeb, Summers & Pugesek, 2000; Lee, 2000). Both BBN and SEM rely on specification of causal pathways *a priori*, and in the case of BBN, the relative strengths of these pathways are evaluated by estimating conditional dependence relationships between sets of variables in the model. A BBN is developed for data regarding the "biotic integrity" of macro-invertebrates in watersheds in the Mid-Atlantic region using Hugin Expert © (http://www.hugin.com/). Channel sediment conditions and mine drainage levels appear to influence macroinvertebrate biotic integrity more strongly than either stream nutrient quality or riparian condition.

This presentation is available as a pdf file at:

gitelman.pdf

Parametric Model-Assisted Survey Methods for Environmental Surveys

Virginia M. Lesser, Paul Murtaugh, Breda Munoz-Hernandez,

Ruben A. Smith-Cayama, and Bev Benham
Department of Statistics, Oregon State University

This PowerPoint presentation is available as a pdf file at:

lesser.pdf

Design-Based Orthogonal Empirical Functions

Breda Munoz-Hernandez and Virginia M. Lesser
Department of Statistics, Oregon State University

This PowerPoint presentation is available as a pdf file at:

munoz.pdf

Handling Missing Data in Environmental Surveys

Ruben Smith-Cayama and Virginia M. Lesser
Department of Statistics, Oregon State University

This PowerPoint presentation is available as a pdf file at:

smith.pdf

OVERVIEW OF STARMAP: THE PROGRAM AT COLORADO STATE UNIVERSITY

N. Scott Urquhart
Department of Statistics, Colorado State University

The Space-Time Aquatic Resources Modeling and Analysis Program (STARMAP) in the Statistics Department at Colorado State University is organized around five projects: Combining Environmental Data Sets, Local Estimation, Indicator Development, Outreach, and Administration-Coordination. Because each of these Projects will be discussed subsequently in the meeting by their principal investigators, this talk will focus on the whole Program, relations among the Projects, collaboration with interested parties, mainly external to the University, involvement of students, and a Program-wide modeling effort. In order to do probability-based sampling now encouraged by EPA's Office of Water, the States and Tribes will need a sampling frame. The existing frame materials for streams contain substantial errors, identified in recent surveys. The Program will develop statistical models and methods to support prediction of stream flow and status relative to perennial flow. The resulting methods and models will be made available to developer of programs for probability-based selection of water monitoring sites.

This PowerPoint presentation is available as a pdf file at:

starmap.overview.pdf

Combining Environmental Data Sets

Jennifer Hoeting
Department of Statistics, Colorado State University

This PowerPoint presentation is available as a pdf file at:

csu.project1.pdf

Optimal Lattice and Geo-referenced Sample Designs for EMAP

Molly Leecaster, Idaho National Engineering and Environmental Laboratory
Jennifer A. Hoeting, Department of Statistics, Colorado State University
Kerry J. Ritter, Southern California Coastal Water Research Project

EMAP data are often used to make management decisions, using plots and maps. The usefulness of these maps is determined by the adequacy of the models used to create the maps and the accuracy and precision of the resulting predictions. Precision and accuracy depend, in part, on distance between sample sites. We plan to investigate optimal sample designs (sample spacing) for two **models** to produce maps of binary EMAP data; kriging for binary geo-referenced data and the autologistic model for lattice data. We then plan to compare the two models based on precision, accuracy, and sample size. Today we present the models, the issues, and our outlined approach.

This PowerPoint presentation is available as a pdf file at:

leecaster.pdf

Predicting Acid Neutralizing Capacity

Brett Kellum, Jennifer A. Hoeting, and N. Scott Urquhart
Department of Statistics, Colorado State University

This PowerPoint presentation is available as a pdf file at:

kellum.pdf

Composition Models for Benthic Invertebrate Data

Devin Johnson
Department of Statistics, Colorado State University

This PowerPoint presentation is available as a pdf file at:

johnson.pdf

Local Inferences via Nonparametric Model-Assisted Methods

F. Jay Breidt, Department of Statistics, Colorado State University, and
Jean D. Opsomer, Department of Statistics, Iowa State University

This PowerPoint presentation is available as a pdf file at:

csu.project2.pdf

Nonparametric Survey Regression Estimation in Two-Stage Spatial Sampling

Siobhan Everson-Stewart
Department of Statistics, Colorado State University

Nonparametric model-assisted survey estimation is adapted to use auxiliary information for status estimation in two-stage, unequal-probability, spatial sampling designs. An example of such a design is EMAP's Northeast Lakes Survey, to which the methodology is applied.

This presentation is available as a pdf file at:

everson.pdf

Estimating Distribution Functions from Survey Data Using Nonparametric Regression

Alicia A. Johnson
Department of Statistics, Colorado State University

This presentation is available as a pdf file at:
johnsona.pdf

Semiparametric Mixed Models in Small Area Estimation

Mark J. Delorey,
Department of Statistics, Colorado State University

This PowerPoint presentation is available as a pdf file at:

delorey.pdf

Quantifying Taxonomic Richness in Terms of the Level of Rarity Assessed by a Fixed Count

Kerry J. Ritter, Southern California Coastal Water Research Project
David Birkes, Department of Statistics, Oregon State University and
N. Scott Urquhart, Department of Statistics, Colorado State University

A common measure of taxonomic richness used in macroinvertebrate inventories is numerical taxonomic richness (NTR), i.e. the number of taxa observed for a fixed (usually small) number of individuals subsampled from a larger collection. However, rare taxa tend to be excluded from the metric for smaller subsamples due to their low inclusion probabilities. NTR, then, essentially targets the number of non-rare taxa, rather than total number of taxa. The level of rarity assessed by the metric depends on the size of the subsample. By specifying the target parameter that NTR is "really" estimating, researchers may understand more precisely what is being assessed and compared. Furthermore, determining a target parameter allows the consideration of alternatives for estimation that may be used to increase resolution or to reduce cost. This paper provides a means for determining the parameter targeted by NTR. Specifically we show that for a fixed count of size n, NTR is nearly unbiased for the number of taxa which occur in the collection with relative frequency $\geq 1/2n$. Further, the jackknife adjustment to observed taxa count is shown to enhance the level of rarity assessed by NTR for the same-sized subsample.

This PowerPoint presentation is available as a pdf file at:

ritter.pdf

STARMAP OUTREACH

N. Scott Urquhart
Department of Statistics, Colorado State University

This PowerPoint presentation is available as a pdf file at:

csu.project4.pdf

NEEDS ASSESSMENT OF TRIBAL REQUIREMENTS FOR INSTRUCTION IN THE USE OF STATISTICALLY-BASED AQUATIC WATER QUALITY MONITORING TECHNIQUES

Stephen W. Johnson, Principal Investigator
Water Quality Technology, Inc.
Fort Collins, Colorado

This report is written to document the need for statistically-based aquatic water quality monitoring techniques by selected American Indian Tribes for use by the U.S. Environmental Protection Agency (USEPA) in testing electronic instructional materials under development. The goal is to document the needs of ten American Indian Tribes for the design of aquatic water quality monitoring plans and the statistical analysis of the data. The selected American Indian Tribes are situated on reservations within USEPA Region IX (Arizona, California, and Nevada) except the Southern Ute Indian Tribe which is situated in USEPA Region VIII. A Tribal needs assessment form was developed to document Tribal responses to a set of specific questions. At present, only three of ten American Indian tribes have been contacted for completion of the individual Tribal needs assessment forms.

Tribal needs for the design of monitoring plans reflects environmental programs in their infancy and at a crisis-management level. Criteria for selecting monitoring sites predominantly addresses the detection and quantification of water quality contaminants. Current tribal aquatic water quality monitoring plans appear to be based on responses to specific tribal needs as opposed to comprehensive plans. Future tribal monitoring plans appear to also be based on specific (not comprehensive) tribal needs other than the San Carlos Apache tribe which is pursuing water quality baseline establishment and developing tribal water quality standards. The Tribes are not required (states are required) by the USEPA to produce Section 305(b) Water Quality Assessment Reports, which would provide useful guidelines for the design of water quality monitoring plans. The Tribal needs for monitoring cultural uses of Tribal waters is very definitive and reflects the Tribes differences with state programs in that the Tribes actively pursue protection of culturally-sensitive water uses. The Tribes rely heavily on contractual consultants and continued federal funding is needed to maintain Tribal accessibility to outside expertise and cross-training of Tribal employees. Tribal needs for statistical analyses include database management software and statistical programs that are user-friendly without the need for extensive workshop training.

This PowerPoint presentation is available as a pdf file at:

johnsons.pdf