



FINAL TECHNICAL REPORT: STARMAP

PERIOD COVERED BY FINAL REPORT: October 1, 2001 - September 30, 2006

DATE OF FINAL REPORT: December 15, 2006

EPA COOPERATIVE AGREEMENT NUMBER: CR 829095

PROGRAM NAME (TITLE): SPACE-TIME AQUATIC RESOURCES MODELING AND ANALYSIS PROGRAM (STARMAP)

INVESTIGATORS: N. Scott Urquhart, Program Director; Richard A. Davis, Co-Principal Investigator; F. Jay Breidt, Jennifer A. Hoeting, Hariharan K. Iyer, David M. Theobald, Jean D. Opsomer (ISU), Don L. Stevens, Jr., Alix I. Gitelman (OSU), Kerry Ritter (SCCWRP), and Devin Johnson (UAF)

INSTITUTIONS: Colorado State University (CSU), Oregon State University (OSU), Iowa State University (ISU), Southern California Coastal Water Research Program (SCCWRP), University of Alaska, Fairbanks (UAF), and Water Quality Technology, Inc. (WQTI)

RESEARCH CATEGORY: Research Program on Statistical Survey Design and Analysis for Aquatic Resources

PROJECT PERIOD: 10/1/2001 - 9/30/2006

OBJECTIVE OF RESEARCH: Development and implementation of statistical methods for aquatic surveys, including tools need to evaluate landscape covariates

SUPPLEMENTAL KEYWORDS: Geostatistical modeling, computational statistics, latent processes, spatial covariance functions, model selection, sampling design remote sensing, path analysis, kernel regression, thin plate splines, small area estimation, GIS, tessellation stratified sampling, water quality, land cover, land use, accuracy, precision, outreach, distance learning, web-based learning, management, efficiency

RELEVANT WEB SITES:

STARMAP WEB SITE <http://www.stat.colostate.edu/starmap/>
<http://www.stat.colostate.edu/~jah/software/matern.R>

WEB SITE for GIS Tools: FloWS, FunConn and RRQRR
<http://www.nrel.colostate.edu/projects/starmap/>

0. STRUCTURE OF THIS FINAL TECHNICAL REPORT

STARMAP has five projects which produced a large number of outputs. As there had been an annual report for each project, the Project Officer recommended that there be a final technical report for each project, as well as an overall report. This is the overall report. The entire final report consists of 14 distinct pieces, enumerated in Section 6.

1. ACCOMPLISHMENTS of STARMAP

The Environmental Protection Agency (EPA)'s Science to Achieve Results (STAR) Program announced a competition for Research Programs on Statistical Survey Design and Analysis for Aquatic Resources through a Request for Applications (RFA) on October 1, 2000. The Space-Time Aquatic Resources Modeling and Analysis Program (STARMAP) was conceived, funded, and developed under this governing RFA. That RFA states three major objectives within the context of cooperating with EPA's Environmental Monitoring and Assessment Program (EMAP):

- ◆ Identify and perform cutting-edge research in environmental statistics that will further the application of statistics to the environmental sciences;
- ◆ Facilitate the professional development of future environmental statisticians; and
- ◆ Develop and extend the expertise on design and analysis to States and Tribes.

STARMAP has achieved these three objectives. The RFA clearly anticipated a substantial involvement of landscape ecology in this research, and that was accomplished.

Although STARMAP had five distinct Projects, each of which made notable contributions, one sustained cooperation among all of the Projects produced tools of immediate utility for the analysis of aquatic monitoring data, regardless of what organization collected the data. Water quality indicators gathered at a point in a stream network depend on myriad contributions from the landscape above that collection point, as well as on the same indicator at collection points higher in the network. Three sets of tools are required to appropriately analyze such data: Geographic Information System (GIS) tools to delineate the upstream area contributing to any collection point; extensions of spatial statistics tools to accommodate the special features of stream networks; and methods for handling missing data. Project 3 developed, test, and distributed the needed GIS tools. Projects 1 and 2 have extended relevant statistical tools in two directions. The cooperating Program, DAMARS at Oregon State University, adapted the needed tools for missing data. The tools have been exercised on state-wide and county-wide data from the state of Maryland, gathered using the sort of sampling and collection protocols advanced by EPA's EMAP. The results have been and are being communicated to diverse audiences (Project 4). Project 5 provided the necessary administration and coordination to allow the other Projects functioned effectively. Results have been and are being published in relevant outlets.

The RFA described the need for two cooperating programs with differing perspectives. The two Programs funded from that solicitation were the Space-Time Aquatic Resources Modeling and Analysis Program (STARMAP) at Colorado State University and Designs and Models for Aquatic Resource Surveys (DAMARS) at Oregon State University. The first had a model-based perspective while the second had a design-based perspective, as specified in the governing RFA. The two Programs cooperated extensively, eventually to the extent that they had two joint Projects: The tools of local estimation can be applied in design-based, model-assisted, or model-based ways; STARMAP's Project 2 on local estimation was implemented as a joint project. Two approaches to outreach were proposed: the development of CD-ROM based learning materials, and direct interaction with members of the client community. Although the development of the learning materials occurred primarily under the guidance of STARMAP's Project 4, it was funded by both Programs. Likewise even though a substantial part of the direct interaction with the client community was done under the leadership of DAMARS, it was funded by and participated in by participants of both Programs.

STARMAP facilitated the professional development of future environmental statisticians in several ways. The RFA anticipated a substantial involvement of postdoctoral fellows in the research program, and their development of interests in environmental statistics. The discipline of statistics does not have a tradition of postdoctoral fellows. Students completing doctorates in statistics are in such short supply that no domestic applicants with real potential for becoming environmental statisticians applied for the available positions. One postdoctoral fellow from Italy spent two periods of time with Project 2 and made substantial contributions, but she returned to Italy. She definitely is advancing the cause of environmental statistics there, as evidenced by her talks, publications, and mentoring of students there. A second postdoctoral fellow spent nine months conducting research under Project 1, but has returned to Korea.

Once it became clear that the planned postdoctoral fellows would not be a viable path for developing future environmental statisticians, STARMAP, with the concurrence of its Science Advisory Committee, focused some resources on helping early career statisticians develop their environmental interests. This was done for Devin Johnson at the University of Alaska, Fairbanks, and Kerry Ritter at the Southern California Coastal Water Research project. Other resources were invested in masters-level students. Overall three students completed doctorates under STARMAP sponsorship; all are engaged in collaboration with the environmental sciences. Six more are well along in their doctoral research as this report is written. Four of these students likely will finish this academic year; one already is employed in the environmental sciences; the other three are seeking employment which will allow them to continue their collaboration with the environmental sciences. Eight students completed masters degrees under STARMAP sponsorship and one is nearing completion.

It is clear that STARMAP's activities have had substantial impact on the career outlook of several Program investigators. Although STARMAP funding has ended, these investigators have developed aquatic environmental interests which will influence their research and teaching for the rest of their careers.

The following material describes the accomplishments of STARMAP's five Projects:

◆ PROJECT 1 - Combining Environmental Data Sets:

This Project expanded the analysis and interpretation tools available to aquatic scientists, and statisticians who assist them, especially with tools which utilize spatial or temporal models, as well as ones which utilize hierarchical (Bayesian) methods. Specifically:

- Adapted spatial statistical models to accommodate the branching nature of stream networks; implemented statistical computing tools to support the selection of predictor variables in an aquatic context;
- Developed hierarchal methods for the analysis of categorical responses, of the sort resulting from macroinvertebrate studies in streams;
- Developed hierarchal methods for the analysis of ordered categorical responses, of the sort resulting from studies to monitor stream health;
- Developed and demonstrated hierarchical analysis methods for assigning causes of effects in aquatic systems like stream networks;
- Produced a textbook that surveys modern methods in computational statistics;
- Expanded temporal methods for identifying structural breaks;

- Investigated the uncertainty associated with contour curves developed from spatial statistical models;
- Developed sampling plans for near-coastal systems;
- Expanded components of variance tools for characterizing both temporal and spatial variability; and
- Trained future statisticians, some of whom are already working in environmental fields.

1a. Summary of work

Project 1 involved a substantial number of researchers working in a large variety of areas. This section summarizes the output of Project 1.

Methodology for statistical modeling of spatially-referenced, aquatic, and other environmental data

PI Hoeting produced a textbook, *Computational Statistics*, under this cooperative agreement, co-authored by G. H. Givens. This graduate level textbook surveys a wide variety of topics in modern statistical computing and computational statistics, including optimization, integration (including MCMC methods), bootstrapping, and smoothing. The book has been adopted widely for teaching and is being used by statisticians and non-statisticians alike as a reference book on methods for computational statistics. The book includes a webpage with related software and numerous examples and homework problems. *Computational Statistics* has become a bestselling textbook for Wiley (publisher) and is currently in its fourth printing.

Investigator Gitelman's "Isomorphic Chain Graphs for Modeling Spatial Dependence in Ecological Data" by Gitelman and Herlihy (to appear in *Environmental and Ecological Statistics*) was an important contribution toward developing causal inference models. In this paper, they extend Bayesian belief network models (also called acyclic directed graphs) to accommodate correlation through space.

In conjunction with this work, Investigator Gitelman was invited to serve as the guest associate editor for a special forum on the application of Bayesian belief networks to natural resource management problems organized by the *Canadian Journal for Forest Research*. She was invited to participate in a section on models for multi-scale analysis at the 2005 annual meeting of the Ecological Society of America in Montreal, and to review a collection of papers for inclusion in *EcoHealth*, a Springer publication in environmental science (to appear).

PI Hoeting, PI Davis, STARMAP-funded student Andrew Merton and S. E. Thompson (Pacific Northwest National Laboratory) developed important methodology for model selection in regression-like spatial models (called geostatistical models). In a paper that appeared in *Ecological Applications* these co-authors show that ignoring spatial correlation can result in models being selected that are not reflective of the variables that generated the data. This work also produced freely available software (<http://www.stat.colostate.edu/~jah/software/matern.R>) that has already been used in several publications to select explanatory variables for geostatistical models.

In closely allied work, Davis and Merton derived the limiting behavior for the maximum likelihood estimator of the range parameter in an exponential covariance function under various scenarios including infill and increasing domain asymptotics. Limit behavior for the case when the sampling strategy within blocks was clustered, regular, or random was also considered. It was shown that for the exponential case in one dimensional, all three samplings paradigms produced asymptotically equivalent estimates.

Investigator Gitelman, in collaboration with PI Hoeting and STARMAP-funded student Irvine (at OSU), has submitted a manuscript in which the properties of spatial covariance are examined in situations with varying amount of spatial dependence, under different sampling designs and for different sample sizes. This work demonstrated that large sample sizes are needed to accurately model spatially-dependent data and that sampling pattern can impact the quality of parameter estimates.

Investigator Gitelman and STARMAP-funded student Megan Dailey developed new methodology reported in the paper “Habitat selection models to account for seasonal persistence in radio telemetry data,” to appear in *Environmental and Ecological Statistics*. In this work, the authors build a flexible hierarchical model for seasonal persistence and seasonal changes in habitat selection, all fit using Bayesian modeling. The model was applied in an attempt to understand habitat selection behaviors in fish in Oregon streams.

PI Hoeting worked with ecologists to investigate various factors affecting the accuracy of predicted species distributions in a paper that appeared in *Ecological Applications* in 2005 (co-authors G. Reese, K Wilson, and C. Flather) . In related work on species distributions, Hoeting served as the discussant for ground-breaking work on Bayesian models for species distributions (conference and journal article in *Bayesian Analysis* in 2006). In another paper that appeared *Ecological Applications* in 2006, PI Hoeting (and co-authors M. Farnsworth, N.T. Hobbs, and M. W. Miller) developed models to examine how animal movement can be linked to the spread of disease. In a manuscript that appeared in *Biometrics* in 2003, PI Hoeting and STARMAP funded student D. Johnson developed new models for capture-recapture data. While not directly related to aquatic resource monitoring, the models that appeared in these publications will help ecologists further understand where species live, why they live there, and how diseases impact these organisms. This work, in conjunction with aquatic resource monitoring, will help us further understand entire ecosystems.

STARMAP-funded student D.S. Johnson, PI Hoeting, and N.L. Poff (a researcher on EPA-funded STAR project R8286301), developed new models for monitoring aquatic resource data. The new models are for multiple variable, categorical data where researchers are interested in modeling the proportion of observations in each group as well as the relationships between these groups and various explanatory variables. These models were used to investigate the relationship between fish traits and environmental predictors of the presence of fish with these traits, using the EPA EMAP MAHA data set. This work appeared in a recent book *Bayesian Statistics and its Applications* (2006).

Confidence Bounds for Map Contours

Josh French, a graduate student formerly supported by STARMAP, is working with Richard Davis devising confidence bands around level curves for a spatial field. The idea is that from a map

produced via kriging one can display level curves of the predicted conditional mean. However, calculation of error bounds for these level curves are not so easy to construct or even to define in a probabilistic sense. Davis and French are developing procedures that allows one to put confidence bands, displayed as small rectangular boxes that have a preset confidence probability. In other words, these boxes can be constructed so that we are 95% confident the process takes the desired threshold value somewhere in the box. The boxes are then linked together to give a “confidence set” for the level curve. This work is still ongoing, illustrating that the effect of STARMAP’s funding from EPA is extending past its funding period.

Structural Breaks in Time Series

PI Davis worked with colleague Thomas Lee and postdoctoral fellow Gabriel Rodriguez-Yam, exploring the problem of detecting structural breaks in a time series. The key idea was to assume that the nonstationary time series can be well represented by piecewise autoregressions. The principle of minimum description length (MDL) was used to assess the quality of fit for various structural break locations and the genetic algorithm was used to find near optimal minima of the MDL. The number of structural breaks, their locations, and the orders of the respective piecewise AR models were assumed unknown. This paradigm seemed to work well in a variety of applications. This research was published in the Journal of the American Statistical Association.

Methodology for sampling

Investigators Ritter and associates of the Southern California Coastal Water Research Project (SCCWRP) developed spatially distributed sampling plans suited to wastewater oceanic outfalls. Their project investigated cost effective ways to distribute sample points in a near-coastal system to support estimation of the spatial pattern of various analytes and macroinvertebrate indices. The work is summarized in a paper that appeared in *Environmental and Ecological Statistics*. Ritter applied these methods to produce a design which was implemented by the San Diego Metropolitan Wastewater District.

Some of Director Urquhart’s investigations of components of variance concerned temporal and spatial matters so are reported under this Project. When field visits to aquatic sites are widely distributed in space, and occasionally in time, components of variance can be used to capture most of the spatial and temporal variation, but not to model its form. Nevertheless such characterizations have proved useful in evaluating the likely (statistical) power to detect trend. The developed methodology was used to compare temporal or revisit designs relative to their power to detect trend. Some are decidedly better than others, a fact which was communicated to various interested clients.

1b. Training future environmental statisticians

A significant outcome of this project included a large number of students trained in environmental statistics. A number of these students are now working in fields directly related to environmental statistics in the U.S. The students involved in Project 1 include:

Ones who have completed their degrees:

- Devin Johnson, PhD from CSU in Statistics, 2003. Thesis title: *Bayesian Analysis of State-Space Models for Discrete Compositions*. After working for two years on the faculty at the University of Alaska, Fairbanks, Johnson is now

- a statistician at the National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA. Advisor: PI Hoeting.
- Steve Jensen, MS from OSU in Statistics, *Brief Introduction to Reversible Jump MCMC for Bayesian Networks and an Application* (Advisor: Investigator Gitelman). Jensen current work for a small graphical modeling software company which includes environmental applications.
- Brett Kellum, MS from CSU in Statistics, 2003. *Analysis and Modeling of Acid Neutralizing Capacity in the Mid-Atlantic Highlands Area*. Advisor: PI Hoeting.
- Andrew Merton, PhD from CSU in Statistics, 2006, *Geostatistical Models: Model Selection and Parameter Estimation under Infill and Expanding Domain Asymptotics*, Advisors: PIs Hoeting and Davis. Merton is currently working as a contract employee for the National Park Service. He is developing sampling plans for inventory and monitoring of vegetation, water quality and a number of other factors for a number of National Parks in the Western U.S.
- Julia Smith, MS from CSU in Statistics, 2005, *Modeling and Predicting Median Substrate Size in Oregon and Washington Streams Utilizing Geographic Information Systems*. This work was joint work with Brian Bledsoe, another EPA STAR-grant researcher. Ms. Smith teaches high school advanced placement statistics in Anchorage, AK, transmitting ideas of environmental statistics to future generations of environmental scientists.
- Sarah Williams, MS from CSU, 2006. *A Comparison of Variance Estimates of Stream Network Resources*. (Advisor: Director Urquhart).

Degrees currently in progress:

- Stephanie Fitchett, current MS student, CSU. Anticipated graduation date: 2007. Fitchett is extending work by Erin Peterson (STARMAP project 3) to develop user-friendly guidelines for GIS users as well as geostatistical models to be used to analyze stream monitoring data. (Advisors: Director Urquhart and PI Hoeting).
- Joshua French, current PhD student, CSU. Anticipated graduation date: 2008. French is developing models for uncertainty in contour lines for spatial maps, including contours for oceanic pollutants. (Advisor: PI Davis).
- Megan Dailey Higgs, current PhD student. Anticipated graduation date: 2007. Higgs is currently working on Bayesian models for ordered categorical spatial data and categorical habitat data. These models are being developed for analysis of a number of EPA data sets and will be particularly useful to model data on stream health and to predict health at unobserved sites. Advisor: PI Hoeting.
- Kathryn Irvine, current PhD student, OSU. Anticipated graduation date: 2007. Irvine's PhD work is in two separate, but related topics. She examined behavior of parameter estimates in geostatistical models in a paper currently submitted to a peer reviewed journal. For a related paper, Irvine was awarded 2nd place honorable mention in the student paper competition sponsored by the Section on Statistics in the Environment of the American Statistical Association in 2006. Her current Ph.D. work extends the work of Gitelman and Herlihy on several features of graphical models that will allow simple-to-understand graphical model diagrams explaining relationships for aquatic data.

In a position similar to Merton's described above, Irvine is also currently working as a contract employee for the National Park Service developing survey and monitoring plans for National Parks in the Western U.S.

Another outcome of the project is the post-doctoral training offered to Man Sik Park, who worked on the project from late 2005 through fall 2006. This work produced two manuscripts which have all been submitted to peer-reviewed journals and one working paper. Two of the manuscripts propose new ways to examine spatio-temporal data and one manuscript proposes new models for spatially-referenced data. The last manuscript, related to Gaussian Markov random fields, can be used to analyze very large data sets and is an improvement over existing methods.

The investigators on this project were widely scattered throughout the Western U.S. The project and the related STARMAP and DAMARS workshops allowed for additional interactions that would not have been possible without EPA funding. These links have and will continue to lead to the development of new methodology for to collect and analyze aquatic data. Interactions between Ritter and Urquhart; Hoeting and Gitelman; Merton, Theobald (Project 3), Urquhart, Ver Hoef and Peterson are just several examples of the new linkages that developed under this funding. All these interactions produced new methodology for the analysis of aquatic resource data.

- ◆ PROJECT 2 - Local Estimation: This Project had an overall goal of developing hierarchical spatio-temporal models for local inferences about aquatic resources. The project was conducted jointly with a DAMARS project on development of nonparametric model-assisted estimators for data obtained in probability surveys of aquatic resources. Accomplishments include
 - Extension of nonparametric model-assisted and model-based estimators for standard survey problems and for small area estimation problems;
 - Adaptation of deconvolution methods for spatial distribution function estimation;
 - Development of new state-space models and estimation methods for stream networks; and
 - Development of a novel algorithm (*spatial Lasso*) for selection of covariates and neighborhoods from GIS data in spatial regression problems.

The extensions of the nonparametric model-assisted methodology allow for a variety of complex designs and for incorporation of the major smoothing techniques in use today (including spline-based regression, additive models and semi-parametric models). The methods were applied to the general problems of estimation of population means, totals, and distribution functions. In many surveys, estimators are desired for small domains within the overall population. Because a survey often is not designed to provide reliable estimators for such small domains, the estimation requires the assumption of a model for the population. Investigators in this Project adapted the nonparametric methodology used in the model-assisted context to this situation, and showed how this approach generalizes existing small area estimation methods.

In addition to studying estimation of the distribution function in the design-based setting using nonparametric model-assisted estimators, Project personnel also investigated deconvolution, which is the estimation of the cumulative distribution function (cdf) of a variable given noisy measurements of that variable and distributional information about the measurement noise. They treated this problem as one of constrained Bayes estimation, which they extended to hierarchical Bayesian spatial models and studied under aggregation of small areas.

Because of the natural flow of water in a stream network, characteristics of a downstream reach may depend on characteristics of upstream reaches. The flow of water from reach to reach provides a natural time-like ordering throughout the stream network. Investigators in this Project developed a state-space model to describe the spatial dependence in this tree-like structure with ordering based on flow. The model formulation is flexible, allowing for a variety of spatial and temporal covariance structures in the state and measurement equations. They also derived a variation of the Kalman filter and smoother to allow recursive estimation of unobserved states and prediction of missing observations on the network, as well as computation of the Gaussian likelihood. The state-space formulation is extensible to non-linear and non-Gaussian processes. The Project investigators also developed several models of within-stream dependence, including network analogues of autoregressive-moving average models and of structural models, and fitted those models to real and simulated data.

Geographic Information Systems (GIS) organize spatial data in multiple two-dimensional arrays called layers. In many applications, a response of interest is observed on a set of sites in the landscape, and it is of interest to build a regression model from the GIS layers to predict the response at unsampled sites. Model selection in this context then consists not only of selecting appropriate layers, but also of choosing appropriate neighborhoods within those layers. Project investigators formalized this problem and proposed the use of Lasso (Least absolute shrinkage and selection operator) to simultaneously select variables, choose neighborhoods, and estimate parameters. They incorporated spatial smoothness in selected coefficients through use of *a priori* spatial covariance structure, leading to a modification of the Lasso procedure. The LARS algorithm, which can be used in a fast implementation of Lasso, was also modified to yield a fast implementation of spatial Lasso. The spatial Lasso performed well in numerical examples, including an application to prediction of soil moisture. The work reported in this paragraph was done in cooperation with investigators working under Project 3.

A number of graduate students were involved in this research, including Ji-Yeon Kim and Curtis Miller (Iowa State University), Alicia Johnson, Siobhan Everson-Stewart, Mark Delorey, and Bill Coar (Colorado State University). Work also involved a postdoctoral fellow, Giovanna Ranalli (Colorado State University), plus two junior researchers (Hsin-Cheng Huang, Academica Sincia and Nan-Jung Hsu, National Tsing-Hua University).

- ◆ PROJECT 3 - Development and Evaluation of Aquatic Indicators: This Project developed, tested, and distributed GIS-based tools that would facilitate computation of useful watershed

metrics for statistical analysis of aquatic indicator variables. Project investigators met these objectives through three key accomplishments by:

- Developing ArcGIS-based toolsets called FLoWS, FunConn and RRQRR;
- Conducting series of demonstrations and application of those tools to a range of key EPA constituents; and
- Collaborating with other scientists and training of students.

FLoWS (Functional Linkage of Waterbasins and Streams) is a set of tools that operate within ArcGIS v9 (written in Python). These tools allow users to rapidly generate a stream network, identify and correct topological errors in a network (fairly common in GIS data), extract watershed characteristics derived from other ancillary data such as topography, land cover, road density, etc. in a way that allows ecologically-relevant processes to be developed. For example, discharge volume (flow volume) is estimated as a function not only of waterbasin area, but also of the precipitation regime and the watershed topographic characteristics, including solar insolation and slope. Project investigators also developed a novel approach to identify catchments around stream reaches by identifying water basin boundaries using a cost-weight method, rather than relying on strictly local conditions (slopes) identified in a Digital Elevation Model. The goal was to ensure that these tools work with very large datasets (basins to nationwide) and in a variety of situations. For example, the traditional approach to identify watersheds failed in the Central Shortgrass Prairie that contains low relief, many intermittent streams, inconsistent network topology, as many local ridges caused the algorithms to “stop short” and caused “holes” or portions of watersheds that would not converge or connect with the larger network. Not only is the investigators’ new method robust and fast, preliminary analysis suggest that it is more accurate as well. For a small test study area in the Fraser River, Colorado water basin (high topographic relief), we found a higher mean accuracy using our novel methods as compared to the traditional method (i.e. 85% vs. 78% using Jaccard’s coefficient as compared to expert-based delineation of watershed boundaries at 1:24,000 scale).

The goal of the functional connectivity model, whose GIS implementation is named FunConn, is to allow landscape connectivity to be examined from a functional perspective. Functional connectivity recognizes that individuals, species or processes respond functionally (or behaviorally) to the physical structure of the landscape. From this perspective, landscape connectivity is specific to a landscape and species/individual/process under investigation.

Project investigators also strove to develop metrics and approaches that are more robust to possible data quality issues. For example, a well-known problem with “blue-line” hydrography is that the identification of streamlines can abruptly change at a topographic quadrangle boundary. Traditional metrics that rely on Strahler stream order, for example, are very sensitive to these issues, whereas water basin-area computations are more robust.

A key to the investigators’ accomplishment was close interaction and collaboration with a variety of constituents. Two major collaborations were with Oregon Department of Fish & Game (through collaborations with DAMARS personnel) and the Alaskan Department of Fish & Game. Project investigators participated in a variety of workshops and provided technical assistance throughout the STARMAP project.

Two unanticipated products of this Project were the result of synergistic activities. The key to each of these was an informal (initially) exchange of ideas, enthusiasm injected by graduate students involved (esp. Peterson), and the STARMAP Director's support for risk taking. For example, the new geostatistical method for stream networks developed by van Hoef, Peterson, and Theobald was the result of informal discussion at workshops, identification of an important research question, and eagerness of key individual (Peterson) who provided a key trans-disciplinary role. A second example is the development of a robust spatially-balanced sampling design algorithm implemented in ArcGIS, called the RRQRR tool (Reversed Randomized Quadrant-Recursive Raster). This is built fundamentally around Stevens and Olsen's GRTS algorithm, but development within a GIS framework provides the ability to develop a raster of sample locations, as well as extends a tool to a different (and broader) user base.

This project's investigators collaborated closely with investigators working on Projects 1 and 2. Specifically there was close cooperation between Erin Peterson and Andrew Merton, graduate students funded, respectively, under Projects 3 and 1. Merton, under the guidance of Hoeting (Project 1) and Davis (Project 2), developed computer software used extensively by Peterson, and adapted it to several of her special situations. This collaboration produced a jointly authored publication and presentations illustrating how the collaboration functioned. Peterson also collaborated with postdoctoral fellow Ranalli who conducted research under the auspices of Project 2. Another interaction involved Breidt (Project 2), Theobald (Project 3) and international visitors unfunded by STARMAP; this is described in more detail under Project 2. Further, Project 3 investigators developed or assisted in the developing covariate data sets for at least five other graduate student projects.

Erin Posten Peterson earned a PhD under the sponsorship of this Project. She is part way through a postdoctoral fellowship on aquatic monitoring in Queensland, Australia.

- ◆ PROJECT 4 - Extension and Outreach: The client community for the results of these two research programs consisted of aquatic monitoring scientists in state, tribal, federal, and more local agencies charged with monitoring aquatic resources in compliance with the Clean Water Act. Such aquatic scientists will be assisted by affiliated statisticians and landscape ecologists. Thus the outreach and extension efforts extended to each of these groups of scientists, using a variety of means. Program personnel
 - Engaged in direct interaction with personnel in relevant state, tribal, and more local environmental agencies.
 - Interacted with EPA personnel who have direct contact with aquatic monitoring personnel in the states and more local agencies.
 - Communicated with diverse members of the client community in a wide variety of conferences and other settings. Program personnel gave more than 200 talks and displayed at least 20 posters. The Program Directors organized and conducted five specialty conferences, and made substantial contributions to the organization and execution of several other conferences; most of these conferences included international participants. One of the Directors edited two issues of the *Journal of Environmental and Ecological Statistics* as part of this communication effort.

- Presented short courses and tutorials in several settings, some based on a textbook coauthored by one of the Program's principal investigators.
- Developed a set of browser-based learning materials suitable for self-study that has been submitted as part of this report. EPA is free to utilize those materials as it sees fit.
- Gave presentations to teachers and students in high school advanced placement statistics courses to interest students in possible careers in environmental statistics.

This Project had three main modes of outreach and extension: Direct implementation of project tools in the design and execution of active state, regional, tribal, and more local monitoring efforts, primarily executed by Stevens (DAMARS) and Theobald (STARMAP); a CD-ROM containing relevant learning materials, directed by Urquhart (STARMAP); and communication with various communities by publications, oral presentations, and posters in diverse contexts, executed by all members of both Programs.

- Most of the personnel of both STARMAP and DAMARS participated in interactions with various members of the client community, but major and sustained efforts in this area were made by Don Stevens, the Director of DAMARS, and David Theobald, a STARMAP Principal Investigator. Don Stevens interacted throughout the Programs' lives with the San Francisco Estuary Institute and other California agencies, and several Oregon state agencies. Toward the end of the Programs, he and, to a lesser extent, David Theobald and collaborators, interacted with the Alaska Department of Fish and Wildlife on the design of monitoring activities related to Pacific salmon. David Theobald, a STARMAP landscape ecologist, developed GIS tools relevant to aquatic monitoring. The nature of these tools are documented in the report for Project 3. Theobald and collaborators made these tools available to diverse potential users and gave a number of presentations to the client community on their use; the requesters for these tools span a wide range of aquatic interests, including a surprising number from outside the US. In the case of Alaska salmon they actually installed the software and demonstrated its use on computers of the Alaska Department of Fish and Game. Scott Urquhart, the STARMAP Director, had continuing interactions with the National Park Service's Inventory and Monitoring Program, and with the San Diego Metropolitan Wastewater treatment activity in collaboration with Kerry Ritter of the Southern California Coastal Water Research Project.
- STARMAP results have received international exposure, both by current investigators and former investigators who have gone on to other positions, but continue to report on results originating from their work with STARMAP: Germany (Jennifer Hoeting, 2002; Richard Davis, 2005), New Zealand (Jennifer Hoeting, 2003), France (Jean Opsomer, 2003), Australia (Jean Opsomer, 2003; Richard Davis, 2003 & 2004; Don Stevens, 2005 & 2006; Erin Petersen, 2005 & 2006; Loveday Conquest, 2005), Denmark (Richard Davis, 2003 & 2004), Greece and Cyprus (Richard Davis, 2004), India (Jay Breidt, 2004; Jennifer Hoeting, 2005), Sweden (Jean Opsomer, 2004; Richard Davis, 2004 & 2006; Don Stevens and Loveday Conquest, 2006), Singapore

(Jay Breidt, 2004), Spain (Jean Opsomer, 2004), Portugal (Richard Davis, 2004 & 2006; Don Stevens, 2006), Japan (Richard Davis 2004 & 2006), Italy (Giovanni Ranalli, 2005 & 2006), Czech Republic (Richard Davis, 2005 & 2006), Pacific Rim Nations (Don Stevens, 2005), and Republic of China (Richard Davis, 2006). Most of this international exposure came at the expense of cooperating organizations; STARMAP funded, with prior approval, only three of the trips to give these talks.

◆ **PROJECT 5 - Integration and Coordination:**

The investigators planned, coordinated and supported research under STARMAP. The Director

- Monitored the progress of Projects 1 - 4, including oversight of their budgets;
- Oversaw staffing and coordination;
- Assured the availability of needed computing resources and data;
- Monitored subcontracts to
 - Oregon State University;
 - The Southern California Coastal Water Research Project;
 - The University of Alaska Fairbanks;
- Assembled and submitted quarterly and annual reports;
- Coordinated various matters with DAMARS, the Oregon State University Program, the Southern California Coastal Water Research Project, and with the Program's EPA Project Officer.

The STARMAP Director organized and executed the first and third Annual Conferences on Statistical Survey Design and Analysis For Aquatic Resources with assistance from Don Stevens, the Director of the DAMARS Program. The Director also assisted Don Stevens in organizing and executing the second and fourth Annual Conferences on Statistical Survey Design and Analysis for Aquatic Resources. These conferences were prescribed in the Request For Proposals which led to funding for STARMAP and DAMARS.

2. SIGNIFICANCE OF ACCOMPLISHMENTS

- ◆ **PROJECT 1 - Combining Environmental Data Sets:** A major accomplishment of Project 1 was the training of statisticians in environmental problems. As described above, a number of these students are already using the knowledge gained from working on the STARMAP project in various government agencies and private businesses.

The model selection component of the research accomplishments serves as a warning to scientists in selecting covariates in geospatial models. This process needs to be conducted in concert with the modeling of the error term. Often the error term can be used as a proxy for missing covariates or can be used as a correction factor for incorrectly selected covariates.

The asymptotic theory for the exponential covariance function is useful to the scientist for selecting optimal sampling strategies with the goal of producing the most efficient parameter estimates. In the exponential case, it is difficult to beat a uniform sampling plan. This might change for the Matern covariance function, which is a topic of future study.

The structural break detection research has shown great promise in segmenting a time series into stationary segments. The strategy developed, called AutoPARM for Automatic Piecewise AR Modeling, is a general procedure that overcomes many of the limitations and defects of previously proposed procedures. In addition, very few assumptions are made in this framework. We intend to explore versions of AutoPARM that would apply more directly in the geospatial context.

Ritter of the Southern California Coastal Water Research Project (SCCWRP) investigated cost effective ways to distribute sample points in near-coastal systems to support estimation of the spatial pattern of various analytes and macroinvertebrate indices. The resulting design was implemented by the San Diego Metropolitan Wastewater District. This design will likely serve as a prototype for many similar studies along the California coast, developed in collaboration with the Southern California Coastal Water Research Project.

The work on power to detect trend in aquatic surveys by Urquhart has been and is being used by designers of both aquatic and terrestrial surveys to make effective use of limited resources. This work is widely cited in publications concerned with the design of environmental surveys.

- ◆ PROJECT 2 - Local Estimation: This Project developed statistical analysis tools of relevance to aquatic scientists concerned with surveys, and spatial-temporal modeling. The client community for the results of this research program consisted of aquatic monitoring scientists in state, tribal, federal, and more local agencies charged with monitoring aquatic resources in compliance with the Clean Water Act. Such aquatic scientists will be assisted by affiliated statisticians and landscape ecologists.
- ◆ PROJECT 3 - Development and Evaluation of Aquatic Indicators: The final report for this Project lists known adopters of the methodology developed under this Project. The diversity of adopters speaks eloquently to the current significance of the accomplishments of this Project. Adopters ranged from local environmental agencies to national environmental agencies, from governmental agencies to nonprofit agencies to academics, from across the United States to across the world. The acceptance of the products of this Project was due to two factors: The relevancy of the products developed, and the scale of the outreach activities associated with them. The specific outreach activities are documented under Project 4, but it needs to be noted here that this Project's investigators made a major effort in this area.
- ◆ PROJECT 4 - Extension and Outreach: This Project has effectively communicated the results of the STARMAP and DAMARS Programs to diverse members of the client community in a variety of effective ways. Most of the communications have been contemporary, but the CD ROM based learning materials can be used in several ways, including being available on the internet.
- ◆ PROJECT 5 - Integration and Coordination: These accomplishments of Project 5 were critical for the success of STARMAP.

3. STAKEHOLDERS AND USERS OF RESULTS

Most analyses of aquatic environmental data will involve several professional specialties, ordinarily an environmental scientist familiar with the nature of the aquatic indicators, a statistician, and a landscape ecologist or GIS specialist at a minimum. Thus we need to distinguish between immediate and eventual stakeholders. The immediate stakeholders will be responsible for the analysis and technical interpretation. The eventual stakeholders will be aquatic managers and through them all beneficiaries of our Nation's aquatic resources. Furthermore, some of STARMAP's research contributions are immediate; others will take longer to be utilized. The results of Project 3 are already being used extensively by diverse stakeholders. Likewise some of the results of Projects 1 and 2 have already been utilized in published research. On the other hand, some of the results of Projects 1 and 2 will form the basis for applications yet to be recognized by the aquatic community, but probably will become used in the future. (Statistic abounds with examples of things which seemed very abstract and of no immediately recognized utility 20 years ago, but are widely used now.)

The results of STARMAP research will be of direct use primarily by the professional specialists mentioned above. Thus this section focuses on those stakeholders.

- ◆ PROJECT 1 - Combining Environmental Data Sets: Since Project 1 covered a wide variety of topics, the stakeholders and users of the results also vary widely.

The textbook *Computational Statistics*, has been adopted by a number of universities including Stanford University, The Ohio State University, University of Minnesota, Bowling Green State University, and others. The book is also sold internationally. It is anticipated that as more universities adopt the textbook, this work will be disseminated throughout the U.S. and used to educate statisticians, with particular emphasis on the education future environmental statisticians due to the ecological examples used in the book.

This textbook has resulted in four short courses, to date. These courses have served a wide audience of U.S. based statisticians, including a number of statisticians working for Federal and State government agencies, including several from environmental agencies. The courses include:

- Statisticians attending the Joint Statistical Meetings, August 2006, Seattle, WA.
- Statisticians affiliated with the Alaska chapter of the American Statistical Association (Participants were mainly employed by U.S. Fish and Wildlife Service, Alaska Fish and Game Department, and affiliated with the University of Alaska, Fairbanks).
- Statisticians attending the Joint Statistical Meetings, [August 2005, Minneapolis](#).
- Statisticians attending the American Statistical Association Section on Statistics and the Environment, and University Chicago Center for Integrating the Statistical and Environmental Sciences. [October 2004, Chicago](#).

We expect that the modeling tools that we have developed will be useful to a wide range of users of statistics in the environmental sciences, geosciences, biology and medical professions. A current example of the structural break work deals with record sounds in the National Parks. There is interest in segmenting a long audio stream of recordings obtained

from microphones strategically placed in many of the National Parks. One of the goals of this data collection process was to allow the US Park Service to measure and monitor noise pollution in the national parks. One measure of the pollution is the proportion of unnatural (man-made) sound heard in the parks. Auto-PARM, software developed by the investigators, can be used to help segment the audio tapes into homogeneous pieces of natural and unnatural (e.g., snow mobile and jet plane noise) sounds. After segmenting the sounds, we attempt to classify the individual pieces into various categories of known sound types. This research is of value to various government agencies, including the National Park Service.

With regards to Ritter's work, the San Diego Metropolitan Wastewater District used Ritter's design. Other stakeholders include the people and other living organisms who utilize near-coastal waters and adjacent beaches. Her work shows how to design cost-effective studies of the consequences of oceanic wastewater outfalls.

The immediate users of Urquhart's work are agency personnel who have to design ecological surveys; these include state-level environmental scientists in many states, including Oregon, Wisconsin, and Maryland. Longer term we all benefit from better and more defensible information being gathered in cost-effective ways.

- ◆ PROJECT 2 - Local Estimation: There are many potential users of the methods developed by STARMAP and DAMARS under this joint Project. The two Programs have organized and presented a number of conferences or parts of conferences directed specifically at potential users. Program personnel also have participated in a number of conferences at the invitation of potential users. Some of the conferences are explained in more detail under Project 4, Outreach and Extension.

- ◆ PROJECT 3 - Development and Evaluation of Aquatic Indicators: This project has written and made web-available three sets of GIS tools oriented toward the design and statistical analysis of data resulting from studies in aquatic systems. The tools are programmed in Python and accessible as ArcGIS tools in v9. Each is further documented and available through this web site:
<http://www.nrel.colostate.edu/projects/starmap/>
 - *FLoWS v1: Functional Linkage of Watersheds and Streams tools for ArcGIS v9:*
The goal of the functional linkage of watersheds and streams tools is to allow aquatic and terrestrial landscapes to be hydrologically-linked. In this sense, relationships between sites can be represented through functional distance measures. For many hydrological processes (not all!) downstream flow direction is an important ecological process, so that distance is not symmetric. Also, including important landscape attributes that modify the degree to which nearby locations are connected is important. This would include topographic considerations such as stream gradient and slope, as well as features that might impede the movement of a species or process such as waterfalls, dams, or certain vegetation types. As of December 1, 2006, 32 agencies and organizations requested and received a copy of the FLoWS software. This includes 18 domestic and 14 international users, ranging from EPA

organizations to state agencies to university users to nonprofit agencies to commercial firms. A detailed list appears in the final report for Project 3.

- *Functional Connectivity tools (FunConn)*: There is a large and critical difference between simple hydrological datasets that “look” correct on a map, but must have correct topology and attribution to run network-based algorithms correctly. The goal of the functional connectivity model is to allow landscape connectivity to be examined from a functional perspective. Functional connectivity recognizes that individuals, species or processes respond functionally (or behaviorally) to the physical structure of the landscape. From this perspective, landscape connectivity is specific to a landscape and species/individual/process under investigation. As of December 1, 2006, 47 agencies and organizations requested and received a copy of the FunConn software. This includes 22 domestic and 25 international users, ranging from EPA organizations to state agencies to university users to nonprofit agencies to commercial firms. A detailed list appears in the final report for Project 3.
- *Spatially-balanced sampling using RRQRR*: The goal of the Reversed Randomized Quadrant-Recursive Raster (RRQRR) algorithm is to provide environmental managers a practical, useful GIS tool to generate simple, efficient, and robust survey designs for natural resource applications. RRQRR generates a rigorous probability-based survey design that is spatially-balanced and allows surfaces to be used to specify the inclusion probability. As of December 1, 2006, 9 domestic agencies and organizations have requested and received a copy of the RRQRR software. These users range from US governmental organizations to state agencies to university users to commercial firms. A detailed list appears in the final report for Project 3. The relevance of this software is indicated by the fact that Environmental Systems Research Institute (makers of ArcGIS software) – are currently implementing the RRQRR algorithm into their core software package.
- In addition to building tools for statistical analysis of hydrology, this project generated a database (called the FLoWS database) that builds on USGS National Elevation Data and National Hydrography Dataset (1:100K). This effort provided two important benefits:
 - A nationwide, pre-processed and pre-packaged dataset that will support many types of hydrological analysis and provides a significant “head-start” for EPA clients; and
 - Nationally-consistent, hierarchical, and high-resolution catchment boundaries at a variety of scales – from basins (HUC 2s) to roughly the HUC 14 level.

- ◆ PROJECT 4 - Extension and Outreach: There are many potential users of the methods developed by STARMAP and DAMARS. The two Programs have organized and presented a number of conferences or parts of conferences directed specifically at potential users. Program personnel also have participated in a number of conferences at the invitation of potential users. Some of the conferences listed below are explained in more detail in the final Report for STARMAP Project 4. The conferences listed immediately below had a major involvement of potential users of developed methods.

- Don Stevens (DAMARS) and Tony Olsen (EPA, WED) presented the short course **Spatial Sampling** at the Joint Statistical Meetings, Seattle, WA, August, 6-10, 2006
- David Theobald (STARMAP), John Norman (STARMAP), Erin Peterson (formerly of STARMAP), and Don Stevens (DAMARS) were the major presenters at the Workshop on Modeling Salmon Habitat, hosted by the Alaska Department of Fish & Game (ADF&G) and The Nature Conservancy, Anchorage, Alaska, May 17-19, 2006.
- ◆ Conference on Statistics for Aquatic Resources Monitoring, Modeling, and Management (SARMMM), hosted by DAMARS: This was the Fourth Annual Conference on Statistical Survey Design and Analysis for Aquatic Resources, Corvallis, OR, September 7 - 9, 2005. It was well attended by both statisticians and natural resource managers, especially ones from the Pacific Northwest, but other attendees came from across the US, and several from other countries. This included a well-attended day-long short course (9/7/05) for natural resource managers on Designing Aquatic Resources Surveys.
- Monitoring Science and Technology Symposium, Denver, CO, September 21-24, 2004: The STARMAP Director organized two and a half days of sessions on the statistical aspects of natural resource surveys. Stevens (DAMARS) and Urquhart (STARMAP) made major presentations as a part of the short course. Three STARMAP graduate students also made presentations in other conference sessions. This short course was well attended by resource managers from throughout the US and elsewhere in the Americas.
- The Third Annual Conference on Statistical Survey Design and Analysis for Aquatic Resources, Fort Collins, CO, September 10 - 11, 2004, hosted by STARMAP: This conference was well attended by Program participants, statisticians and natural resource managers, especially ones from EPA Region 8, but other attendees came from across the US, and included at least one international participant. The keynote address was given by an EPA resource manager.
- Graybill Conference, June 16 - 18, 2004: This conference constituted outreach to the statistical community, providing STARMAP and DAMARS investigators with an opportunity to present some of their results to both established and young statisticians. It was attended by about 80 statisticians, and was officially sponsored by STARMAP. The STARMAP Director served as editor of a special edition of *Environmental and Ecological Statistics* which served as the conference proceedings, and provided an outlet for several young environmental statisticians. The keynote address was given by a member of the Programs' Science Advisory Committee.
- EMAP Symposium 2004, Newport, RI, May 3 - 7, 2004: The STARMAP director organized a session on the statistical aspects of the linkage of Clean Water Act sections 305(b) and 303(d) and he and another STARMAP PI made presentations.

- Temporal Sampling Workshop, Port Angeles, WA, November 12 - 14, 2003, sponsored by the National Park Service: Urquhart (STARMAP), and Stevens (DAMARS) gave invited talks, and Ranalli (STARMAP) participated.
 - The Second Annual Conference on Statistical Survey Design and Analysis for Aquatic Resources, Corvallis, OR, August 11 - 13, 2003, hosted by DAMARS: This conference was well attended by Program participants, statisticians and local natural resource managers, including ones from EPA, WED and Oregon Department of Fish and Wildlife. The keynote address was given by an EPA resource manager.
 - The First Annual Conference on Statistical Survey Design and Analysis for Aquatic Resources, Fort Collins, CO, September 20 - 11, 2002, hosted by STARMAP: It was well attended by Program participants, statisticians and local natural resource managers, including ones from EPA Region 8. The keynote address was given by an EPA resource manager.
- ◆ PROJECT 5 - Integration and Coordination: The immediate stakeholders were the investigators and students involved in the Program. In a larger sense, though, all potential and eventual users of the results of STARMAP were the stakeholders and users.

4. HOW PRODUCTS WILL FURTHER SCIENCE/ MANAGEMENT OF RESOURCES

The products developed and disseminated by STARMAP will support the analysis of aquatic responses in diverse contexts, but will be especially useful in analyzing aquatic data associated with specific sample points. These tools and demonstrations will support the more accurate and defensible analysis of diverse environmental indicators.

- ◆ PROJECT 1 - Combining Environmental Data Sets: The statistical methodology, software, textbook, papers, short courses and talks, all have and will continue to support the design of aquatic environmental studies and the analysis of the resulting data in diverse contexts. These tools and demonstrations will support the more accurate and defensible analysis of diverse environmental variables.

The training of statisticians versed in methods for modeling aquatic data have already resulted in new statisticians working in positions related to ecology and aquatic resources.

The design results of Ritter's work will allow better management of wastewater outfalls. The associated current work of French demonstrates the uncertainty associated with the sorts of contours often computed by GIS applications, and used in regulatory statements. The latter eventually will be improved by ongoing work in setting confidence intervals for map contours.

- ◆ PROJECT 2 - Local Estimation: The statistical analysis tools (products) developed and disseminated by this Project provide aquatic scientists and affiliated statisticians with expanded and more defensible ways to draw inferences to local concerns from wide-area surveys than were available prior to this Project. These tools extend previously available spatial-temporal methods to accommodate the branching nature of streams and rivers.

- ◆ PROJECT 3 - Development and Evaluation of Aquatic Indicators: These products will support the analysis of aquatic responses in diverse contexts, but will be especially useful in developing landscape indicators associated with specific aquatic sample points. These tools and demonstrations will support the more accurate and defensible analysis of diverse environmental variables.
- ◆ PROJECT 4 - Extension and Outreach: The products of this project are both tangible (the manuscripts, presentations, software, and learning tools) and intangible (the insight and knowledge of environmental sampling that has been passed on to non-statistician environmental scientists and managers). The tangible products developed by STARMAP and DAMARS provide an expanded tool kit for designing, monitoring, and analyzing the resulting data resulting from studies of aquatic resources at a variety of levels, from national surveys to studies near a single oceanic sewage outfall. The intangible products have the potential for greater impact on how the resources of this Nation are managed. DAMARS and STARMAP have demonstrated the utility of rigorous statistical design and analysis of environmental monitoring program to diverse parts of the client community for which these tools were developed and evaluated. Thus, we have not only provided tools, but we have provided the tools to people who are in a position to apply them, and given them the knowledge to do so. There is substantial evidence that the use of tool kit is spreading... more states, tribes, federal agencies, even other countries, are recognizing its utility.
- ◆ PROJECT 5 - Integration and Coordination: The products of this Project furthered the science/management of resources by enabling Projects 1 - 4 to do this.

5. LISTING OF STARMAP PUBLICATIONS

Note that this list is a subset of the entire output list from STARMAP; the complete list, supplied as a separate document as a part of this final report, provides internal links to most of the presentation materials, technical reports, and some manuscripts. Published material ordinarily was copyrighted by the publisher, so access to it usually is restricted to subscribers of that publication. The subset below includes only those outputs which have been published, or accepted for publication.

- ◆ Breidt, F.J., J.D. Opsomer, A.A. Johnson, and M.G. Ranalli. Semiparametric model-assisted estimation for natural resource surveys. To appear in *Survey Methodology* (Accepted November 29, 2006).
- ◆ Theobald, D.M., D.L. Stevens, Jr., D. White, N.S. Urquhart, A.R. Olsen, and J.B. Norman. Using GIS to generate spatially-balanced random survey designs for natural resource applications. *Environmental Management* (Accepted July 17, 2006).
- ◆ Peterson, E.E., D.M. Theobald, and J.M. Ver Hoef (2007). Support for geostatistical modeling on stream networks: Developing valid covariance matrices based on hydrologic distance and stream flow. To appear in *Freshwater Biology* (November, 2006).
- ◆ Andrews, B., R.A. Davis, and F.J. Breidt (2007). Rank-based estimation for all-pass time series models. To appear in the *Annals of Statistics* April, 2007
- ◆ Ritter, K.J. and M.K. Leecaster (2007). Multi-Lag cluster enhancement of fixed grid sample designs for estimating the variogram in near coastal systems. *Environmental and Ecological Statistics* **14**: .

- ◆ Gitelman, A.I. and A.T. Herlihy (2007). Isomorphic chain graphs for modeling spatial dependence in ecological data. *Environmental and Ecological Statistics* **14**: .
- ◆ Dailey, M.C., A.I. Gitelman, F.L. Ramsey, and S. Starcevich. Habitat selection models to account for seasonal persistence in radio telemetry data. *Environmental and Ecological Statistics* **14**: .
- ◆ Theobald, D.M. and J.B. Norman (2007). FLoWS (Functional Linkage of Water basins and Streams): ArcGIS tools for network-based analysis of freshwater ecosystems. *HydroLine*, ESRI Press (in press).
- ◆ Breidt, F.J., N.-J. Hsu, and S.M. Ogle (2006). Semiparametric mixed models for increment-averaged data with application to carbon sequestration in agricultural soils. To appear in the *Journal of the American Statistical Association*
- ◆ Brockwell, P.J., R.A. Davis, and V. Yang (2006). Continuous-time Gaussian autoregression. To appear in *Statistica Sinica* .
- ◆ Wang, H. and M.G. Ranalli (2006). Low-rank smoothing splines on complicated domains. To appear in *Biometrics* **62**:
- ◆ da Silva, D.N. and J.D. Opsomer (2006). A kernel smoothing method to adjust for unit nonresponse in sample surveys. To appear in the *Canadian Journal of Statistics* **34**:.
- ◆ Francisco-Fernandez, M., M. Jurado-Exposito, J.D. Opsomer, and F. Lopez-Granados (2006). A nonparametric analysis of the distribution of *Convolvulus arvensis* in wheat-sunflower rotations. *Environmetrics* **17**:849-860.
- ◆ Breidt, F.J., N.-J. Hsu, and W.J. Coar (2006). A diagnostic test for autocorrelation in increment-averaged data with application to soil sampling. To appear in *Environmental and Ecological Statistics* **13**:.
- ◆ Opsomer, J.D., F.J. Breidt, G.G. Moisen, and G. Kauermann (2006). Model-assisted estimation of forest resources with generalized additive models. To appear in the *Journal of the American Statistical* **101**: as a Discussed Paper in Applications and Case Studies.
- ◆ Ver Hoef, J.M., E.E. Peterson, and D.M. Theobald (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological Statistics* **13**: 449-464.
- ◆ Peterson, E.E. and N.S. 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**:613-636.
- ◆ Peterson, E.E., A.A. Merton, D.M. Theobald, and N.S. Urquhart (2006). Patterns of spatial autocorrelation in stream water chemistry. *Environmental Monitoring and Assessment* **121**: 569-594
- ◆ Pratesi, M., N. Salvati, and M.G. Ranalli. P-splines M-quantile small area estimation: assessing the ecological conditions of lakes in the Northeastern US. Proceedings, Conference on Spatial Data Methods for Environmental and Ecological processes, Foggia, Italy, September 14-15, 2006
- ◆ Thomas, D.E., D.S. Johnson, and B. Griffith (2006). A Bayesian random effects discrete-choice model for resource selection: Population-level selection inference. *Journal of Wildlife Management* **70**: 404-412.
- ◆ Farnsworth, M.L., J.A. Hoeting, N.T. Hobbs, and M.W. Miller (2006). Linking chronic wasting disease to mule deer movement scales: A hierarchical Bayesian approach. *Ecological Applications* **16**: 1026–1036.

- ◆ Andrews, B., R.A. Davis, and F.J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis* **97**: 1638-1659.
- ◆ Hoeting, J.A., R.A. Davis, A.A. Merton, and S.E. Thompson (2006). Model selection for geostatistical models. *Ecological Applications* **16**: 87-98.
- ◆ Theobald, D.M. (2006). Exploring the functional connectivity of landscapes using landscape networks. To appear in: Crooks, K.R. and M.A. Sanjayan (eds.). pp416-444 in **Connectivity Conservation: Maintaining Connections for Nature**. Cambridge University Press. 726 pages.
- ◆ Davis, R.A., T.C.M. Lee, and G. Rodriguez-Yam (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association* **101**: 223-239.
- ◆ Johnson, D.S., J.A. Hoeting, and N.L. Poff (2006). Biological monitoring: A Bayesian model for multivariate compositional data. pp 270-289 in *Bayesian Statistics and its Applications*, Eds. S. K. Upadhyay, U. Singh and D. K. Dey. Anamaya publishers: New Delhi.
- ◆ Hoeting, J.A. (2006). Some perspectives on modeling species distributions. Discussion of article by A.E. Gelfand, J.A. Silander, S. Wu, A. Latimer, P.O. Lewis, A.G. Rebelo, M. Holder. *Bayesian Analysis* **1**: 93-98.
- ◆ Breidt, F.J., G. Claeskens, and J.D. Opsomer (2005). Model-assisted estimation for complex surveys using penalized splines. *Biometrika* **92**: 831-846.
- ◆ Davis, R.A., T.C.M. Lee, and G.A. Rodriguez-Yam (2005). Structural break estimation for nonstationary time series signals. Proceedings of IEEE/SP 13th Workshop on Statistical Signal Processing, Bordeaux, France, July, 2005.
- ◆ Montanari, G.E. and M.G. Ranalli (2005). Nonparametric methods for sample surveys of environmental populations. Proceedings of the Meeting of the Italian Statistical Society on Statistics and the Environment. September 21-23, 2005, Messina, Italy **CLEUP**: 147-158.
- ◆ Davis, R.A. and G.A. Rodriguez-Yam (2005). Estimation for state-space models based on a likelihood approximation. *Statistica Sinica* **15**: 381-406.
- ◆ Breidt, F.J. and N.-J. Hsu (2005). Best mean square prediction for moving averages. *Statistica Sinica* **15**: 427-446.
- ◆ Givens, G.H. and J.A. Hoeting (2005). **Computational Statistics**. John Wiley & Sons, New York, 418 pages.
- ◆ Francisco-Fernandez, M. and J.D. Opsomer (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian Journal of Statistics* **33**: 279-295. {
- ◆ Opsomer J.D. and C.P. Miller (2005). Selecting the amount of smoothing in nonparametric regression estimation for complex surveys. *Journal of Nonparametric Statistics* **17**: 593-611.
- ◆ Theobald, D.M. , J.B. Norman, E.E. Poston, and S.B. Ferraz (2005). Functional linkage of watersheds and streams: ArcGIS FLoWS tools. *Proceedings, ESRI User Conference 2005*.
- ◆ Montanari, G.E. and M.G. Ranalli (2005). Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association* **100**: 1429-1442.

- ◆ Reese, G.C., K.R. Wilson, J.A. Hoeting, and C.H. Flather (2005). Factors affecting the accuracy of predicted species distributions: A simulations experiment. *Ecological Applications* **15**: 554-564.
- ◆ Hall, P. and J.D. Opsomer (2005). Theory for penalised spline regression. *Biometrika* **92**: 105-118.
- ◆ Kauermann, G. and J.D. Opsomer (2004). Generalized cross-validation for bandwidth selection of backfitting estimators in generalized additive models. *Journal of Computational and Graphical Statistics* **13**: 66-89.
- ◆ Opsomer, J.D., F.J. Breidt, G. Claeskens, G. Kauermann, and M.G. Ranalli (2004). Nonparametric small area estimation using penalized spline regression. *Proceedings of the Section on Survey Research Methods* [CD-ROM], Alexandria, VA. American Statistical Association: 4127-4134.
- ◆ Kahl, J.S., J.L. Stoddard, R. Haeuber, S.G. Paulsen, R. Birnbaum, F.A. Deviney, J.R. Webb, D.R. DeWalle, W. Sharpe, C.T. Driscoll, A.T. Herlihy, J.H. Kellogg, P.S. Murdoch, K. Roy, K.E. Webster, and N.S. Urquhart (2004). How have U.S. surface waters responded to the 1990 Clean Air Act Amendments? Feature article, *Environmental Science and Technology* **38**: 484A-490A.
- ◆ Courbois, J.P. and N.S. Urquhart (2004). Comparison of survey estimates of the finite population variance. *Journal of Agricultural, Biological and Environmental Statistics* **9**: 236-250.
- ◆ da Silva, D.N. and J.D. Opsomer (2004). Properties of the weighting cell estimator under a nonparametric response mechanism. *Survey Methodology* **30**: 45-55.
- ◆ Kincaid, T. M., D.P. Larsen, and N.S. Urquhart (2004). The structure of variation and its influence on the estimation of status: Indicators of condition of lakes in the Northeast, U.S.A. *Journal of Environmental Monitoring and Assessment* **98**: 1-21.
- ◆ Larsen, D.P., P.K. Kaufmann, T.M. Kincaid, and N.S. Urquhart (2004). Detecting persistent change in the habitat of salmon-bearing streams in the Pacific Northwest. *Canadian Journal of Fisheries and Aquatic Sciences* **61**: 283-291.
- ◆ Waite, I.R., A.T. Herlihy, D.P. Larsen, N.S. Urquhart, and D.J. Klemm (2004). The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: an example from the Mid-Atlantic Highlands, U.S.A. *Freshwater Biology* **49**: 474-489.
- ◆ Urquhart, N.S. and J.C. Moore (2004). Statistics in EPA's STAR Program: Learning materials for surface water monitoring. *OPPTS Tribal News* **4(3)**: 45 - 46. (EPA 745-N-00-001)
- ◆ Opsomer, J.D., C. Botts and, J.Y. Kim (2003). Small Area estimation in a watershed erosion assessment survey. *Journal of Agricultural, Biological and Environmental Statistics* **8**:139-152. {
- ◆ Opsomer, J.D. and F.J. Breidt (2003). Nonparametric estimation in complex surveys with auxiliary information. *Proceedings of the 54th Session of the International Statistical Institute*. Available only on conference CD
- ◆ Johnson, D.S. and J.A. Hoeting (2003). Autoregressive Models for Capture-Recapture Data: A Bayesian Approach. *Biometrics* **59**: 341-350.
- ◆ Davis, R.A., W.T.M. Dunsmuir, and S.B. Streett (2003). Observation-driven models for Poisson counts. *Biometrika* **90**: 770-790.

- ◆ Hoeting, J.A. (2002). Methodology for Bayesian model averaging: An update. Proceedings of invited paper presentations, International Biometric Conference, Freiburg, Germany, pp 231-240.
- ◆ Larsen, D.P., T.K. Kincaid, S.E. Jacobs, and N.S. Urquhart (2001). Designs for evaluating local and regional scale trends. *Bioscience* **51**: 1069-1078.

6. COMPONENTS OF STARMAP's FINAL REPORT

- ◆ This overall FINAL TECHNICAL REPORT, electronically submitted December 15, 2006, and named STARMAP.Final.Report.pdf.
- ◆ FINAL TECHNICAL REPORT: PROJECT 1 - COMBINING ENVIRONMENTAL DATA SETS, electronically submitted and named STARMAP.Project1.Final.Report.pdf.
- ◆ FINAL TECHNICAL REPORT: PROJECT 2 - LOCAL INFERENCE, electronically submitted December 15, 2006, and named STARMAP.Project2.Final.Report.pdf.
- ◆ FINAL TECHNICAL REPORT: PROJECT 3 -DEVELOPMENT AND EVALUATION OF AQUATIC INDICATORS, electronically submitted December 15, 2006, and named STARMAP.Project3.Final.Report.pdf.
- ◆ FINAL TECHNICAL REPORT: PROJECT 4 - EXTENSION and OUTREACH, electronically submitted December 15, 2006, and named STARMAP.DAMARS.Outreach.Final.Report.pdf.
- ◆ FINAL TECHNICAL REPORT: PROJECT 5 - INTEGRATION and COORDINATION for STARMAP, electronically submitted December 15, 2006, and named STARMAP.Project5.Final.Report.pdf.
- ◆ Compilation of Outputs with links to presentations and posters. This was submitted on two CD ROMs which have these materials and internal links to them. (Three copies, mailed with the following entry). The content, but not linked materials, were submitted electronically December 15, 2006, and named STARMAP.Output.Summary.pdf.
- ◆ Learning Materials. This was submitted on two CD ROMs which have these materials and internal links to them. (Three copies, mailed with the preceding entry)
- ◆ Web summary of this overall FINAL TECHNICAL REPORT, electronically submitted December 18, 2006, and named STARMAP.Web.Summary.Report.pdf.
- ◆ Web summary of FINAL TECHNICAL REPORT: PROJECT 1 - COMBINING ENVIRONMENTAL DATA SETS, electronically submitted December 18, 2006, and named STARMAP.Project1.Web.Summary.Report.pdf.
- ◆ Web summary of FINAL TECHNICAL REPORT: PROJECT 2 - LOCAL INFERENCE, electronically submitted December 18, 2006, and named STARMAP.Project2.Web.Summary.Report.pdf
- ◆ Web summary of FINAL TECHNICAL REPORT: PROJECT 3 -DEVELOPMENT AND EVALUATION OF AQUATIC INDICATORS, electronically submitted December 18, 2006, and named STARMAP.Project3.Web.Summary.Report.pdf.
- ◆ Web summary of FINAL TECHNICAL REPORT: PROJECT 4 - EXTENSION and OUTREACH, electronically submitted December 18, 2006, and named STARMAP.DAMARS.Outreach.Web.Summary.Report.pdf.
- ◆ Copies of published papers (54) and book (1) mailed to the Project Officer December 15, 2006.

Submitted electronically 12/15/2006