24th Annual National Conference on Managing Environmental Quality Systems

8:30 – 12:00 TUESDAY, APRIL 12TH - A.M. Stockholder Meetings

12:00 – 4:30 TUESDAY, APRIL 12TH
Opening Plenary (Salons A-H)

- Opening Address
  - Reggie Cheatham, Director, OEI Quality Staff, EPA
  - Linda Travers, Principal Deputy Assistant Administrator, OEI, EPA
- Invited Speakers
  - Tom Huetteman, Deputy Assistant Regional Administrator, EPA Region 9
  - John Robertus, Executive Officer of San Diego Regional Water Quality Control Board, Region 9
- Keynote Address
  - Thomas Redman, President, Navesink Consulting Group
- Panel Sessions
- **Value of the Data Quality Act—Perspectives from OMB, Industry, and EPA (VDQA)**
  - Nancy Beck, OMB
  - Jamie Conrad, American Chemistry Council
  - Reggie Cheatham, Director, OEI Quality Staff, EPA
- **Wadeable Streams: Assessing the Quality of the Nation's Streams (WS)**
  - Margo Hunt, Panel Moderator
  - Mike Shapiro, Deputy Assistant Administrator, Office of Water
  - Steve Paulsen, Research Biologist, ORD

8:30 – 10:00 WEDNESDAY, APRIL 13TH
Environmental Measures (EM) (Salons A-C) *Chair: L. Bradley, EPA*

- Data Error Reduction by Automation throughout the Data Workflow Process (A. Gray, EarthSoft, Inc.)
- Analytical Approaches to Meeting New Notification Levels for Organic Contaminants in Calif. (D.Wijekoon, Calif. DHS)
- Streamlining Data Management and Communications for the Former Walker AFB Project (R. Amano, Lab Data Consultants, Inc.)

Quality System Implementation in the Great Lakes Program (QSI-GLP) (Salon D) *Chair: M. Cusanelli, EPA*

- GLNPO's Quality System Implementation for the New "Great Lakes Legacy Act for Sediment Remediation"(L. Blume, EPA)
- Black Lagoon Quality Plan Approval by GLNPO, MDEQ, ERRS, and USACE (J. Doan, Environmental Quality Management, Inc.)
- Remediation of the Black Lagoon Trenton Channel . . . Postdredging Sampling & Residuals Analysis (J. Schofield, CSC)

Quality Systems Models (QSM) (Salons F-H) *Chair: G. Johnson, EPA*

- Improving E4 Quality System Effectiveness by Using ISO 9001: 2000 Process Controls (C. Hedin, Shaw Environmental)

Applications of Novel Techniques to Environmental Problems (ANTEP) (Salon E) *Chair: B. Nussbaum, EPA*

- On Some Applications of Ranked Set Sampling (B. Sinha, University of Maryland)
- Combining Data from Many Sources to Establish Chromium Emission Standards (N. Neerchal, University of Maryland)
- Estimating Error Rates in EPA Databases for Auditing Purposes (H. Lacayo, Jr., EPA)
- Spatial Population Partitioning Using Voronoi Diagrams For Environmental Data Analysis (A. Singh, UNLV)

Ambient Air Session I (Sierra 5&6) *Chair: M.Papp, EPA*

- Changes and Improvements in the Ambient Air Quality Monitoring Program Quality System (M. Papp, EPA)
- Guidance for a New Era of Ambient Air Monitoring (A. Kelley, Hamilton County DES)
- Environmental Monitoring QA in Indian Country (M. Ronca-Battista, Northern Arizona University)
- Scalable QAPP IT Solution for Air Monitoring Programs (C. Drouin, Lake Environmental Software)

10:30 – 12:00 WEDNESDAY, APRIL 13TH

Environmental Laboratory Quality Systems (ELQS) (Salons A-C) *Chair: L. Bradley, EPA*

- A Harmonized National Accreditation Standard: The Next Step for INELA Field Activities (D. Thomas, Professional Service Industries, Inc.)
- Development of a Comprehensive Quality Standard for Environmental Laboratory Accreditation (J. Parr, INELA)
- Advanced Tracking of Laboratory PT Performance and Certification Status with Integrated Electronic NELAC-Style Auditing Software (T. Fitzpatrick, Lab Data Consultants, Inc.)

Performance Metrics (PM) (Salon D) *Chair: L. Doucet, EPA*

- Formulating Quality Management Metrics for a State Program in an Environmental Performance Partnership Agreement (P. Mundy, EPA)
- How Good Is "How Good Is?" (Measuring QA) (M. Kantz, EPA)
- Performance-Based Management (J. Santillan, US Air Force)

Quality Assurance Plan Guidance Initiatives (QAPGI) (Salons F-H) *Chair: A. Batterman, EPA*

- A CD-ROM Based QAPP Preparation Tool for Tribes (D. Taylor, EPA)
- Military Munitions Response Program Quality Plans (J. Sikes, U.S. Army)

Ask a Statistician: Panel Discussion (Salon E) *Moderator: B. Nussbaum, EPA* Panelists:

- Mike Flynn, Director, Office of Information Analysis and Access, OEI, EPA
- Reggie Cheatham, Director, Quality Staff, OEI, EPA
- Tom Curran, Chief Information Officer, OAQPS, EPA
- Diane Harris, Quality Office, Region 7, EPA
- Bill Hunt, Visiting Senior Scientist, North Carolina State University (NCSU)
- Rick Linthurst, OIG, EPA

Ambient Air Session II (Sierra 5&6) *Chair: M. Papp, EPA*

- National Air Toxics QA System and Results of the QA Assessment (D. Mikel, EPA)
- Technical System Audits (TSAs) and Instrument Performance Audits (IPAs) of the National Air Toxics Trends Stations (NATTS) and Supporting Laboratories (S. Stetzer Biddle, Battelle)
- Interlaboratory Comparison of Ambient Air Samples (C. Pearson, CARB)
- Developing Criteria for Equivalency Status for Continuous PM2.5 Samplers (B. Coutant, Battelle)

1:00 – 2:30 WEDNESDAY, APRIL 13TH

Environmental Laboratory Quality (ELQ) (Salons A-C) *Chair: L. Doucet, EPA*

- Environmental Laboratory Quality Systems: Data Integrity Model and Systematic Procedures (R. DiRienzo, DataChem Laboratories, Inc.)
- The Interrelationship of Proficiency Testing, Interlaboratory Statistics and Lab QA Programs (T. Coyner, Analytical Products Group, Inc.)
- EPA FIFRA Laboratory Challenges and Solutions to Building a Quality System in Compliance with International Laboratory Quality Standard ISO 17025 (A. Ferdig, Mich. Dept. of Agriculture)

Performance—Quality Systems Implementation (P-QSI) (Salon D) *Chair: A. Belle, EPA*

- Implementing and Assessing Quality Systems for State, Tribal, and Local Agencies (K. Bolger, D. Johnson, L. Blume, EPA)

1:00 – 2:30 WEDNESDAY, APRIL 13TH (continued)

Quality Initiatives in the EPA Office of Environmental Information (QI-OEI) (Salons F-H) *Chair: J. Worthington, EPA*

- Next Generation Data Quality Automation in EPA Data Marts (P. Magrogan, Lockheed)
- The Design and Implementation of a Quality System for IT Products and Services (J. Scalera, EPA)
- Data Quality is in the Eyes of the Users: EPA's Locational Data Improvement Efforts (P. Garvey, EPA)

A Win-Win-Win Partnership for Solving Environmental Problems (W3PSEP) (Salon E) *Co-Chairs: W. Hunt, Jr. and K. Weems, NCSU*

- Overview of Environmental Statistics Courses at NCSU (B. Hunt, NCSU Statistics Dept.)
- Overview of the Environmental Statistics Program at Spelman College (N. Shah, Spelman)
- Student presentations: H. Ferguson and C. Smith of Spelman College; C. Pitts, B. Stines and J. White of NCSU

Ambient Air Session III (Sierra 5&6) *Chair: M. Papp, EPA*

- Trace Gas Monitoring for Support of the National Air Monitoring Strategy (D. Mikel, EPA)
- Comparison of the Proposed Versus Current Approach to Estimate Precision and Bias for Gaseous Automated Methods for the Ambient Air Monitoring Program (L. Camalier, EPA)
- Introduction to the IMPROVE Program's New Interactive Web-based Data Validation Tools (L. DeBell, Colorado State University)
- The Role of QA in Determination of Effects of Shipping Procedures for PM2.5 Speciation Filters (D. Crumpler, EPA)

3:00 – 4:30 WEDNESDAY, APRIL 13TH

Topics in Environmental Data Operations (TEDO) (Salons A-C) *Chair: M. Kantz, EPA*

- Ethics in Environmental Operations: It's More Than Just Lab Data (A. Rosecrance, Laboratory Data Consultants, Inc.)
- QA/QC of a Project Involving Cooperative Agreements, IAGs, Agency Staff and Contracts to Conduct the Research (A. Batterman, EPA)
- Dealing with Fishy Data: A Look at Quality Management for the Great Lakes Fish Monitoring Program (E. Murphy, EPA)

Quality System Development (QSD) (Salon D) *Chair: A. Belle, EPA*

- Development of a QA Program for the State of California (B. van Buuren, Van Buuren Consulting, LLC)
- Integrating EPA Quality System Requirements with Program Office Needs for a Practical Approach to Assuring Adequate Data Quality to Support Decision Making (K. Boynton, EPA)
- Introducing Quality System Changes in Large Established Organizations (H. Ferguson, EPA)

Auditor Competence (AC) (Salons F-H) *Chair: K. Orr, EPA*

- Determining the Competence of Auditors (G. Johnson, EPA)

To Detect or Not Detect—What Is the Problem? (TDND) (Salon E) *Chair: J. Warren, EPA*

- A Bayesian Approach to Measurement Detection Limits (B. Venner)
- The Problem of Statistical Analysis with Nondetects Present (D. Helsel, USGS)
- Handling Nondetects Using Survival Anal.(D. Helsel, USGS)
- Assessing the Risk associated with Mercury: Using ReVA's Webtool to Compare Data, Assumptions and Models (E. Smith, EPA)

Ambient Air Session IV (Sierra 5&6) *Chair: M. Papp, EPA*

- Status and Changes in EPA Infrastructure for Bias Traceability to NIST (M. Shanis, EPA)
- Using the TTP Laboratory at Sites with Higher Sample Flow Demands (A. Teitz, EPA)

5:00 – 6:00 PM WEDNESDAY, APRIL 13TH

EPA SAS Users Group Meeting Contact: Ann Pitchford, EPA

8:30 – 10:00 THURSDAY, APRIL 14TH

Evaluating Environmental Data Quality (EEDQ) *(Salons A-C) Chair: M. Kantz, EPA*

- QA Documentation to Support the Collection of Secondary Data (J. O'Donnell, Tetra Tech, Inc.)
- Staged Electronic Data Deliverable: Overview and Status (A. Mudambi, EPA)
- Automated Metadata Reports for Geo-Spatial Analyses (R. Booher, INDUS Corporation)

Satellite Imagery QA (SI-QA) (Salon D) *Chair: M. Cusanelli, EPA*

- Satellite Imagery QA Concerns (G. Brilis and R. Lunetta, EPA)

Information Quality Perspectives (IQP) (Salons F-H) *Chair: J. Worthington, EPA*

- A Body of Knowledge for Information and Data Quality (J. Worthington, L. Romero Cedeno, EPA)
- Information as an Environmental Technology – Approaching Quality from a Different Angle (K. Hull, Neptune and Co.)

To Detect or Not Detect—What Is the Answer? (TDND) (Salon E) *Chair: A. Pitchford, EPA, Co-Chair: W. Puckett, EPA*

- Using Small Area Analysis Statistics to Estimate Asthma Prevalence in Census Tracts from the National Health Interview Survey (T. Brody, EPA)
- Logistical Regression and QLIM Using SAS Software (J. Bander, SAS)
- Bayesian Estimation of the Mean in the Presence of Nondetects (A. Khago, University of Nevada)

Ambient Air Workgroup Meeting (Sierra 5&6) *Contact: Mike Papp, EPA*
NOTE: This is an all-day, closed meeting.

10:30 – 12:00 THURSDAY, APRIL 14TH

Environmental Data Quality (EDQ) (Salons A-C) *Chair: V. Holloman, EPA*

- Assessing Environmental Data Using External Calibration Procedures (Y. Yang, CSC)
- Groundwater Well Design Affects Data Representativeness: A Case Study on Organotins (E. Popek, Weston Solutions)

Information Quality and Policy Frameworks (IQPF) (Salons F-H) *Chair: L. Doucet, EPA*

- Modeling Quality Management System Practices to an Organization's Performance Measures (J. Worthington, L. Romero Cedeño, EPA)
- Development of a QAPP for Agency's Portal (K. Orr, EPA)
- Discussion of Drivers and Emerging Issues, Including IT, That May Result in Revisions to EPA's Quality Order and Manual (R. Shafer, EPA)

Office of Water; Current Initiatives (OW) (Salon D) *Chair: D. Sims, EPA*

- Whole Effluent Toxicity--The Role of QA in Litigation (M. Kelly, EPA, H. McCarty, CSC)
- Review of Data from Method Validation Studies: Ensuring Results Are Useful Without Putting the Cart Before the Horse (W. Telliard, EPA, H. McCarty, CSC)
- Detection and Quantitation Concepts: Where Are We Now? (Telliard, Kelly, and McCarty)

Sampling Inside, Outside, and Under (SIOU) (Salon E) *Chair: J. Warren, EPA*

- VSP Software: Designs and Data Analyses for Sampling – Contaminated Buildings (B. Pulsipher, J. Wilson, Pacific Northwest National Laboratory , R. O. Gilbert)
- Incorporating Statistical Analysis for Site Assessment into a Geographic Information System (D. Reichhardt, MSE Technology Applications, Inc.)
- The OPP's Pesticide Data Program Environmental Indicator Project (P. Villanueva, EPA)

1:00 – 2:30 THURSDAY, APRIL 14TH

Information Management (Salons A-C) *Chair: C. Thoma, EPA*

- Achieve Information Management Objectives by Building and Implementing a Data Quality Strategy (F. Dravis, Firstlogic)

UFP Implementation (Salon D) *Chair: D. Sims, EPA*

- Implementing the Products of the Intergovernmental DQ Task Force: The UFP QAPP (R. Runyon, M. Carter, EPA)
- Measuring Performance: The UFP QAPP Manual (M. Carter, EPA, C. Rastatter, VERSAR)

Quality Systems Guidance and Training Developments (QSG) (Salons F-H) *Chair: M. Kantz, EPA*

- A Sampling and Analysis Plan Guidance for Wetlands Projects (D. Taylor, EPA)
- My Top Ten List of Important Things I Do as an EPA QA and Records Manager (T. Hughes, EPA)
- I'm Here---I'm Free----Use Me! Use Me!—Secondary Use of Data in Your Quality System (M. Kantz, EPA)

Innovative Environmental Analyses (IEA) (Salon E) *Chair: M. Conomos, EPA*

- Evaluation of Replication Methods between NHANES 1999-2000 and NHANES 2001-2002 (H. Allender, EPA)
- Assessment of the Relative Importance of the CrEAM Model's Metrics (A. Lubin, L. Lehrman, and M. White, EPA)
- Statistical Evaluation Plans for Compliance Monitoring Programs (R. Ellgas, Shaw Environmental, Inc.; J. Shaw, EMCON/OWT, Inc.)

Implementing the Products of the Intergovernmental Data Quality Task Force
The UFP QAPP

In the summer of 2005, the Intergovernmental Data Quality Task Force (IDQTF), consisting of EPA, DoD and DOE, adopted as "IDQTF Final" a Uniform Federal Policy Quality Assurance Project Plan Manual. Formal adoption by the agencies who are part of the IDQTF is anticipated in the Spring of 2005.

The purpose of the UFP QAPP is to act as a single national consensus guidance document for implementing the requirements of ANSI/ASQ E4, *Quality Systems for Environmental Data and Technology Programs – Requirements for guidance with use, Section 6 (Part B)* consistently and systematically across the Federal agencies involved in the IDQTF. The UFP QAPP is consistent with EPA's G-5, and G-8 and meets EPA's Quality System Requirements.

The intent of the UFP QAPP is to integrate quality requirements for environmental sampling and analysis through a Systematic Planning Process. It places renewed emphasis on proper scoping of projects, and will require project teams to spend more time on this phase in order to save significant time and money during later phases.

Activities designed to help insure the implementation of the UFP QAPP will be described in this paper. These activities include:

- Publication of directives and instructions in EPA and DoD
- Development of metrics (the subject of another paper)
- Communication through FFRRO web site
- Training
    - Pilot testing
    - ISEERB Certification (DoD)
    - Revised hands-on training, including blended learning approach with internet based modules

Measuring Performance
The Uniform Federal Policy Quality Assurance Project Plan Manual

In the summer of 2004, the Intergovernmental Data Quality Task Force (IDQTF), consisting of EPA, DoD and DOE, adopted as "IDQTF Final" a Uniform Federal Policy Quality Assurance Project Plan Manual. Formal adoption by the agencies who are part of the IDQTF is anticipated in the Spring of 2005.

The purpose of the UFP QAPP is to act as a single national consensus guidance document for implementing the requirements of ANSI/ASQ E4, *Quality Systems for Environmental Data and Technology Programs – Requirements for guidance with use, Section 6 (Part B)* consistently and systematically across the Federal agencies involved in the IDQTF. It is anticipated that conformance to the UFP QAPP will result in:

- Improvements in the quality associated with environmental sampling projects by ensuring that all projects systematically address specific quality requirements.
- Improvements in the cost and time of preparing QAPPs by creating a single template that is in use across the country and that is likely to be both easier to prepare initially, as well as easier for regulators to review and approve.
- Improvements in the cost and time of completing environmental sampling projects by ensuring that the right people are involved in planning the projects and by building in quality up front. This attention to quality up front is expected to reduce rework, as well as the number overall samples lost to the project due to poor sampling and analytical practices or insufficient attention to the decision requirements.

The IDQTF has been in the process of developing metrics to measure implementation and effectiveness of the UFP QAPP. It is anticipated that these metrics will be phased in over a number of years (as implementation is phased in), and that data will be collected in a variety of ways. The purpose of this presentation is to outline:

- The overall process for identification of metrics
- The goals and objectives that the measurements will address
- The specific inputs and outputs that were evaluated, and
- Current thinking on the measurements recommended for data collection and analysis.

A CD-ROM Based Quality Assurance Project Plan Preparation Tool for Tribes

David R. Taylor

US Environmental Protection Agency Region 9 Quality Assurance Office

A Quality Assurance Project Plan (QAPP) preparation tool for use by tribes who are performing basic water monitoring programs has been developed. The tool is CD-ROM based and incorporates several components and a wealth of reference material. It is hoped that this document will facilitate the preparation of QAPPs by tribal and other organizations faced with the preparation of a QAPP for the first time.

1.0 Introduction

Tribal environmental staff have long been challenged by the preparation of Quality Assurance Project Plans to meet EPA grant requirements as outlined in 40 CFR 31.45. A number of factors have contributed to this problem. Tribal staff are often new to environmental work and have had limited prior experience in preparing QA documents such as QAPPs and field sampling plans; few staff have had an opportunity to attend national or regional training in QA or in sampling and analysis; the technical background of tribal staff varies considerably and very few personnel in Indian Country are chemists; the remoteness of many tribal lands and EPA's limited travel budgets restrict the Agency's ability to visit and train tribal personnel directly; the retention of staff in Indian Country is often difficult, resulting in a loss of institutional knowledge concerning sampling and analysis and QA, the use of consultants by some tribes (as a way of overcoming some of the problems mentioned earlier) often means that tribal capacity in the development of plans is developed by consultants rather than the tribe itself. These problems are complex, and their solutions are beyond the scope of this paper, but it was felt that some of these problems could be ameliorated through the development of a CD-ROM based tool which could be used by tribes to facilitate the writing of QA Plans.

2.0 Approach

The result of these discussions was the creation of a small workgroup in early 2004. Originally the workgroup consisted of a diverse group of individuals containing members from two tribes, an Alaskan Native village, five regions (1, 3, 6, 9, and 10), the Region 9 Indians Program Office, and the Office of Environmental Information. Two individuals were members of the Tribal Science Council. Unfortunately, as time went on, workgroup members were drawn to other job responsibilities and most of the development work fell to a core group of five EPA people, four from regional QA Offices and one from a regional drinking water office, all of whom had experience in working with tribes or other similar small grantees. The effort was supported conceptually by the Tribal Science Council and EPA Environmental Science Directors, and financially by the Office of Environmental Information's Quality Staff and by the Region 9 Indian Programs Office.

The first issue to be evaluated was the vehicle to be used to provide the information. It was decided that a CD-ROM approach would be used. Some tribes in more remote areas do not have good quality or fast internet access which made a web based tool less optimal. A web site also requires on-going maintenance, upkeep it was not clear anyone on the workgroup would be able to provide All EPA web sites also require a yearly maintenance fee, and since funding for the project was only provided on a one time, not an on-going, basis, it was thought that this might also present a problem. CD-ROMs are inexpensive, and a replacement could be burned if staff turnover resulted in the tribes copy being lost. Thus, a CD-ROM based approach to the tool was considered to be the best.

With respect to the software form of the tool, a "Turbo Tax®" type program was first considered. Using this approach an interactive questionnaire would be used. When completed the program would generate the final version of the document. There were several other software programs which would permit a similar approach, but all were ultimately felt to be less than optimal. It was felt such an approach would be too inflexible to capture the potential diversity of projects in Indian Country for which the tool might be used, and also that the challenges and expense associated with this would present in programming. Of special concern was how changes to the document might be handled, since the goal was never to necessarily create a QAPP that would be 100% approved by all QA Offices in all regions on the first iteration. It was realistically expected that most QAPPs submitted by tribes who used the tool would still receive comments and require revision, instead the goal was to create a QAPP that was 80-90% of what was required so that revisions would be relatively minor and quick to make. However, it was hoped that having representatives of two of the regions with the largest number of tribes (Regions 9 and 10), as well as Region 1 with 26 tribes and input from Region 6 with over 60, would help make the tool acceptable across EPA. This hopefully would help address an often expressed tribal concerns about a lack of consistency in the expectations of EPA Regions.

A conscious decision was made to focus the tool on water monitoring and water sample, specifically on surface water. This was done for several reasons. First, the majority of tribes who receive EPA grants other than General Assistance Program (GAP) funds, receive money under the Clean Water Act 106 program. This money most often is applied to surface water (and sometimes groundwater source water) sampling and analysis projects. Thus, it was felt by the workgroup that a focus on water would make the tool useful to the widest audience. It was also felt that once developed, expanding the scope of the tool to other programs, such as Brownfields projects which might involve a soil component, would be relatively straightforward. It was decided that trying to make a universal QAPP guidance would be too difficult becasue of the often differing data quality objectives, analytical methods, action levels, etc., that would need to be incorporated into all of the different components of the tool, from the guidance to the standard operating procedures (SOPs). A comparable document for Brownfields remains of interest, but will have to wait for a new workgroup to be formed.

All the modules are in portable document format (pdf) with the exception of the template which is provided in both Word and WordPerfect word processing formats.

3.0 Components of the Tool

The final tool consists of an introduction and six modules, which will be discussed separately below. These consist of:

Introduction

1. Guidance
2. QA Project Plan Template
3. Model QA Plan
4. References and Links
5. Standard Operating Procedures
6. How to select a laboratory

3.1 Module 1:Guidance

This section is based on the EPA's "Guidance for the Preparation of Quality Assurance Project Plans" (G-5), and follows that format. However, it is written in a briefer more directive style. The idea is to inform the plan preparer what information needs to be obtained that is most critical to preparing the organization's QAPP. Thus, it identifies questions that will need to be thought through before the plan can be prepared, such as decisions that will be made with the data and who will make those decisions. One of the area most often found to be deficient in grantee plans submitted for review is that the intended purpose(s) of the data are not described adequately; the guidance attempts to avoid this problem nor are project action levels known. In addition, much of the information related to sampling and analysis must be assembled, be it laboratory QA Plans, quality control (QC) criteria for field or laboratory methods, method detection limits, sampling locations and rationale for their selection, chain of custody forms, and so forth. The guidance attempts to ensure that all necessary material is assembled before the plan is written.

The guidance uses a directive approach in instructing the plan preparer concerning what information should be put into the plan. This approach hopefully lessens the possibility of confusion on the preparer's part in that what is required and essential in each section is stated explicitly. Every effort has been made to limit the amount of QA "jargon" and to describe requirements in less technical language. One problem expressed by tribes who have attempted to use G-5, is that it contains many terms with which they have limited or no familiarity. Although in some cases the terminology cannot be avoided, its use can be minimized, or alternative explanations for terms provided. An attempt has also been made to minimize redundancy among sections. This has been found by some grantees to be a problem with G-5 as the same information is seemingly requested in multiple sections with no clear differentiation as to whether the information should be repeated.

3.2 Module2: Template

The template also follows the format of the G-5 guidance. In general it does not contain a lot of text, although a brief prompt is given in some cases to help identify material relevant to the section. The two word processing formats hopefully cover most of the programs tribes might to using to prepare their plans. The hope is that the material gathered or the questions considered in reading through the guidance will now be applied to this section. The output of the template is designed to be the QA Plan that is submitted for review, both internally within the tribe as well as by the appropriate Project Officer or QA Manager or designee in the region.

3.3 Module 3; Model QA Plan

The model QA Plan reflects both the guidance and the template, and describes the type of project that typically is submitted by a tribal grantee to a region. The example is not overly complicated, encompassing about a dozen samples, but it does show how a QAPP might look for a tribe writing up a description baseline surface water monitoring activities. The main parameters included were general chemistry (anions, total dissolved solids, etc.) and metals, since these are the analytes most often represented in tribal water monitoring plans. Some limited field measurements are also included. The model is based on a fictitious tribe in northern California with a river and several tributary streams.

3.4 Module 4: References and Links

it was thought useful to provide tribes as much reference material as could be contained on the CD-ROM. Since this vehicle contains some 670 Mb of capacity, there was room for considerable information. In many cases links are provided rather than the documents themselves, but in some cases the actual reference is provided. Links to several regional web sites are given and the standard EPA QA references are given. There are analytical methods or links to sources of analytical methods. EPA's Handbook to Quality Control in Water and Wastewater Laboratories is included. EPA's major QA guidances, R-2, R-5, and G-5 are reproduced. Although some of the information may not be useful to tribes starting their environmental programs, it may be useful at a later date as the program matures.

3.5 Module 5: Standard Operating Procedures

Because many tribes are just beginning their water monitoring programs at the time they prepare their first QAPP, it was thought useful to provide SOPs primarily on collection of water samples and operation of field equipment related to water sampling. SOPs that could be used to collect samples, ship samples, fill out chain of custody forms, etc. Several sources of these documents were used. The 29 Palms Band of Mission Indians provided its collection of SOPs. In addition, the environmental laboratories/sampling teams for Regions 1, 6, and 9 provided SOPs related to sampling and field activities. The SOPs from the Emergency Response Team in Edison New Jersey were also included. The result is a plethora of methods from which a trip can choose, at least for the common activities, such as collecting a surface water sample. For other activities, for example use of a global positioning system (GPS) apparatus, there is only one SOP. Since it is quite acceptable under most circumstances to reference and attach SOPs for many of the

activities which must be described in a QAPP, it was felt that tribes could take the SOPs provided, modify them to reflect their own approach, as appropriate, and include them with their QAPPs. Hopefully this would both ensure a minimal level of detail in the SOPs and simplify the process of documenting activities when writing the QAPP.

Module 6A: Selecting an Environmental Laboratory

Since many tribes have not had to contract for analytical services prior to beginning their environmental programs, it was thought that it would be beneficial to include a discussion about selecting and evaluating an environmental laboratory. The discussion is relatively generic, but also includes a discussion of certification/accreditation as questions often arise in this regard. Although not all questions will be answered, it was thought that this information might be useful, especially as few tribal environmental departments have chemists or their own laboratories.

4.0 Conclusion

Although not a universal solution for all tribes who need to prepare a QAPP, the CD-ROM tool that has been developed is designed to facilitate the challenges of tribes and other small organizations who must prepare a QAPP to cover their environmental measurement activities. The tool contains a wealth of information and has put QAPP requirements into a simpler more straightforward format to make it easier to use and understand. The tool should soon be in distribution, and will be available soon, if it is not already available at the time of the 24th QA Conference.

MY TOP TEN LIST OF IMPORTANT THINGS I DO AS AN EPA QA AND RECORDS MANAGER

Thomas J. Hughes, QA and Records Manager, Experimental Toxicology Division (ETD), National Health and Environmental Effects Research Laboratory (NHEERL), Office of Research and Development (ORD), U.S. EPA, Research Triangle Park, NC 27709

I have been the QA and Records Manager for ETD since September 26, 1999 (5.5 years). ETD is the largest health research division in the U.S. EPA, with more than 70 Federal full-time employees (FTE) and 40 Cooperators (students, Post Docs and Seniors). I am responsible for the QA and Records oversight and training of these 110 scientists. I conduct the usual data audits, surveillances, and technical systems audits (TSA) on the QA Category 1 and 2 studies, and monitor the QA Category 3 and 4 research studies (in total, 55 yearly research studies). I track the comings and goings of the Federal Staff (fairly easy because they are a relatively stable group) and the Cooperators (fairly difficult because they can come or go at any time of the year). This presentation will NOT discuss the usual things an EPA QA and Records Manager does. It will, however, discuss the top ten ADDITIONAL things that I do on a yearly basis, that I believe make my Division an Agency leader in defensible and verifiable scientific records.

Therefore, without further ado, here is *"My Top Ten List of Important Things I Do as an EPA QA and Records Manager*" (I will start at Number Ten because David Letterman always starts with Number Ten):

(10) Say "Thank You" to everyone who assists me, including Management, Scientists, and QA Assistants;

(9) Developed and employ a data transfer form for transfer of scientific records;

(8) Deploy a scientific data management system (SDMS) in my Division (NuGenesis);

(7) Conduct Annual Surveillances on the 24 Labs in my Division with the ETD Surveillance Checklist and the ETD QA Core Team (see #2);

(6) Conduct QA briefings on a regular basis (e-mails, Staff Meetings, scheduled Management briefings, and through the use of ETD Preaudit and ETD QC Checklists);

(5) Conduct formal Staff QA and Records training (ETD Newsletter articles, e-mails, seminars);

(4) Attend Leadership and QA Training on an annual basis, and bring national speakers to ETD to discuss QA, Records and Leadership issues;

(3) Plan the major QA and Records activities for the following year in December, and send the schedule to all the Staff in my Division in January;

(2) Obtain QA and Records assistance from scientists in my Division. I have eight scientists on the ETD QA Core Team, who assist me with technical systems audits (TSAs), data audits, and surveillances. I train them throughout the year on QA techniques and procedures, and they draft QA reports. The ETD QA Core Team has succeeded beyond my wildest expectations (the eight-member QA Team was started in 2003). I could not do my QA and Records job effectively in such a large Division without them.

(1) The Number ONE "Important Thing I Do as an EPA QA and Records Manager" is: Perform the other nine important things listed above exceptionally well. This enables me to sustain strong management support from my Division Director and three Branch Chiefs for my QA and Records Management activities.

This is an abstract for presentation which has been reviewed by the U.S. EPA; views expressed do not necessarily represent EPA policy.

Marcus E. Kantz
US EPA Quality Staff and Region 2 732-321-6690 kantz.marcus@epa.gov
Talk at EPA National QA Conference
April 14, 2005

I'm here. I'm free./Use me! Use me!
(Secondary Use of Data in Your Quality System)

Imagine the temptation: the data you need for your project are sitting right there in front of you on your PC screen. Somebody else already collected them and published them. All you need to do is download them. But what if there's something wrong with those data? What if that problem with the data would make your final product worthless? But, of course, you just don't know. Would you yield to temptation? Maybe YOU wouldn't, but others would. People in your organization are using existing data right now, without any concern for the appropriateness of that use. They are probably making wrong decisions right now, as a result. And they don't even realize it? It would be nice if your organization's Quality System had safeguards to help you resist the temptation, and find out whether the data are suitable for your use. It can be done. You can understand what Secondary Use of Data really means, and you can manage it in your organization. Find out how.

Imagine the temptation: the data you need for your project are sitting right there in front of you on your PC screen or in a journal article. Somebody else already collected them and published them. All you need to do is download them. But what if there's something wrong with those data? What if that problem with the data would make your final product worthless? But, of course, you just don't know. Would you yield to temptation? Maybe YOU wouldn't, but others would. People in your organization are using existing data right now, without any concern for the appropriateness of that use. They are probably making wrong decisions right now, as a result. And they don't even realize it?

You think this is exaggeration? Some of you think that you would **never** use data that you weren't sure were absolutely suitable for your intended use. But what about this? For your project you need historical data on the ambient temperatures in town over the past 10 years. You ask at the airport, and they tell you. You use the data. You trust the data implicitly. Did you ask for the measurement method, for their SOP, calibration data, QAPP, and log books to make sure they didn't have any down time during which they used data from the neighboring town, or that they moved the monitoring site, or switched the daily time of measurement? Probably not. Why not? If you were going to start a project today in which YOU were going to take the measurements, You would require all of that. Why not require it of the person at the airport?

For your project you need rainfall data for the same 10 years. The airport has that too. No problem. But wouldn't the rainfall have been different at the wide-open airport than in the hills on the other side of town where the run-off from the factory is affecting the river? But the only data available are from the airport. Are they good enough? They're there. They're free.

This paper is intended to be a wake-up call - a call that says "Good morning Mr. and Ms. DataUser. You and your decisions are seriously vulnerable unless you take better care of your Secondary Use of Data. You need to put mechanisms into your Quality System to protect you from your almost unconscious desire to use all of the found data you can without sufficient regard for their appropriateness for your Secondary Use."

Let's start with a quick definition.

We are talking about the Secondary Use of Data for a purpose that is different from the purpose for which those data were originally collected. These are data that are already out there. Somebody - maybe even you - collected them for a given purpose that is different from yours. For some reason, you want to use those data instead of generating data yourself now for your specific purpose. Maybe you want to use them: because you can't afford to collect data yourself; because you don't have the expertise or special equipment or resources to collect them; because you are not legally allowed to collect them; or because the situation on which you need data no longer exists - like the first 9 years of the 10 year trends. The point is this: The data were collected for a primary purpose. You now want to use them for a secondary purpose. That doesn't mean that your purpose is less important, or valuable than the primary purpose. It only means that yours was not the reason those data were collected or generated, and maybe they don't quite fit your needs.

We can call these data found data, available data, existing data, old data, external data, outside data, or historical data. But for you, they are just data you want to use.

There are three different kinds of data that are of interest and available for use in environmental projects carried out by or for EPA.

> The first and most obvious is the set of major environmental data bases from EPA and other US agencies (such as NOAA, USGS), including AIRS, STORET, PCS, RCRIS, CERCLIS, etc. (by extension, this list also includes data bases maintained by the states, etc. that mimic and/or extend these data bases with similar data types).
>
> The second is the set of the results of all of the environmental studies conducted by EPA, our states, grantees, contractors, regulatees, and everyone else. These data are documented in peer reviewed journals, in magazines and newspapers, in presentations, in published reports, required submittals to Agencies, and in lab notebooks in someone's drawer.
>
> The third collection of data is all of the data that are not traditional environmental data. These data, some of which are maintained in large published data bases, and some all the way down to research notebooks, contain the information about the world that we need to support our environmental measurements but that are "beyond the scope of the course," information that we and our environmental colleagues have not and in most cases could never collect on our own. They include, for example: all of the demographic data collected and reported by the US Census Bureau; all of the meteorological information collected and reported by the National Weather Service; all of the housing, financial, land use, and health information collected and reported by HUD, HSS, USGS, CDC, NIOSH, etc.; 'routine' aerial

and satellite images, both federal and private; urban, rural, and farm information reported by national and local governments; manufacturing and industrial data like resource use and production rates; state and local government information on transportation, zoning.

While all three kinds of data are routinely subjected to secondary use, the first and the third are the most common, and the most dangerous. They are used because: they cover so many situations; they carry the imprint of a major, quality-conscious organization; they are fully vetted for their original use; and they are usually the only readily available source of information for those particular situations. Where else but in EPA data bases can one find 20 years of nationwide ambient air quality data, or the discharges from every sewage treatment plant. Where else can one find a 10-year wind rose for a given town than from the Weather Service, or the demographics of that town other than from the Census Bureau? Their Secondary Use is dangerous because the raison d'être for the data was often very narrow, and that narrowness is not at all obvious when seeking the data. Even if the wind rose is perfect for the airport where it was generated, how far from the airport is it still valid for your purpose. Can you tell whether the Census demographics cover the illegal aliens who actually catch and eat fish from the river?

Now we know what kinds of data are out there calling us, and the potential dangers. Next, we need to determine whether we may actually yield to the call and use the available data, and if so, under what conditions. As you might expect, it depends. It depends on a comparison between the original intent of the data and your needs, and on your ability to figure it all out.

Whenever data generators collect data *properly*, they have specific goals in mind that are stated in their systematic planning documentation, like a QA Project Plan (QAPP), including specific measurement quality objectives (MQOs, which may also be Data Quality Objectives (DQOs)). These MQOs, along with certain other information associated with the data themselves, are conveniently termed the meta-data associated with the data. When the measurement project is completed, the data and the meta-data are assessed in a Data Usability Assessment (DUA, which may also be a Data Quality Assessment (DQA)). The data are then accepted for their intended use and stored in some way - in some cases in one of EPA's major data bases. Rarely do project managers ensure the usability of the data for purposes beyond their own. Even more rarely do they accompany the stored data with a narrative explanation of those uses for which the data have been judged by the original project organizers to be appropriate. And even in those few cases, any future user would be left to trust that judgement. In other words, there are almost no cases in which a particular data set comes pre-certified as usable for your new application.

Let's go back for a minute to the town's demographics from Census to examine the difficulties in matching found data to your needs. The meta-data are very well documented, but extreme care must be taken in how you USE the data. The demographic data can tell very different stories when you change what is called the aggregation level. When you look at the block group instead of the block aggregation of exactly the same data, you can completely miss high concentration clusters of certain characteristics that might be of interest to you. In QA-speak, the representativeness of the data may shift dramatically as you raise the aggregation level, and some organizations have only subscribed to the less expensive broader levels.

All of this is not to say that there is or should be a general prohibition against the Secondary Use of Data. Of course there are some situations in which the existing data are completely inappropriate and should not be used in any way. But there is a myriad of situations in which the data may be considered for use, with or without caveats, etc. These possible situations can be summarized in five scenarios facing a potential Secondary User of Data, each with its own conditions under which the data may be used:

1. If your application is explicitly included by the generator as one for which the data may be used, and the generator and its QA reviewer are of sufficient reputation as to be trustworthy by default, you may use the data without concern, and even without a QAPP if they are the only data to be used in the project (If other new data are to be used, an inclusive QAPP is needed. If more than one set of data satisfying these conditions is to be used, an inclusive QAPP is recommended.) Examples of such uses of data might be:

    NAMS and SLAMS ambient air quality data to be used for attainment designations.

2. If your potential use is very similar to the original use of the data, and all of the meta-data that you would routinely utilize are available, you may include the data and the meta-data in your QAPP or subsequent report, along with your project-specific MQOs (or DQOs), which need to be in your new QAPP.

    If the meta-data satisfy your MQOs, you may use the data without further explanation.

    If the meta-data are close to your MQOs, you may consider using the data, but will need to include in your QAPP or in a subsequent report, sufficient documentation of an analysis confirming the usability of the data for your purpose, including whatever caveats, conditions, and/or estimates of resulting uncertainty are appropriate.

    If the meta-data are not close to your MQOs, you may still decide to use the data, but will need to stress in the documentation estimates of the resulting uncertainty.

3. If your potential use is related to the original intent of data generator but is not similar, it is less likely that your MQOs are even close to the data quality achieved by the generator's project. You may consider using the data but will need an even more rigorous assessment and documentation in your QAPP and/or in subsequent reports of the rationale for using the data. and of the expected uncertainty of that use.

4. If your potential use is not related to the original intent of the data generator, you may consider using the data, but only with an assessment and documentation in your QAPP and/or in subsequent reports of the rationale for using the data.

    If the data are from traditional environmental measurements, the assessment should include consideration of the usual MQOs, etc.

    If the data are not from traditional environmental measurements - such as Census or meteorological data - extreme care must be applied because environmental users are frequently not highly conversant in the language, techniques, variabilities, uncertainties, and limitations of such data. Data usability assessments may well require the assistance of an expert in that area. Examples of potential troubles are:

        For Census data, aggregation levels (block, block group, etc.) can have a profound effect on the implications to be drawn from the data.

        For meteorological data, averaging (time period (monthly, hourly), compass sector (quadrant or 10 degree increment) as well as monitoring height, etc. can have a profound effect on the implications to be drawn from the data.

        For epidemiological data, the universe of potential subjects (walk-ins, survey respondents, doctor referrals, emergency room visitors) can have a profound effect on

the implications to be drawn from the data.
For other anthropological data (vehicle patterns, pedestrian movement, product use, etc.) the myriad of potential variables seemingly unrelated to the measurement of interest can have a profound effect on the implications to be drawn from the data.

5. If you plan to use data derived by calculation or the results of modeling or other estimation procedures, the effects of the calculation, estimation, and/or modeling can have profound and often unpredictable or un-estimable effects, along with the effects of the uncertainty brought on by the underlying data or assumptions. The implications for your QAPP and results are impossible to state in general terms, but will require even more careful, and likely expert assistance. You definitely need to study the QAPP(s) from the model development and use.

The attached Table summarizes these five scenarios of Secondary Use of Data. Or even shorter:

The Secondary Use of Data is acceptable, and can be very valuable,

BUT

It needs to be done carefully, in recognition of the potential uncertainties resulting from using data for a purpose different than the purpose for which the data were originally intended.

A Brief Aside Concerning YOUR Specific Use

You may have noticed that we have not talked at all about your specific need. For example, we have not differentiated among enforcement decisions, trends analyses, risk assessments, records of decision, attainment analyses, Environmental Justice investigations, methods development, or pure or applied research or development. The fact is that it doesn't matter. You already have specific policies and guidance for how you deal with data quality within the context of project development and implementation. Whether you are using original data or available data really makes very little difference, within the context described here. In other words, as long as you determine how your use of the data relates to the originally intended use and follow the guidelines shown in the scenarios, your underlying motive for the work doesn't matter. Your measurement quality needs are determined by why you are doing it, independently of where you are getting the data. A generic enforcement project would generally have tighter MQOs and documentation requirements than would a trends analysis, but that is all factored into your analysis before you get to the point of deciding where to get the data. Once you know what you need, it really doesn't matter to the data why you need it.

Thus, whether you are a researcher or an enforcer, follow your normal procedures but make sure you prove the available data will satisfy your needs, document that proof , and then - and only then - use the data.

So what can we do institutionally?

Wake up and spread the word. We don't need new rules, and new guidance or training would only help a little. The difficulty will be in getting people to pay attention. The only way I see to make people pay attention is for each organization to institute measures in their basic decision-making process that will treat Secondary Use of Existing Data just like they treat new data. Quality Management Plans, grants tracking programs, Quality Assurance Check-off Boxes, all need to say those magic words "including the Secondary Use of Existing Data". A few people will see it, and pay attention, and try to comply. A few auditors will begin to ask questions about Secondary Use of Data. Decision-making across the community will start to improve.

Still, for yourselves, when you hear existing data calling you, tempting you, "I'm here. I'm free. Use me! Use me!" Be careful. Treat those data like you would your own. It will take time and cost money, but in the end you will know what your results mean. No data are perfect, not even your own. But some data are appropriate for your use, and some are not. Make sure you prove the data will satisfy your needs, document that proof , and then - and only then - use the data. Do what you can to make sure that your organization sets up formal procedures to ensure that you and your colleagues treat found data like you would new data. Then tell all your friends to do the same.

Excerpt from EPA Order 5360.1 A2 [emphasis added]
showing longstanding EPA policy on the Secondary Use of Data

5. SCOPE AND FIELD OF APPLICATION.
   a. Scope. This Order defines the minimum requirements for quality systems supporting EPA environmental programs that encompass:
      (1) the collection, **evaluation, and use** of environmental data by or for EPA,
   b. Applicability to Environmental Programs. This Order applies to (but is not limited to) the following environmental programs:
      (3) **the use of environmental data collected for other purposes or from other sources (also termed "secondary data")**, including literature, industry surveys, compilations from computerized data bases and information systems, results from computerized or mathematical models of environmental processes and conditions; and
6. QUALITY SYSTEM REQUIREMENTS AND IMPLEMENTATION.
   a. Quality System Requirements. EPA organizations covered by the scope of this Order shall develop, implement, and maintain a quality system that demonstrates conformance to the minimum specifications of ANSI/ASQC E4-1994 and that additionally provides for the following:
      (8) **Assessment of existing data, when used to support Agency decisions or other secondary purposes, to verify that they are of sufficient quantity and adequate quality for their intended use.**

Scenarios for the Secondary Use of Data

#	Your new use is	QAPP	Conditions
1	Explicitly included in original QAPP, etc.	No	None
2	Very similar to intended use	Yes	If meta-data satisfy your MQOs, none. Otherwise, dependent on meta-data
3	Related to intended use	Yes	Rigorous assessment , with uncertainty estimates
4	Not related to intended use	Yes	Detailed assessment of usability and uncertainty needed, especially for non-environmental data (consult experts)
5	Using calculated or modeled results	Yes	Detailed assessment of usability and uncertainty needed (consult subject matter experts and modeling experts)

Evaluation of the Different Replication Methods Between NHANES 1999-2000 and NHANES 2001-2002

Hans D. Allender, Ph.D., P.E.

Introduction

To evaluate pesticide exposure to US population, the Office of Pesticides Programs (OPP) at EPA considers data in the Center for Disease Control's (CDC) data-base known as NHANES, National Health and Nutrition Examination Survey (http://www.cdc.gov/nchs/nhanes.htm). The survey produces an extensive and comprehensive information data-base on more that 20,000 individuals. Because the survey is designed to permit statistical statements to be made about certain minority groups within the population, the survey over-samples these subgroups and uses sampling weights to subsequently adjust for this over-sampling. To extend the survey results to the rest of the population, special statistics are necessary; this includes the utilization of replications for each data point.

The recent CDC release of NHANES 2001-2002 data uses a different approach to replication methods than that used in the earlier NHANES 1999-2000 survey. In the earlier 1999-2000 survey, the CDC data files provided the users with replications generated by CDC's statisticians. The purpose of these replications is to facilitate the calculation of point estimates from the survey data. For that purpose, the NHANES 1999-2000 survey data files include 52 replicates for each subject in the survey. Since the 1999-2000 data covers about 10,000 individuals and that there are two types of replicates(one for the interview (INT) and another for the examination (MEC)), this generates about 1,004,000 (52 x 2 x 10,000) additional entries.

With the release of the new NHANES 2001-2002 data which adds an additional 10,000 individuals to the survey, CDC has recognized that this may increase the number of entries to an unmanageable file size. To prevent this problem, CDC has released new variables that describe the survey design (DSMVPSU and DSMVSTRA); they are the Primary Sampling Unit (PSU) and the survey strata (STRA). These variables in conjunction with the survey weights allow for the computer-generation of the replications by the user. Also, CDC went back to the 1999-2000 survey data and retrofit it with design variables that allow users to generate their own replicates. And finally, CDC provided a 4-year weight that allows for the combination of the two two-year surveys in a continuous fashion. With the introduction of these new PSU and STRAT variables to the survey data, the files become manageable, and it is responsibility of the user to generate the replicates required to compensate for the non-random design of the survey.

Objectives of the Presentation

The objectives of this paper are to analyze the new way data is released, to demonstrate the generation of replicates, and to compare the different methods used in the biannual and continuous 4-year NHANES data.

Methodology

Before entering a detailed discussion, a brief explanation of weighted data is required. The NHANES is a large data collection that needs to make sure that certain subgroups in the population are properly represented in the survey. To assure proper representation it is necessary to over sample these subgroups. The result of over-sampling implies that samples are not purely collected at random and in order to compensate for the bias introduced by the sample design it is required to balance the survey. The balance is reached by given a representative weight to each element in the sample. Through a complex mathematical procedure involving the Jackknife design option, NHANES estimates the proportion of subjects, and weights are assigned to each subject. The NHANES 1999-2000 provided the weights and also created 52 replicates for each element that facilitates the calculation of point estimates.

In this paper four different ways to estimate a variable will be used for comparison; they are

1) NHANES 1999-2000 with the original CDC provided 52 replications
2) NHANES 1999-2000 with self-generated replications
3) NHANES 2001-2002 with self-generated replications, and
4) NHANES 1999-2004 the combined 4-year with self-generated replications.

Because of the goal of the comparison is directed to find stability in the replication methodologies it is best to select a variable that has minimum variation through time; the analyst selected a variable that is predictable stable over time. For example, one of the first variables under consideration was weight, but this variable is somehow subdued to eating habits of the population that may change over the span of 4 years, the duration of the combined survey. Other variables, as for example pesticide residue concentration, are also subjected to changes on application of pesticide, pest pressure, etc. A more stable variable is height and to further stabilize it, the analyst refined the variable limiting the comparison study to white males 21 and older.

In order to deal with replicates special software is required. Two commercially available computer programs are leading the field; they are SUDAAN and WesVar. The calculation is this paper were done using WesVar version 4.2.

The Inner Workings of WesVar

WesVar supports both balanced repeated and Jackknife approaches and provides the following replication methods

- Balanced Repeat Replications (BRR) for designs with two primary sampling units (PSUs) per stratum
- Fry's BRR variant
- Jackknife for un-stratified designs (JK1)
- Jackknife for stratified designs with two PSUs per stratum (JK2), and
- Jackknife for stratified designs with two or more PSUs per stratum (JKn)

Other methods of replication such as the Bootstrap (Efrom 1982) can be handled in WesVar, but it's required to input the replicate weights and factors appropriated for that method.

Replication methods are implemented in WesVar using the following steps:

1. Divide the sample into sub-samples replicates that mirror the design of the sample by specifying the variance strata (VarStrat) and PSU (VarUnit) variables.
2. Calculate weights for each replicate, using the same procedures used for the full-sample weight.
3. Calculate replicate estimates for each of the replicates using the same methods used for the full-sample estimate.
4. Estimate the variance of the full-sample estimate, using the resulting full-sample and replicate estimates.

If replicate weights are already available, there is no need to create new ones. However, the replication method used when they were created outside of WesVar must be identified.

The formation of replicates in any give replication method is determined by the Hadamard matrices stored in WesVar. Hadamard matrices are constructed from the combination of a basic "4 x 4" Hadamard (H) matrix through the algorithm

$$H^* = \begin{vmatrix} H & H \\ H & -H \end{vmatrix}$$

For example

Basic Hadamard Matrix (4 X 4)

```
1 1 1 1
0 1 0 1
0 0 1 1
1 0 0 1
```

Composed Hadamard Matrix (12 x 12)

```
1 1 1 1 1 1 1 1 1 1 1 1
0 1 0 1 1 1 0 0 0 1 0 1
0 0 1 0 1 1 1 0 0 0 1 1
1 0 0 1 0 1 1 1 0 0 0 1
0 1 0 0 1 0 1 1 1 0 0 1
0 0 1 0 0 1 0 1 1 1 0 1
0 0 0 1 0 0 1 0 1 1 1 1
1 0 0 0 1 0 0 1 0 1 1 1
1 1 0 0 0 1 0 0 1 0 1 1
1 1 1 0 0 0 1 0 0 1 0 1
0 1 1 1 0 0 0 1 0 0 1 1
1 0 1 1 1 0 0 0 1 0 0 1
```

In some cases these matrices dictate the number of replicates produced on each replication method. If

L = number of strata in the sample, and
G = number of replicates produced by WesVar

Then:
In the Balanced Repeat Replication (BRR), the number of replications G is the smaller integer divisible by 4 which is greater of equal to L. For example, if L = 13, then G = 16; if L = 28, then G = 28; if L = 15, then G = 16.

For Jackknife 1 (JN1), the number of replications is G = L, where G is identified using the VarUnit of the program. For Example if L = 13 then G = 13.

For Jackknife 2 (JN2), is the same, G = L

For Jackknife n (JNn), because there are more than 2 PSU, G is calculated by

$$G = \sum_{h=1}^{L} n_h$$

This is the addition of all strata existing in each PSU. For example if there are 3 PSUs two of them with 12 strata, and the third has only 6 strata. Then the number of replications will be, G = (2 x 12) + (1 x 6) = 30.

Calculations

Replications in each NHANES case are as follow:

- In the original HNANES 1999-2000 CDC produced 52 replicates and recommended the use of Jackknife 1. Calculation of point estimates will be produced with 52 replications.
- The same NHANES 1999-2000 was later retrofitted with variables for the PSU = 2, and 13 strata. Calculation of point estimates using the BRR method will be produced with 16 replicates.
- NHANES 2000-2001 has PSUs = 2, and 15 strata, so using the BRR method, point estimates will be calculated with 16 replicates
- For the combined 4 years NHANES 1999-2001 there are 2 PSUs and 28 strata, the use of BRR method produces 28 replication

The outcome of a Summary Table for the calculations of the combined 4 years NHANES 1999-2002 is shown on Table 1.

Table 1. Summary Information of Table Request "4Year 1999-2001"

```
 WESVAR VERSION NUMBER: 4.2
 TIME THE JOB EXECUTED: 15:11:23 02/22/2005
 INPUT DATASET NAME:
 C:\Program Files\Dr. Allender\WesVar 4.2\nhanes1999_2002white_21ht.var
 TIME THE INPUT DATASET CREATED:    15:02:04 02/22/2005
 FULL SAMPLE WEIGHT:    WTMEC4YR
 REPLICATE WEIGHTS:     RPL01...RPL28
 VARIANCE ESTIMATION METHOD:  BRR

 OPTION COMPLETE: ON
 OPTION FUNCTION LOG:   ON
 OPTION VARIABLE LABEL: OFF
 OPTION VALUE LABEL:    OFF
 OPTION OUTPUT REPLICATE ESTIMATES: OFF
 FINITE POPULATION CORRECTION FACTOR:     1.00000
 VALUE OF ALPHA (CONFIDENCE LEVEL %):     0.05000 (95.00000 %)
 DEGREES OF FREEDOM:    28
 t VALUE :  2.048
 OPTION BIN QUANTILE:   OFF
 ANALYSIS VARIABLES:    BMXHT

 COMPUTED STATISTIC:

      MeanHT = MEAN(BMXHT)
      Q10 = QUANTILE(BMXHT,0.10)
      Q25 = QUANTILE(BMXHT,0.25)
      Q50 = QUANTILE(BMXHT,0.50)
      Q75 = QUANTILE(BMXHT,0.75)
      Q90 = QUANTILE(BMXHT,0.90)
      Q95 = QUANTILE(BMXHT,0.95)

TABLE(S):   None Specified.

 FACTOR(S): 1.00

 NUMBER OF REPLICATES:  28
 NUMBER OF OBSERVATIONS READ: 2108
 WEIGHTED NUMBER OF OBSERVATIONS READ:    65100145.427
```

Overall Results

STATISTIC	ESTIMATE	STDERROR	LOWER 95%	UPPER 95%	CELL_n
SUM_WTS	63796791.84	2476582.069	58723743.24	68869840.44	2041
MeanHT	177.16	0.189	176.77	177.55	2041
Q10	168.00	0.398	167.05	168.67	2041
Q25	172.45	0.252	171.90	172.93	2041
Q50	177.00	0.266	176.51	177.60	2041
Q75	182.00	0.286	181.34	182.51	2041
Q90	186.16	0.311	185.66	186.93	2041
Q95	188.96	0.405	188.25	189.91	2041

* Warning: 67 observations were excluded from the preceding table. These observations were excluded because they contained one or more requested variables with missing values.

With the variable defined as “Height” for white males over 21 years old using the software WesVar 4.2, the attached Table No.2 was assembled. The Table shows four ways to calculate the variable “height” and calculates the mean value with its standard error, and selected percentiles i.e., 10th, 25th, 50th (the median), 75th, 90th, and 95th; all with 95% Confidence Interval. The table also shows the sample size N and the number of replications produced by the software.

Conclusions/Findings

1.- Regardless of the replication methodology used, the point estimates are all within the 95% confidence interval of prediction. This holds even in the extreme 95th percentile of the distribution.

2.- Replication methods proved to work within the predicted accuracy.

3.- As expected, the standard error of point estimates decreases with an increase in sample size.

4.- Difference in results are well within the variability predicted by CDC’s statisticians.

5.- It is recommended to use the 4-year 1999-2002 continuous NHANES data base whenever possible.

Table No. 2 Comparative Table for Different NHANES Replications Methods for the Variable: Height (in cm) of White Males 21 and Older.

Survey Replications (Rep)	Mean (95% CI) Standard Error	Selected Percentiles (with 95% Confidence Interval)						Sample N
		10%	25%	50%	75%	90%	95%	
NHANES 1999-2000 CDC Given Rep (52 Rep)	177.01 (176.3 -177.7) 0.355	168.11 (166.6 -169.1)	172.43 (171.4 – 173.1)	176.64 (176.1 – 177.8)	181.86 (180.9 –182.8)	186.34 (185.4 –187.1)	188.97 (187.7 –190.3)	930
NHANES 1999-2000 Self-Generated Rep (16 Rep)	176.98 (176.4 – 177.6) 0.271	167.90 (166.7 – 169.0)	172.40 (171.6 – 173.0)	176.57 (176.1 – 177.5)	181.66 (180.8 – 182.8)	186.31 (185.5 – 187.1)	189.00 (187.9 – 190.2)	864
NHANES 2001-2002 Self-Generated Rep (16 Rep)	177.44 (176.9 –178.0) 0.262	168.15 (166.9 – 169.1)	172.82 (172.0 – 173.3)	177.52 (176.3 – 178.2)	182.16 (181.4 – 183.0)	186.15 185.57 – 187.4)	188.97 (188.0 – 191.1)	1177
NHANES 1999-2002 Self-Generated Rep (28 Rep)	177.16 (176.8 – 177.6) 0.189	168.00 (167.1 – 168.7)	172.45 (171.9 – 172.9)	177.00 (176.5 – 177.6)	182.00 (181.4 – 182.5)	186.16 (185.7 – 189.9)	188.96 (188.3 – 189.9)	2041

An Assessment of the Relative Importance of the CrEAM Model's Metrics

Mary L. White, Ph.D., Arthur N. Lubin, Ph.D., and Lawrence Lehrman, PE., EPA Region 5, Chicago, IL

ABSTRACT

The Critical Ecosystem Assessment Model (CrEAM) employs a geographic information system (GIS) analysis platform to aggregate multiple geographically referenced data sets in order to conduct landscape scale analysis. Three important criteria were integrated: ecological diversity, ecological sustainability and rarity. The Model was applied to the six upper Midwestern states of the United States (Illinois, Indiana, Michigan, Minnesota, Ohio and Wisconsin) which comprise U.S. Environmental Protection Agency Region 5. Using the National Land Cover Database nine land cover categories were identified and aggregated into composite cells of 300m x 300m that were analyzed against the five landscape fragmentation, seven stressor data sets, four diversity data sets and four rarity data sets. A numerical score ranking was generated for each cell. This paper discusses a simple but novel way to conduct sensitivity analysis on model components in a case of a spatial prediction model generated from many very large data sets.

This analysis was done to determine which of the individual data sets within the three composite criteria (ecological diversity, ecological sustainability and rarity) and which of the components -- the model is most sensitive to. The sensitivity analysis was done by correlating: 1.) composite criteria scores; 2.) Model results for each cell derived using the original scores generated by the model with the corresponding model results derived without one of the composite criteria scores; 3.) Individual data layer scores within the composite criteria scores; and 4.)The entire individual composite metric score with the corresponding scores when one sub-metric is deleted. The relative importance of individual data layers within each of the aggregate criterion is further shown via maps which display the relative differences between the original score generated by the model compared to those resulting when single data layers within aggregate components are deleted. The results of this analysis will be used to facilitate future modifications of the model.

An Assessment of the Relative Importance of the CrEAM Model's Metrics

Introduction

The purpose for the development of the CrEAM model was to facilitate decision making by program office managers concerning natural resource management. These professionals make decisions every day based upon the perceived relative quality of areas and those decisions could be better informed if they had a consistent way to rank the ecological significance of an area. This information is generally not available because the evaluation of ecological significance is a difficult task at the landscape scale. Not only are there various proposed ways to evaluate ecosystem quality (O'Malley and Wing 2000), (Xu et al. 2001), (Campbell 2001), (Day et al. 1997), (Costanza and Mageau 1999), (Jenson et al. 1996), there is the additional problem of consistent and comparable data collected over large scales (Gaston 2000), (Levin et al. 1997), (Verburg et al. 2002)

We defined ecosystem ecological significance as the integration of three important conditions: 1) ecosystem diversity, 2) ecological self-sustainability and 3) species and land cover rarity. A brief discussion of each criterion follows:

1. *Ecological diversity* - The presence of population, community, and/or ecosystem diversity (Chapin et al. 2000), (Ehrlich and Wilson 1991).
2. *Ecological sustainability* - The potential for an ecosystem to persist without loss or decline, preferably without external assistance or management (Dale et al. 2000), (Gunderson et al. 2002).
3. *Rare Species and Land Cover* - The occurrences of rare native species, or communities and land cover types of special ecological interest (Dobson et al. 1997), (Pimm and Lawton 1998) .

The primary objective of this model is to identify the most significant ecosystems so that EPA staff could use the information as supplemental information in identifying and targeting protection and restoration efforts.

The CrEAM model is based upon 20 data layers which comprise three aggregate metrics. It is generally recommended that any model should be as parsimonious as possible. Certain data layers because they correlate highly with the other data layers and aggregate metrics provide redundant information and could be possibly deleted from the model. However, if none of the data layers are highly correlated, the importance of each would be supported. Another fundamental idea in this analysis is that the components of the model which have the lowest associations should have the greatest impacts upon the criterion scores. This paper focuses on the results of the assessment process.

Methods

The data used for this product were CrEAM results of more than 3.6 million sub-areas, or cells representing the undeveloped areas of the study area. Relevant existing data sets were sought out

as indicators for the three criteria. We required data sets to be spatially and temporally consistent covering the entire six state area and representative of conditions that existed in the early 1990's. Two strategies are used to offset the paucity of detailed Region-wide data sets. First, the analysis is conducted at a regional landscape scale in order to reduce the significance of minor, small scale heterogeneity. Second, data sets that spatially quantify environmental conditions which can be tied back to ecological theory are used as indicators (e.g.. measures by proxy) for the three criteria consistent covering the entire six state area and representative of conditions that existed in the early 1990's. Two strategies are used to offset the paucity of detailed Region-wide data sets. First, the analysis is conducted at a regional landscape scale in order to reduce the significance of minor, small scale heterogeneity. Second, data sets that spatially quantify environmental conditions which can be tied back to ecological theory are used as indicators (e.g.. measures by proxy) for the three criteria (Verburg et al. 2002).

As previously noted, the total of 20 component data sets were used as indicators for the three criteria: four for ecological diversity (C1.1-4), 12 for self sustainability (C2.1-12), and four for rarity (C3.1-4). (See Table 2). The individual components or data layers within each of the aggregate criteria were equally weighted. Thus, the individual data layers in C1 and C3 each represented one-fourth each of the respective composite criteria scores and those in C2 were one-twelfth of the composite. However, each of the aggregate data layers within the entire model was equally weighted. In other words, each of the individual data layers within the aggregate criteria were equally weighted but the individual data layers within the aggregate criteria were not equally weighted in the final ecosystem score.

All of the individual data layers were normalized from 0 to 100 by taking the maximum individual cell value minus the actual cell value divided by the observed range of values ((maximum score - actual cell score)/range). Zero always indicates the lowest quality, the greatest stress or the least valuable observation and 100 indicates the highest quality, least stress or most valuable observation. Each set of composite scores also were normalized to range from 0 to 100. The final scores for each cell were generated by summing the three composite scores. Thus, each undeveloped land cover cell across Region 5 was assigned a relative rating potentially ranging between 0 and 300 with each overall composite score being equally weighted in the model.

The statistical analysis was done to determine which of the individual criteria within the entire model (ecological diversity, ecological sustainability and rarity) and which of the data layers within the individual composites the model and composites are most sensitive to. The assumption is that relatively low associations demonstrate that the measures are not providing redundant information. This sensitivity analysis was done by correlating the individual data layer scores within each composite (C11-C14, C21-C212 and C31-C34) (all of these correlations are available upon request) , the individual composite scores (C1 - C2, etc.), the model results for each cell derived using the entire model with the corresponding model results derived when one aggregate component is deleted (Call_totalstand - (Call_totalstand - C1), etc.) as well as the individual composite metric results versus the composite results when single sub-metrics were

deleted (C1_totalstand – C11, etc.) to verify that low correlations among sub-metrics actually impact individual metric scores. This latter step is especially critical for the C2 sub-metrics which have lower correlations. The correlation results are shown in Table 2. Both the Pearson

Table1. Data Sets Used as Indicators of Ecological Significance.

ECOLOGICAL DIVERSITY C1	SELF-SUSTAINABILITY C2		RARITY C3
	Landscape Fragmentation	Stressor Presence	
C1.1 patch size of undeveloped land	C2.1 perimeter to area analysis	C2.5 airport buffers	C3.1 land cover rarity
C1.2 land cover diversity	C2.2 patch size by land cover type	C2.6 NPL Superfund sites	C3.2 species rarity
C1.3 temperature & precipitation maxima	C2.3 weighted road density	C2.7 RCRA corrective action sites	C3.3 rare species abundance
C1.4 temporal continuity of land cover type	C2.4 waterway impoundment	C2.8 water quality summary	C3.4 rare species taxa abundance
	C2.12 land cover suitability	C2.9 watershed obstruction	
	C2.11 development disturbance buffer	C2.10 air quality summary	

and Spearman coefficients were calculated because the data were not necessarily normally distributed and some of the individual data layers have discontinuous data (scored either 0 or 100).

The Call_totalstand variable is the score derived using the original CrEAM model. The C1 through C3 variables are the scores using the individual original CrEAM model composite metrics. The C1.1, C2.1, C3.1, etc. variables are based on the original CrEAM data layers listed in Table 1. The C1_totalstand through the C3_totalstand variables are the scores obtained using the listed composite only. For example, the C1_totalstand variable is the score using only the initial composite.

Scoring differences were mapped to verify that lower correlations actually were associated with substantial impacts on actual cell scores. The differences in the area scoring resulting from

deleting individual data layers within each of the aggregate components were calculated using the following process: (Call_totalstand - (Call_totalstand - individual metric)) = difference score. The variable listed in the title is the deleted metric. Maps of selected results (those results where the previously discussed correlations are relatively low) are shown in Figures 1-3. The maps demonstrate the locations of the substantial scoring differences. Additional maps corresponding to all of the correlation results as well as the difference scores are available upon request.

RESULTS

The correlation results are shown in Table 2. Both Pearson and Spearman coefficients are shown in the table. Unless substantially different or the variables are discontinuous the Pearson results only will be discussed.

The correlation results (C1 - C2, etc.) demonstrate that none of the aggregate data layers (ecological diversity, self-sustainability and rarity) within the entire CrEAM model duplicate each other. The correlations among the scores of the aggregate metrics are relatively low with the highest (.55) being between C1 (ecological diversity) and C2 (self-sustainability). The C3 metric is rarity. The other correlations are minimal (an absolute value of $\leq$ .10). This demonstrates that none of the aggregate metrics are redundant and, thus, are useful in determining ecosystem condition. Table 2 also displays the correlations between the entire model scores and those obtained after deleting one composite where for example the (Call__totalstand - (Calltotalstand - C1) is the correlation between the scores obtained using the entire model and the scores obtained after deleting the first aggregate metric. Because all of the correlations are less than .75 the significance of each of the aggregate metrics is further substantiated because the scoring results are impacted by the deletion of each of the aggregates. The highest correlation (.74) is obtained with C1 (ecological diversity) is omitted.

In addition, correlations among the data layer scores within the individual composites were calculated. Those correlations mentioned in the discussion are shown in Table 2. The other correlations among the data layer scores are available upon request. Within the diversity composite (C1), the highest correlation is .81, between C1.1 and C1.2 (land cover diversity and patch size of undeveloped land). Thus, each of the C1 layers appears to be adding information to the scale. Within the sustainability composite most, but not all, of the correlations among the scores are above .90 with the exception of the associations of C2.11 and C2.12 with the other second aggregate metric's components.

The correlations among each of the sustainability composite data layers and C2.12 (land cover suitability) were approximately .5 or less and those among C211 (development disturbance buffer) were from .70 to .80. Consequently, C2.11 and C2.12 should not be deleted. The Spearman coefficients provide similar results. However, it should be noted that these low

correlations may be derived from the discontinuity of the two variables. Thus, a revision of the model to make these two data layers continuous may alter the associations. Finally, within the rarity composite (C3), the correlations among the scores were .86 or higher. It should be noted that some of the corresponding Spearman coefficients are substantially lower. Given the relatively high correlations among the model's data layers, it may be possible in a future revision of the model to omit certain rarity sub- metrics without substantially modifying the final scores.

Next the cell scores using each of the individual aggregate metrics versus the individual aggregate model metric score without one of the data layers were correlated (see Table 2). The results provide further evidence that the data layers with relatively low associations tend to not be redundant. The lowest correlation (.77) for the first metric was obtained when C14 (temporal continuity of land cover type) was deleted from the C1 composite (ecological diversity). The other correlations were from .91 to .84. These correlations appear to be sufficiently low to support the importance of each of the C1 sub-metrics. The analogous results for the C2 composite tend to show higher correlations. Except when C2.11 (development disturbance buffer) and C2.12 (land cover suitability) are removed, the correlations are .93 or higher. The correlations between the C2 score and the C2 score when C2.11 and C2.12 are deleted are .81 and .47 respectively. This was expected given the relatively low correlations between C2.11 and C2.12 and the other C2 data layers. The Spearman coefficients provide similar associations. As expected with the relatively high correlation coefficients among the C3 (Rarity) data layers the correlations among the C3 composite cell scores and the corresponding scores obtained when a single C3 data layer is deleted were .94 or higher.

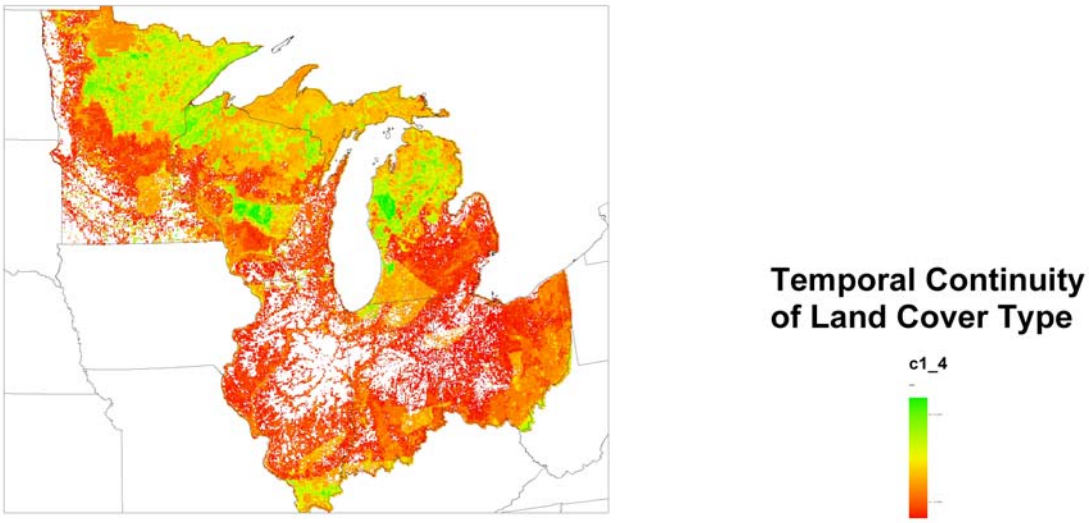
Table 2 Pearson and Spearman Correlation Coefficients Among the Scores from the Entire CrEAM and the Model Without the C1 (ecological diversity), C2 (self sustainability) and C3(rarity) Aggregate Metrics, the Coefficients for Among Scores Obtained Using the Individual Aggregate Metrics, the Coefficients for Among Scores Using the Individual Data layers and the Coefficients for Among Scores Obtained Using the Individual Aggregate Metrics With and Without the Data layer Listed Second (Difference Scores)

	Pearson	Spearman
C1 - C2	.55	.56
C1 - C3	-.02	- .03
C2 - C3	-.10	-.10
Call_totalstand - (Call_totalstand - C1)	.74	.74
Call_totalstand - (Call_totalstand - C2)	.65	.66
Call_totalstand - (Call_totalstand - C3)	.55	.49
C11 - C12	.81	.73
C11 - C13	.63	.72
C11 - C14	.44	.58
C12 - C13	.74	.71
C12 - C14	.55	.51
C13 - C14	.62	.65

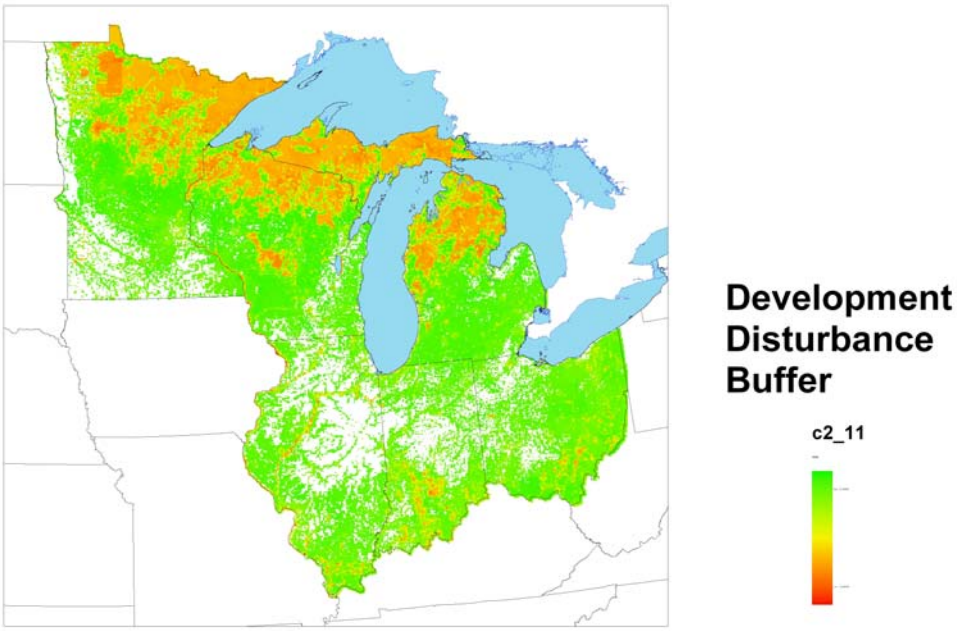
C21 - C211	.70	.67
C21 - C212	.49	.51
C22 - C211	.80	.78
C22 - C212	.40	.45
C23 - C211	.80	.80
C23 - C212	.48	.49
C24 - C211	.76	.76
C24 - C212	.49	.51
C25 - C211	.76	.77
C25 - C212	.48	.50
C26 - C211	.81	.81
C26 - C212	.47	.48
C27 - C211	.81	.80
C27 - C212	.47	.49
C28- C211	.78	.76
C28 - C212	.47	.49
C29 - C211	.79	.78
C29 - C212	.50	.49
C210 - C211	.81	.79
C210 - C212	.48	.49
C211 - C212	.51	.55
C1_totalstand - C11	.84	.88
C1_totalstand - C12	.91	.86
C1_totalstand - C13	.89	.90
C1_totalstand - C14	.77	.78
C2_totalstand - C211	.81	.80
C2_totalstand - C212	.47	.49

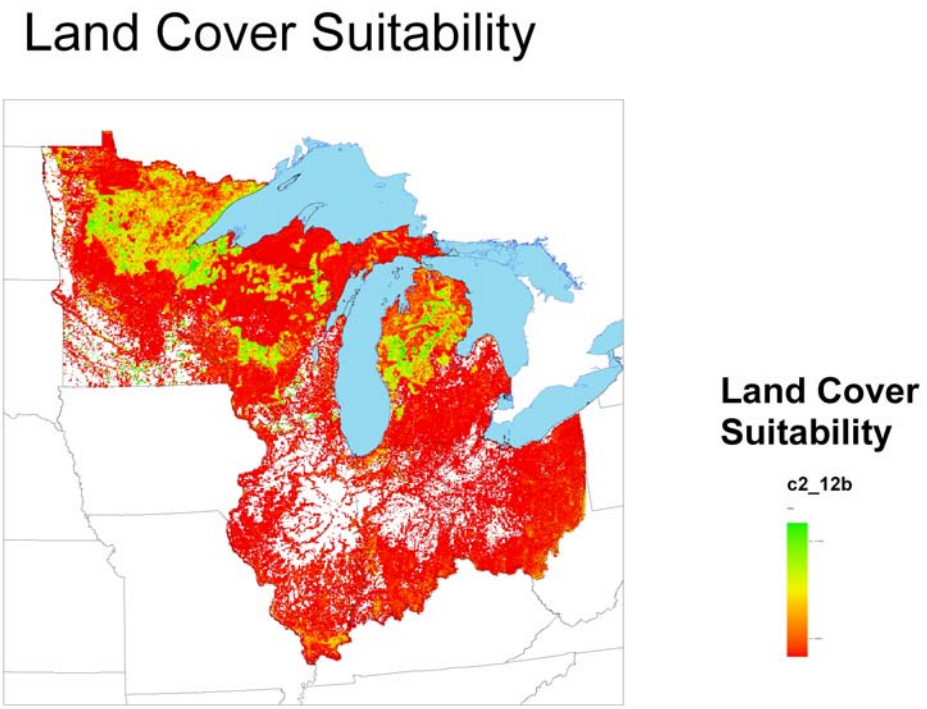
* The statistical significance status of the coefficients is not shown because the large sample size (>3.6 million cells) causes all of the coefficients to be statistically significant.

Temporal Continunity & Land Cover Type



Development Disturbance Buffer





Discussion

What is important, and was investigated at this time, is that some of the data layers *within* the composite criterion substantiallly duplicate each other. Thus, some of the data layers possibly may be deleted. However, it should be noted that none of the data layers were totally redundant. Of course, if any data layers are deleted, sensitivity analysis must be redone to assess how the modified model corresponds with the original CrEAM results given that results in one temporal and/or areal application may not be mirrored by those in another. The particularly low correlations within the second aggregate metric tend to be present when the discontinuous variables are assessed. Thus, if the model is modified to make those metrics continuous, the associations should be reassessed. In addition, a statistical clustering procedure may be done in the future to determine which subareas in the Region are similar to each other. This may allow for the modification of the ecoregions.

References

Campbell, D. E. 2001. Proposal for including what is valuable to ecosystems in environmental assessments. *Environ. Sci. Technol.* 35: 2867-2873.

Chapin, F. S., E. S. Zavalet, V. T. Evine, R. L. Naylor, P. M. Vitouse, H. L. Reynold, D. U. Hoope, S. Lavore, O. E. SalaI, S. E. Hobbie, M. C. Mack, and S. Díaz. 2000. Consequences of changing diversity. *Nature* 405: 234-242.

Costanza, R., and M. Mageau. 1999. What is a healthy ecosystem? *Aquatic Ecology* 33: 105-115.

Dale, V. H., S. Brown, A. Haeuber, N. T. Hobbs, N. Huntly, R. J. Naiman, W. E. Riebsame, M. G. Turner, and T. J. Valone. 2000. Ecological principles and guidelines for managing the use of land. *Ecological Applications* 10: 639-670.

Day, J. W. J., J. M. Rybczyk, G. Garson, and W. H. Conner. 1997. The development of functionally based ecosystem indices of biotic integrity for Southeastern bottomland hardwood forests. Coastal Ecology Institute, Louisiana State University and the Baruch Forest Science Institute.

Dobson, A.P., J.P. Rodriguez, W.M. Roberts, and D.S. Wilcov e. 1997. Geographic distribution of endangered species in the United States. *Science* 275: 550-553.

Ehrlich, P.R. and E.O. Wilson. 1991 Biodiversity studies: Science and policy. Pages 758-761. *Science.*

Gaston, K. J. 2000. Global patterns in biodiversity. *Nature* 405.

Gunderson, L.H., L. Pritchard, Jr., C.S. Holling, C. Folke,a nd G.D. Peterson. 2002. A summary and synthesis of resilience in large-scale systems. Pages 249-266 in L.H. Gunderson and L. Pritchard, Jr., eds. *Resilience and the Behavior of Large-Scale Systems.* Island Press, Washington, D.C.

Jenson, M. E., P. Bourgeron, R. Everett, and I. Goodman. 1996. Ecosystem Management: A Landscape Ecology Perspective. *Water Resources Bulletin* 32: 203-216.

Levin, S.A., Bryan Grenfell, Alan Hastings, and A.S. Perelson. 1997. Mathematical and computational challenges in population biology and ecosystems science. Science 275: 334-343.

O'Malley, R., and K. Wing. 2000. Forging a new tool for ecosystem reporting. Environment 42: 20-31.

Pimm, S.L., and J.H. Lawton. 1998. Planning for biodiversity. *Science* 279:2068-2069.

Verburg, P.H., W. Soepboer, A. Veldkamp, R. Limpiada, and V. Espaldon. 2002. Modeling for spatial dynamics for regional land use: The CLUE-S model. *Environmental Management* 30: 391-405.

X u, F.L., R. Sawson, S. Tao, J. Cao, and B.G. Li. 2001. A method for lake ecosystem health assessment: An ecological modeling method (EMM) and its application. *Hydrobiologia* 443: 159-175.

STATISTICAL EVALUATION PLANS FOR COMPLIANCE MONITORING PROGRAMS

Robert A. Ellgas, Ph.D. (Shaw Environmental, Inc., Concord, CA, USA) and Julian C. Isham, R.G., C.E.G., C.HG. (Shaw EMCON/OWT, Inc., Sacramento, CA, USA)

Abstract:
Guidance on developing statistical evaluation plans for compliance unit groundwater analytical data was recently issued in California. The guidance lists many statistical methods that can be used to evaluate data; however, it does not recommend specific methods and permits use of other methods with sufficient justification. An intrawell tolerance limit method is proposed for establishing statistically significant evidence of releases and for updating concentration limits for parameters monitored in point-of-compliance wells at a facility in California. Summary and descriptive statistics (moments, time series and box and whiskers plots); the Wilcoxon rank sum, Shapiro-Wilk, Kruskal-Wallis and Mann-Kendall tests; and Sen's slope estimator are used to evaluate data. Concentration limits calculated using the current method on a small set of early background data are compared to limits calculated using the methods listed above on a much larger set of complete background data. Many limits using the complete set of background data are higher, and a process is proposed for updating limits in the future. Statistical evaluation plans can be developed and applied to update concentration limits in many types of environmental compliance monitoring programs.

INTRODUCTION

To assure compliance with regulations, environmental monitoring programs must often compare the statistically established "background" value of a monitored parameter to temporal data for that parameter. If a newly acquired measurement exceeds the background value, the parameter might be considered "out of compliance", and additional actions might be required under the regulations.

The United States Environmental Protection Agency (USEPA) has issued guidance for evaluating groundwater monitoring data on regulated units at facilities permitted under the US Resource Conservation and Recovery Act of 1976 (RCRA; USEPA 1989, 1992). This guidance presents statistical methods for comparing upgradient (background) monitoring well data to downgradient (compliance) monitoring well data for regulated units to establish statistically significant evidence of a release from the regulated unit.

The California EPA Department of Toxic Substances Control (DTSC) has recently issued guidance that broadens and systematically formalizes the USEPA guidance into statistical evaluation plans (SEPs; DTSC, 2001). The guidance describes numerous statistical methods that could potentially be used for evaluating data, but it does not make recommendations for use of specific

methods. The guidance also permits the regulated facility owner/operator to propose other methods with sufficient justification.

The California guidance on SEPs is based on California Code of Regulations Title 22, Division 4.5, Chapter 14, Article 6, *Water Quality Monitoring and Response Programs for Permitted Facilities*, Sections 66264.90 through 66264.100. These Title 22 sections require the development of a water quality protection standard consisting of: 1) a list of constituents of concern (COCs), 2) concentration limits (not to exceed background values) and 3) the point of compliance and all monitoring points. These Title 22 sections also require that the statistical methods chosen shall comply with certain performance standards, such as adequacy of distributional assumptions, specified Type I error levels (where appropriate), interval coverage and confidence coefficients, accounting for data below the practical quantitation limit (PQL), and controls or correction for seasonal or spatial variability. Additional Title 22 requirements include tabulated data summaries, site and contour maps, well construction details, time series graphs, box and whiskers plots, flow charts for statistical protocol and background calculation method, and a demonstration that the proposed methods will be protective of human health and the environment.

EXAMPLE CASE

The SEP was developed for evaluating eight point-of-compliance (POC) monitoring wells and two background wells, and for calculating the concentration limits of 22 parameters, for metal cleaning waste impoundments at a facility in central California (Figure 1).

The 22 parameters are monitored quarterly and consist of COCs (barium, cadmium, cobalt, lead, mercury and molybdenum) and monitoring parameters (MPs; ammonia, arsenic, bromide, copper, fluoride, hexavalent chromium, hydrazine, iron, magnesium, nickel, potassium, sodium, sulfate, total chromium, vanadium, and zinc). Six water quality parameters and field parameters are also monitored at the site, but concentration limits are not calculated for these parameters.

Average groundwater flow velocities at the site result in travel times of about 4 to 12 feet between each quarterly monitoring event. One quarter's groundwater samples therefore constituted distinct water quality in each well when compared to the previous quarter's samples and consequently are concluded to be temporally independent for statistical evaluation in the SEP. Quarterly samples are also concluded to be spatially independent for statistical evaluation because the method of purging wells before sampling does not cause a hydraulic influence greater than a few feet, but wells are generally separated by a minimum distance of 60 feet.

The current statistical method used for calculating concentration limits at the site is an intrawell "modified" X-bar control chart. Depending on the percentage of non-detects in the data, the PQL or the maximum background concentration is the concentration limit. This method is approved for use only on a limited set of early analytical data, although many additional quarterly rounds of data have been acquired since the early data. The longer data records using the later data would more accurately reflect minor variations in sampling technique and natural variations in groundwater quality due to subsurface geochemical effects and long-term regional hydrogeological influences. Also, the method strongly assumes that data distributions are normal or transformed-normal, when in fact this might not be the case for many data sets. It was therefore decided that new statistical methods would be proposed in accordance with the recent DTSC guidance which also provides options for nonparametric testing and, if possible, the full set of data would be used.

STATISTICAL METHODS

Site analytical data are tabulated for each parameter for each well and summary statistics (total and detection counts, percentage non-detects, means, standard deviations, minima and maxima) are calculated. In cases with a sufficient number of detected concentrations, a non-detect result had one half the reporting limit substituted for the result to calculate summary statistics for the purpose of updating concentration limits (Gilbert, 1987). Time series graphs and box and whiskers plots are presented, and it is observed that there have likely been some changes in groundwater quality, but these changes are not attributed to site activities or in particular to releases from the regulated units. It is therefore concluded that concentration limits can be updated.

Intrawell Tolerance Limit Method.

The SEP proposes this method to establish statistically significant evidence of a release. The method is widely accepted and is described in detail in Montgomery (1985) and USEPA (1989, 1992). The method estimates the highest value that each monitoring parameter could have that could still be reasonably attributed to natural temporal variability, and it establishes a concentration limit below which is contained a specified proportion of the population, or coverage, with a specified confidence (both 95 percent, in accordance with DTSC guidance). For the 95 percent upper tolerance limit, it is important to point out that there is an approximate five percent chance that observations above the limit are in fact from the background distribution of well data (false positives, when it is concluded that such observations provide statistical evidence of a release).

Because of significant spatial variation in water chemistry at the site and no documentation of previous contamination, intrawell tolerance limit methods are more appropriate than interwell methods, which unreasonably assume that all natural spatial variability can be represented by upgradient background wells. Moreover, using only the spatial and temporal variability characteristics of the background wells to represent site conditions could lead to erroneous conclusions regarding statistically significant evidence of a release from a regulated unit.

Parametric tolerance limits require that the background data are normally or transformed-normally distributed, that at least three (preferably eight or more) observations are available, and that no more than 50 percent of the background data set are non-detects.

For censored data, if less than 15 percent of the background well observations are non-detects, these are replaced with one half the method reporting limit prior to calculating tolerance limits (USEPA, 1989). If more than 15 percent, but less than 50 percent, of the background data are less than the reporting limit, the data's sample mean and sample standard deviation are adjusted according to the method of Cohen as described in USEPA (1989). Assumptions for use of this technique are that the data are normally distributed and that the detection limit is always the same. If multiple detection limits exist, then they are all replaced with the highest detection limit.

A nonparametric tolerance limit method is used if the background data set contains greater than 50 percent but less than 90 percent non-detect values and/or its distribution is not normal or transformed normal. This method requires at least 19 samples to achieve a false positive rate of 5 percent (USEPA, 1992). When the background data set contains 90 to near 100 percent non-detect values and if the facility is in a detection monitoring program, the California Nonstatistical Method is used (Section 4.3.3, DTSC, 2001).

Wilcoxon Rank Sum Test for Data Consolidation Potential.
If interventions (i.e., contaminant releases) have not caused significant shifts in the mean of the later data set, the earlier and later data sets should be able to be consolidated. This would result in most statistical testing procedures having the requisite minimum number of samples to be applied without unacceptable errors, and would also result in concentration limits that are more representative of long-term natural site variations. The Wilcoxon rank sum test is a nonparametric test for a shift in location (mean) between two independent populations. This test has an advantage over the parametric independent-sample Student's *t* test in that it can accept a moderate number of non-detect values by treating them as ties. The test assumes independence and equal variances in the two data sets. There is apparently no modification to the test for unequal variances (Gilbert, 1987).

Data Screening for Outliers.
Once a decision is made that the earlier and later data sets can be consolidated, the consolidated data set is screened for statistical outliers. An outlier is not removed from the data set unless it can be documented to be erroneous. Outliers that cannot be explained by error might call for further investigation (USEPA, 1989). The outlier test used assumes that all data values, except for the suspect observation(s), are normally or lognormally distributed. However, once outliers have been screened from the data set, the test for normality described below can be applied to establish whether the screening process has been appropriately

applied. A minimum of three observations is required for the test, but a minimum of eight observations is recommended.

Data Distribution Evaluation by Shapiro-Wilk Test.
The consolidated background data sets (or transformed data sets, when applicable) for each constituent in each well, once screened for outliers, are tested for normality using the Shapiro-Wilk test. The Shapiro-Wilk test for normality is an approved method for distributional analysis when the sample size is >8 and #50, as referenced in EPA (1989) and DTSC (2001). If the sample size is greater than 50, the Shapiro-Francia test for normality can be used (USEPA, 1992), among other alternatives.

Data Seasonality Evaluation by Kruskal-Wallis Test.
After consolidated data sets have been screened for outliers, the data are evaluated for the significance of seasonality by the nonparametric Kruskal-Wallis test at the five percent significance level (USEPA, 1989). Because the test is nonparametric, it does not depend on the results of the Shapiro-Wilk test described above. This test should still be performed first, before any actions are taken on data that demonstrate seasonal effects. Application of the Kruskal-Wallis test for seasonality requires a minimum sample size of four data points in each season, or a minimum of four years of quarterly data.

Data Deseasonalization.
When seasonal variation is shown in a concentration time series according to the Kruskal-Wallis test, the data are deseasonalized before evaluating trends. In this way, seasonality is noted and eliminated rather than being mistaken as evidence of contamination. A deseasonalization correction is performed on data (transformed if required) before correcting for censoring by Cohen's adjustment.

Trend Evaluation by Mann-Kendall Test and Sen's Slope Estimator.
Once data have been deseasonalized as applicable, they are evaluated for temporal trends. The Mann-Kendall test for temporal trend (Hollander & Wolfe, 1973; USEPA, 1989) and Sen's slope estimator (Gilbert, 1987) are used to evaluate the correlation of selected constituent concentrations over time. The Mann-Kendall test is a nonparametric rank correlation test that uses only the relative magnitudes of data rather than actual values. As such, missing values are allowed for the test. When a trend is concluded to exist based on the Mann-Kendall test, Sen's slope estimator can be used to estimate the true slope of the trend. Sen's slope estimator is a simple nonparametric procedure developed by Sen (1968) that measures the change in constituent concentration per unit time. The advantage of this method over linear regression to estimate slope is that it is not greatly affected by gross data errors or outliers, and because it uses the relative magnitudes of the data rather than their actual values, it can be used when data are missing. Only data detected at least four times above the minimum detection limit (MDL) are evaluated for trend, although trace values can be incorporated into the test. For

wells having fewer than 41 data points, an exact test is performed. If 41 or more data points are available, the normal approximation test is used (Gilbert, 1987).

Data evaluation.
Data evaluation using the statistical methods presented above was automated using the software package Sanitas™ (Intelligent Decision Technologies, Ltd [IDT; now the Environmental Division of NIC], 1999). Sanitas performed all pre- and post-analysis tests required so that the data did not violate size and distribution assumptions required to apply the corresponding statistical methods.

RESULTS AND DISCUSSION

Data can be consolidated. In cases where test fails at 5 percent level, justification is given for consolidation. Primary reason is change in DLs. No outliers documented as errors, so outliers not removed. All data set sizes less than 50, S-W concludes many non-normal distributions. Use nonparametric tests. No seasonality effects shown, so no data deseasonalization. Trend analysis influenced by reduction in DLs, but concluded that no declining water quality or evidence of prior impacts. In cases where there is a relatively low number of non-detects and parametric tolerance limits are calculated, the proposed concentration limits are generally slightly higher than the existing limits.

An important data analysis issue in relatively long records is reduction of detection limits due to improvements in laboratory techniques.
REPORT SECTION: UPDATING LIMITS IN FUTURE

CONCLUSIONS

Expense of groundwater monitoring programs and high false positive rates.

REFERENCES

California EPA, Department of Toxic Substances Control. 2001. *DTSC Hazardous Waste Management Program, Monitoring Guidance Document, Appendix C.*
Gilbert, R.O. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Reinhold, New York, NY.
Hollander, M. and D.A. Wolfe. 1973. *Nonparametric Statistical Methods*. John Wiley Publishing, New York, NY.
Intelligent Decision Technologies, Ltd. 1999 *Sanitas™ for Ground Water, User's Guide*. Longmont, CO.
Montgomery, D.C. 1985. *Introduction to Statistical Quality Control*. John Wiley Publishing, New York, NY.
Sen, P.K. 1968. "Estimates of the Regression Coefficient based on Kendall's Tau", *J. Am. Stat. Assn.* 63: 1379-1389.
U.S. Environmental Protection Agency. 1989. *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities, Interim Final Guidance Document.* Report EPA/530-SW-89-026, Washington, D.C.

U.S. Environmental Protection Agency. 1992. *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities, Addendum to Interim Final Guidance.* Report EPA/530-R-93-003, Washington, D.C.