Water Quality Analysis for the Heartland Inventory and Monitoring Network (HTLN) of the US National Park Service:

LESSONS LEARNED

June 2006

Donald G. Huggins
Robert C. Everhart
Debra S. Baker
Robert H. Hagen

Central Plains Center for BioAssessment
Kansas Biological Survey
University of Kansas
Takeru Higuchi Building
2101 Constant Avenue, Room 35
Lawrence, KS 66047-3759

This report was produced under the cooperative agreement between the University of Kansas Center for Research and the National Park Service (Coop. Agreement #H6067B10031).

Table of Contents

Overview

Aquatic and water dependent environments are fundamentally important to the ecological, biological, chemical, and physical integrity of significant portions of lands and waters protected by the U.S. National Park Service (NPS). As part of its mission to preserve and protect these aquatic and water dependent resources, NPS has recognized the need for a plan to ensure the integrity of water quality within park service units over the long term.

Accordingly, the Heartland Inventory and Monitoring Network (HTLN) of the National Park Service began the aquatic component of the Servicewide Inventory and Monitoring Program in FY 2001. As part of this aquatic component, HTLN entered into a cooperative agreement with the University of Kansas' Central Plains Center for BioAssessment (CPCB) to aid in the development of a comprehensive Network Monitoring Plan. In this endeavor, CPCB's tasks were: 1) to summarize existing relevant state, national, and tribal water quality standards applicable to waterbodies within each of the HTLN parks, 2) to summarize existing relevant state, national, tribal, and NPS protocols for biological monitoring and to determine, when applicable, whether appropriate methods exist for use by each HTLN park, 3) to recommend monitoring designs based on these findings, 4) to update water quality monitoring data collected for each HTLN park since publication of that park's NPS Water Resources Division (WRD) Baseline Water Quality Report, where possible, 5) to develop an accessible relational database to house these data, 6) to re-analyze and summarize the collected data to reflect current knowledge of the condition of aquatic resources within each of the HTLN park service units, and 7) to make recommendations regarding future research and monitoring efforts based on the information and experience garnered from the previous steps. The first three steps reside in documents previously submitted by CPCB (Goodrich and Huggins 2003; Goodrich et al. 2004; Huggins and Dzialowski 2005). Water quality analysis reports, prepared in order of HTLN priority, for Ozark Scenic National Riverways (Huggins et al. 2005b), Buffalo National River (Huggins et al. 2005a), Hot Springs National Park (Huggins et al. 2006b), and Cuyahoga Valley National Park (Huggins et al. 2006a) fulfill the latter four steps of this process.

As part of the cooperative agreement, a document was also requested to summarize any important information gathered during this process that does not appear in any of the previously listed documents. This "Lessons Learned" report (Huggins et al. 2006c) is intended to fulfill this role.

General Comments

Through the course of the background research on water quality standards, biological and physical habitat monitoring, and reference condition characterization, as well as the development of relational databases for water quality data and analyses of existing data by park, the following general points became apparent:

Cooperative agreements are useful, but require clear goal setting, consistent communication, flexibility, and contributions from all partners involved. Throughout the development and execution of this project, many objectives were discussed and pursued. Often, considerable amounts of time and effort were expended to develop information or methodologies to approach a given goal, only to realize that a different goal might be more appropriate. In our experience, this is the nature of cooperative research. Therefore, it is important to maintain consistent communications as efforts continue. Periodic meetings between relevant personnel seem to be much more effective than written correspondence in this regard. With consistent communication, development issues and changing goals can be addressed in a timely fashion, and efforts to date can be recognized by all parties. Future efforts can also be coordinated in this manner. The initiation of meetings between relevant HTLN and CPCB staff in the spring of 2005 helped to reestablish this link. When goals are unclear, contacts are difficult to maintain, or efforts are duplicated, and cooperative agreements tend to suffer.

Also, clear goals and specific questions are a pre-requisite for meaningful analysis. Without a clear understanding of the questions to be answered and what it will take to answer them, efforts may be lost in development of ultimately peripheral information. In specific terms, the development of lists of priority concerns for each park in the Heartland Network, as well as the priority rank of each park relative to the other Heartland parks, was a significant aid in developing water quality analysis reports that are geared toward the actual concerns of the park managers. Such direction has hopefully led to the development of products that are ultimately more user friendly than those previously available. Similar direction from HTLN concerning data collection and database development was also helpful when available.

Once clear goals are defined, be sure to pursue them. Although this seems intuitively obvious, it is apparent from legacy data sets that the main issues of concern to park managers are often neglected in data collection. For example, Hot Springs National Park's (HOSP's) highest priority concern is the drinking water quality of its springs, yet few data are available to characterize this quality. Similarly, stream pathogens such as fecal coliforms are a high concern for the Ozark National Scenic Riverways (OZAR), but data regarding a major source of these pathogens – the park guests themselves – is not currently associated with water quality data. The ability to characterize use with impairment could be particularly helpful for park managers, especially in parks with heavy seasonal use such as OZAR.

Legacy data are often filled with problems. This was especially true for STORET data. Essentially, there are four common problems with legacy data. First, the context under which the data were taken is often difficult to ascertain. Metadata associated with legacy data are often incomplete or altogether missing. This can lead to serious problems when attempting to use legacy data for any meaningful analysis, since information necessary for combining similar parameters (units, recording limits, sampling methods, etc.) is frequently insufficient. Second, sampling locations are often poorly characterized in legacy data. Though common knowledge at the source of the data may easily discern between sites, formal, recorded knowledge is often insufficient to match sites with study areas, observations, regulatory units, or even to themselves over relatively long periods of time. For example, gray literature such as masters' theses, park specific reports, and small-scale research papers often do not have sufficient spatial information to relate their data to existing sampling locations or even to establish new sampling locations. Third, legacy data often require significant efforts to confirm their quality. For the most part, quality assurance is absent in legacy data, and data often have to be verified by inspection. The final common problem with legacy data is that data owners are often unaware of the amount (or lack thereof) of statistically viable data in their possession. Data that cannot be tied to specific stations or that have relatively few (< 5) observations at any location are essentially worthless for analytical purposes. The previously mentioned paucity of drinking water data for springs in HOSP is a perfect example of this issue.

Database development is time consuming and filled with unexpected issues. Regardless of the quality of data, combining electronic databases from multiple sources is often surprisingly complicated. Issues such as software compatibility, file structure, file consistency, and formatting and naming conventions often render automated or semi-automated processes impractical. When planning to combine data from multiple sources into a novel database, provide additional funding and time beyond initial estimates as a hedge against unforeseen difficulties.

Compounding data questions, water quality criteria, standards, and benchmarks are complicated in and of themselves. Multiple criteria often apply to waterbodies within park service units, with either federal or state criteria being more restrictive, depending on specific circumstances. As a further complication, criteria often specify specific conditions (e.g. geometric means of 5 samples within 30 days, or outside mixing zone maximum concentrations for a specific hardness value) that require fairly high levels of data analysis for any meaningful commentary. Often, the relevant criteria must be derived from manual inspection of the standards for each waterbody within a park unit, since these waterbodies are typically unrelated to state regulatory units, at least in park service records. For example, in order to assess whether metal toxicity is an issue at a particular station, that station must be associated with a waterbody, the waterbody's designated use determined, the relevant federal criteria identified, the relevant state criteria identified, and the available data sufficient for meeting relevant criterion requirements must be compared to the relevant criterion values. In some cases, additional, concurrent physical and chemical data (e.g. water temperature, hardness, pH) must also be used to determine toxicity of the parameter being examined. This additional information must further be spatially and temporally linked to the parameter data, which is often complicated as well. Specifically, if total cadmium is a suspected problem, both federal and state criteria for aquatic life (both chronic and maximum), human health, maximum contaminant loading and secondary drinking water, agricultural or industrial use, and primary or secondary contact recreation must all be determined for comparison in order to quantify the relative impact of the *in situ* total cadmium concentration, and to compound the issue, the toxicity and criteria for total cadmium both change with total hardness of the waterbody. However, without a relevant criterion for comparison, it is difficult to assess the impact of a particular water quality parameter. Obviously, this is a very involved process, and previous experience or aid from knowledgeable state or federal personnel could significantly reduce the time and effort required to assess such questions.

Finally, constructive feedback and follow-ups are necessary to improve outcomes. This is essentially a corollary of the cooperative agreement process. Significant efforts have been put into developing the products (literature reviews, databases, water quality analysis reports) of this agreement, and any lessons (such as those included in this document) that may apply to similar projects in the future should be retained along with the products themselves.

General Recommendations from Water Quality Analysis Reports

For convenience, the following recommendations based on the water quality analyses in the reports generated for this cooperative agreement (Huggins et al. 2005a, b, 2006a, b) and based on the various processes required for the development of these reports were also included in this document:

Be sure to document and standardize metadata. Institutional knowledge of station locations, sampling methods, and data handling must be clearly recorded. The implementation of this recommendation is likely a service- or network-wide concern, but each park should be prepared to contribute. The National Research Council has characterized some of the relevant concerns and applications associated with metadata. An executive summary of their recommendations is available at http://www.nap.edu/execsumm_pdf/4896.pdf.

Standardize database files. Rigorous adherence to standards is the key to automating data extraction and analysis, especially with data field names (e.g. DissolvedOxygen, DO, DissOxy) and field types (e.g. text, number, date). NPS has begun to institute servicewide data standards and has been developing new data entry and management tools that should aid in this regard.

Be sure to uniquely define sampling locations. For proper statistical analysis and relational database development, every observation must be linked to a unique place and time. Based on the scale of analysis (e.g. riffle versus watershed), sites may be grouped for analysis, but their original identity should be clearly maintained by documentation.

Relate sampling locations to relevant regulatory waterbody segments. In order to make decisions within a regulatory context, it is important to know which stations are within which regulatory unit. Without easily accessible or fairly precise spatial information, it is difficult to ascertain whether stations contribute to a particular subbasin in a watershed or not. Without this knowledge, it is also difficult to assess the relative water quality at a given site in comparison to its relationship with the watershed, both upstream and downstream. For example, Ohio EPA currently has a system of spatial identification for the Cuyahoga and other river and stream segments, based both on river miles and assessment units. Current CUVA sampling stations should be related to these spatial indices for more meaningful analyses and future data comparisons.

Establish a sampling design for long-term trend analysis. In order to appropriately characterize trends over the long term (decades to centuries), comparable sampling procedures and efforts should be maintained throughout the period of comparison. The goals of trend analysis should be clearly defined to address the scale of the trends being monitored (e.g. differences between months, seasons, years, stations, tributaries, waterbody types), so that sampling design can accommodate the statistical requirements for discerning the desired scale of change.

Take hardness measurements concurrently with metals. Several priority pollutant metals vary in toxicity with hardness. Future studies should include hardness as part of the sampling regime.

Take pH and temperature measurements concurrent with ammonia. Ammonia toxicity is dependent on both pH and water temperature. Since pH and water temperature fluctuate on multiple scales (hourly, daily, weekly, etc.), it is difficult to assess ammonia toxicity without concurrent measurements of these parameters.

Develop study areas along watershed boundaries. Watershed boundaries provide more meaningful study areas than arbitrary upstream or downstream distances. By placing future assessments within a watershed framework, park managers will be better able to recognize the limits of their control over water quality within their parks, and they will also be able to identify the relevant entities whose cooperation will be required for water quality conservation and improvement.

Correlate water quality parameters with actual park use. Some water quality concerns, especially turbidity and fecal coliforms, may be highly correlated to the number of users and the type of use of park waterbodies. Without an understanding of the amount and pattern of park usage, it is difficult to assess its impacts on water quality. In some cases, these impacts may be very significant to waters within and downstream from the park.

Identify springs and study areas clearly and consistently. Due to the nature of springs in areas of karst topography, it can be difficult to accurately locate springs as stations. However, without knowledge of which data are associated with which station, meaningful analysis can be difficult to achieve. A rigorous accounting of springs and development of a geospatial database (i.e. a database using Geographic Information Systems (GIS) to identify the geographic location and attributes of each spring) is recommended.

Periodically sample springs for drinking water quality parameters. Since drinking water quality of springs is a major concern of HOSP, waters from relevant springs should be periodically sampled to ensure relevant drinking water standards are being met. Anecdotal evidence (e.g. the fact that no water-related illnesses are being reported) is sufficient for general information, but actual testing would provide the park a valuable scientific and legal background.

Exchange information with other agencies that study park-related waters. For example, Ohio EPA and USGS also perform water quality research in the Cuyahoga basin. Since major portions of the Cuyahoga River are currently listed as impaired by Ohio EPA, it is apparent that significant water quality research efforts and literature are already associated with waters relevant to CUVA. Information sharing, including data exchange, circulation of published reports, and contact between data managers and relevant staff, could greatly benefit all parties involved.

Documents Generated through this Cooperative Agreement

Goodrich, C. and D. Huggins (2003). "National Park Service Inventory and Monitoring Program: Review of state water quality standards applicable to Heartland Network Parks." Central Plains Center for BioAssessment, Kansas Biological Survey.

Goodrich, C., D. Huggins, et al. (2004). "Summary of State and National Biological Assessment Methods, Physical Habitat Assessment Methods, and Biological Criteria." Central Plains Center for BioAssessment, Kansas Biological Survey.

Huggins, D. G. and A. Dzialowski (2005). "Identification and Quantification of Reference Conditions Associated with Lotic Ecosystems of the Mid-Continent Region of the US: A summary of approaches and factors applicable to Heartland Network facilities." Central Plains Center for BioAssessment, Kansas Biological Survey, University of Kansas.

Huggins, D. G., R. C. Everhart, et al. (2005a). "Water Quality Analysis for the Heartland Inventory and Monitoring Network (HTLN) of the US National Park Service: BUFFALO NATIONAL RIVER." Central Plains Center for BioAssessment, Kansas Biological Survey, University of Kansas.

Huggins, D. G., R. C. Everhart, et al. (2005b). "Water Quality Analysis for the Heartland Inventory and Monitoring Network (HTLN) of the US National Park Service: OZARK NATIONAL SCENIC RIVERWAYS." Central Plains Center for BioAssessment, Kansas Biological Survey, University of Kansas.

Huggins, D. G., R. C. Everhart, et al. (2006a). "Water Quality Analysis for the Heartland Inventory and Monitoring Network (HTLN) of the US National Park Service: CUYAHOGA VALLEY NATIONAL PARK." Central Plains Center for BioAssessment, Kansas Biological Survey, University of Kansas.

Huggins, D. G., R. C. Everhart, et al. (2006b). "Water Quality Analysis for the Heartland Inventory and Monitoring Network (HTLN) of the US National Park Service: HOT SPRINGS NATIONAL PARK." Central Plains Center for BioAssessment, Kansas Biological Survey, University of Kansas.

Huggins, D. G., R. C. Everhart, et al. (2006c). "Water Quality Analysis for the Heartland Inventory and Monitoring Network (HTLN) of the US National Park Service: LESSONS LEARNED." Central Plains Center for BioAssessment, Kansas Biological Survey, University of Kansas.

Data Products Generated through this Cooperative Agreement

Relational databases were compiled from legacy data provided by the park service, combined with available data from federal and state agencies relevant to each park service unit. While these data were often inclusive of observations made after the original NPS Water Resources Division Baseline Reports, additional data may reside in gray literature or other documents published by NPS or other agencies. Significant efforts were made, however, to include data from as many available sources as possible.

Relational Databases

Water Quality Database

A relational database based on the NPS National Relational Database Template with input from HTLN was developed in MS Access 2000. Water quality data for the following park service units are included, and are listed in order of Heartland Network priority rank. Data screening and station resolution were performed for the first four listed.

Ozark National Scenic Riverways (OZAR) (2005)
Buffalo National River (BUFF) (2005)
Hot Springs National Park (HOSP) (2006)
Cuyahoga Valley National Park (CUVA) (2006)
Effigy Mounds National Monument (EFMO) (2004)
Herbert Hoover National Historic Site (HEHO) (2004)
Tallgrass Prairie National Preserve (TAPR) (2004)
Pipestone National Monument (PIPE) (2004)
Arkansas Post National Memorial (ARPO) (2004)
George Washington Carver National Monument (GWCA) (2004)
Pea Ridge National Military Park (PERI) (2004)
Homestead National Monument of America (HOME) (2004)
Lincoln Boyhood National Memorial (LIBO) (2004)

Data for Wilson's Creek National Battlefield (WICR) were made available in the fall of 2005, but were not compiled, since they will be included in the NPS Water Resources Division Baseline Report still under development for this park service unit. Similarly, Baseline Reports for George Washington Carver National Monument (GWCA) and Hopewell Culture National Historic Park (HOCU) had not been completed or made available at the time of publication of this report.

Analytical Databases

Two additional relational databases were developed during the course of this agreement.

The first, “NPSRecommendations.mdb” (2004), contains information relating waterbodies, designated uses, sampling regimes, and priority concerns to the various park service units that make up the HTLN.

The second, “EPApriorityPollutants_CompletedScreens.mdb” (2005) summarizes USEPA Water Quality Criteria & Recommended Drinking Water Standards (based on 2002 documents), including hardness-dependent criterion calculations for various hardness values and representative hardness values for the four park service units with full water quality reports (i.e. OZAR, BUFF, HOSP, and CUVA).

Data Tools

Early in the water quality database development process, it became apparent that data conversion would require significant efforts. An attempt was made to automate this process, but legacy data proved too inconsistent for full automation. However, tools were also developed to aid in manual conversion of data. These automated and manual tools are housed in a front-end/back-end model database called “ConversionTools.mdb” (2004). A basic manual of use was also developed for this database.

Additional analysis tools (primarily data screening queries and process documents) were developed through the course of the cooperative agreement and are available upon request.