Coryell A. Ohlander
6048 So. Lakeview Street
Littleton CO 80120
cohlander@polnow.net

April 14, 2000

To: Content Analysis Enterprise Team
USDA Forest Service
Bldg. 2 Suite 295
5500 W. Amelia Earhart Dr.
Salt Lake City, UT 84116 (cleanwater/wo_caet-slc@fs.fed.us)

FROM: Coryell A. Ohlander, hydrologist

RE: Good Science aspects in UFP
Part 4 regarding evidentiary standards

Goals "... in a unified and cost-effective manner."
Guiding principles: "...consistent and scientific approach...."
Agency Objectives: "...common science-based approach...."
"...test watershed assessment procedures...."
"...implement ... consistent with applicable legal authorities."
"...will base watershed management on good science."
"...science-based total maximum daily loads (TMDLs)."
"...sharing of scientific and technical resources;"
"...monitoring and evaluation...."

Summary Recommendation

The application of the Clean Water Act is a legal exercise, as such, efforts at monitoring and evaluation under the auspices of good science needs to meet minimum evidentiary standards. This is particularly true for TMDL because States are looking at what constitutes "creditable evidence" for listing - or delisting - stream segments on the 303(d) list. This will not be an academic exercise for UFP. Because agency efforts rarely use evidentiary standards to scope their efforts, UFP needs to articulate a structure for determining if monitoring/evaluation efforts actually meet legal requirements. Since the structure and quality control are also determined by monitoring objectives, I suggest UFP define a standard set of monitoring plan objectives as well.

The UFP needs to emphasis the use of Standard Methods and to generate a committee of scientists to review field procedures if they are not already part of Standard Methods. There are far too many roll-your-own methods in use that can not be supported in a legal setting and are, in some cases, just plain sloppy work. The CWAP will demand huge resources to make it come about; UFP can't afford to waste its time and resources.

Given UFP objectives, you need to declare a process and perhaps certification for training levels to define competence. I hate to recommend certification, but we don't seem to get anywhere unless there is some form of accountability attached to job performance. Sooner or later, both staff and managers must be held accountable for the hydrologic condition of watersheds and related aquatic health. Without accountability, the UFP exercise is a nonevent.

APR 14 2000
CAET RECEIVED

The UFP also needs to correct the current convention of assigning a lesser risk to environmental damage than to restricting land use. Under the current convention, as expressed in statistics with Type I and II errors typically set to 0.05/0.20, we are effectively saying that we are 4 times more willing to accept environmental damage than to restrict an economic usage. I recommend that UFP set Type I and II errors equally based on a front end - and daylight - process that balances the risks associated with the project or land use. In the T-Walk rationale, Type I and II errors are set at 0.05/0.05 and sampling is established to meet these statistical levels.

Support

1) Under CWA S505, Citizen Suits, the issue is a factual one of whether there has been compliance or not. The burden of demonstrating compliance is on the agency (as actor) and must be supported by an objective evidentiary standard of "considerable record" [3745 & 3746 USCC&AN 1972]. I would argue that failing to meet evidentiary standard also would fail to meet the UFP's need for good science.

   I believe that UFP needs to incorporate criteria for what constitutes evidentiary quality data so there is no doubt about what UFP considers good science and essential elements of routine monitoring/evaluation. And, sad to say, this can not be left up to the judgement of the field people doing the monitoring. I recommend the following questions, obtained from Charles Lennahan (retired OGC) as a foundation:

   - Have data collection methods been standardized and applied consistently?
   - Have analytical methods been standardized and applied consistently?
   - What type of data, accuracy, and detection limits are required?
   - What time and space is represented?
   - What frequency and distribution is represented?
   - What is the physical, chemical, or electrical basis of the equipment?
   - Does equipment measure what is required?
   - Does it have the required accuracy under field conditions?
   - How many units of equipment are required?
   - Are maintenance and calibration provided for?
   - How and when will data be collected?
   - How will data be recorded and stored?
   - What is the introduced error?
   - Is the analysis statistically, mathematically, and procedurally valid?
   - Is interpretation logically correct and appropriate to the objective?
   - Are the people involved with equipment and collection competent?
   - Are the people involved with analysis and interpretation competent?

   In spite of the fact that the list is pretty much common sense, I suggest that UFP address the questions directly and not take it for granted that such questions are used as a monitoring/evaluation foundation. While these questions were used as a foundation for T-Walk, I know of no other effort that makes a similar review. Even the vastly expensive ($10 million) Colorado Water Division I Court case in the late 1980's did not track these questions - much to the detriment

OACT RECEIVED

APR 14 2000

of the Forest Service case. Opposing lawyers were able to discredit 3/4 of all FS data sets and analyses because of sloppy or incomplete field data, incorrect analyses, and plain failure to follow instructions. Part of the problem was the frequent rotation of field hydrologists as short-term details expired and new people started up. Nevertheless, what disturbed me the most was the serious lack of field skills needed by rank and file hydrologists to carry out detailed site investigations - even to the point of not knowing how to make discharge measurements or spin test equipment.

The UFP makes a big point that the policy will be implemented without increases in people assigned to do the job. Absolute nonsense. Agency management will either voluntarily re-allocate or be forced to by the Court system that will dictate changes the federal agencies must make. All one has to see is the TMDL lawsuits to know the Courts have the power and vehicle RIGHT NOW; I expect CWAP to further set the stage.

Given UFP objectives, the most difficult task will be to address the training - and certification - fundamental to these four questions:

- Have data collection methods been standardized and applied consistently?
- Have analytical methods been standardized and applied consistently?
- Are the people involved with equipment and collection competent?
- Are the people involved with analysis and interpretation competent?

As you might expect, both "consistent and competent" require some hard-nosed training **and testing** to insure competence. The lack of this kind of training - plus the FS bias supported over many years in promoting GS-13's and above based on managerial skills without technical skills - has resulted in a degraded and reduced technical gene pool. The technical pool has never been lower than it is now. Consider, for example, that in 1964 the Watershed Director (WO) had enough technical knowledge to help author a Chapter on Forest Hydrology in the Handbook of Applied Hydrology. Now, that quality of technical expertise resides at GS-12 and below. The implication is that, for UFP to accomplish its goals, the entire effort at training has to run counter to current downward trends of expertise development and retention. I would also guess that many, if not most, FS managers see no problem with further loss of technical ability, which makes a hefty effort at training a lot like going up the down staircase at rush hour.

2) Because of legal implications, impact studies need to stay pretty close to officially recognized methods such as those in Standard Methods for the Examination of Water and Wastewater. While it is true that each project brings its own special needs, UFP can set the guidance for the use of Standard Methods or list equivalents if Standard Methods does not treat the subject. Considering that there are five different levels or kinds of monitoring, one UFP role should include identifying methods including technical committee reports and published critical reviews that could be used for each monitoring level.

CAET RECEIVED

APR 14 2000

The overwhelming problem is that the quantity and quality of technical expertise itself is in trouble. Being a technical or subject matter expert is not rewarded by the FS and promotions to GS-13 and above are typically program management related. Few program managers know anything about Standard Methods or suitable equivalents and the technical staff are often inexperienced. For example, during the national T-Walk review two macroinvertebrate productivity studies were presented: the first used data based on weighing the samples still wet; and the second used data based on a physical sample of 0.1 square meter in large cobble and boulder substrate. Both studies were accepted by the review group without challenge - in spite of the fact that Standard Methods (and common sense) requires drying macroinvertebrate samples first before weighing and taking large enough physical samples (0.25 square meters) to actually encompass the large substrate.

3) Good science dictates that carefully identified technical objectives be specified in advance so that monitoring efforts are effective. In spite of EPA's technical guidance, however, it is common for monitoring to be implemented without any objectives, or at best, only vague ones. Current regulatory and legal mandates can be achieved by using a combination of 5 monitoring purposes; I suggest that UFP adopt them:

> Trend monitoring -- a long term data collection effort designed to detect statistical trends and to compare current events with long term patterns.

> Reference monitoring -- designed to describe and quantify physical, chemical, and biological parameters of pristine (or best available) conditions in representative ecological situations.

> Advance warning system monitoring -- a continuous, on-going effort to identify conditions that are or could lead to environmental deterioration, litigation, or unnecessary restoration costs.

> Compliance monitoring -- comparison of project oriented activities and resource conditions with established standards or requirements.

> Effects monitoring -- determination of the effects of human activities on the relationships among climate, geology, topography, vegetation, biota, and soil variables; measured in terms of selected physical, chemical, and biological factors of the target ecosystem.

4) Starting with the bottomline ...

Maintenance of ecological integrity requires that any changes in the environment resulting in a physical, chemical or biological change ... be of a temporary nature, such that by natural processes, within a few hours, days, or weeks, the aquatic ecosystem will return to a state functionally identical to the original. [3742 USCC&AN 1972].

CAET RECEIVED

APR 14 2000

... creates the statistical playing field that draws our attention to what to measure, the time increments, and what needs to be part of the analyses. The legislation gives us a list of elements to monitor:

* *key species, natural temperature and flow patterns.... [72:3717; 33 USC 1314]*
* *Concentration and dispersal of pollutants (w/ by-products) through biological, physical, and chemical processes and any related changes in the diversity, productivity, or stability.... [72:3716; 33 USC 1314]*
* *Description of factors related to rates of eutrophication; organic material accumulation; and inorganic sediment accumulation.... [72:3717; 33 USC 1314]*
* *Dissolved oxygen conditions needed by location, species, and activity (hiding cover, propagation, food supply, reproduction).... [77:4366; 33 USC 1311]*
* *Effects of Road construction, use, and maintenance on the biological character or flow, reach, and circulation.... [77:4401; 33 USC 1344]*
* *Factors needed for restoration of the natural chemical, physical, and biological integrity.... [72:3715; 33 USC 1314]*
* *Effects on hydrologic cycle and storm runoff [72:3719; 33 USC 1314]*
* *Accurate assessment and comparison of existing condition to water quality objectives to be met.... [72:3681 33 USC 1314]*

In generic terms, an impact study is best designed when it judges impact effects against previously collected baseline data, and it is best used when the results provide a structure for subsequent monitoring to detect future impacts of the same type. That certainly suggests some effort at standardizing data definition, collection, analysis, and interpretation. But to accomplish this, UFP will have to rule with an iron hand all the renegades and wanna-be renegades.

Determining how much information to get is also a function of how willing you are to be wrong; in particular, sample size is contingent on assignments made for Type I and Type II errors. (Type I is the probability of declaring a false difference; Type II is the probability of missing a true difference). Type I and Type II are commonly set at 0.05 and 0.20, respectively (aka 95% Confidence at 80% Chance).

However, Lee (*Lee, Kai. 1993. Compass and gyroscope: integrating science and politics for the environment. Island Press. 243p.*) argues that for environmental issues, the consequences from Type II errors are as bad, and often worse, than Type I errors and that the application of Type I = 0.05 and Type II = 0.20 creates a bias that guards against "true that turn out false" but not "false that turns out true." Falsely concluding that nothing bad will happen, the Type II error, is the crux of NEPA documents as well as the Court injunction problem.

(The basis for injunctive relief is irreparable injury and the fact money will not make it better (857 F.2d 1307; 107 S.Ct.1396). The presumption is that environmental injury, by it's nature, can seldom be adequately remedied by money damages and is often of permanent or long duration; i.e. irreparable. (840 F.2d @722; 107 S.Ct. @1404). Thus, when environmental injury is sufficiently likely, the balance of injury will usually favor the issuance of an injunction to protect the environment. The court must expressly consider the public interest and the record must be "adequate". (851 F2d @1157)).

RECEIVED
APR 14 2000

5) If UFP accepts the emphasis on judicial use under CWA S505, then the corresponding need is for monitoring programs to do 3 things:

1) make efficient use of time and resources in data collection;
2) define the significance of impacts in social/political terms; and
3) train/qualify people to make reasoned and consistent judgements.

These objectives must function together. The exploding spectrum of environmental issues forces evaluations to be done in compressed time frames of hours or days rather than weeks or months. The need to rely on the subjective judgement of professionally competent people is as great day as it ever has been. That, together with the current downsizing and re-engineering focus on removing experienced people as rapidly as possible, blazingly accelerates the need to train inexperienced people on how to cope. UFP, to carry out its mandate, will need to address the question with bite and perseverance.

Given that statistical power (the probability of rejecting the null hypothesis when it is false) is constrained by the variability, the magnitude of the impact, and the number of independent sampling events, then the major limits on time and resources creates a need for efficient statistical designs that deal explicitly with the lack of spatial replication and randomization. UFP might want to look into the range of approaches that satisfy this criteria. T-Walk is structured as a "Before After Control Impact Paired" (BACIP) approach, but there might be others that would fit as well.

A primary purpose of BACIP is to determine spatial and temporal variance of natural systems for, at least, the variables expected to change as a result of the activity; and this can be a long list. The key is to select the variables and features that are relatively sensitive to the earliest effects, yet can be obtained and analyzed without undue capital investment or complex theoretical arguments.

To be efficient, monitoring has to specify the parameter differences worth detecting and the Type I and II errors. What is worth detecting is a matter of judgement, and sometimes policy, about the relationships among variables and expected impacts. The question needs to be explicitly answered early in the monitoring process, so an effective study plan can emerge. As a matter of impact assessment policy, CWA S404 creates an obligation to "fully consider" the views of the USF&WS; UFP could enlarge upon the S404 directive and declare that a 20% change in the HSI index would be worth detecting.

The use of Signal/Noise Ratios, such as CohenD, daylights the assignment of what constitutes important resource differences and related risk (Type I and II) beforehand. The purpose of monitoring then stops being an inventory and becomes an effort to see if pre-selected biological increments are exceeded. UFP could make this happen with very little effort.

CAET RECEIVED

APR 14 2000

6) Professional judgement is the answer given to justify nearly all land management decisions and is at UFP's foundation. But in a judicial setting, professional judgement is wide open to challenge regarding the expert's qualification, past training, the authority to make the call, and the quality of the supporting information. To accomplish UFP's objectives for consistency in professional judgement, there has to be a lot more than several bureaucrats agrees to a set of words. To prepare and use this expertise requires that UFP routinely provide for:

Testing consistency of definitions used for categories:

Given a cadre of experts and a series of comprehensive data sets, to what extent will the experts select the same categories? Goes to the issue of clear and concise definitions and the ability of knowledgeable people to apply them uniformly. If the experts independently come to the same conclusions, then the definitions and procedures are consistent.

Testing trainees against a panel of experts:

Assuming clear and concise definitions and consensus by a panel of experts for a series of suitable data sets, to what extent would a trainee match the experts? Goes to the issue of developing an expert status and being able to support professional judgements. Also goes to the issue of the quality of training against some measure of competence.

Using professional judgement to select categories about the past from an assortment of past and current data, photographs, and field visits:

Assuming qualified people, to what extent will they agree on past conditions? This goes to the issue of using experts to make category judgements about conditions given the typical mismatch of data sets. Also goes to the issue of comparing trends for the same site using old data, historical notes, and photographs against current conditions.

Using professional judgement to predict future categories and trends given current data, photographs, and field visits:

Assuming qualified people, to what extent will they agree on future conditions? This goes to the issue of using experts to make category judgements about future conditions given a plan of action.

End of Part 4 letter.

CAET RECEIVED
APR 14 2000