Analytical Reliability in the Decision Making Process—The Numbers Game

RICHARD MCQUISTEN AND KARL A. GEBHARDT

Abstract

This paper is not intended to give the pros or cons of conventional sampling methods (generators of numbers). Rather, our intent is to emphasize the importance of the credibility and integrity of the generated number, by what ever methodology selected, and point out some problems that can be encountered when numbers are poorly generated or analyzed. As generators of numbers we are responsible for the end-products of data analysis for the decision maker.

In response to a 1977 Presidential directive to the Council on Environmental Quality, CEQ, an Interagency Task Force was developed to review present federal environmental monitoring and data collection programs and to recommend effective improvements. The Federal Government depends upon the analyses that result from such programs to direct sound policy and decisions. These directives affect the public health and welfare and result in large annual expenditures of funds by all levels of government and the private sector. During the past few years, several major federal environmental and data collection programs were found to be inadequate.

Although concern about the reliability of analytical results has always existed, this concern has come more into focus because of the growing interest and activity in environmental pollution control, with its heavy reliance on analytical results for enforcement, regulation, and litigation. This concern primarily is due to the inherent limitations in the conduct and analysis of observed measurements. Certainly, since there are no absolutes in analytical results, in terms of a particular protocol, some indication of value or reliability of the results is needed.

Consistent with environmental pollution control, range management deals with precious resources and cannot be ignored when implementing monitoring and data collection programs. Vegetation data, as do other natural resource data, relate the condition of a system at a particular place and time. It is not possible to perform verifying remeasurement because of the inherent special and temporal variability of the system. Additionally, if a sample is collected and analyzed, the nature of the sample is often changed in the measurement process (i.e. clipping, water quality dynamics, etc.). Because of the impossibility of verifying past measurements, quality assurance can only rest on *documented* application of *proven methodology* by *qualified personnel* following accepted *guidelines on sampling design.* The accuracy of measurement and analysis are all too often placed at the bottom of range management priorities. Inherently, the lack of such accuracy can result in litigation, incorrect decisions, or even worse, poor range management.

Anyone active in the environmental field appreciates the importance of accurate measurements. Many important environmental issues require that difficult decisions be made on the basis of data taken at numerous monitoring stations over long periods of time. Such data, include many different parameters (utilization, composition, cover, water quality, erosion, etc.) that are simultaneously fitted into decisions which manage the range resource. The importance of the data used in rangeland management cannot be overlooked. Simply, we must be prepared to provide accurate quantitative answers to questions such as: "How much variability, both in space and time, is there from a measurement system?", "Is the variability equivalent in other similar locations?", "What is the significance of the differences observed in the two locations?", and "How well do the analyses represent what should be done for the range resource?".

Authors are environmental engineer, USDOE, Laramie Energy Technology Center, P.O. Box 3395, University Station, Laramie 82071; and P.E. state hydrologist, USDI, Bureau of Land Management Idaho State Office, 83724.

The authors wish to express their appreciation to Jim Westhoff, DOE-Laramie Energy Technology Center, and to Patrick Green, Bruce Van Haveren, Rich Herbert, Cliff Franklin, and Dave Green, USDI, Bureau of Land Management, for their valued assistance and review.

Manuscript received July 7, 1981.

Reliance on Numbers

Our present day society has increased its reliance on numbers to staggering degrees. This can generally be attributed to our increase in technological capability and refinements in data collection and processing. Computer technology has increased our ability to utilize quantitative techniques to the extent that once complex, time-consuming calculations can now be done in a matter of seconds. With this increased capability has come an almost sacred reliance on the *number*, (as an absolute value, |X|). Many of us use numbers without regard to their origin, reliability, or limitations. Numbers can be useful, or they can be quite devastating—depending upon how they are generated and interpreted or misinterpreted. It is incumbent, then, upon the resource specialist (the originator of the "number") to identify and quantify all the factors which affect the precision and accuracy of measurements, identify inherent variability and uncertainties in the sampling and analyses process, and to establish a monitoring system which will detect important environmental changes.

Precision, Accuracy, and Significance

Technology is advancing so rapidly that many basic concepts are not adequately understood nor applied to the depth required by present day needs. Some research effects have become so acutely focused on details that application to the basic problem becomes lost. So it is with numbers. To many, numbers are presented without regard to the user needs (those who must interpret or understand them). This often results in confusion, wasted effort, and loss of credibility. Nothing confuses a concept like the person who uses but does not fully understand the generated number. Numbers must be understood in terms of precision, accuracy, and significance.

Precision measures mutual agreement among individual measurements, usually under prescribed similiar conditions (EPA 1980). It describes reproducibility, independent of accuracy. For example, if a range technician takes a large number (n) of samples (of N observations) on a range and each sample comes out as 14% cover, his sample is very precise (Schultz et al. 1961). Many factors affect the precision of an analytical measurement. Included are sensitivity to sample contamination, knowledge and skill of observer, error contributed by the sampling device, and natural influences in the system. Equally important is the ability to obtain a representative sample aliquot for each replication.

Accuracy is the degree of agreement of a measurement with an accepted reference or true value (EPA 1980). In the above example on precision, accuracy of the observations is unknown. The application of "accuracy" can be better understood if the types of data classes are understood. Generally three basic types of data exist (Lide 1981).

Class A—repeatable measurements on well-defined systems. Such data are subject to verification by repeated measurement.

Class B—observational data. Time and space dependent. Generally cannot be checked by remeasurement.

Class C—statistical data—not necessarily scientific, such as, demographic data, energy consumption, health statistics, etc.

All too often, accuracy and precision are confused. Schultz (1961) points out that there are too many papers in range literature referring to "best" methods, implying most accurate, using one favorite method as a standard. Unfortunately, it seems we must continue to relearn the basics; i.e., sampling procedures and methods of measurement must be compared with known values.

Significance is a difficult, complex, and frequently misused concept. Significance, whether in terms of difference or similarity, can exist only in terms of a relationship to some base datum, e.g., a previous measurement. Knowing the precision and accuracy of a measurement leads to an understanding of significance, whether it refers to a change in vegetal cover or complex; a violation of a given standard, e.g., water quality; or some other acceptable base line. Often times, a level of significance exists at too low a level of confidence to make a justifiable decision on the basis of the reported data. Thus, significancy determination (of a number) usually does require basic statistical analyses, i.e., sensitivity, reliability, correlation, and error, to determine either similarity or dissimilarities. Statistics in range management should be considered a necessity rather than a luxury.

Validity of a Number

A number's usefulness depends upon some given level of confidence. Unfortunately, in entirely too many instances it can be shown that our methodologies are such that a very minor change (e.g., a reagent concentration, a "professional opinion of the amount of vegetal cover", etc.) would result in a major change in the value of the reported number. Conversely, other methods are so inexact that a relatively large change in input data leads to little, if any, change in the final result. In the extremes of both these cases there is a valid question as to what is really being measured.

A measurement system can be inaccurate and imprecise, having both random and systematic errors. It can be precise but inaccurate, having systematic errors. It can be accurate and precise, having neither random nor systematic errors. Before placing a great deal of confidence in it, the use must consider the makeup of the number. Validation of the number requires several types of information. A few examples are:

- Was appropriate and adequate sampling equipment used?
- Were sampling procedures described in detail?
- Were precautions established to avoid sample contamination?
- Were appropriate techniques used in selecting sampling sites?
- Were replicate samples taken in order to define sampling variablity?
- Were background samples taken?
- Were the sampling frequency and duration adequate and appropriate to the purposes of the project?
- Was the sampling program designed to assure all samples are representative of the source?
- Was the sampling method commensurate for intended application; i.e., is it too accurate or not accurate enough?
- Did the analytical methods measure the appropriate variable and result in data of adequate detection limit (accuracy and precision commensurate with project requirements)?
- Were the results reported correctly, (with no typographical errors and with correct units assigned)?
- Were samples preserved to prevent changes between the times of collection and analysis?
- Were the analytical methods able to overcome suspected interferences in the samples?
- Were the personnel making the measurement qualified to do so?

In principle, a generated number can be accurate and imprecise; however, if the imprecision is high, there is no reasonable way to test its accuracy. Ideally, a number will be both accurate and precise. However, since this ideal situation is seldom achieved, it becomes necessary to define an acceptable level of confidence in the data. A conceptual design for hiearchial parameter confidence is expressed in Table 1. (These are arbitrary values, meant only to provide an exemplary data use categorization). Ultimately, the goal should be to combine all parameters to determine a confidence in all data generated for a specific analysis report in order to define the usefulness of the report's conclusions.

Table 1: Data use categorization for each parameter.

Level of Confidence	Use
>95%	General purpose, litigation, etc.
>85%	General purpose, land use decisions, etc. except litigation
>75%	Reports, with qualifications
<75%	Information purposes or discard

Combined Numbers

Typically, use of numbers is not single-purpose oriented. In land management practices, it is common to make various biological measurements, which generally are not used until they are combined with other data to provide information ultimately leading to a land management recommendation—livestock grazing capacity, for example. A determination of this type relies on vegetation measurements, livestock use, and climatic data. Rarely do the final capacity figures include an indication of uncertainty (measurement error) of the numbers used in making the determination. This places the decision maker in a tenuous position of not knowing the significance of the capacity figure so that reasonable adjustments to the carrying capacity can be made. Can he go higher or lower? How important is the climate? Should the allocation include some adjustments that consider occurrence frequency of drought? And so on.

Another example involves determination of soil loss using soil loss equations. Most of these equations were empirically developed for use on cultivated crop lands and are now being applied to other areas such as rangelands. These equations are made up of a number of parameters including precipitation, erodability, length of slope, steepness of slope, and ground cover. The individual parameters measured generally do no have a known degree of associated error specified. Any one of the parameters can have considerable variability, which may result from multiplicative, divisional or exponential error, causing the production or estimate from the equation to have a high degree of unreliability.

It is vitally important that a minimum confidence limit for the factors involved in a combined number, and the sensitivity of the combination for any factor with a low confidence limit, be established before a generated "number" is used. Otherwise, the number user may obtain an unequivical value upon which his demonstrated conclusion is based.

Clearly then, use of combined data must be carefully assessed and presented, relative to its end use, or a correct final result may not be reached.

Over-Extension

Over-extension is the practice of developing conclusions based on projected data of limited range and/or application. While many over-extended conclusions may be correct, over-extension becomes serious when critical decisions are based upon incomplete, inaccurate, or limited analyses. Usually it presents a far more credible description than is actually the case, for example, extrapolation beyond the range of values in a regression equation. Application of over-extension causes a false sense of security to those who depend on the information developed to make a decision. To avoid over-extension, one must, at a minimum, qualify assumptions and conclusions developed from such data.

Some of the best examples of over-extension can be found in environmental analyses and impact statements. Often these documents are required to quantify impacts on the environment using qualitative data. To be colloquial, they often use "space-age analysis with horse and buggy data."

Traceability

The word "traceability" causes much confusion because it is one of those words that mean different things to different people. Citing of traceability requirements by regulatory agencies, for environmental measurements, is common. For example, the Environmental Protection Agency invokes traceability requirements in its "Hazardous Waste Guidelines and Regulations" [40 CFR Part 250], December 18, 1978.

Interpretations of traceability have sometimes focused on the documentation of instrument calibration rather than on the accuracy of the measurements. For example, one may have a "widgit" detector properly and accurately calibrated, but if it is improperly used, the data it produces will be worthless. In brief, calibrated instruments are usually a necessary, but often insufficient, condition for demonstrating traceability. Many factors besides the quality and accuracy of the instruments used will affect the accuracy of the measurements, including the procedures used, skill and training of the operator, environmental conditions, etc. Traceability cannot be achieved until it can be shown that the actual measurements produced by the measurement process are, on a continuing basis, accurate relative to national or other established (specified) standard (Belanger 1978).

Faulty Analysis

Faulty analysis, like over-extension, is an error on the part of the person generating the number. However, unlike over-extension, faulty analysis usually has acceptable data but poor application of analysis techniques. If the method selected is inappropriate for the situation, results will be inaccurate. Without thorough knowledge and evaluation of the factors affecting various methods, an inappropriate method may be selected. Since most measurements and/or samples represent a unique set of conditions a universally applicable method is difficult to establish. Lack of descriptive controls, resulting in misinterpretation of good data, is an additional cause of faulty analysis. For example, in many research projects, even though proper study designs and analysis techniques were used, the project conclusions are invalid because good and bad data were used interchangeably.

Problem Solutions

To this point, this discussion has attempted to define the problem of using imprecise, inaccurate, unvalidated, unreliable, misinterpreted data and, to some degree, limitations in using derived numbers. As a means of solution, it is suggested that utilizing good quality assurance/quality control procedures will provide the sought-for level of acceptability. Quite likely in the near future, as a result of the CEQ Task Force report on Quality Assurance (mentioned in the introduction), each Federal agency responsible for monitoring programs will be required to establish its own quality assurance program. Such programs should be adopted by the natural resource profession in order to promote reasonable and scientifically sound management decisions.

Conclusions

The importance of measurements required by environmental analyses and regulations is becoming increasingly recognized. The inherent concern is detecting significant changes resulting from a proposed action. As investigators, results of our analyses play an extremely crucial role in the entire environmental process. Therefore, we need to be more concerned with using representative data in determining changes in environmental conditions, emphasizing identification of natural variability, monitoring of real changes and traceability to accepted standards. As generators of numbers, we are responsible for the end-products of data analyses, and we must be cognizant of the many problems that can be encountered when numbers are poorly generated or inadequately explained. Briefly stated, we must be assured of the credibility and integrity of the number generated for the decision making process.

Literature Cited

Belanger, Brian C. and Kieffer, Lee J. 1978. The NBS Role in Quality Assurance for Physical Measurements. Paper presented at the Nat. Conference on Quality Assurance; Denver, Colo. Nov. 27-29, 1978.

Environmental Protection Agency (EPA). Draft April 29, 1980. Guidelines and Specifications for Preparing Quality Assurance Program Plans.

Lide, David R. Jr. 1981. Critical data for critical needs. Science 212:1343-1349.

Schultz, Arnold M. Gibbens, Robert P., and DeBano, Leonard, 1961. Artificial populations for teaching and testing range techniques. J. Range Manage. 14:236-242.