INTRODUCTION

The collection and laboratory analyses of samples needed to characterize environmental quality are already expensive. Further, as society expresses increasing concern for environmental protection and as instrumentation technology can detect contaminants at ever lower concentrations, expenditures for monitoring environmental quality will rise with time.

As a direct consequence of the rising costs of environmental monitoring, it is essential to use available environmental quality data effectively. Effective utilization involves answering questions such as, "Is the environmental quality acceptable?" and, "Is the environmental quality improving or deteriorating?" Responding to these types of questions requires interpretation of data, and this stage of assessment is beset with difficulties. Some difficulties with interpreting environmental-monitoring results:

(i) Since the data are frequently expensive to accumulate, the data sets being interpreted are usually very modest in size.

(ii) The data may involve a vector of chemical and biological constituent measurements because consideration must typically be given to a range of constituents. Correlation between the constituents may help the infilling of missing data or the identification of outlier data.

(iii) Early detection of any deterioration in environmental quality is highly desirable because early detection may provide the opportunity for controlling the problem at a lower cost before the problem magnifies. Any procedure for identifying early warning signals must not, however, falsely identify a problem of apparent environmental deterioration when one does not actually exist; nor should it fail to identify a problem when one does exist.

(iv) The vagaries of nature introduce significant noise and sources of variability such as seasonal effects. This can make the identification of trends more difficult.

(v) The derivation of quantitative risk assessments is in many ways data dependent. But will the information returned by these risk estimates be worth additional data collection efforts?

The net result of difficulties such as the five mentioned is that making sense of environmental quality data necessarily involves statistical interpretation. Statistical interpretation procedures must be sensitive to small changes in environmental quality and yet recognize the potentially substantial costs of any additional data collection requirement.

A multitude of statistical analysis tests are available, but each of the tests possess assumptions that may or may not be appropriate for specific circumstances. Computer programs now becoming widely available facilitate use of various procedures. The difficulty remains for the student and the practitioner to learn which conditions dictate a particular procedure and which conditions render it highly inappropriate.

For example, look at the following analysis:

LET $A = B$

THEN

$$AA = AB$$
$$AA - BB = AB - BB$$
$$(A + B)(A - B) = B(A - B)$$
$$A + B = B$$
$$2A = A$$
$$2 = 1$$

The black box here tells us that 2 = 1. The problem in this procedure is when the (A – B) was “cancelled” from both sides. What is actually being done here is that both sides are being divided by (A – B). But this can only be done if A is not equal to B because A = B gives division by zero.

Two among the many reasons for increased focus on environmental quality are perhaps most basic. First, population increases and urban densification have led to locally concentrated pollutant discharges and deteriorated environmental quality. Second, enhanced laboratory technologies have enhanced the ability to measure chemical concentrations at levels not previously quantifiable. With the intensive monitoring efforts now being carried out, instances of deteriorated environmental quality that might otherwise have remained undetected may now be identified.

Statistical analyses are not an interpretation of the facts, but when properly used, these analyses make the facts easier to see and allow other evidence to enter into judgments about environmental phenomena (Unwin et al, 1985).

CHARACTERISTICS OF ENVIRONMENTAL QUALITY DATA

In general, the ability of a sample of environmental quality data to characterize the population from which it is drawn is related to (a) the size of the sample, (b) the degree to which it was selected at random, and (c) the degree of independence among the observation that make up the sample.

TABLE 1 .1 Indications of features that contribute to the variability and problems of data analysis in ground water quality

- Sample collection may involve drilling, sampling, and laboratory analysis for many water quality constituents. The expense argues for utilization of brief records.
- Sample collection and laboratory analyses have inherent difficulties resulting in uncertainties in subsequent interpretation of the findings.
- Many important groundwater phenomena take years to evolve, making the available timeframe for sampling programs only statistical “windows” of temporally varying processes. Example, DNAPL.

A further complicating factor is that many of the data records are highly variable or “noisy” due to, for example, phenomena. An additional consideration arising in part with improved instrument technology is that we can measure features that previously were reported only as censored (less than) data. Further, a number of chemicals have maximum concentration levels (MCLs) to which humans and the environment are exposed while not incurring injury, where these MCLs are very close to the technological instrumentation capability. The result is that problems associated with statistical analyses of censored data sets are increasing.

EXAMPLE: Lake Pontchartrain – salinity, temperature, wind speed, etc.

COPY OF ARTICLE

EXAMPLE 1.1—EXAMPLE OF STATISTICAL HYPOTHESIS TESTING

Consider the question of whether a landfill is leaking leachate which will contaminate the underlying ground water, as depicted on Fig. 1.1.

One way to consider this question is to monitor the ground water quality both upgradient (point A) and downgradient (point B) from the landfill. We might then compare the quality at B and A to determine if there is a difference.. Thus, an hypothesis might be the following:

Hypothesis: There is no statistically significant difference between the quality at the two locations

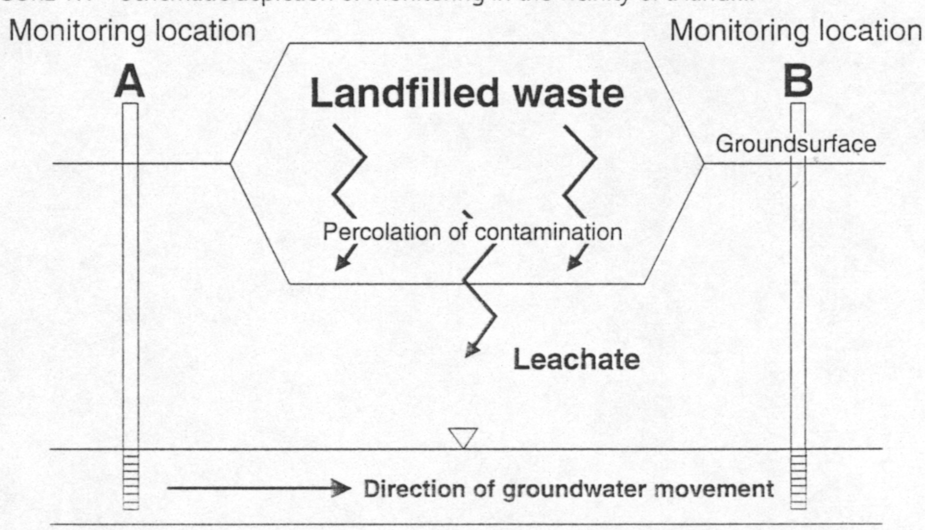
Outcome 1. If we accept the hypothesis, then we are concluding there is insufficient information to indicate the landfill is leaking.

Outcome 2. If we reject the hypothesis, evidence exists to indicate that the landfill is leaking.

The question of hypothesis testing in environmental phenomena is a recurring one. The details of hypothesis testing will be a recurring question throughout this book, and the quantitative aspects of hypothesis testing will be left to later chapters.

NOTE: OUTCOME 1 DOES NOT SAY THAT THE LANDFILL IS NOT LEAKING.

FIGURE 1.1 Schematic depiction of monitoring in the vicinity of a landfill



UNCERTAINTIES AND ERRORS IN ENVIRONMENTAL QUALITY DATA

Errors in sampling procedures, inadequate sample storage and preservation techniques, and laboratory analytical errors are examples of errors in environmental data sets. As a demonstration of the multifaceted initiation points for such errors to exist in a data set, further examples of the sources of error in the collection and analysis of groundwater quality data are listed in Table 1.2.

There is always a degree of uncertainty associated with each discrete measurement of environmental quality. In interpreting data, each discrete measurement is really a range of statistically probable values instead of a single value. There are two subdivisions of reproducibility criteria, namely replication and repeatability. Replication is when two or more results are obtained by the same operator in a given laboratory using the same apparatus of successive determinations on identical test material within a short period of time. Often this is done during quality assurance and quality control testing of a laboratory, to ensure that the lab results are trustworthy. Alternatively, repeatability is a quantitative expression of the random error associated in the long run with a single operator in a given laboratory obtaining successive results with the same apparatus under constant operating conditions on identical test material. Obviously, the requirements for quality assurance and quality control can be substantial.

Many of the statistical analyses described in this book are concerned with sampling errors and the estimation of population characteristics from samples of data. The fact that sampling errors are inherent in random data does not mean, however, that statistical manipulation and sophistication can in any way over come faulty data. The quality of any statistical analysis is no better than the quality of the data utilized. Furthermore, statistical considerations should not be used to replace judgment and careful thought in analyzing data. Statistics must be regarded as a tool or an aid to understanding, but never as a replacement for careful thought.

TABLE 1.2 Examples of sources of error in the sampling and analysis of ground water quality data

- Sampling of a nonhomogeneous region in which wells and springs intersect more than one chemical type in water can lead to misinterpretation.
- Piezometers and wells that are inadequately flushed out prior to sampling of groundwater may render a sample of the groundwater unrepresentative of conditions in the adjacent soil environment.
- Sampling stations that are subject to temporal variations of chemical concentrations can exhibit significant sampling error.
- Cross-contamination of a sample may occur at a time of sampling as a result of an unclean container into which the sample is placed.
- An error in the laboratory protocol of the experiment can occur during the laboratory analysis.
- Improper preservation techniques can alter the sample. For example, groundwater samples are often particularly susceptible to changes in the pressure of oxygen and carbon dioxide and improper sample storage. Improper preservation techniques can result in a chemical alteration of the sample as it adjusts to new equilibrium conditions. In the case of pH levels, groundwater samples have been shown to increase as much as LU pH units due to CO2 escape to the atmosphere during storage.

Statistics is concerned with scientific methods for collecting, organizing, analyzing, summarizing and presenting data, as well as with drawing valid conclusions and making reasonable decisions. Statistical analyses do not consist of a standard set of rules.

Even in the best circumstance, however, statistical data analysis can provide only evidence, never proof.