Technical Guidance for the Long-Term Monitoring of Natural Attenuation Remedies at Department of Energy Sites

Office of Environmental Restoration

October 8, 1999

INTRODUCTION

The purpose of this guide is to provide Department of Energy (DOE) Remedial Project Managers (RPMs) with technical direction on:

- The role of monitoring for effective implementation of a natural attenuation remedy;
- The key considerations for designing a natural attenuation monitoring network; and
- Statistical approaches for interpreting monitoring data and refining conceptual site models.

This guide should be used in conjunction with other Departmental guidance, specifically the *Decision-Making Framework Guide for the Evaluation and Selection of Monitored Natural Attenuation (MNA) Remedies at Department of Energy Sites* and the *MNAtoolbox*. The *Decision-Making Framework Guide* provides RPMs with a framework for evaluating the efficacy of MNA as a remedial alternative within the bounds established by applicable regulations and EPA's MNA policy directive.[1] The *MNAtoolbox* is a DOE software tool *(http://www.sandia.gov/eesector/gs/gc/na/mnahome.html)* that assists site managers in determining whether their sites are good candidates for the implementation of MNA and, therefore, worth the cost of more in-depth characterization and analyses to support selection of MNA as the remedy of choice.

[1] EPA, Office of Solid Waste and Emergency Response (OSWER), Directive 9200.4-17P, "Use of Monitored Natural Attenuation at Superfund, RCRA Corrective Action, and Underground Storage Tank Sites," April 21, 1999.

[**NOTE:** Although these guides are intended to help focus the evaluation and implementation of MNA remedies, they (or any references to any other guidance therein) are not intended to be prescriptive. Rather, project managers should follow the guidance within the context of site-specific circumstances and best professional judgement.

This guide is organized into the following sections:

- Role of Monitoring in MNA: Describes the use of contingency planning to help manage uncertainty and the key role monitoring plays in warning of deviations from assumed conditions.
- Design Considerations: Discusses the three types of monitoring; the where, when, what, and how to monitor for each, and describes the triggering of contingency plans.
- Data Interpretation: Discusses statistical methods for comparing data to baseline, discusses methods for comparing trends to conceptual site model predictions, and describes criteria for determining when a contingency trigger has been exceeded.

ROLE OF MONITORING IN MNA

As defined by EPA, monitored natural attenuation is *the reliance on natural attenuation processes, within the context of a carefully controlled and monitored site cleanup, to achieve site-specific remedial objectives within a time frame that is reasonable as compared to those offered by more active measures*. In this context, attenuation processes may include biodegradation; sorption; dilution; dispersion; volatilization; and chemical or biological stabilization, transformation, or destruction. By definition, *monitoring* is the critical component of any natural attenuation remedy ensuring performance objectives are being achieved and, when they are not, identifying when contingency measures are necessary to prevent any unacceptable risks to human health and the environment.

As delineated in the EPA's policy directive, an MNA monitoring system should be adequately designed to:

- Demonstrate that natural attenuation is occurring at an acceptable rate;
- Determine if the contaminant plume is expanding either laterally, vertically, or downgradient;
- Ensure no impact to downgradient receptors;
- Detect new contaminant releases into the environment that could alter the effectiveness of natural attenuation processes;
- Detect changes in environmental conditions that may reduce the efficacy of natural attenuation processes;
- Identify potentially toxic transformation products resulting from the degradation of organics or the decay of radionuclides; and
- Verify the attainment of cleanup objectives.

In addition, where institutional controls are relied upon to prevent exposures prior to remedial objectives being achieved, site managers need to demonstrate the efficacy of those controls.

It is often assumed that to accomplish these objectives under an MNA remedy, more extensive monitoring -- as compared to other more active remedial approaches -- is required; however, this is not necessarily the case. For example, optimization of pump and treat systems frequently involves extensive monitoring. If initial hypotheses on attenuation rates in the conceptual site model are supported by monitoring data, and the location or frequency of sampling is reduced in accordance with such confirmation, the amount of required monitoring may actually be less.

On the other hand, monitoring for an MNA remedy has the potential to be prohibitively expensive unless the frequency of monitoring is progressively scaled back over time and the suite of variables to be monitored is minimized. This is especially true in those situations where the time required to reach remedial objectives is extensive. This tiered approach to monitoring is the key to cost effectively addressing MNA performance uncertainties, while ensuring no unacceptable contaminant migration occurs.

As set forth in the *Decision-Making Framework Guide*, sites pursuing an MNA approach should work closely with their regulators and the public when designing monitoring networks to ensure the system is considered adequate to all stakeholders.

Monitoring to Manage Uncertainties

A basic principle of environmental restoration is that uncertainty is inherent in any cleanup activity and must be managed through a balance of uncertainty reduction (data collection) and uncertainty mitigation (*i.e.*, using monitoring data, probabilistic modeling, and contingency planning to counteract the impacts that may arise from unexpected conditions).[2] If all uncertainties could be eliminated prior to the implementation of a remedy, one could accurately predict the outcome of applying that remedy and there would be no need for post-implementation monitoring. In reality, cleanup decisions are made with incomplete data, and uncertainties always exist in the conceptual site models used to predict performance.

Examples of uncertainties include:

- The exact nature of the viable transport and attenuation processes at the site and their variability over space and time (*e.g.*, permeability and degradation rates vary from point to point across the site, recharge rates and depth to the water table vary from year to year);

- The specific rates by which these transport- and attenuation-specific processes are operating (*e.g.*, sorption effects at any point in space are uncertain, effective flow velocity is uncertain); and

- The operable geometry of the site (*e.g.*, For the scale of this problem, are the geologic properties effectively homogeneous or heterogeneous? Does the system predominantly operate as two-dimensional or three-dimensional?).

Once consensus is reached by the core team (DOE, EPA, and State Remedial Project Managers) and other stakeholders that MNA represents an appropriate remedial strategy, a site-specific monitoring program will need to be developed. The monitoring program will need to be based on the conceptual site model, which presumably provides an adequate understanding of contaminant geochemistry and transport to explain past contaminant movement and predict future trends. As site-specific monitoring data are collected, trends are quantified and serve as further input to support the continual updating and calibration of the conceptual site model. Any significant deviation from predicted trends must be detected and may result in abandonment of MNA and the implementation of agreed upon contingency measures to ensure protection of human health and the environment.

Ultimately, the magnitude of required monitoring activities is directly dependent on the nature and magnitude of uncertainties. In that regard, greater site characterization at the outset may help to reduce the extent of required monitoring during implementation, *i.e.*, sampling frequencies and number of monitoring locations. Therefore, the costs of site characterization and the probability of significantly reducing uncertainty in the conceptual site model should be weighed against the potential reduction in the monitoring costs and the potential cost of implementing a contingency should one be needed. Initially, the significance of specific uncertainties on monitoring requirements can be evaluated through sensitivity analyses, using numerical models to predict natural attenuation's effectiveness. As uncertainties in key decision-making parameters are reduced through site characterization (or through monitoring once MNA is being implemented), the sampling frequency and number of monitoring locations should be reduced accordingly as confidence in the predictive capabilities of the conceptual site model increases.

DESIGN CONSIDERATIONS

[2] U.S. Department of Energy, *Uncertainty Management: Expediting Cleanup Through Contingency Planning*, DOE/EH/(CERCLA)-002, February 1997.

Monitoring can be considered as a continuation of site characterization activities in an iterative decision-making process. In this context, monitoring offers an opportunity to bound uncertainty in key decision-making parameters. As a continual feedback mechanism, monitoring provides a means for assessing whether conditions differ from those assumed in the conceptual site model enough to cause a change in the selection of a remedy. Accordingly, selecting MNA as the remedy of choice for any site requires the design, construction, and operation of an efficient and effective monitoring network.

Reliance on conceptual site modeling and the explicit treatment of uncertainty and variability (*e.g.*, probabilistic modeling) form the basis for the development of monitoring strategies as described in this document. This approach is critical to the monitoring system design, the interpretation of data, and the evaluation of options should performance results differ from original expectations. This approach to monitoring system design also relies heavily on site-specific process modeling and quantification of uncertainty to define where to monitor, when to monitor, what to monitor, and how to analyze monitoring results. In addition, focusing on the conceptual site model and process modeling provides a foundation for interpreting monitoring data and making sound contingency decisions.

MNA Management Zone

All monitoring should occur within or at the boundary of the MNA management zone (Figure 1). The MNA management zone is defined by the core team and encompasses the maximum projected plume boundary based on understanding of flow, transport, and attenuation processes and quantitatively accounts for all remaining uncertainties. This region can be thought of as providing a natural reaction bed of sufficient size to accommodate the anticipated performance of the natural attenuation processes.

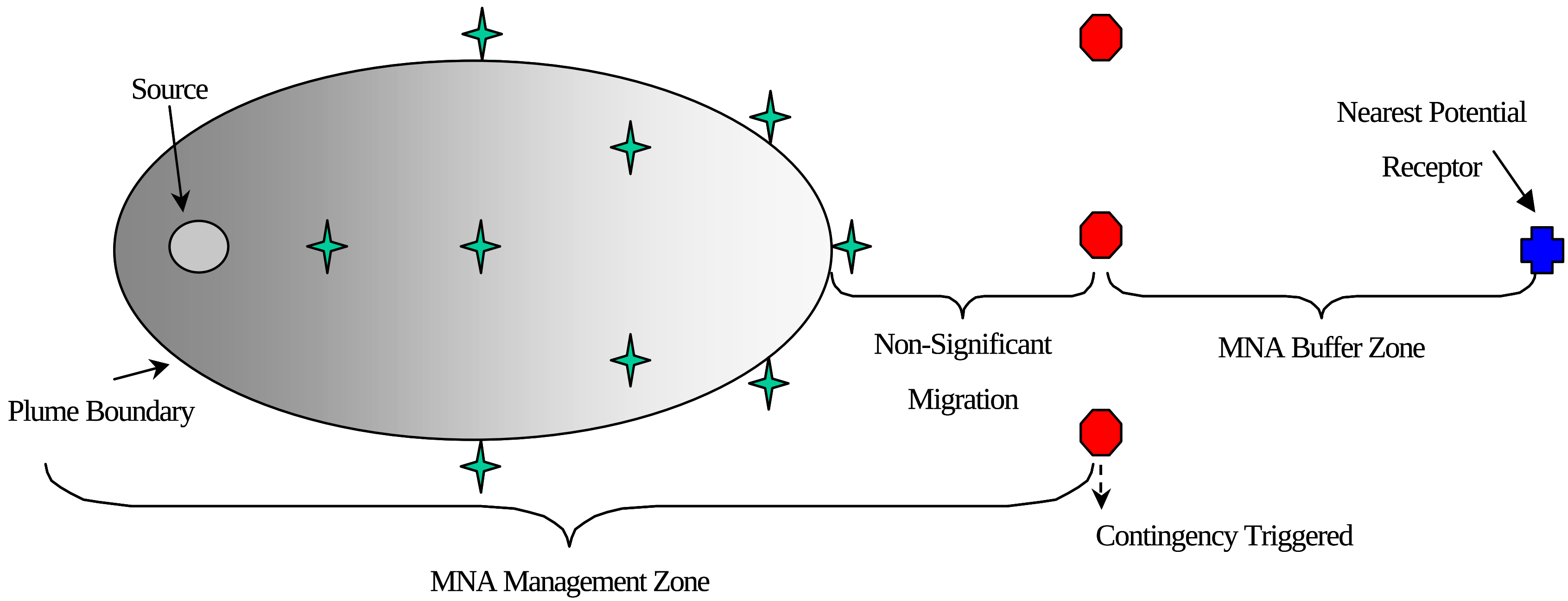
The MNA management zone is not necessarily the boundary of the plume as it currently exists. A larger zone may be established by the core team for plumes that have not become static but are expected to do so in a reasonable time frame without posing unacceptable risks to human health or the environment. Such a situation may occur in an area at a DOE site where the plume is still distant from the facility boundary, but within an area where access controls are in place and easily maintained.

MNA Buffer Zone

The MNA buffer zone is the area extending from the MNA management zone boundary to the nearest potential receptor(s). The MNA buffer zone must be of sufficient size so that if contaminants are detected at the sentinel wells, appropriate measures can be taken (*i.e.*, implement the contingency plan) before any receptors are adversely affected.

The conceptual site model should be used to calculate the potential maximum distance of plume migration between the time the plume crosses the MNA management zone boundary and the time its advance can be arrested by implementation of a contingency. As a general rule, the extent of the MNA buffer zone should

Figure 1
Conceptual Monitoring Network



= "ambient" wells - designed to provide hydrogeologic information from wells upgradient of the original source and contamination plume as a baseline of pre-contamination conditions

= "performance" wells - designed to trace contaminant concentrations within the plume and to measure other indirect parameters to determine if attenuation mechanisms are functioning as predicted in the site-conceptual model

= "sentinel" (detection) wells - designed to alert site managers that contaminants have migrated to sentinel wells indicating that natural attenuation processes are not performing as expected and that contingency measures should be implemented not be shorter than this distance. In addition, any area that may be impacted in the event MNA fails to perform as expected, must be institutionally controlled to prevent potential exposures.

The MNA monitoring program designed for a site should specify the location, frequency, parameters, and methods necessary to evaluate whether performance objectives of the remedy are being met. Conceptual details of the monitoring program should be sufficiently developed by the core team prior to final remedy selection so all stakeholders are generally aware of the anticipated costs and associated requirements.

Three types of monitoring (Figure 1) are typically required for the implementation of an MNA remedy:

- Performance monitoring (within and immediately adjacent to the plume);
- Detection monitoring (at the boundary of the MNA management zone); and
- Ambient monitoring (upgradient of the plume).

The objectives and the where, when, what, and how for each type of monitoring are described below; potential contingencies are also discussed.

Performance Monitoring

The purpose of performance monitoring is to track contaminant concentrations and key parameters or indicators of attenuation performance at a given site (*e.g.*, degradation products, redox potential, etc.). Performance monitoring data are used to quantify the rate at which natural processes are attenuating contaminant concentration. These data are used to calibrate or revise the conceptual site model as a means of continually reducing uncertainties and improving the capability of numeric models to predict the time required for natural attenuation processes to achieve media-specific concentration goals.

The parameters to be monitored, and the sampling frequencies and locations, are based on the conceptual site model and the relative uncertainty reduction expected to be achieved. Hence, conceptual site modeling and site-specific monitoring are integrated to provide a continuous feedback mechanism during operation of the remedy to demonstrate that attenuation processes are sufficient to meet performance objectives and are functioning within the envelope of acceptable time frames to achieve those objectives.

Data collection designed to provide several corroborating lines of evidence is advantageous because it gives increased confidence that attenuation processes are occurring and that conceptual site models are likely to yield defensible results for decision making. In the event that data analyses indicate attenuation processes are likely to be insufficient to reach remediation goals within the desired time frame, the performance monitoring network can also be thought of as an early warning system that signals the potential need to implement a contingency action.

An initial set of conceptual site model assumptions, reflecting the current state of knowledge and limited site-specific information, can be used to evaluate the potential efficacy of natural attenuation as a remediation alternative at the site.[3] However, it is expected that conceptual site models will be refined as additional site-specific information is collected and used for detailed evaluation of MNA pursuant to existing MNA policy and guidance. In turn, the conceptual site model plays a key role in determining where, when, what, and how to monitor.

<u>Where</u>

Performance monitoring takes place within and immediately surrounding the existing contaminant plume (Figure 1) utilizing, to the greatest extent possible, existing monitoring locations (*e.g.*, wells, boreholes). Monitoring locations within the plume are needed to track the evolution of plume behavior as well as to assess the efficacy of attenuation processes. Since some site characterization will have been conducted prior to selection of MNA as the remedy of choice, monitoring system design often begins with an assessment of the adequacy of the existing monitoring network. Some monitoring locations will provide more valuable data than others. For example, monitoring locations that are located where evidence suggests natural attenuation processes are presently occurring may provide more useful information for developing and calibrating the conceptual site model.

Several considerations are relevant for determining the location and number of performance monitoring stations. First and foremost, locations should provide a complete representation of the impacted area. Locations should span the vertical and lateral planes through which the plume is expected to occur. Additionally, multiple locations along the longitudinal axis of the plume may help demonstrate decreasing trend of contaminant concentrations towards the MNA management zone boundary.

The density of sample locations should be determined on the basis of the scale of heterogeneities in the conceptual site model, *i.e.*, spacing should be on the same order of the dimensions of significant features of the conceptual site model. Additionally, spacing should provide for a sufficient number of samples to meet the acceptable level of uncertainty for data on critical characteristics.

If there is any uncertainty with regard to the effectiveness of source control measures or the presence of active source residues, some sample locations should be selected immediately downgradient from suspect areas in order to detect new releases should they occur. To the extent that the boundary of the management zone is set at a distance from the current downgradient boundary of the plume, some monitoring locations should be established between the two boundaries to detect any movement of contaminants. Data from those monitoring locations would be incorporated into the conceptual site model to determine whether such movement is explicable within the current set of site assumptions.

If specific anomalies are present within the management zone that may have unusual or unknown effects on attenuation or transport (*e.g.*, faults, unique geologic features, potential sources of other chemical releases), additional sample points should be added proximate to those anomalies. It is also important to install monitoring stations at the interface between

[3] *MNAtoolbox* is used to answer the question, "Is natural attenuation technically feasible at a given site for a particular contaminant?" The technical support document to the *MNAtoolbox* identifies a number of conceptual site process models that may be appropriate for the contaminant/site pair.

media on a pathway (*e.g.*, the point where groundwater discharges to surface water and losing reaches of water ways).

Ultimately, core team decisions concerning the density of the monitoring network will involve balancing the costs of collecting data against their expected value in supporting decision-making. In addition to the more quantitative and objective computational techniques for assessing the sensitivity or importance of data, as a practical matter, expert technical judgement and the interests of stakeholders will always play a role.

It will often be the case that models will be applied to generate the expectations against which monitoring results will be compared. In those cases, use of probabilistic modeling will also be instructive in selecting optimum locations for performance monitoring wells. Clearly, evaluation of monitoring data is simplified when predictions are available for the specific locations at which data are collected and when those predictions include an estimate of the statistical distribution of readings that may be anticipated. Even if modeling has not been conducted to establish performance expectations, the output from runs made to help locate monitoring sites can provide a better basis for data interpretation.

<u>When</u>

Selecting the time interval between performance monitoring sampling events should be based on the anticipated rate of plume evolution as predicted by the conceptual site model and the degree of confidence with these predictions. The interval between monitoring events should be consistent with the rate at which varying conditions affect system behavior. That is, data should be collected often enough to assess system variability but not so often as to be redundant. Initially, the interval between monitoring events will be relatively short, as this is the period in which data for conceptual site model calibration are likely to have the greatest impact. Over time, the conceptual site model is successively updated using monitoring data. At the point where there is good agreement between its predictions and incoming data (demonstrating a progressively better understanding of the system), the interval between sampling events may be increased.

<u>What</u>

When conducting performance monitoring, data should be collected that contribute to developing one or more lines of evidence that demonstrate attenuation processes are occurring. Candidate lines of evidence include:

- Documented reduction of transportable mass over time, *i.e.*, periodic measurements of contaminant concentration and distribution can be used to document mass reduction over time by allowing observation of trends in plume size and contaminant mass;

- Presence of geochemical or biological conditions necessary or favorable for attenuation processes to occur (*e.g.*, favorable redox conditions);

- Presence of indicators of geochemical or biological attenuation (*e.g.*, degradation products);

- Direct field evidence of attenuation processes (*e.g.*, chemical analysis of soil samples demonstrates precipitation or adsorption of contaminants onto aquifer materials).

In addition to providing evidence of attenuation, data are also needed to determine the direction and rate of any contaminant migration. Data may also be needed to indicate changes in environmental conditions over time, especially changes that may indicate diminished system performance or an inability to effectively monitor the system.

Common contaminants found at DOE sites that may be considered for an MNA approach can be separated into six groups: fuel hydrocarbons, chlorinated organics, high explosives, metals, inorganic anions, and tritium (radionuclides are divided into the latter three groups). These groups differ in the factors that must be considered when designing a monitoring system for natural attenuation. Contaminant-specific considerations for monitoring each group are discussed below.[4]

Fuel hydrocarbons (FHC's) may be associated with light non-aqueous phase liquids (LNAPL's) that typically degrade rapidly in the subsurface. Although many DOE sites typically have plumes that contain FHCs, these contaminants do not represent a sizeable fraction of the contaminant chemistry found in most plumes throughout the DOE complex. Several states have passed, or are in the process of passing, laws which guide implementation of MNA for FHC's. Typically, these statutes call for removal of free product (to the extent practicable) followed by monitoring. The details of the monitoring plans that are acceptable vary from state to state. The reader is referred to existing guidance on indicator parameters for monitoring FHC attenuation.[5]

Chlorinated organics are often associated with dense non-aqueous phase liquids (DNAPLs). DNAPLs undergo attenuation to various degrees, the primary attenuation mechanisms being biodegradation and dispersion. Perchloroethylene (PCE), trichloroethylene (TCE), vinyl chloride (VC), and chlorobenzene are examples of chlorinated organics commonly found at DOE sites. Only the mobile, toxic, soluble fraction that dissolves in groundwater is a candidate for MNA.

Because the biodegradation of chlorinated organics is typically slow (relative to fuel hydrocarbons), determination of the biodegradation rate constant is critical to the development of the conceptual site model for attenuation. Furthermore, the biodegradation rate may change over time, and therefore, reappraisal is periodically necessary to ensure any change is detected early. Vinyl chloride, a hazardous breakdown product of TCE and PCE, is extremely toxic and must be monitored in the groundwater when TCE and PCE plumes are present. Complete breakdown of chlorinated organics occurs most rapidly when reducing conditions exist at the source and when oxidizing conditions prevail downgradient. Long-term performance monitoring should focus on gathering data on these conditions. Again, the reader is referred

[4] Relevant indicator parameters associated with attenuation mechanisms are contaminant specific and can be identified using *MNAtoolbox*.

[5] Wiedemeier T. H., J. T. Wilson, D. H. Kampbell, R. N. Miller, and J. E. Hansen, *Technical Protocol for Implementing Intrinsic Remediation with Long-Term Monitoring for Natural Attenuation of Fuel Contaminant Dissolved in Groundwater*, Vols I & II, Air Force Center for Environmental Excellence, 1995.

to existing guidance on indicator parameters for monitoring chlorinated solvent attenuation.[6]

With respect to **high explosives**, the mobile, toxic, soluble compounds, for example RDX (hexahydro-1,3,5-trinitro; 1,3,5-triazine) and TNT (2,4,6-trinitrotoluene), which dissolve in groundwater, are viable candidates for MNA. MNA should also be considered as a remediation alternative for related products such as dinitrotoluene and amine derivatives.

Metals can either behave as **cations** (*e.g.*, lead, cadmium, zinc, copper, cesium, strontium, nickel, uranium, plutonium, americium, and beryllium) or **anions** (*e.g.*, chromium and technetium) depending upon the subsurface geochemical conditions. In general, cationic metals sorb strongly to soil surfaces, except at low pH (< 5). One exception is uranium, which sorbs poorly at high pH (>8) as well.[7] Cesium, and to an extent, strontium, form ion exchange complexes with clay minerals, and their binding is less affected by pH. Primary sorbing phases in soils are iron hydroxides, clays, organic matter, and carbonate minerals. Given the right conditions, all of the cationic metals (with the exception of cesium) can form poorly soluble minerals, and for many, there exists a sizeable natural background.

The natural attenuation processes for cationic metals are sorption (reversible and irreversible), dispersion, and in the case of the radioactive isotopes, decay. Because of the strong component of sorption, surface processes must be considered when designing an MNA monitoring system. Sequential soil extractions are used in the conceptual site model development stage to determine the bioavailability of sorbed metal cations and to identify the host phase of the soil matrix. This allows the non-bioavailable fraction, which may be bound up in the solid phase, to be quantified. Long-term monitoring for cationic metals will involve subsequent sequential leach tests to assess further uptake.

For MNA of most cationic metals, it is important that pH, and to a lesser extent soil redox conditions, periodically be assessed during performance monitoring; pH because it controls sorption and often the stability of the host phase and soil redox potential because it often controls the stability of iron hydroxide hosts.

Natural attenuation of chromium and technetium anions typically involves reduction by soil organic matter from hexavalent chromate to trivalent chromium and heptavalent technetium to quadrivalent technetium, respectively, both of which are poorly soluble. Therefore, long-term monitoring for MNA of chromium and technetium must be designed to monitor the redox potential and the valence state of the chromium and technetium, as well as the electron donor availability of the aquifer.

Other **inorganic anions** of interest include iodide, cyanide, fluoride, and nitrate. Iodine-129, a radio isotope, has a very long half life (approximately 1.7×10^7 years) and sorbs poorly to most soil minerals. Dilution is the primary process for the attenuation of iodide. Cyanide, fluoride, and nitrate also typically exist as anions in groundwater and are, therefore, not

[6] Wiedemeier T. H., M. A. Swanson, D. E. Moutoux, J. T. Wilson, D. H. Kampbell, J. E. Hansen, P. Haas, and F.H. Chapelle, *Technical Protocol for Evaluating Natural Attenuation of Chlorinated Solvents in Groundwater*, prepared for the Air Force Center for Environmental Excellence, Technology Transfer Division, Brooks Air Force Base, San Antonio, Texas, 1997.

[7] Americium and plutonium may also fall into this category.

expected to sorb strongly to subsurface solids. For this reason, dilution may often be the primary attenuation process for decreases in their respective concentrations over time. Nitrate can be transformed by redox processes (*e.g.*, denitrification) that are operative in the subsurface. By the same token, high calcium levels might decrease fluoride transport due to the formation of calcium-fluoride salts.

Tritium is another common contaminant present at DOE sites. Currently, there is no technology to effectively remove tritium contamination from the subsurface. Because tritium has such a short half-life (~12.3 years) and its daughter product is non-radioactive helium-3 (a stable isotope of helium that will not decay), it is an excellent candidate for remediation by MNA. Tritium, however, does not generally sorb to subsurface soils and is readily transported by groundwater. Distance to the nearest potential receptor(s), and the time required for groundwater to travel this distance, will therefore, be the critical factors in determining when evaluating MNA remedies for tritium contamination. Vapor phase tritium, and volatile organic chemicals, can be attenuated through physical processes under the right conditions. For example, in arid climates with fractured or porous geology, exchange between the subsurface and the atmosphere can result in dilution of the subsurface contamination.

How

It is important that accepted, applicable analytical methods be used for all MNA monitoring activities. For many parameters, approved methods can be found in EPA Publication SW 846.[8, 9] However, it is not a requirement that only those methods be used. Other comparable methods, such as those provided in ASTM standards, Standard Methods, or specific state guidance, may also be used.[10] Current EPA and ASTM documents on sampling and analytical methodologies, however, may not be adequate when assessing a site that contains either radionuclide or a combination of radionuclide and chemical contamination. RPMs should refer to existing DOE guidance for sampling procedures and analytical methodologies when assessing a site with these contaminants.[11] The key to monitoring for performance is to select methods that are valid and that apply to the matrix being analyzed. When there is no compelling reason to deviate from SW 846, use of these methods is advised.

Contingencies

A primary tenet of MNA is that contingencies are identified and ready for implementation should monitoring data indicate that conditions differ significantly from those assumed when the remedy was being selected and designed. As a consequence, each monitoring activity should be accompanied by contingencies and

---

[8] U.S. Environmental Protection Agency. Compendium of Superfund Field Operations Methods. OSWER Directive 9355.0-4. 1987.

[9] U.S. Environmental Protection Agency. Data Quality Objectives for Remedial Response Activities. OSWER Directive 9335.0-7B. 1987.

[10] American Public Health Association. Standard Methods for the Examination of Water and Wastewater, 18th ed. 1992.

[11] Goheen, S., M. McCulloch, R. Riley, D. Sklarew, A. Sharma, and S. Fadeff. DOE Methods for Evaluating Environmental and Waste Management Samples. Columbus, OH: Battelle Press, 1997.

decision criteria indicating when those contingencies should be triggered.

Three possible outcomes from performance monitoring can be anticipated:

- Data support the conceptual site model and indicate attenuation rates are adequately predicted,

- Data do not support the conceptual site model and indicate attenuation rates are not adequately predicted, and

- Data are ambiguous, insufficient, or defy meaningful interpretation.

In the first case, results will bolster confidence that MNA will meet remedial objectives. The only change in operating procedures that might be contemplated would be a reduction in the frequency and/or number of sampling activities. Reductions may be taken across all parameters or may be limited to specific parameters for which previous results indicate particularly strong agreement with the conceptual site model.

In the second case, there is reason to question whether MNA can meet remedial objectives. The data must be interpreted carefully to determine if the difference between data and predictions results from mis-specification of attenuation mechanisms in the conceptual site model, use of the wrong rate data, or assumption of the wrong baseline geochemical characteristics for the site. In order to help identify the cause of differences between predictions and results, it may be necessary to modify the monitoring program by adding parameters, increasing the frequency of measurements, adding additional sample locations, and/or conducting additional laboratory studies. Selection of a course of action should be based on identification of the data that would be needed to improve or update the conceptual site model for future predictions.

A presumption for MNA remedies is that no active source remains. Therefore, unexpected fluxes in contaminants generally would not be expected. If higher than expected fluxes of contaminants into the plume are at the root of the inaccuracies, one course of action would be to implement additional source control measures, if possible. It is not likely that abandonment of MNA as the remedy of choice would be a predetermined contingency for performance monitoring. That course of action is generally reserved as the contingency for detection monitoring. However, if performance monitoring data indicate key attenuation mechanisms are not operable, but could be brought to bear with the addition of important amendments (*e.g.*, addition of missing nutrients), such additions could be included as a contingency. Since MNA assumes no active additions or direct intervention, implementing a contingency would constitute conversion from MNA to an alternate remedy (*e.g.*, "enhanced bioremediation").

In the case of ambiguous results, the contingency may consist of the modification of sample locations, increases in sampling frequency, or addition of parameters to provide a clearer picture of contaminant response to attenuation mechanisms.

Detection Monitoring

The purpose of detection monitoring is to ensure protection of human health and the environment while the MNA remedy is being implemented. Detection of significant contamination at the boundary of the MNA management zone will trigger implementation of an appropriate contingency action. The detection monitoring network is designed to provide an additional level of assurance that in the event attenuation processes turn out to be insufficient to maintain concentration goals within the MNA management zone, there will be adequate time to take the necessary actions to ensure no unacceptable exposures occur.

Where

Detection monitoring locations should be placed between the edge of the contaminant plume and nearest potential receptors, on the boundary line of the MNA management zone (Figure 1). This line should extend far enough laterally and vertically to capture all potential/plausible plume migration paths. It also delineates the detection boundary between the end of the management zone and the beginning of the MNA buffer zone which is the area extending from the detection boundary to the nearest potential receptor(s). Detection monitoring locations should be spaced at separation distances that could be reasonably expected to intercept migration of the contaminant plume. Expert judgement or optimization techniques can be used to place monitoring sites along the detection boundary. Once again, probabilistic modeling techniques may be helpful in identifying optimum locations for sentinel wells, although, preference should be given to locations directly upgradient of the nearest receptor.

When

Initially, the timing of detection monitoring will be a direct function of the potential transport velocities for site contaminants and the distance between the extant plume and the edges of the MNA management zone. This approach eliminates the potential for migration to proceed past the detection boundary before being detected. Over time, intervals between sampling events should be based on the current version of the conceptual site model that reflects the most recently collected performance monitoring data. If performance monitoring data continue to be consistent with the conceptual site modeling results, and natural attenuation appears to be effective, the likelihood of detection decreases. In this case, it may be appropriate to decrease the sampling frequencies for detection monitoring accordingly.

What

Detection monitoring should include sampling for the contaminant(s) of concern, as well as any toxic degradation byproducts that may be produced or any radioactive or hazardous daughter products of radioactive decay. The identity of likely degradation and decay products can be obtained from the *MNAtoolbox*. Another important consideration is the potential for contaminants to be transported on natural colloids or as radiocolloids (*e.g.*, cesium-137, thorium). Colloids have been found to move small quantities of contaminants much farther than predicted by reaction/transport models.[12] A number of sites have observed rapid transport of detectable levels of radionuclides over long distances by colloids, but there are as yet no known cases where colloid-bound contaminants have posed a significant health threat. Nevertheless, it is important that detection monitoring include contaminants known to be transported by colloids.

How

[12] McCarthy, J.F. and J.M. Zachara. "Subsurface Transport of Contaminants." Environmental Science Technology, 23:496-502, 1989.

Detection monitoring analysis should be performed using the same methods as prescribed for the performance monitoring of like parameters to assure comparability of data.

Contingencies

A key element in the acceptability of MNA is the stated assurance that if detection monitoring data clearly indicate a trigger level is exceeded, an agreed upon contingency measure will be implemented. Therefore, it is critical that the trigger concentration be defensible and embedded in a decision rule that removes any potential for differing interpretation of when a contingency should proceed. Trigger levels for contaminant concentrations should not be so restrictive that normal variability in analytical data cannot be accommodated. The DOE recommends trigger levels be set such that they detect the edge of a plume at the detection boundary, and that the edge of a plume be defined as a concentration for the contaminant at which an exposure would present an unacceptable risk.

Contingencies will be site and contaminant specific. In general for saturated zones, pump and treat systems or similar means of providing control within the MNA management zone (*e.g.*, recirculating wells, in situ sparging, permeable treatment walls) are the most likely contingency actions. Pump and treat systems can be quickly installed and, if properly located, can target the contamination responsible for triggering the action. Presumably, failure of an MNA remedy will mean that attenuation rates are inadequate to completely control contaminant migration but only to the extent that the need for active measures is limited. Thus, further mass reduction or possible enhancements could bring attenuation fluxes back in line and allow for the possible resumption of the MNA remedy in the near future.

Ambient Monitoring

Ambient monitoring data typically are collected to provide a baseline against which to compare results of performance and detection monitoring and to provide boundary conditions for numerical modeling. Ambient monitoring is used to monitor baseline conditions outside of the source with respect to those parameters that affect attenuation mechanisms. As such, ambient monitoring data provide a baseline for identification of contamination through comparisons with data collected during performance and detection monitoring and should be included in the conceptual site model.

If upgradient sources exist that are being transported towards the management zone, ambient monitoring can be expected to reveal the presence of these sources of contamination and provides notice that trigger levels for detection monitoring may need to be revisited. For instance, should levels exceed the target cleanup criteria (*e.g.*, maximum contaminant levels) or background concentrations for the contaminants of concern, attaining these levels within expected the time frame may no longer be possible. Therefore, it is important to monitor background and identify when or if contaminant concentrations upgradient of the site increase to values above initial target criteria.

Additionally, ambient monitoring is used to detect trends in geochemical conditions that may impact the rate at which attenuation mechanisms will operate. To the extent that trends suggest changes in the geochemical baseline, the conceptual site model is altered to reflect those changes and provide a more realistic metric for evaluation of performance. In the extreme, measures may be required to mitigate adverse trends that would jeopardize attainment of the remedial action objectives.

Where

Ambient monitoring stations must be located outside of the contaminant plume (Figure 1). Where baseline levels are highly variable or geologic media are particularly heterogeneous, more locations will facilitate statistically meaningful assessments of baseline conditions. Additionally, if alternate contaminant sources are suspected, ambient monitoring stations should be located downgradient from those sources/areas to prevent misinterpretation of performance monitoring data.

When

Ambient monitoring is performed in the initial phase of the monitoring program to allow characterization of baseline and calculation of the conceptual site model. The ambient monitoring data are used in comparison with performance and detection monitoring data to assess whether significant contamination exists. Periodic ambient monitoring is performed to confirm baseline conditions are stable, and can provide a check for unanticipated sources of contamination. Thus, ambient monitoring need not be performed at the same frequency as performance or detection monitoring. However, ambient monitoring might need to be designed to accommodate temporal or spatial patterns in environmental conditions. For example, if the conceptual site model includes a significant seasonality component, the monitoring network should be designed to adequately characterize this seasonality.

What

Ambient monitoring should include the contaminant(s) of concern and known degradation or decay products, contaminants associated with upgradient source terms, and other parameters designated for performance and detection monitoring.

How

Ambient monitoring analysis should be performed with the same methods as prescribed for the performance monitoring of like parameters to assure comparability of data.

DATA INTERPRETATION

The monitoring network is designed to collect data either to show that natural attention processes are acting as predicted by the conceptual site model (performance monitoring) or to trigger the implementation of contingency plans (detection monitoring) when they are not. The challenge for data interpretation in support of MNA is to carefully link these objectives, the resources available to collect data, the anticipated variability or uncertainty in the data, and the allowable decision errors so that cost-effective monitoring and contingency decisions can be made. This section provides general guidance on an approach for meeting this challenge.

A successful monitoring network will allow the following questions to be addressed in a framework within which uncertainty can be managed effectively:

- Are monitoring data consistent with predictions?
- Do monitoring data clearly indicate that a trigger level has been exceeded?
- Have baseline conditions changed?

To formally and rigorously address these questions, decision rules and acceptable decision errors must be developed by the core team and assessed within a statistical hypothesis testing framework. For example, in detection monitoring, decision rules represent a merging of trigger levels and decision errors. In general, two types of decision errors are of concern: 1) incorrectly concluding a trigger level has been exceeded and 2) incorrectly concluding a trigger level has not been exceeded. The possibility of making one of these types of decision errors is never eliminated since collected data represents a subset of the full environmental system. However, these decision errors can be explicitly accounted for, at pre-specified levels of uncertainty, through the collection of a statistically specified number of samples. The acceptable level of uncertainty is identified based on the consequences of each type of decision error. Thus, decision rules, in conjunction with decision errors, help guide design of the monitoring network and provide a statistically rigorous means of assessing the efficacy of MNA.

Exploratory Data Analysis

Once monitoring data have been collected, the first step in data interpretation should be exploratory data analysis (EDA).[13] EDA refers to a collection of graphical, and largely qualitative, techniques for examining and understanding data. EDA can provide a visual tool for identifying spatial and temporal patterns and interrelationships between parameters. Classical methods of statistical inference often depend heavily upon assumptions that the data being analyzed are normally distributed (either directly or after transformation), not serially correlated (independent), and do not contain outliers. EDA should be used to qualitatively assess the validity of these assumptions, thus providing a basis for quantitative assessment. Examples of EDA tools include box plots, histograms, quantile plots, normal quantile-quantile plots and estimated (probability) density plots.[13] Extreme concentration results that are identified as outliers by visual inspection as part of the EDA process should be evaluated for potential data errors (laboratory analysis or reporting errors, or data recording errors).

Hypothesis testing, the second step in MNA data interpretation, is fundamentally a decision process concerned with testing of proposed hypotheses. These hypotheses are assessed using statistical tests that provide evidence, as defined by decision rules, on which hypothesis is most likely. For example, competing hypotheses might consist of an assertion that trigger levels are exceeded (which, if accepted, would require that contingency measures be implemented), and a counter assertion that trigger levels are not exceeded (which, if accepted, might suggest natural attenuation is performing as expected). An incorrect decision can be made due solely to large variability in the data. This highlights the need to carefully structure the hypotheses to be tested through the decision rules development process. EDA and hypothesis testing performed in conjunction provide a means of verifying intuition based on visualization of the data against results that are based on mathematical assumptions and theory.

Hypothesis Testing

Statistical hypothesis tests can be described as parametric and non-parametric. Parametric tests typically are based on the assumption that the data are normally distributed (either directly

[13] Hoaglin, D.C.,F. Mosteller, J.W. Tukey. Understanding Robust and Exploratory Data Analysis. John Wiley and Sons, Inc., New York, 1983.

or after transformation). Additionally, parametric tests often rely on data being identically distributed (equal variances) and independent (serially uncorrelated). In general, non-parametric tests do not rely on the normality assumption. Robust tests refer to tests, both parametric and non-parametric, that are insensitive to the test assumptions. A variety of parametric and non-parametric tests are available for addressing each of the MNA questions. The following discussion provides general guidance on suitable statistical tests to address each of the general MNA questions.

Performance Monitoring - Determining if Monitoring Data Are Consistent with Predictions

Central to the acceptance of MNA as an effective remedy is the ability to accurately predict trends in contaminant concentration levels and distributions. If conceptual site model predictions closely match subsequent performance monitoring results, the conceptual site model is adequate and need not be modified. Moreover, it may be possible to decrease both performance and detection monitoring frequency and locations with no loss of assurance that the remedy is protective. If significant deviations are seen between predictions and subsequent observations, the conceptual site model must be revised to reflect more accurate attenuation rates. The key is being able to determine when performance monitoring data are sufficiently different than predictions such that conceptual site model revision is necessitated. This model verification process is a decision process that can be performed using hypothesis testing.

A variety of qualitative and quantitative approaches are available for verifying model performance. As a first step an EDA comparison of model predictions and performance monitoring data will provide a qualitative assessment of the model goodness-of-fit to the data. EDA can also provide insight into the results of hypothesis testing.

The *t*-test is a parametric test that can be used to test a difference between observed and predicted data. It is somewhat robust to violations of normality and equality of variances, but is sensitive to serial correlation and outliers. Care should be taken to check that the observations are independent. In addition, chemical analysis results reported as non-detects are not easily accommodated by the *t*-test.

The Wilcoxon rank sum test is a non-parametric test that can be used to test whether data and model predictions originate from the same distribution. While the *t*-test is used to test the difference between two normal distributions based on their means, the Wilcoxon test is based on the ranks of the data from each dataset and does not require the normality assumption. As with the *t*-test, however, the Wilcoxon test is sensitive to violation of the independence assumption. The Wilcoxon test can be used effectively when the data include outliers and can accommodate non-detects through relatively simple modifications.

The Kolmogorov-Smirnov (KS) test involves a comparison between the cumulative distribution functions (CDF) of the data and the model predictions. Since the test is based on deviations between two CDFs, the restrictive normality assumption is not required and the data and model predictions do not need to have the same number of observations.[14] The KS

[14] Keith, Lawrence H. Environmental Sampling and Analysis: A Practical Guide. Chelsea, MI: Lewis Publishers, 1991.

test statistic has the added advantage that the statistic and the data can be graphically displayed facilitating visual interpretation of the analysis.

Another approach for model verification is to compare trend estimates of the data and of the model predictions. A parametric approach, such as regression analysis, can be used to estimate trends in the data and model predictions. Confidence intervals of the estimated trends can be compared with model predictions based on the specified decision rules to assess the assumptions of the conceptual site model. However, since regression analysis is a parametric method and is subject to normality, equality of variance, and independence assumptions, non-parametric alternatives may be more appropriate. Non-parametric alternatives include the Mann-Kendall test, the seasonal Kendall test, and the Sen slope estimator.[14] These tests are based on the relative magnitudes of the data rather than the actual values and do not require distributional assumptions. As in regression analysis, non-parametric confidence intervals for trend estimates from data and model predictions can be compared based on the decision rules. Non-parametric smoothing approaches can also aid in the visual comparison of data and model predictions, especially if strong seasonality is present.

Detection Monitoring - Determining When a Trigger Has Been Exceeded

Any statistically defensible detection of a primary analyte above prescribed thresholds at the edge of the MNA management zone (*e.g.*, sentinel wells) will trigger a contingency plan. Because of the ramifications of that finding, it is essential that such a detection be accurate and verified. A specific, unambiguous verification procedure should be selected in advance, presumably involving rapid resampling and criteria defining the required level of agreement between samples. Both accuracy and precision are important when analyzing data used to determine when a trigger has been exceeded. Accuracy and precision are dependent upon the ability to control sampling and analytical practices.[14]

A fundamental issue with detection monitoring is the chance of triggering a contingency plan unnecessarily. Decision rules based on trigger level exceedences dictate when contingency measures will be implemented. Implementing contingency measures could be an expensive decision. These decisions should be evaluated carefully, and must account for variability in the data and the potential for incorrect decisions to be made. Statistical methods can be used effectively to properly account for variability and to manage the decision error rates.

Ambient monitoring data can be used to calculate an interval that will contain future compliance monitoring data with a given level of confidence.[15] This parametric prediction limit is developed based upon the tolerable decision errors developed by the core team. Alternatively, a non-parametric prediction limit can be specified based on an upper order statistic, such as, the maximum of the ambient monitoring data. Prediction limit decision rules specify the number of prediction limit exceedences that are allowable under sampling at detection monitoring sites.[16] These decision

[15] Charles B. Davis and McNichols, Roger J. "One-sided Intervals for at Least *p* of *m* Observation From a Normal Population on Each of *r* Future Observations," Technometrics 29:3:359-370,1987.

[16] Gibbons, R.D. Statistical Methods for Groundwater Monitoring. John Wiley & Sons, Inc., New York, 1994.

rules are based on the hypergeometric distribution of the number of allowable "hits" in sequential sampling given the tolerable decision errors. This approach allows for specification of an acceptable number of "hits" before triggering contingency measures.

Comparisons between contaminant levels from detection monitoring locations and trigger levels can be made based on the magnitude of the difference in concentration values or through an indicator variable that is classified as a "hit" if a detection monitoring value exceeds a trigger level. The data can be evaluated using EDA to identify values that are greater than trigger levels. Decision rules can be developed that specify the magnitude of exceedence that would trigger implementation of contingency measures. When such exceedences occur, samples should be collected as soon as possible from detection monitoring locations and compared to prediction limits to confirm whether contamination has indeed migrated to the MNA management zone boundary.

Ambient Monitoring - Determining When Baseline Conditions Have Changed

Ambient monitoring data are evaluated as a means of determining when baseline conditions have changed. Change is important in that it may impact how the trigger levels are defined, or it may indicate when attenuation mechanisms will not operate at rates on which initial predictions were based. Specifically, in the former case, if background concentrations for contaminants of concern increase to levels above the initial target criteria, it may not be possible or appropriate to attain those criteria. Rather, the end state definition should be modified to reflect completion when the plume has reached concentrations matching those in background geochemistry. In the latter case, geochemical changes are evaluated to determine quantitative impacts on attenuation mechanisms and rates relative to adjusting predictions. Each of these uses of ambient monitoring data require different data interpretation techniques.

Rarely, if ever, will monitoring result in a single value for a measured parameter. More typically, results will display a distribution across a range of values. As a consequence, interpreting ambient monitoring requires the ability to recognize when the baseline range represents conditions greater than risk-based criteria and then comparison of that baseline range to the range obtained from performance monitoring.

The comparison of ambient conditions with risk-based criteria often begins with a simple <u>screening</u> comparison in which the maximum concentrations for each analyte from the ambient monitoring are compared to the criteria. Once it is determined that ambient concentrations exceed criteria, subsequent comparisons with performance data are based on use of the range or distribution of data obtained from ambient monitoring.

If sufficient ambient and performance monitoring data are available, distributional shift tests can be used to compare the distribution of performance monitoring data to the distribution of baseline concentrations. Shifts in distribution provide a more comprehensive and defensible analysis than simple threshold comparisons (*e.g.*, tolerance limit comparisons). A distributional shift test is used to determine whether performance monitoring data are systematically different than baseline data. In addition to the previously described student *t*-test and Wilcoxon rank sum test, distributions can be tested with the Gehan test, which is a modified ranking of sample results to handle non-detects in the data that are then subjected to the Wilcoxon rank sum test. The Gehan test, hence, handles multiple detection limits in a statistically robust manner and is recommended so long as the number of non-detects is not too large.[17]

For detecting partial distribution shifts between sets of data (shifts in a subset of the data, the largest concentrations), the following statistical tests can be used:

- The Quantile test, a non-parametric test, indicates whether most of the high concentration data are from the performance or detection monitoring locations in comparison to the ambient monitoring locations.[18] The test is capable of detecting a statistical difference when only a small number of performance or detection monitoring concentrations are elevated. This test accommodates data in a robust manner that includes non-detects.

- The slippage test is based on the number of performance or detection monitoring concentrations that exceed the maximum observed concentration in the baseline data set.[19] The result of the slippage test is the probability that the number of exceedences occurred by chance alone. The slippage test is similar to a tolerance limit comparison in that it evaluates the largest measurements; however, the test is more useful than a tolerance limit comparison because it is based on a statistical hypothesis test and not simply a statistic of a distribution.

The ability to use the distributional shift tests is dependent upon the number of samples available for comparison. In general, at least 10 sample concentrations for comparison with ambient data are needed to provide adequate assurance that a shift in distribution is discernible. When planning in advance of sample collection, the number of samples should be based on the decision rules and the consequence of decision errors. For those analytes that are rarely detected in ambient data (*e.g.*, mercury, antimony, and thallium), an increase in the detection rate at the monitoring sites may be evidence of a release. The following test is recommended.

- The Chi-square Goodness of Independence can be used to compare detection rates for performance or detection monitoring data with ambient monitoring data, so long as the level of censoring (detection limits) is the same for both data sets.[20]

When conducting any of the previously described tests, care should be taken that false positives or false negatives resulting from the large number of comparisons being made do

[17] Millard, W.P., and S.J. Deverel. "Non-Parametric Statistical Methods for Comparing Two Sites Based on Data with Multiple Non-Detect Limits," Water Resources Research, 24:12:2087-2098, 1988.

[18] Gilbert, R.O., and J C. Simpson. Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media. Pacific Northwest Laboratory, Richland, Washington, 1992.

[19] Gilbert, R.O., and J.C. Simpson. "Statistical Sampling and Analysis Issues and Needs for Testing Attainment of Background-Based Cleanup Standards at Superfund Sites," in Proceedings of The Workshop on Superfund Hazardous Waste: Statistical Issues in Characterizing a Site: Protocols, Tools, and Research Needs. Environmental Protection Agency, Arlington, Virginia, 1990.

[20] Box, G.E.P., W.G. Hunter and J. S. Hunter. Statistics for Experimenters. John Wiley and Sons, New York, 1978.

not inappropriately impact decisions. For example, a typical chemical analysis for a suite of metals requires comparison of about 23 analytes with baseline concentrations. This number of comparisons can inflate the chance of at least one of these analytes exceeding baseline beyond the significance levels specified in the hypothesis test. This artifact can be mitigated by appropriately adjusting the testing procedure or by constructing decision rules and decision errors with these issues in mind.

Further evidence of differences between ambient and performance or detection monitoring data can be assessed by considering correlations between parameters of interest. If ambient conditions are the same as those in the MNA management zone, then the correlations should also be similar. Correlations between naturally occurring chemicals can be evaluated on a site-specific basis using ambient monitoring data and can be compared with correlations between the same chemicals using data collected from within the MNA management zone. If the correlations are not the same, then MNA management zone conditions can be concluded to be different than ambient conditions. The correlation analysis provides a further method for determining differences that are often difficult to determine through distributional comparisons.

When ambient monitoring results are used to detect shifts in geochemistry that may impact attenuation, it may be appropriate to apply trend tests such as those Mann-Kendall, seasonal Kendall and Sen slope estimator. These tests provide evidence of a trend and the direction of the trend. The latter feature allows for determination of the likelihood that attenuation will increase or decrease with the changes in geochemistry. If a detected trend indicates reduced attenuation, it may be important to quantify the rate of change and extrapolate to assess if MNA can still meet remedial objectives.