ATTACHMENT 6: STATISTICAL PROCEDURES

1.0 INTRODUCTION

As part of the post-closure permit for the waste disposal cell (WDC) at the former Pennzoil refinery, Roosevelt, Utah, the Utah Department of Environmental Quality, Division of Waste Management and Radiation Control (UDEQ/DWMRC) requires a statistical evaluation of groundwater monitoring data collected from wells adjacent to the WDC. This document, *Attachment 6 – Statistical Procedures*, is the portion of the permit that explains the statistical methods employed to evaluate the groundwater data.

The objective of statistically evaluating the data is the timely detection of possible groundwater degradation due to the WDC, while at the same time reducing the probability of falsely concluding that groundwater quality has degraded when it has not (false positive). To satisfy this objective, groundwater quality data are collected from 6 (six) monitoring wells at the WDC; MW-7 and MW-12 are background wells, whereas MW-11, MW-19, MW-20, and MW-21 are compliance wells. Groundwater samples were collected from these wells semi-annually from 1992 to 2008, except MW-7 where sampling began at 1996. From 2009 to 2018, groundwater samples were collected annually from all six wells. This current permit modification proposes to reduce the sampling frequency for general chemistry parameters and dissolved metals to every five years, while maintaining the annual sampling frequency for volatile organic compounds (VOCs).

The Shewart-CUSUM control charting technique (USEPA, 2009; Gibbons et al., 2009; ASTM, 1998) is the main method proposed for future statistical evaluation of data from the WDC. This is an intra-well approach in which the current measured concentration of a groundwater constituent within a well is compared with the historic records of the constituent in the same well. No inter-well comparisons are performed between compliance wells and background wells, or between compliance wells, thereby avoiding the high false positive rate inherent to inter-well comparisons (USEPA, 2009; Gibbons et al., 2009). The use of the intra-well Shewart-CUSUM technique is justified because monitoring at the WDC suggests that there are naturally occurring spatial differences in groundwater chemistry between the background and compliance wells, and the historical records since 1992 indicate no significant impacts to groundwater quality that could be ascribed to the WDC. The use of an intra-well approach eliminates the confounding results due to spatial variability (USEPA, 2009).

The remainder of this document describes the Shewart-CUSUM methodology in more details, including the underlying assumptions of the method, how data will be evaluated to determine whether the assumptions are satisfied, and how violation of these assumptions will be addressed. The steps in the data evaluation procedure are then explained using a flow chart in which decision points are clearly identified and the data analysis procedures are indicated. The data analysis procedures are based on U.S. Environmental Protection Agency guidance documents (USEPA, 2009; 2006) and are also well documented in the environmental statistics literatures (Gilbert, 1987; Gibbons et al., 2009). The described procedures are consistent with the Utah Hazardous Waste Rules, specifically R315-8-6, Groundwater Protection.

2.0 STATISTICAL EVALUATION METHODOLOGY

2.1 Shewart-CUSUM Control Charts

Shewart-CUSUM control charts were developed for statistical quality control of manufacturing processes (Bowker and Lieberman, 1972; Alwan, 2000) and have been adapted to groundwater monitoring at landfills (Gibbons et al., 2009). The basic idea is to use a time series record of a process, groundwater quality measurements in this case, to evaluate the current and very recent behavior of the process. If the current and very recent behavior is within the limits of natural random fluctuations consistent with the past behavior of the process, the process is considered to be in-control. If the current and very recent behavior is beyond the limits exhibited in the past, the process is considered to be out-of-control. The key point is that the historical background (baseline) data used as the basis for comparisons are obtained from the well itself (i.e., an intra-well approach). In the context of groundwater monitoring, “in-control” behavior indicates no adverse impact to groundwater quality, whereas “out-of-control” behavior suggests the potential for adverse impact to groundwater.

Accordingly, the two key objectives of the statistical evaluation are to: 1) establish upper control limits for each constituent in each well, based on the past history of the well; and 2) compare current monitoring results to the upper control limits.

There are two components to this control chart approach. The Shewart methodology focuses solely on the current measured concentration of a monitored groundwater constituent (arsenic, for example) and its relation to the mean measured concentration of the constituent within the same well as computed from past measurements. It is sensitive to large and sudden changes, but less sensitive to slow, upward trending changes in measured concentrations. The CUSUM methodology measures cumulative deviations from the mean, and thus incorporates information from previous measurements in the recent past; it is sensitive to small, gradual changes in the mean relative to the historical records. The Shewart and CUSUM statistics are presented below, within the context of the overall data evaluation procedure (Section 2.3). USEPA (2009), ASTM (1998), and Gibbons et al. (2009), describe the Shewart-CUSUM approach in detail.

2.2 Requirements and Assumptions

The Shewart-CUSUM method requires that enough historical measurements are available to obtain reasonable estimates of the mean and variance for the groundwater constituent within a particular well. The method further assumes that the data are independent, meaning that they are uncorrelated and do not exhibit a trend, and that they are identically distributed samples from a Gaussian (normal) distribution. Thus, although the Shewart-CUSUM method is the key to evaluating the effect on groundwater quality of the WDC, other statistical methods are needed to assess whether these basic conditions are satisfied. These supporting methods are described in the next section, which presents a step-by-step narrative of the entire data evaluation process together with a flow chart illustrating the decision logic of the process. The narrative also describes the procedures to be followed when the conditions are not met.

Before proceeding, however, it is important to note that some aspects of the statistical evaluation rely on time series analysis techniques (Box et al., 1994; Chatfield, 2004), which usually require at least 50 measurements at equally spaced intervals. The WDC groundwater monitoring data do not satisfy this requirement. At most, around 30 to 40 measurements were available through May 2016 (the last time baseline data were updated) for the majority of the groundwater constituents. The data were acquired on a semi-annual basis earlier and on an annual basis later, and thus, the measurements are not strictly evenly spaced. In spite of these shortcomings, standard time series analysis techniques will be utilized.

Furthermore, in many instances, a measurement in which a constituent is not detected is replaced by the median of the reporting limit for the given well-constituent for evaluation purposes. This practice amounts to replacing censored data with a fixed number, and can distort statistical inferences because qualitative data (non-detects, meaning concentrations somewhere below or “less-than” the reporting limit) are combined with quantitative data (concentrations above the reporting limit). For example, if 50% of the measurements for a constituent are non-detects, the data histogram will exhibit a large spike on the left, since an unusually large fraction (50%) of the measurements occur at the median reporting limit for the given well-constituent.

Throughout the following narrative, the sample variance is assumed to be constant through time. As previously noted, the time series involved are short and do not provide the historical record needed to thoroughly assess the issue of heteroscedasticity. In the future, when the record is longer (>50), it may be possible to address heteroscedasticity by comparing the variance computed from different segments of the record.

2.3 Evaluation Procedure

The description of the data evaluation procedure pertains to a single constituent in an individual well. It must be applied to each constituent in the well, and to each well, including the up-gradient wells. For purposes of explanation, denote the measured concentrations for the constituent in a well as $\{x_1, x_2, \ldots, x_N\}$, where the subscript i represents sampling event (and thus time) and N is the number of sampling events under consideration.

There are two stages in the data evaluation. In the first stage, a background (baseline) period is specified, the data within the background period are evaluated to determine whether the Shewart-CUSUM assumptions are met, the data are adjusted if necessary to meet the assumptions, and the Shewart-CUSUM upper control limits are established. This stage constitutes modeling of the historical records to establish the basis for evaluating whether future measurements are consistent with past behavior. It is performed at the outset, prior to evaluation of measurements from subsequent monitoring events, and is periodically updated as the historical records increase in length. This update is not performed for each individual monitoring event.

In the second stage, a current measurement is evaluated in relation to the upper control limit established using the data from the background period. Much of the analysis used to evaluate the background data is not repeated for each monitoring event, in particular for the assessment of whether the Shewart-CUSUM assumptions are satisfied, and the determination of the upper control limits. Determinations that are made during the development of the historical model in the first stage (for example, that the data exhibit a particular trend) are kept fixed and are not reevaluated until the historical model is updated.

Data from approximately 30 to 40 sampling events were available through May 2016. These historic data are proposed to be used for development of the historical model (first phase), with all subsequent events analyzed according to the procedures for the second stage of data evaluation, and the historical model is proposed to be updated every eight (8) new sampling data points after 2016.

2.3.1 Development of the Historical Model

Each of the subsections below corresponds to a decision element shown in Figure 6-1, a flowchart of the procedure for developing the Shewart-CUSUM upper control limits.

2.3.1.1 Detection Frequency

To ensure that a reasonable number of historical background data is available, the Shewart-CUSUM method is applied only to a constituent with a detection frequency greater than 25% (less than 75% non-detects). When there are more than 75% non-detects, the time series is plotted and the prediction limit is set to the maximum of the $\{x_i\}$. If the measured concentration in a future sampling event exceeds the prediction limit (i.e., the maximum value), the exceedance is verified by the next round of sampling. If the exceedance is confirmed, UDEQ/DWMRC will be notified and the prediction limit will be updated.

When the detection frequency is greater than 25%, non-detects are replaced by the median of the reporting limit.

2.3.1.2 Descriptive Statistics and Data Plots

The basic descriptive sample statistics of the $\{x_i\}$ (mean, median, standard deviation, minimum, maximum, absolute range, 25th percentile, 75th percentile, and interquartile range) are computed and tabulated. The time series plot is prepared for each well and each constituent. This is a standard, basic step of data analysis and characterization.

2.3.1.3 Identification of Monotonic Trends

Trending data violate the independence assumption of the Shewart-CUSUM method. Prior to the Shewart-CUSUM calculation, the existence of a monotonic trend is investigated using the non-parametric Theil-Sen trend test, as recommended by USEPA (2015), at the 99% confidence level for a null hypothesis of no trend against a two-tailed alternative (i.e., existence of either increasing or decreasing trend). The Theil-Sen trend test is a non-parametric technique that is less sensitive to the presence of outliers and the violation of normality assumption than the standard linear regression. Non-monotonic trends are not investigated.

Both the slope $\hat{m}$ and intercept $\hat{b}$ are estimated by the Theil-Sen method (USEPA, 2015). The trend line is given by $\hat{x}_i = \hat{m} \cdot i + \hat{b}$, where $\hat{x}_i$ represents an estimate of the measured concentration from the *i*th sampling event.

If a trend is detected, it will be assumed to be linear and removed from the time series, except that the overall mean of the time series is maintained in order to keep the evaluation in terms of the magnitude of the measured concentrations. Thus, subsequent analyses are performed with a modified time series $\{y_i = (x_i - \hat{x}_i) + \bar{x}\}$, where

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

is the sample mean of the data. The quantity $(x_i - \hat{x}_i)$ is called the *i*th residual.

2.3.1.4 Identification of Correlation

Correlated data also violate the independence assumption of the Shewart-CUSUM method. Following the evaluation, and removal if necessary, of a trend, correlation between the $\{x_i\}$ (or $\{y_i\}$) is investigated. To evaluate data correlation, the sample autocorrelation function (acf) (Alwan, 2000; Chatfield, 2004)

$$r_k = \frac{\sum_{i=1}^{N-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})}$$

is computed. The quantity r_k is called the autocorrelation coefficient at the *k*th lag. The number of lags *k* is equal to $N/2$ if the number of data N in the time series is even, and $(N+1)/2$ if N is odd. To test for the significance of autocorrelation, the Ljung-Box *Q* and *p*-value are computed using a standard commercial statistical software (JMP, 2010). The *Q*-statistic is used to test whether a group of autocorrelations is significantly different from zero or to test that the residuals from a model can be distinguished from white-noise. The test is performed at 1% significance live (i.e., 99% confidence level), which is appropriate to guard against falsely inferring correlation because the time series involved are quite short ($N < 50$ for the purpose of developing the historical model), fewer than what are generally required as a minimum (50 to 100 samples) for time series analysis.

If the *p*-value is less than 0.01, the time series $\{x_i\}$ (or $\{y_i\}$) is assumed to be a realization of a autoregressive process or order 1, denoted as AR(1), (Alwan, 2000; Chatfield, 2004), for which $\hat{x}_i = w_0 + w_1 \cdot x_{i-1}$ (or $\hat{y}_i = w_0 + w_1 \cdot y_{i-1}$). The coefficients w_0 and w_1 are obtained by performing a linear regression of $(x_2, x_3, \ldots, x_{20})$, considered as the dependent variables, against $(x_1, x_2, \ldots, x_{19})$, considered as the independent variables. Subsequent analyses are performed using the time series of residuals, $\{u_i = x_i - \hat{x}_i + \bar{x}\}$ (or $\{u_i = y_i - \hat{y}_i + \bar{y}\}$).

2.3.1.5 Testing for Normality

Testing for normality of the $\{x_i\}$ (or residuals $\{y_i\}$ or $\{u_i\}$) is performed using the Shapiro-Wilk W test, which is appropriate when $N \leq 50$ (Gilbert, 1987; USEPA, 2006; USEPA 2015). In the future, when $N > 50$, D'Agostino's test (Gilbert, 1987) or Lilliefors Test (USEPA, 2015) will be used.

Normality of the $\{x_i\}$ is one of the assumptions of the Shewart-CUSUM method. The purpose of testing for normality is simply to evaluate whether this assumption is met. When the assumption is not met, a common approach is to transform the data by taking the logarithm of each of the $\{x_i\}$ or by raising them to a power so that the transformed data are Gaussian. However, as pointed out by Gibbons (2009), the normality assumption is less important than the independence assumption to the robust performance of the Shewart-CUSUM method. Therefore, we propose to proceed with the Shewart-CUSUM method regardless of whether the distribution of the $\{x_i\}$ is consistent with a hypothesis of normality, except that a non-parametric upper control limit is established in those cases where normality is rejected.

The main reason for choosing to use non-parametric control limits is to keep the statistical evaluation in terms of the units of measured concentration (mg/L). Data transformation generally obscures this meaning, since it is not clear that the control limits in terms of concentration units are obtained by simple inverse transformation of the control limits established using transformed data.

2.3.1.6 Shewart-CUSUM Control Limits

The Shewart-CUSUM method relies on control limits that are based on background statistics for a well that are computed using past measurements from within the well itself. Eight (8) initial samples are often used as a background sample when a monitoring program first gets underway. At the WDC, however, monitoring has been ongoing since 1992, and for most constituents, approximately 30 to 40 samples are available up to May 2016. Therefore, all data collected on or before May 2016 (the last time background statistics were updated) are used to represent background in each monitoring well at the WDC. With these background samples, first compute the sample mean:

$$\bar{x}_{bkgrnd} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and the sample standard deviation:

$$s_{bkgrnd} = \left[\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}_{bkgrnd}\right)^2\right]^{\frac{1}{2}}$$

With these statistics, compute the upper control limit $UCL = \bar{x}_{bkgrnd} + 4.5 \cdot s_{bkgrnd}$, which serves as the control limit for both the Shewart and the CUSUM methods. In the non-parametric case, the control limit is specified as the maximum detected value, analogues to the non-parametric prediction limit suggested by USEPA (2009).

The remaining steps in the Shewart-CUSUM procedure are described in the next section.

2.3.2 Comparison of Current and Future Data Against Background

The steps outlined above in Section 2.3.1 establish a control limit based on the historical records. The evaluation of current and future sample results is not as exhaustive as the evaluation of the historical data, until the historical model is updated (after every 8 new future data points). The entire time series (historical data plus the post-baseline data) is plotted, and the post-baseline data modified according to the model identified in the evaluation of background. For example, if the historical data indicated a historical trend, then the post-baseline data are de-trended as described above. If normality was not rejected during the historical evaluation, then normality is assumed to apply to the post-baseline data. Then, the comparison using the Shewart-CUSUM method is as follows:

1) Beginning with the first sample after the baseline period, in the time series, denote the new measurement taken at time $j = i$ - (background sample size, N) as x_j. (Recall that the first N samples were used to compute background and are not subsequently evaluated individually. In general, $j = 1$ in the first sampling event following the historical evaluation period.)

2) At each time t_j, compute $S_j = \max[0,\ (x_j - 0.75s_{bkgrnd}) + S_{j-1}]$, the cumulative deviations from the mean, where $\max[A, B]$ denotes the maximum of A and B and $S_0 = 0$. (Note: S_j is often compute using the standardized quantity, $z_j = (x_j - \bar{x}_{bkgrnd}) / s_{bkgrnd}$. However, in doing so the relationship of S_j to measured concentrations is obscured. The above formula is equivalent to the formula computed using z_j given in Gibbons (2009) and ASTM (1998). The *k* factor (proportion of background standard deviation to be deviated) is suggested to be 0.75 by USEPA (2009) when $N \geq 12$.)
3) Plot both x_j and S_j versus j (or the date corresponding to j) on a time chart, thus constructing the combined Shewart-CUSUM control chart. The control chart also has a horizontal line drawn at a value (*y*-axis) of $\bar{x}_{bkgrnd} + 4.5s_{bkgrnd}$ corresponding to both the Shewart and CUSUM control limits (in the parametric case). As noted by Gibbons (2009), this control limit corresponds approximately to the upper limit of a 95% confidence interval. When either x_j or S_j exceed the control limit, a potential impact to groundwater may have occurred and is reported to UDEQ/DWMRC. However, it is also possible that the control limit is exceeded due to laboratory error, transcription error, or some other anthropogenic cause (a review of past data from the WDC indicates that this has sometimes occurred). Thus, the report of exceedance to UDEQ/DWMRC will include a statement regarding the possible causes of the suspected measurement. The out-of-control condition is then verified on the next round of sampling before further action is initiated. If the verification sample is also out-of-control, the UDEQ/DWMRC will be notified and the cause of the exceedance ascribed to the WDC.

REFERENCES

Alwan, L. C., 2000. *Statistical Process Analysis.* McGraw-Hill, Boston.

ASTM. 1998. "Standard Guide for Developing Appropriate Statistical Approaches for Ground-Water Detection Monitoring Programs." ASTM Guide D6312-98.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel, 1994. *Time Series Analysis, Forecasting, and Control.* Prentice-Hall, Englewood Cliffs, NJ.

Chatfield, C., 2004. *The Analysis of Time Series: An Introduction.* Chapman & Hall/CRC, Boca Raton, FL.

Gibbons, Robert D., D. K. Bhaumik, and S. Aryal. 2009. *Statistical Methods for Groundwater Monitoring.* Second Edition. John Wiley & Sons, Inc., New York, New York.

Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring.* John Wiley and Sons, New York.

JMP. 2010. SAS Institute Inc., Cary, NC. Version 9.

U.S. Environmental Protection Agency (USEPA). 2009. *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities – Unified Guidance.* USEPA 530-R-09-007.

U.S. Environmental Protection Agency (USEPA). 2006. *Data Quality Assessment: Statistical Methods for Practitioners (USEPA QA/G-9S).* USEPA/240/B-06/003. Office of Environmental Information, Washington, D.C.

U.S. Environmental Protection Agency (USEPA). 2015. ProUCL Version 5.1.002 Technical Guide. Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC. Report No. EPA/600/R-07/041.

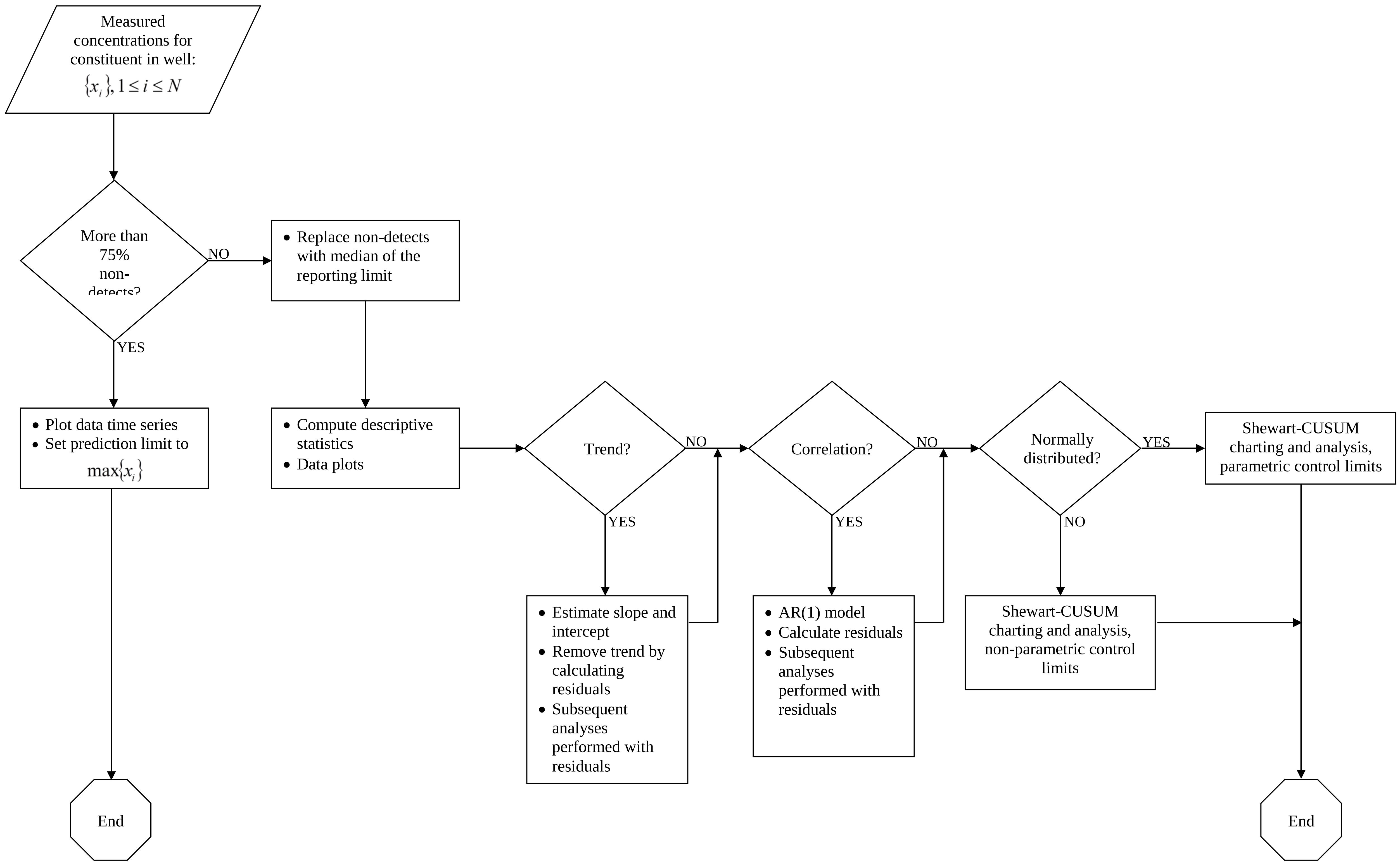

Figure 6-1. Flowchart showing main steps used in developing the historical model of groundwater monitoring data from the WDC.