# SECTION 7

## TANK TESTING DATA REDUCTION AND QUALITY ASSURANCE (RETEST) RESULTS

This section describes the statistical data reduction process whereby the raw data generated by physical tank tightness tests in the field were converted into estimates of volume change rates under test conditions for tank systems (vessels plus piping). It then gives results of the quality assurance retests which help in judging the overall accuracy of the physical test and data reduction process. Note that throughout this Section of the report, quantitated volume change rates are given as measured under test conditions and are not adjusted from test pressures to operating pressures.

### I. DATA COLLECTION AND REDUCTION

#### A. Data Collection and Transmission

Raw data of volume and temperature change at five-minute intervals were collected for a two-hour period during the physical test, as described in Section 6. These data were collected in handwritten form on the data sheets normally used by Petro-Tite, and were also entered onto a spreadsheet using a mini-computer at the field site. Data were transmitted from the spreadsheet to MRI by telephone for timely analysis. The diskettes and hard-copy data sheets were shipped to MRI on a weekly basis. The telephone transmission was checked against the diskette, and the diskette against the hard copy to ensure that the correct raw data were entered in the working spreadsheet file, which was then used to estimate volume change rates and their (within-test) variability.

B. Standard Data Reduction

Several statistical analysis methods for reduction of the tank test data were considered for use on the national survey. The test method produced a volume change measurement at five-minute intervals. The other measurement recorded at five-minute intervals was a temperature measurement. The temperature was recorded as a cumulative reading -- the tank temperature -- while the volumes were recorded as differences. In order to make the temperature and volume data comparable, they had to be put in the same form. Either both must be changes or both must be cumulative.

As a result of the considerations of the types of analyses available (see Part III of Appendix D for discussion of possible methods) and the advantages and disadvantages of each, a standard analysis was designed. For the standard analysis, the estimated volume change due to temperature change and the observed total volume change were both expressed in cumulative form, beginning at zero for the start of the test. A straight line through the origin was fit to the temperature-related volume change data by least squares. The predicted values of this line were calculated and used as a smoothed temperature correction for the observed volume changes. The data were plotted and inspected visually for outliers or deviations of the temperature data from linearity. Any questionable data were checked in detail or considered for special analysis.

If no problems with the data were found, the predicted values from the smoothed temperature line were used as the temperature correction. This smoothed temperature correction was subtracted from the observed volume data for each time point. The resulting differences were divided by the time interval to obtain a series of volume change rates expressed in gallons per hour, typically based on a five-minute interval. The arithmetic mean of these rates was calculated and used as the estimate of the volume change rate. In calculating the variance n-1 was used as the divisor, where n is the number of terms in the mean. The result was divided by n to form the variance of the mean. The square root of this is the within-test standard error reported before adjusting for between-test variation. See Part V of Appendix D and Part I of Section 8 for a discussion of the between-test variance.

## C. Special Analyses

A number of data set features called for a different or more detailed analysis than that described above. These were dealt with on an individual basis. Occasionally apparent outliers were found. These were checked against the raw data and the test log to see if there was any physical reason for them. A few tests had thermistor boxes fail during the test for some reason (rain, FM interference). These generally gave temperature data that appeared as outliers. When outliers were found and a physical reason identified, the aberrant data were removed from the analysis. This generally required smoothing over the missing data by interpolation. If errors were identified, they were corrected and the analysis redone.

The typical data showed a consistently increasing temperature, generally linear. A smaller proportion of the data sets showed linearly decreasing temperature. Some data sets showed evidence of temperature increase that was curvilinear. If this curvilinearity appeared or was suspected, a test for curvilinearity was done by fitting both a linear and quadratic to the temperature data by least squares (through the origin). If the quadratic improved the fit significantly, the curvilinear fit (using both linear and quadratic terms) was used for smoothing.

A few cases were found where both temperature and volume were not only non-linear, but also not moving consistently in one direction. Provided that they showed the same pattern, analysis proceeded. In this event, a five point moving mean was used to smooth the temperature data. Equal weights were used. This resulted in the loss of four data points; two at the start and two at the end of the test.

Some tests showed volume change rates that were initially increasing rapidly in curvilinear fashion, while the temperature changes were quite linear. The volumes typically increased rapidly for the first few observations, then slowed. This was interpreted as relaxation or tank deformation. The apparent relaxation appeared to follow an exponential curve and to approach the temperature change rate as an asymptote. However, the constant of this differed by tank. The rate of relaxation may be related to the nature of the soil in backfill and water conditions. When this was identified, the initial points exhibiting this relaxation of the tank deformation were deleted before analysis.

## D. Criteria for Invalid Data

A few of the data sets from the tank tests were judged invalid based on the analysis of the data. This occurred infrequently (in 6% of test results).

There were a number of criteria for declaring a data set to be invalid. The most common was that the data showed a volume increase even after adjusting for temperature. Since the test method places pressure on the tank, a volume increase cannot occur from inflow of water. Data that showed volume increases after temperature adjustment that exceeded levels that could be reasonably attributed to the variability of the measurement process were judged to be invalid tests. The reason for this is that such an apparent volume increase with no explanation could be eclipsing a small actual volume loss or leak. Generally any tank that showed a volume gain rate of more than 0.1 gallons per hour after temperature adjustment was judged to be an invalid test. The most likely explanation for such tests is that those tanks had trapped vapor pockets.

As described in Section 8 (Part III) and Appendix D (Part VII), at the next stage of analysis, some additional tests were judged to be invalid due to a measured inflow that was excessive when compared with its estimated total standard error, even though the inflow was not as large as 0.10 gallons per hour.

A variety of other data features led to the conclusion that the test was invalid. A few instances were found where the temperature as recorded fluctuated erratically during the test while the volume measurements were relatively stable. If the temperature data were so erratic as to preclude a temperature adjustment, then the test was declared to be invalid.

One or two tests showed both temperature and volume measurements that were erratic and did not appear to track together. These tests were also judged invalid. Such behavior may have been caused by incomplete tank deformation, followed by relaxation, combined with mixing problems. No valid volume change rate could be estimated.

## II. RETEST RESULTS

Three types of retests were conducted as part of the national survey of underground storage tanks. One was a back-to-back retest, conducted immediately after the original test used to estimate the volume change rate. The second was a leak simulation test also conducted immediately after the original test. The third type was a complete retest conducted on a different day and generally by a different crew. Each of these types of second testing estimates a different source of variation possible in the tank tests. The leak simulation and back-to-back retests estimate variation of the overall measurement procedure from one two-hour period to the next, with the same set-up, crew, day of the week, and so on, while the complete retests measure variation between tests as well. The initial test result in each case was used as the data for the survey estimate. When the test and retest results differed, the results were examined to discover reasons for the differences. This led to the discovery of the repairs that had been made in two cases. The primary purpose of the quality assurance program was to measure the overall performance of the test, which was accomplished. A list of all of the retests appears in Appendix D, as does a list of the simulated leak retests. A table summarizing the estimates of bias (lack of accuracy) and standard deviation (precision) based on each type of test is presented as Table 7-1. We discuss these three types of quality assurance retests in more detail below.

### A. Leak Simulations

The leak simulation tests were conducted after the original test was concluded. Generally they were only conducted when the original test indicated that the tank was tight or had a small estimated volume change. The volume rate used for leak

Table 7-1. Retest results (volume change rates measured under test conditions, not adjusted to operating pressure)

| Type | Mean difference (gallons per hour) | N | Variance $(gph)^2$ | Mean squared error $(gph)^2$ | Standard deviation (gph) | Root mean squared error (gph) |
|---|---|---|---|---|---|---|
| Leak simulation | -0.00891 | 11 | 0.00066 | 0.00074 | 0.0257 | 0.0272 |
| Back to back | 0.00629 | 14 | 0.00053 | 0.00057 | 0.0231 | 0.0239 |
| Retests | 0.00297 | 34 | 0.00254 | 0.00255 | 0.0504 | 0.0505 |

simulation was on the order of 0.1 gallons per hour, so a large observed volume change would overwhelm it.

The purpose of the leak simulation tests was to document that the testing method could detect leaks of known size in tanks that appeared to be tight. In addition, use of the leak simulation allows for an estimate of the accuracy of the test as well as its precision. The accuracy refers to the ability of the test to measure a known volume change, while the precision of the test refers to its ability to reproduce measured rates.

Thirteen leak simulation tests were conducted. Two of these were conducted on tanks that had estimated volume rates that indicated that the tanks were probably leaking (as evidenced by the observed volume changes). These tests were excluded from the analysis because variability is known to increase for leaking tanks.

Three rates were calculated from leak simulations. The first was a baseline rate for the tank. This was estimated during the regular tank test. While the leak simulation was conducted, a measured rate was estimated. This is the rate observed by the testing method during leak simulation. It is presumed to be composed of the tank rate plus the simulated rate. The simulated rate is calculated by collecting product drawn from the tank at a constant rate, weighing it on a triple beam balance, and converting the weight to volume at the temperature of the product in the tank. The difference between the observed rate during the simulation and the baseline rate provides an estimate of the simulated rate. The difference between this and the actual simulated rate can be used to assess the accuracy of the test.

The average difference between the measured rate and the simulated rate was -0.00891 gallons per hour, based on the 11 leak simulations where the tank was not estimated to be leaking (see Table 7-1). Recall that these rates are reported as observed under test conditions and not as adjusted for test pressure. The difference between the measured rate and the simulated rate is interpreted as an estimate of bias. The variance of the differences about their mean provides an estimate of precision. This variance was estimated to be 0.00066 gallons per hour squared. The mean squared error (MSE) is a measure which incorporates both types of error--accuracy and precision. It is calculated as the sum of the bias squared plus the variance. In this case it was 0.00074 gallons per hour squared.

The bias is clearly not significant in that it does not differ significantly from zero (t = -0.347, 10 degrees of freedom). As a result, the variance and the mean squared error are nearly identical. A measure of variation often used is the standard deviation (or root mean squared error if bias is present), which is the square root of the variance (or MSE). This measure has the advantage that its units are the same as the measurement, gallons per hour. The standard deviation (estimating within-test variation)was estimated to be 0.0257 gallons per hour for these data.

## B. Back-to-Back Retests

Back-to-back retests were conducted on a total of 18 tanks, which includes the 13 tanks with leak simulations. The purpose of the back-to-back retests was to estimate the stability of the test method. That is, to ensure that the volume change estimate did not differ markedly if based on the succeeding two hours after the test.

As with all of these tests, variability is expected to be larger if the initial leak rate or volume change is larger. For this reason, the results of the back-to-back retests are presented here for the 14 tests with volume change rates less than 0.1 gallons per hour in absolute value. Retest results for tanks with larger volume rates were more variable but generally consistent. (See Appendix D, Part IV for a discussion of these retests.)

The average difference between the original and retest for the 14 tests with small volume changes was 0.00629 gallons per hour. The variance estimate was 0.00053 gallons per hour squared, giving a mean squared error of 0.00057 gallons per hour squared. The corresponding standard deviation was 0.0231 gallons per hour and the root mean squared error estimate was 0.0239 gallons per hour (not adjusted for test pressure). The mean difference was not significantly different from zero ($t = 0.272$, 13 df).

C. <u>Complete Retests</u>

The complete retests consist of revisits to the site on a different day. Typically this includes a different crew and involves rescheduling and refilling the tank. The complete retests incorporate all of the features of a tank test and so include all the sources of error including potential difference from crew to crew (including differences between sets of testing equipment) and differences due to weather conditions, nearby traffic, day of the week, etc. In addition, there is a possibility that the tank is different at the time of the retest. In fact, two of the retests originally scheduled were canceled when it was found that the tanks had been repaired between the initial test and the scheduled retest. In addition, two retests were performed and it was then discovered that the tanks had been repaired between the initial test and retest. These data are also not included as they would measure an additional source of variation which is not of interest (i.e., repair). Two other retests were performed on tanks that were initially determined to have large vapor pockets. These two tanks were retested later and on retesting were again found to have large vapor pockets. The results of the test and retest for these tanks with vapor problems agreed qualitatively; however, the numerical agreement was not close. The reason for this may be that the vapor pocket trapped in the tank was of different size. There were also different ambient conditions that would affect the vapor differently. For these reasons, the vapor retests were not included in the estimate of the variance from the retests.

The mean difference from the set of 34 relevant retests was 0.00297 gallons per hour. The variance of the difference was 0.00254 gallons per hour squared, giving a mean squared error of 0.00255 gallons per hour squared. The standard deviation of the differences for these 34 retests is 0.0504 gallons per hour and the root mean squared error is also 0.0505 gallons per hour. The mean difference is not significantly different from zero (t = 0.059, with 33 df).

D. Results

The retest data analysis showed no evidence of bias in the test methods. All three retest schemes had very small estimates of bias which were not significantly different from zero. Given the historical leak cut-off of 0.05 gallons per hour, bias of less than 0.01, as was found in all three data sets (less than

0.005 in the largest set) is not of practical concern, in addition to not being statistically significant.

The variability, or magnitude of the measurement error of the physical test can also be assessed using these data. Both the back-to-back retest and the leak simulations estimated within-test standard deviations on the order of 0.025 gallons per hour. The complete retest data gave a standard deviation of 0.05 gallons per hour for the total variability of volume change rate estimates. As is discussed in Section 8, Part I, the difference is probably due to a between-test component of variation which is measured by the retests but not by the back-to-back or leak simulation tests.

In summary, the physical test is <u>accurate</u> (not significantly biased) and has a known precision (total standard error of an estimated volume change) of 0.05 gallons per hour (measured at test pressure and not adjusted to operating pressure). That this standard error coincides with the historical cut-off value for declaring a leak is an interesting coincidence.