United States Environmental Protection Agency
Office of Underground Storage Tanks

EPA

Standard Test Procedures For Evaluating Release Detection Methods: Statistical Inventory Reconciliation

May 2019

Acknowledgments

The U.S. Environmental Protection Agency's Office of Underground Storage Tanks contracted with Battelle under Contract No. EP-C-10-001 to revise EPA's 1990 *Standard Test Procedures for Evaluating Release Detection Methods*. Individual members of the National Work Group on Leak Detection Evaluations, as well as Ken Wilcox and Associates, reviewed this document and provided technical assistance. A stakeholder committee, comprised of approximately 50 representatives from release detection method manufacturers and various industry associations, also commented on this document.

Contents

Appendices

Figure

Table

List Of Acronyms And Abbreviations

ATGS	automatic tank gauging system
B	bias
CFR	Code of Federal Regulations
CITLDS	continuous in-tank leak detection system
df	degrees of freedom
EPA	U.S. Environmental Protection Agency
°F	degree Fahrenheit
gal/hr	gallon per hour
hleak	height of the leak
hmax	maximum fill height for the period
hproduct level	measured fill height
Li	estimated leak rate
LL	lower confidence limit
Lmax	maximum leak rate
MDL	minimum detectible leak rate
MSE	mean squared error
P(d)	probability of detecting a leak
P(fa)	probability of false alarm
R	leak size
S	induced leak rate
SD	standard deviation
SIR	statistical inventory reconciliation

Sp	pooled standard deviation
Th	threshold
UL	upper confidence limit
UST	underground storage tank

Section 1: Introduction

1.1 Background

The federal underground storage tank (UST) regulation in 40 Code of Federal Regulations (CFR) Part 280 specifies performance standards for release detection methods. UST owners and operators must demonstrate that the release detection method they use meets the U.S. Environmental Protection Agency's (EPA) regulatory performance standards. This document provides test procedures for evaluating the release detection category of statistical inventory reconciliation (SIR) methods.

This statistical inventory reconciliation document is one of four EPA standard test procedures for release detection methods. The test procedures present performance testing approaches to evaluate various release detection method categories against the federal UST regulation in 40 CFR Part 280, Subpart D. To provide context for the four test procedure documents, EPA developed *General Guidance For Using EPA's Standard Test Procedures For Evaluating Release Detection Methods*. The general guidance provides an overview of the federal UST regulation, methods, and testing that may result in release detection methods listed as compliant with the regulatory performance standards. The general guidance is integral; it must be used with the test procedures.

Vendors offer UST owners and operators commercial services for SIR methods. These methods obtain inventory data taken by personnel operating the tanks. A SIR vendor then analyzes the inventory data and reports results to an owner or operator. Many SIR methods are based on proprietary computer programs, which analyze the data. The methods vary in the effects they attempt to detect or control for. This document only provides procedures to evaluate the method's ability to detect releases.

The UST release detection performance standards are specified in terms of the probability of a false alarm (P(fa)) and the probability of detecting a leak (P(d)). A false alarm occurs if the release detection method mistakenly indicates a leak when the tank is, in fact, tight. The P(d) measures the method's ability to detect leaks of specified magnitude.

One level of performance for SIR methods is specified as the ability to detect a leak of 0.1 gallon per hour (gal/hr) with a P(d) of at least 95 percent, while operating at a P(fa) of no more than 5 percent, based on an inventory record of specified length. Tightness testing for piping is more stringent. Because leak rates depend on pressure in the line, EPA established in the federal UST regulation that the minimum performance standards for line leak detection methods are specified in terms of the line operating pressure. Section 280.44(b) requires for line tightness testing – "A periodic test of piping may be conducted only if it can detect a 0.1 gallon per hour leak rate at one and one-half times the operating pressure." Not every UST implementing agency allows use of SIR methods to meet the piping line tightness test requirement. If a SIR method is to be used to meet the line tightness testing requirement, it must detect a corresponding leak rate of 0.08 gal/hr. This figure is based on the relationship of leak rate to pressure for flow from a free orifice in which the leak rate is proportional to the square root of the change in pressure.

A second level of required performance is specified as the ability to detect a leak of 0.2 gal/hr with a P(d) of at least 95 percent, while operating at a P(fa) of no more than 5 percent, based on an inventory record of specified length. This level corresponds to the performance requirements for monthly monitoring release detection methods for tanks and piping.

1.2 Objective And Application

The test procedures address two objectives. They provide procedures to test SIR methods in a consistent and rigorous manner. Also, they allow the regulated community and regulatory authorities to verify compliance with the UST regulation.

Note that these test procedures only evaluate the performance of SIR methods as release detection methods for USTs. Many SIR vendors offer commercial services that provide other information for owners and operators, such as identification of probable theft or short deliveries. These test procedures do not evaluate the adequacy of the method for these capabilities. In addition, they do not address equipment safety testing or operating procedures. The vendor is responsible for conducting the testing necessary to ensure the method is safe for use with the type of product being tested. Safety is a concern in collecting inventory records, but not in the statistical analysis.

The application of these test procedures is to determine whether a vendor's SIR method meets EPA's performance standards for release detection. The test results are used to estimate the P(fa) and the P(d) for leak rates of 0.1 and 0.2 gal/hr. The test procedures analyze the difference between reported and induced leak rates and use the variability of these differences together with a normal probability model for the errors to estimate the performance parameters for SIR methods.

The test procedures also provide a process to estimate the size of a leak a SIR method can detect. The rate is estimated by determining the threshold for a 5 percent P(fa) and then calculating the corresponding leak rate that is detectable with a P(d) of 95 percent.

Ultimately, you can use the results of this evaluation to prove that the SIR method meets the requirements of 40 CFR Part 280, subject to the limitations listed on EPA's standard evaluation form in Appendix B.

1.3 Evaluation Approach Summary

To collect the performance data for SIR methods, the process of the SIR client providing data to the SIR vendor is simulated followed by the evaluator's analysis of the vendor's results. According to the test procedures, the evaluator sends a database of inventory records to the vendor, the vendor performs the statistical analysis using the SIR method, and the vendor reports the results or indicates which inventory records are insufficient and cannot be analyzed. To calculate the level of method performance, the evaluator compares the vendor's results to the actual or simulated tank conditions established in the database. The evaluator compares the method performance to EPA's regulatory performance level and the applicability of the method established.

The evaluator knows and will share with the SIR vendor how to collect and process inventory records that make up the database submitted for the evaluation. The database includes inventory records from multiple operating, tight tanks with a variety of monthly throughputs and various data sources of differing quality. The evaluator should have independent evidence the tank was tight, and the system components were operating correctly. The database includes tight tanks and tanks with simulated leaks. The database also covers a range of seasons and represents various operating conditions during a year.

The evaluator will use a computer program to randomly select a number of inventory records of length specified by the SIR vendor. The evaluator will modify inventory records as necessary to include leaks of known rates. The evaluator will introduce loss of product representing leaks of certain sizes into some of the inventory records. This information will be kept blind to the vendors. The inventory records will build a database for a SIR method to evaluate against an actual leak rate. SIR vendors will then evaluate the database and submit their results to the evaluator. The ability of the method to accurately identify leaks of specified sizes forms the basis for the evaluation.

The evaluator evaluates results against the EPA performance standards, and the applicability and limitations of the tested SIR method are calculated. Applicability of the SIR method may be listed as meeting the requirements for single tank and tanks connected by siphon piping. The SIR method must prove performance at the regulatory level for different tank configurations. In addition, the evaluator assesses the applied scaling of the system and throughput applicability of the method on the characteristics of the inventory records in the database.

1.4 Organization Of This Document

This document is organized as follows:

- Section 2 presents a brief discussion of safety issues.
- Section 3 presents the apparatus and materials needed to conduct the evaluation.
- Section 4 provides step-by-step procedures.
- Section 5 describes the data analysis.
- Section 6 provides interpretation of the results.
- Section 7 describes how you must report results to prove method performance.

Three appendices are included in this document.

- Appendix A provides definitions of some technical terms.
- Appendix B contains standard data collection and reporting forms: reporting evaluation results, describing the detection method, and recording data on individual test logs.
- Appendix C contains a protocol designed to be used to compare the reliability of two versions of the same method; that is, to determine whether or not the two versions of the same SIR program written in different computer languages produce the same results.

Section 2: Safety

The evaluator supplies to the SIR vendor an evaluation that consists of statistical analysis of a database. Thus, the work is in an office and no special safety considerations apply.

The instructions specified by SIR vendors should address safety issues involved in collecting inventory data. These activities include taking manual or automatic tank gauging system (ATGS) product level readings of the tank and reading the meter totalizer. Safety issues are only a concern when near tanks during data collection and not with the evaluation of the SIR method.

Section 3: Apparatus and Materials

3.1 SIR Method Description

The federal UST regulation requires that all release detection methods, except for annual line tightness testing and continuous 3 gal/hr pressurized pipeline monitoring, be conducted at least every 30 days for USTs and associated piping. SIR methods are those based on daily, periodic, or continuous inventory measurements and reconciliation to check for loss of product in a tank system. SIR can detect a leak, but SIR may or may not distinguish between a leak in the tank versus a leak in the piping. SIR methods can be implemented with tanks in any type of fuel service, including biofuels and ethanol-blended fuels.

SIR methods can be applied to a wide range of tanks and conditions; however, there are instances in which SIR methods are not applicable. The scope of SIR methods can cover single tanks and tanks connected by siphon piping, which are composed of two or more tanks joined to form a single UST system. SIR may be used for monthly monitoring on USTs with pressurized piping, if allowed by the UST implementing agency. There may be issues with a SIR method's ability to detect small volume changes in USTs with large sales volumes and fuel turnover rates, such as high throughput tanks. The storage capacity of a large throughput facility is often greater than the maximum capacity listed for a SIR method. Therefore, the actual product storage capacity may be a limiting factor for large facilities.

There are two types of SIR release detection methods: traditional and continuous. Traditional SIR uses an ATGS or takes daily manual liquid level readings of the product in the tank and reconciles them with the amounts of dispensed and delivered product. Continuous SIR performs the same product reconciliation as traditional; however, it can differentiate between line and tank leaks and can compensate for temperature variations with a continuous in-tank leak detection system (CITLDS). For continuous SIR, data are gathered from all designated input devices during tank quiet times when there are no sales and no deliveries and then SIR vendor software programs perform leak-test calculations when enough data is recorded. Most CITLDS methods use an ATGS to gather product-level data; this is considered a hybrid SIR method. Other CITLDS methods gather product-level data from input devices such as dispenser totalizers and point-of-sale records. CITLDS are well suited to facilities that are open 24 hours a day, 7 days a week, as long as the volume of the product sold from USTs does not exceed the throughput limit of the CITLDS method and there is enough quiet time to collect enough data. For more information on test procedures for evaluating these types of release detection methods, refer to *Evaluation Protocol for Continuous In-Tank Leak Detection Systems* revised on January 7, 2000 by Dr. Jairus D. Flora, Jr.

The SIR methods then use these inventory records to perform a statistical analysis of inventory discrepancies. CITLDS methods, in comparison to periodic measurements, provide a larger quantity of data, which compensate for temperature and typically provide better data for SIR analysis. Various components that might contribute to these discrepancies are generally isolated before a leak rate is estimated. In addition to a leak rate estimate, some SIR methods claim to provide information on a variety of sources of inaccuracies such as dispensing meter error, delivery error, manual liquid level measurement error, temperature effects, theft, and vapor loss.

These SIR test procedures in this document evaluate the method's ability to detect releases only. These procedures do not evaluate the performance of SIR methods in all capabilities, such as theft detection or delivery shortages.

Although, a qualitative SIR evaluation option was previously allowed, EPA no longer allows these methods and we removed detailed descriptions and test procedures related to those methods. The vendor reports a numerical leak rate; the leak rate estimated by the SIR method is compared to the induced leak rate in the database. The differences are summarized and used in the normal probability model for the measurement errors to estimate the performance of the method. In addition, the vendor's interpretation of the quantitative results as pass, fail, or inconclusive is compared to the result. Any inconclusive results are discussed with the vendor. The three SIR responses are:

- Pass – The SIR analysis indicates that a leak does not exist at or above 0.2 gal/hr, or at the leak rate of the evaluation, with a P(d) of at least 95 percent and P(fa) of no more than 5 percent.
- Fail – The SIR analysis indicates a loss of product from the UST. A fail result could be related to a leak, miscalibrated dispenser, inaccurately metered deliveries, or stolen product.
- Inconclusive – The SIR analysis cannot determine if the UST passed or failed.

3.2 Testing Materials

Since the release detection method consists of a SIR client supplying data analysis of inventory records, only a computer is necessary for evaluating the method. The materials needed consist of the evaluator supplying a database of inventory records to the SIR vendor for analysis and reporting, together with a code or key that allows the evaluator to identify the actual status of each tank system in the test database. The data must be inspected for completeness, transcription errors, and other factors that may impair their use in an evaluation. See Section 4 for more details on the database requirements. To ensure the quality of the data, the test inventory records must include:

- Tank size, in particular capacity, diameter, and length;
- Piping capacity, diameter, and length when evaluating method for tank and piping;
- Tank type, material of construction, and manufacturer;
- Product type;
- Date each product level measurement was taken;
- Daily opening product level measurement and volume;
- Daily closing product level measurement and volume;
- Daily sales volume;
- Gross deliveries over the course of the month;
- Thirty days of observations; and
- Data from leak and no leak conditions.

Developing a database with altered and unaltered conditions is a time-consuming task, and the database may only be used once. However, inventory records used to generate the database may be used again to develop other databases, as long as each set has its own unique leak rates, leak locations, etc.

Section 4: Test Procedures

The test procedures to evaluate the performance of SIR methods consist of six steps. This section explains these steps. The appropriate statistical analysis procedures are presented in Section 5.

Step 1 Determine the data requirements set by the vendor. Obtain the vendor's data reporting form.

Step 2 Collect inventory records used as the basis for the database agreed to with the SIR vendor during Step 1. Ideally, the database should be based solely on inventory records the evaluator generated from the test facility tanks or from field sites under tight and simulated leak conditions. Alternatively, you can obtain inventory records from tanks and lines that are known to be tight or with known leak rates and use them as the basis for an evaluator-generated database. The database may also contain a mixture of field collected inventory records and evaluator generated inventory records as a way to control for the quality of data within the inventory records. The inventory records come from single tanks and tanks connected by siphon piping.

Step 3 Generate the database. Use collected inventory records from tanks with simulated leaks and use evaluator generated data or add induced leaks mathematically to some inventory records that use altered field-generated data; code the inventory records to prevent identification.

Step 4 Submit the database to the SIR vendor for analysis. Keep the test design blind to the vendor to prevent a biased evaluation of the leak status or leak rate identification.

Step 5 Receive the results from the SIR vendor.

Step 6 Analyze the results and report the test evaluation results to the SIR vendor.

4.1 Determine Vendor's Data Requirements—Step 1

Each SIR method will have unique data requirements. The evaluator will discuss the data requirements with the SIR vendor and obtain a copy of the vendor's data reporting form. The evaluator will then determine exactly what data elements need to be included in the inventory records. For example, as part of the inventory record, a method may require a copy of the tank chart, daily mean ambient temperatures, or meter calibrations. In addition, the length of the record is an important consideration. EPA suggests the evaluator obtain a longer record than the minimum required by the SIR vendor. If the vendor is having a method evaluated for single tanks and tanks connected by siphon piping, the database needs to include the various tank configurations.

4.2 Obtain Inventory Records—Step 2

The evaluator obtains inventory records from operational tanks with evidence the tanks and system components are tight or uses evaluator generated inventory records from test tanks with induced leaks of known rates in the field or at a test facility. The evaluator ensures the variables recorded and the length of the data record meet the vendor's specifications. The evaluator collects the inventory records under a variety of actual tank conditions and characteristics. The number of records required depends on the type of method result as described in Step 4.

Ambient Conditions. Since SIR methods could be applied as a monthly monitoring approach, it is important to ensure temperature effects are included; therefore, the database includes inventory records representative of the different ambient conditions encountered over the course of a year. One way to achieve this variation is to have an equal number of records from each month of the year from geographic areas that experience a large seasonal temperature change, including frost and snow in the winter. In order to make data collection more practical while still covering different conditions, EPA established these requirements:

- Those months for which the average daily high temperature exceeds the ground temperature, which is 5 feet below the surface at the typical tank depth, by at least 15 degrees Fahrenheit (°F) are defined as hot months. Cold months are those months for which the average daily low temperature is at least 15°F below the ground temperature, which is 5 feet below the surface. All other conditions are defined as mild.

- Limit the proportion of inventory records from mild months to no more than one-third of the total. The remaining records are to be from hot and cold months, with at least 10 percent from each condition. That means you could set up the database with one-third of the inventory records from hot, one third from cold, and one third from mild. Another possibility is 30 percent from mild, 10 percent from cold, and 60 percent from hot.

Size And System Configuration. The inventory records should come from a variety of tank sizes and configurations. The results of the evaluation will be limited by the tank sizes actually incorporated in the database and whether from single tanks or two or more tanks connected by siphon piping. The evaluation database must contain between 30 percent and 75 percent of the inventory records from tanks connected by siphon piping to be evaluated for both configurations.

System Throughput. The inventory records should also come from tanks with a wide range of product throughputs. Although larger throughputs are generally associated with larger tank sizes, some relatively small tanks also have high throughput. Therefore, throughput is a consideration and the evaluator will need to determine appropriateness.

Manual And ATG Measurements (For Traditional SIR). Traditional SIR methods, either manual or ATGS measurements, use liquid level measurements without temperature compensation. Consequently, the evaluation database should have both inventory records generated by both liquid level measuring methods. The percentage of records collected by

ATGS within the database should be reported with the results and should be a minimum of 25 percent of the conclusive results. In some cases, the SIR method is designed for all inventory records to be collected with an ATGS. If you use inventory records from tanks with ATGS measurements, you should check the pattern of results to determine whether the SIR method is achieving better results for the tanks with the ATGS release detection results. If so, these results indicate the quality of the product level measurements as a consideration for the appropriateness of the evaluated SIR method.

4.3 Generate The Evaluation Database—Step 3

The evaluator acquires various inventory records to generate the evaluation database containing up to 45 inventory records. Use a minimum of 24 results from 24 inventory record SIR analyses in the evaluation. More records are included in the database to design the evaluation of the SIR method according to the vendor's intended use, that is, with tanks connected by siphon piping, ATG measurements, various tank sizes, and throughput volumes. You can use the inventory records in multiple databases for different method evaluations, but inventory records may only be used once per database and, therefore, once per evaluation. Similarly, a database can only be used for one evaluation. The average leak rates will vary around the EPA performance standard values of 0.1 and 0.2 gal/hr.

Introduction Of Leaks. Regardless of the type of evaluation performed, the manner in which leaks are introduced is the same.

Record sets taken from tanks determined by other methods to be tight are divided into groups of data sharing similar leak characteristics. One group remains unaltered, representing non-leaking tanks. The remaining groups are assigned a leak rate to be tested. Each monthly record that comprises a group is then modified with an average leak rate, which is within ± 30 percent of the performance standard.

The variations in average leak rates for records within a group are determined by introducing random variations with a uniform distribution. A uniform random distribution is characterized by lower and upper bounds. Variables are drawn with equal probability from all values in the range. The simplest way to do this is to use the random number generators that typically accompany spreadsheet programs. If the test performed involves more than one leak rate, each leak rate is to introduce its own random variation. Subsequent databases generated for testing or re-testing vendors' methods should each use their own random distributions.

For example, Table 1 shows the three leak rates usually tested in an analysis. The values shown in each column were calculated by a computer spreadsheet program for the ranges of 0.035 to 0.065 gal/hr (0.05 ± 30%), 0.07 to 0.13 gal/hr (0.1 ± 30%), and 0.14 to 0.26 gal/hr (0.2 ± 30%).

Table 1. Random Variation Centered About 0.05, 0.1, And 0.2 gal/hr Leak Rates

0.05 gal/hr	0.1 gal/hr	0.2 gal/hr
0.043	0.08	0.14
0.054	0.12	0.18
0.063	0.08	0.19
0.047	0.09	0.19
0.056	0.10	0.24
0.048	0.08	0.16
0.054	0.12	0.15
0.048	0.11	0.19
0.055	0.07	0.20
0.061	0.12	0.19

After establishing the leak rates, the evaluator modifies monthly inventory record to reflect the desired changes. Inventory amounts are recalculated to take into account the induced leak rate. Book inventories, overages, or shortages are refigured by the evaluator based on the modified inventory amounts. The difference between the sum of the inventory values for the month and the original inventory values for the month reflects the total loss introduced to the systems during that period of time. The average loss over the course of the month must be consistent with the induced leak rate evaluated.

If the evaluator determines they are appropriate, modified inventory figures for leaks in tanks can be calculated based on the presumption of a free flow model for product loss in which the leak rate is proportional to the square root of the product height above the hole. Modified inventory amounts can be calculated by the evaluator for all product level measurements above the level of the leak. At and below the leak, the inventory figures are unchanged from the original values.

For all product level measurements above the level of the hole, modified inventory numbers can be determined by the following equation:

$$Inventory_{mod} = Inventory - 24 * L_{max} \sqrt{\frac{h_{product\ level} - h_{leak}}{h_{max} - h_{leak}}}$$

where L_{max} is the maximum leak rate in the UST, h_{max} is the maximum fill height for the period, $h_{product\ level}$ is the measured fill height, h_{leak} is the height of the leak, and 24 is the number of hours in a day. L_{max} is adjusted in such a way as to produce the average leak rate desired for the month. This is also achievable using built-in goal seeking functions found in many spreadsheet programs. Evaluator should round the modified inventory amounts to the same degree that the original inventory amounts were rounded. That means, if the original record reported inventory to the nearest 0.1 gallon, report the modified amount to the nearest 0.1 gallon.

Evaluators determine the location of the leak within the UST system in a similar fashion. Each monthly inventory record is randomly assigned a leak location within the UST system using uniform random distributions. The bounds of the distributions in this instance are actually percentages to be applied to the highest product level reading for a tank for the month in question.

4.4 Submit Database To Vendor—Step 4

To simulate a SIR customer submitting data, submit the database developed in Step 3 above to the SIR vendor for analysis. At the discretion of the evaluator, inventory records should be passed to the vendor individually or in small batches, as opposed to sending the entire database for the evaluation at once. To ensure the vendor does not know which inventory records have induced leak rates, a random number generator can be used in Step 3 to generate a code for each tank record. The code should also allow the evaluator to identify the induced leak in the inventory record. However, the code itself should not contain any information about the evaluation conditions. Similarly, if not already presented in the inventory record, this same random constant will be added to all the totalizer values for that record, too.

4.5 Receive Analysis Results From SIR Vendor—Step 5

The vendor submits to the evaluator a report on each inventory record in the database. This report is in the same format that the vendor would submit results to a SIR client. The report does not need to include all of the inventory tracking features or other services that might be supplied to a commercial client. The report must indicate the results of the release detection evaluation for each tank system. Report the results quantitatively with an interpretation of the result as inconclusive, fail which means a leak is indicated, or pass which means the system is tight.

Often the SIR vendor identifies inconsistencies in the data such as wrong tank chart used or incorrect tank dimensions given. The SIR vendor should identify such features following his or her usual procedures and submit the findings to the evaluator. If the vendor determines the discrepancy precludes an adequate SIR analysis of that tank record, then record that the SIR method identified a problem and exclude the record from the analysis. A minimum of 12 usable records is needed for each of the tight and induced leak conditions separately, for a total of 24 records usable for the evaluation calculations.

If a SIR method is used for tanks connected by siphon piping as well as single tanks, the evaluation database must contain between 30 percent and 75 percent results from tanks connected by siphon piping. The data set must contain a minimum of 12 conclusive results from systems with tanks connected by siphon piping, with a minimum of three results from each leak rate group from systems with tanks connected by siphon piping.

If data problems reduce the evaluation database below the minimum conclusive results specified above for each data type, more inventory records may be added to the database and submitted to the vendor.

4.6 Analyze Data And Report Results To Vendor Of SIR Method—Step 6

The evaluator analyzes the data as described in Section 5. The evaluator reports results to the SIR vendor. As part of the reporting process, the evaluator completes EPA's forms in Appendix B and attaches them to the report.

Section 5: Calculations

In this test procedure, leaks are viewed as product lost from the tank. As a convention, leak rates are positive numbers, representing the amount of product loss per unit time. That means a larger leak represents a greater product loss. Parts of the release detection industry report volume changes per unit time; a negative sign indicates whether product is lost from the tank or a positive sign indicates product is coming into the tank. Leaks here refer to the direction out of the tank and the rate to the magnitude of the flow.

In the case where a SIR method reports estimated leak rates for each inventory record submitted, the evaluation database will consist of induced and reported leak rates for each inventory record. The SIR method will produce conclusive results on n records. A minimum number of 24 conclusive results is required, with at least 12 conclusive results in the no-leak condition group and at least 4 conclusive results in each induced leak category, for a total of 12. The estimation of the performance of the SIR method is based on the number of conclusive results, n.

5.1 Basic Statistics

Use n to calculate the mean squared error (MSE), the bias (B), and the variance of the method as presented below.

Mean Square Error, MSE

$$\mathrm{MSE} = \sum_{i=1}^{n} (L_i - S_i)^2 / n$$

where L_i is the estimated leak rate obtained by the SIR method from the i^{th} record at the corresponding induced leak rate, S_i, with i = 1,…, n.

Bias, B

$$B = \sum_{i=1}^{n} (L_i - S_i)/n$$

B is the average difference between induced leak rates and the estimated leak rates over the number of usable results. Being a measure of the accuracy of the release detection method, B can be either positive or negative.

Variance And Standard Deviation

The variance is obtained as follows:

$$\mathrm{Variance} = \sum_{i=1}^{n} [(L_i - S_i) - B]^2 / (n-1)$$

Standard deviation (SD) is the square root of the variance. This calculation gives a measure of precision of the release detection method.

Test For Zero Bias

To test whether the method is accurate, that is, B is zero, the following statistical test on B calculated above is performed.

Compute the t-statistic:

$$t_B = \sqrt{n}B/SD$$

From the Student's t-table in Appendix A, obtain the critical value corresponding to a t with (n-1) degrees of freedom (df) and a two-sided 5 percent significance level. Denote this value by t_c.

Compare the absolute value of t_B, abs(t_B), to t_c. If abs(t_B) is less than t_c, conclude B is not statistically different from zero; B is negligible. Otherwise, conclude B is statistically significant.

5.2 False Alarm Rate, P(fa)

The normal probability model is assumed for the errors in the vendor's reported leak rates. Using this model, together with the statistics estimated above, allows for the calculation of the estimated P(fa) and the P(d) of a leak of 0.1 or 0.2 gal/hr.

The vendor will supply the criterion for interpreting the results of this SIR method. Often, the leak rate reported by the method is compared to a threshold (Th) and the results interpreted as indicating a leak if the reported leak rate exceeds Th. The P(fa) is the probability the method-estimated leak rate exceeds Th when the tank is tight. Note that by convention, all leak rates representing volume losses from the tank are treated as positive.

P(fa) is calculated by one of two methods, depending on whether B is statistically significantly different from zero.

P(fa) With Negligible Bias

In the case that B is not significant, compute the t-statistic

$$t_1 = Th/SD$$

where SD is calculated as in Section 5.1 above and Th is the method's threshold.

Using the notational convention for leak rates, Th is positive. P(fa) is then obtained from the t-table, using (n-1) df. P(fa) is the area under the curve to the right of the calculated value t_1. In general, t-tables are constructed to provide a percentile, t_a, corresponding to a given number of df and a pre-assigned area alpha (α), under the curve, to the right of $t_{a;}$ see Figure 1 and Table A-1 in Appendix A. For example, should 41 records be usable in the calculations, (n-1) = 40 df and $\alpha = 0.05$, $t_\alpha = 1.684$.

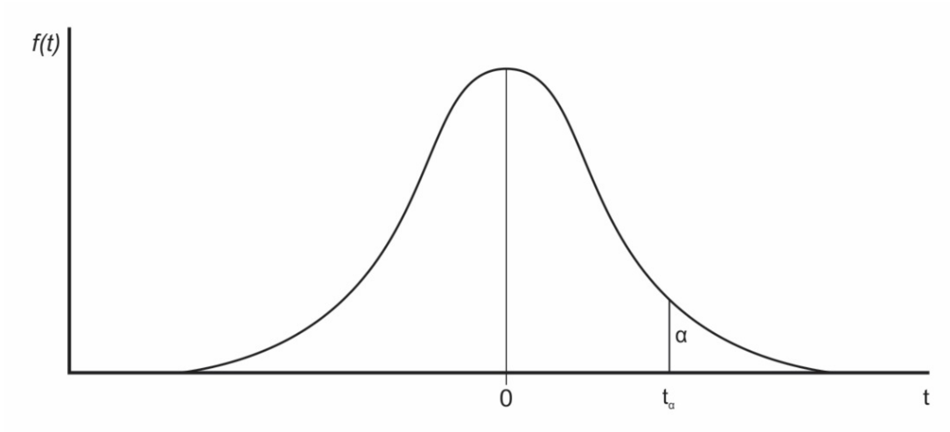


Figure 1. Student's t-Distribution Function.

In this case, however, the area under the curve needs to be determined to the right of the calculated percentile, t_α, with a given number of df. This can be done by interpolating between the two areas corresponding to the two percentiles in Table A-1 on either side of the calculated t-statistic, t_α. This approach is illustrated below.

Suppose the calculated $t_\alpha = 1.7$ and has 40 df. From Table A-1, obtain the following percentiles at df = 40:

t_α	Alpha (α)
1.684	0.05
1.7	X to be determined
2.021	0.025

Calculate X by linearly interpolating between 1.684 and 2.021 corresponding to 0.05 and 0.025, respectively.

$$X = 0.05 - \frac{(0.05 - 0.025)}{(1.684 - 2.021)} \text{ x } (1.684 - 1.7) = 0.049$$

Thus, the P(fa) corresponding to a t_α of 1.7 with 40 df would be 4.9 percent.

A more accurate approach would be to use a statistical software package, for example SAS, SYSTAT, or Microsoft® Excel, to calculate the probability. Another method would be to use a nomograph of Student's t such as the one given by Nelson (1986)[1].

---

[1] Nelson, Lloyd S. 1986. *Technical Aids*, American Society for Quality Control.

P(fa) With Significant Bias

The computations are similar to those in the case of a nonsignificant B with the exception B is included in the calculations, as shown next. Compute the t-statistic

$$t_2 = (\mathrm{Th} - \mathrm{B})/\mathrm{SD}$$

P(fa) is then obtained from the t-table, using (n-1) df. P(fa) is the area under the curve to the right of the calculated value t_2. Note: Th is positive, but B could be either positive or negative.

5.3 Probability Of Detecting A Leak Rate Of Specified Size, P(d)

The P(d) leak rate of 0.1 gal/hr is the probability the estimated leak rate exceeds Th when the true mean leak rate is 0.1 gal/hr. As for P(fa), one of two procedures is used in the computation of P(d), depending on whether B is statistically significantly different from zero.

P(d) With Negligible Bias

In the case that B is not significant or B is zero, compute the t-statistic, t_3 for the specified leak rate as

$$t_3 = (\mathrm{Th} - 0.10)/\mathrm{SD}$$

Next, using the t-table at (n-1) df, determine the area under the curve to the right of the calculated t_3. The resulting number will be P(d).

P(d) With Significant Bias

The procedure is similar to the one above, except that B is introduced in the calculations as shown below. Compute the t-statistic

$$t_4 = (\mathrm{Th} - \mathrm{B} - 0.10)/\mathrm{SD}$$

Next, using the t-table at (n-1) df, determine the area under the curve to the right of the calculated t_4. The resulting number will be P(d).

The P(d) for a leak rate of 0.2 gal/hr can be calculated in the same way, replacing 0.1 by 0.2 in the equations above.

A method will not meet the performance standards if the differences between its reported leak rates and the induced leak rates are too large. It is possible that despite the evidence required to show the database tanks are tight, a tank may actually have a leak. If this is the case, the reported leak rate will presumably include both the underlying leak rate and the induced leak rate, leading to an overestimate of the leak rate and a large B in the direction of the overestimation of the leak rate. In this case, the evaluator might inspect the records that led to large overestimates of the leak rate. It might be possible to check the original inventory record to see if there is evidence of an underlying leak. If so, these data points could be excluded from the analysis or the underlying leak rate added to the induced, if the underlying leak rate can be adequately estimated. After this adjustment to the data, re-analysis might lead to the conclusion that the method is adequate.

Reporting P(fa) And P(d)

In order to meet the EPA performance requirements, the P(fa) must be 5 percent or less. In making this determination, round the calculated P(fa) to the nearest whole percent. Similarly, to meet EPA's performance requirements, the P(d) must be at least 95 percent, again rounded to the nearest whole percent. Depending on the performance level, the P(d) may be calculated for either a leak rate of 0.1 gal/hr or 0.2 gal/hr. If a method meets the requirement for detecting a leak rate of 0.1 gal/hr, it will meet the requirement for 0.2 gal/hr. Thus, the calculations for a leak rate of 0.2 gal/hr would normally be required only if the method did not meet the detection requirement for the smaller leak rate. Appendix B contains the reporting forms and instructions for filling out the forms.

5.4 SIR Method Performance Parameters For Single Tanks And Tanks Connected By Siphon Piping

Calculate the overall P(d) and P(fa) for the entire data set used in the evaluation per the procedures in Sections 5.1, 5.2, and 5.3 to determine whether the combined data meets the 95 percent and 5 percent performance standard. If the combined data does not meet the performance standard, then the SIR method may not be used on single tanks or tanks connected by siphon piping. If the combined data meets the 95 percent and 5 percent performance standard, then calculate the mean and SD separately for the single and siphoned groups. Also, test for zero bias for each group. This can be done by following the same process using the equations in Sections 5.1, 5.2, and 5.3 on each group separately.

5.4.1 Comparison Of Single Tanks Versus Tanks Connected By Siphon Piping Test Results

Standard Deviation Comparison

Use a two-sample F-test to test whether the variances of the two groups are equal. Calculate as

$$F = (SD_1/SD_2)^2$$

where SD_1 and SD_2 are the standard deviations calculated from the two groups. In forming the F ratio, use the SD with the larger calculated value in the numerator. Compare the calculated value of F to the 95th percentile of an F-distribution with $(n_1 - 1)$ df in the numerator (corresponding to SD_1) and $(n_2 - 1)$ df in the denominator (corresponding to SD_2). The sample sizes are n_1 and n_2, respectively. If the calculated value of F is less than the tabled value, there is no significant evidence that the two population variances are different. In this case, there is justification for using the method on both single tanks and tanks connected by siphon piping.

Bias Comparison

If the SDs of the single tanks and tanks connected by siphon piping groups are not significantly different, test to see if the Bs are different for the two groups of tank configurations. Use a two-sample t-test to test whether there is any significant difference in the Bs of the two groups. Compare as

$$t_{bp} = \frac{B_1 - B_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

where Sp is the pooled SD of the two groups and is calculated

$$S_p = \sqrt{\frac{(n_1-1)SD_1^2 + (n_2-1)SD_2^2}{n_1+n_2-2}}$$

Compare t_{bp} to a two-sided 5 percent critical value from a t-distribution with ($n_1 + n_2$-2) df (α = 0.05). If the absolute value of t_{bp} does not exceed the critical value, then there is no evidence that the B is different for single tanks compared to tanks connected by siphon piping. In this case, use of the method for both types of tanks is justified.

If the SDs and Bs of single tanks versus tanks connected by siphon piping are not significantly different, then the SIR method is not affected by siphoning. Therefore, it is not necessary to calculate the P(d) and P(fa) separately for each. It is only necessary to report the overall P(d) and P(fa) for the combined data. There will be only one volume limitation applicable to both single tanks and tanks connected by siphon piping. Volume limitation for SIR methods is determined in Section 6.

However, if either the SDs or Bs of single tanks versus tanks connected by siphon piping groups are significantly different, that is, the calculated value of F exceeds the tabled value or the absolute value of t_{bp} exceeds the percentile from the t-table, then there is evidence that the performance of the method is affected by siphoning. In this case, continue with the computation of the P(d) and P(fa) separately for the single and siphoned tank groups using the following process.

5.4.2 Probability Of A False Alarm, P(fa), For Single Tanks And Tanks Connected By Siphon Piping, Separately

The P(fa) is the probability that the estimated leak rate will exceed the Th for indicating a leak when, in fact, the tank is actually tight. Generally, if the calculated leak rate exceeds a specified Th, for example 0.12 gal/hr, the SIR method determines the tank is leaking. In the below equation, if Th denotes the threshold for indicating a leak, B denotes the bias of the method, and SD denotes the standard deviation, then the P(fa) can be written as:

$$P(fa) = P\{t > (Th - B)/SD\}$$

where the probability is calculated from a one-sided t-distribution with (n-l) df. For example, if there are 28 single tank records and 13 tank records of tanks connected by siphon piping, the degrees of freedom would be 27 for the P(fa) for single tanks and 12 for the P(fa) for the tanks connected by siphon piping. This formula assumes that the errors are approximately normally distributed. If it was determined that the bias is not significantly different from zero, then B is taken to be zero.

5.4.3 Probability Of Detecting A Leak Rate Of R Gallon Per Hour, P(d), For Single Tanks Versus Tanks Connected By Siphon Piping, Separately

The P(d) is the probability that the method will correctly identify a leak of specified size. In general for a leak rate of size R, P(d) is given by:

$$P(d) = P\{t > (Th - R - B)/SD\}$$

where Th, B, and SD are as before. The probability is calculated from the one-sided t-distribution with (n-1) df.

Assume that the method does not perform equivalently on single tanks and tanks connected by siphon piping. If both the single tanks and tanks connected by siphon piping groups meet the 95 percent and 5 percent performance standard, then the method may be used on both single tanks and tanks connected by siphon piping. However, the evaluator should report the difference in performance. Report the P(d) and P(fa) separately for single tanks and tanks connected by siphon piping. The evaluator should not report the overall P(d) and P(fa) for the combined data because the method does not work equivalently on single tanks and tanks connected by siphon piping.

If only one group meets the 95 percent and 5 percent performance standard, then limit use of the method to the group of tanks–either single or connected by siphon piping–which meets the performance standards.

Report the P(d) and P(fa) for the group that meets the criteria. The evaluator should not report the overall P(d) and P(fa) for the combined data because the method is limited to one group of tanks.

5.5 Minimum Threshold And Minimum Detectable Leak Rate

Use test results to calculate the minimum threshold ($Th_{5\%}$) a SIR method can detect. The results of this section are based on the average performance of the method on the data used in the evaluation. This calculation assumes B and precision of the method can be estimated from the evaluation data. Other data sets may exhibit more or less variability, and so the method might do better or worse on individual inventory records.

The vendor will calculate Th for each set of data analyzed and report the results. The evaluation data of induced and reported leak rate data can be used to determine a calculated $Th_{5\%}$ that would result in a P(fa) of 5 percent. Being data specific, this threshold may not be the same as the Th value used by the vendor. Therefore, the vendor may not use the same threshold for all tank records. The following demonstrates the approach for computing $Th_{5\%}$.

Solve below for $Th_{5\%}$.

$$P(fa) = P[t > \left(\frac{Th_{5\%} - B}{SD}\right)] = 0.05$$

If B is not statistically significant, then replace B with zero. From the t-table in Appendix A with (n-1) df obtain the 5th percentile. Denote this value by $t_{5\%,(n-1)}$. Solving the equation above for $Th_{5\%}$ yields

$$\frac{Th_{5\%} - B}{SD} = t_{5\%,(n-1)}$$

or

$$Th_{5\%} = t_{5\%,(n-1)}(SD) + B$$

In the case of a non-significant B this would be

$$Th_{5\%} = t_{5\%,(n-1)}(SD)$$

With the evaluation data, the minimum detectable leak rate, $R_{5\%}$, corresponding to a P(d) of 95 percent and a calculated threshold $Th_{5\%}$, can be calculated by solving the below equation for $R_{5\%}$.

$$P[d(R_{5\%})] = Pr\left[t > \frac{Th_{5\%} - R_{5\%} - B}{SD}\right] = 0.95$$

where $Th_{5\%}$ is the threshold corresponding to a P(fa) of 5 percent, as calculated above, and B is B estimated for the method.

Solving this equation is equivalent to solving

$$\frac{Th_{5\%} - R_{5\%} - B}{SD} = -t_{5\%,(n-1)}$$

or

$$R_{5\%} = t_{5\%,(n-1)}(SD) + Th_{5\%} - B$$

which, after substituting $t_{5\%,(n-1)} \times SD$ for ($Th_{5\%}$-B), is equivalent to

$$R_{5\%} = 2Th_{5\%} - 2B$$

Substitute zero for B in all calculations when B is not statistically significant. Otherwise, use the value of B estimated from the data.

Thus, the minimum detectable leak rate with P(d) of 95 percent is twice the $Th_{5\%}$ determined to give a P(fa) of 5 percent minus twice B.

In summary, based on the evaluation data of induced and reported leak rates, the $Th_{5\%}$ and the $R_{5\%}$ are calculated as shown below:

If B *is not* statistically significant:

For a P(fa) of 5%	$Th_{5\%} = t_{5\%,(n-1)}(SD)$
For a P(d(R)) of 95%	$R_{5\%} = 2Th_{5\%}$

If B *is* statistically significant:

For a P(fa) of 5%	$Th_{5\%} = t_{5\%,(n-1)}(SD) + B$
For a P(d(R)) of 95%	$R_{5\%} = 2Th_{5\%} - 2B$

Note: You can use other significance levels by substituting the appropriate values from the statistical table.

The calculated results represent average results obtainable with data of the quality used in the evaluation. In particular, $Th_{5\%}$ depends on the variability of the inventory records being approximately constant.

Reporting Minimum Detectable Leak Rate

In order to meet the EPA performance requirements, the minimum detectible leak rate must be one half of the performance level, either 0.1 gal/hr for a 0.2 gal/hr evaluation or 0.05 gal/hr for a 0.1 gal/hr evaluation. Appendix B contains the reporting forms and instructions for filling out the forms.

The minimum detectible leak rate – MDL – should not be confused with the threshold leak rate. The threshold leak rate is established based upon a non-leaking tank model in which the average leak rate is zero and the variance is the same as that of the data being evaluated. The threshold leak rate is set so as to provide a Pfa of 5 percent or less, and an UST is said to be leaking whenever the calculated leak rate exceeds that threshold. The curve for the MDL is the mirror image of that for the non-leaking tank, rotated about $Th_{5\%}$; see Figure 2. This provides the minimum leak rate the SIR method is capable of achieving for a Pd of 95 percent. The MDL should, ideally, be less than or equal to the EPA performance standard against which the UST is being tested. If the MDL exceeds that standard, an inconclusive test result is possible.

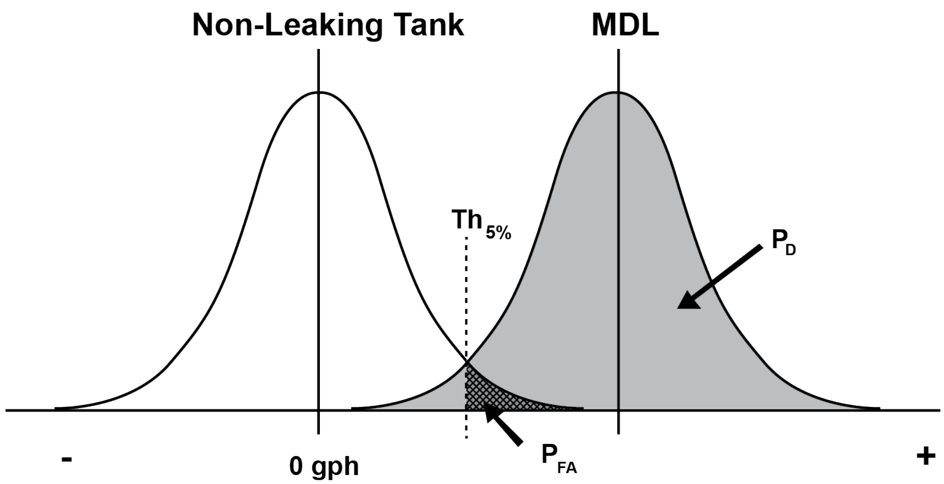


Figure 2. Minimum Detectable Leak Rate

The MDL is calculated as the leak rate that has a Pd of 95 percent (shaded region) for a threshold, $Th_{5\%}$ having a specified false alarm rate of 5 percent (crosshatched region). An UST system test with an MDL that exceeds the EPA performance standard may be judged inconclusive.

Pass – Any SIR test result in which the minimum detectible leak rate is less than or equal to the EPA performance standard against which the test result is being compared, and the calculated leak rate is less than the threshold, is regarded as a pass. The system is judged to be tight.

Section 6: Interpretation

The results reported are valid for the factors considered in the evaluation. These were chosen to represent the factors thought to be most important in influencing performance of the method as a release detection method. Additional factors such as theft could lead to an increased P(fa). Vendors are encouraged to report, and most do, an indication of the variability of the results as well as the estimated leak rates and other factors.

6.1 Performance Parameters Results

SIR methods can be conducted monthly, depending on the release detection option. Each method determines the amount of data needed and is not tied to the frequency of applying the SIR method. The performance standard differs according to the frequency with which the release detection method is applied.

The relevant performance measures for showing that a SIR method meets EPA standards are the P(fa) and P(d) for a leak rate of 0.2 gal/hr in monthly records. The estimated P(fa) can be compared with EPA's standard of P(fa) not to exceed 5 percent. In general, a lower P(fa) is preferable, since it implies the chance of mistakenly indicating a leak on a tight tank is less. However, reducing the P(fa) will generally reduce the chance of detecting a leak. The P(d) for the applicable leak rate standard, generally 0.2 gal/hr, must be at least 95 percent. The P(d) generally increases with the size of the leak. A higher estimated P(d) for a specified leak rate means there is less chance of missing a leak of that size. However, its P(fa) should be no more than 5 percent with 95 percent confidence.

6.2 Limitations On Results

Also include any limiting conditions specified by the vendor for use of the SIR method as limitations on the results form. The applicability of the SIR method for single tanks and tanks connected by siphon piping is further described by the system capacity and throughput. As noted before, system capacity may influence the results of a statistical inventory analysis. Similarly, throughput may affect the performance of the method.

If tanks connected by siphon piping are included, then the SIR program is limited to the number of tanks in the 80^{th} percentile plus one. The tank records are to be ordered by the number of tanks connected by siphon piping from smallest to largest, starting with the single tank records. The 80^{th} percentile is the tank record such that 80 percent of the tank records have less than or equal to this number of tanks in the record. For example, a data set with 41 conclusive records has 28 single tank records, 4 two tank records, and 9 three tank records. Take 80 percent of 41 to get 32.8. Fractions are moved to the next integer, in this case 3. Counting from smallest to largest, the 33^{rd} record has 3 tanks connected by siphon piping. Therefore, limit the method to UST systems connected by siphon piping that have no more than four tanks.

When comparing the SDs of the results from single tanks and tanks connected by siphon piping, if the calculated value of F is less than the tabled value, there is no significant evidence that the two population variances are different. In this case, there is justification for using the method on both single tanks and tanks connected by siphon piping.

When comparing the Bs of the results from single tanks and tanks connected by siphon, if the absolute value of t_{bp} does not exceed the critical value, then there is no evidence that the bias is different for single tanks compared to tanks connected by siphon. In this case, use of the method for both types of tanks is justified.

If the SDs and Bs of single tanks versus tanks connected by siphon piping are not significantly different, then the SIR method is not affected by being connected by siphon piping. Therefore, it is not necessary to calculate the P(d) and P(fa) separately for each. It is only necessary to report the overall P(d) and P(fa) for the combined data. There will be only one volume limitation, which applies to both single tanks and tanks connected by siphon piping.

However, if either the SDs or Bs of single verses tanks connected by siphon piping groups are significantly different, that is the calculated value of F exceeds the tabled value or the absolute value of t_{bp} exceeds the percentile from the t-table, then there is evidence that the performance of the method is affected by tanks connected by siphon piping. In this case, calculate the P(d) and P(fa) separately for the single and tanks connected by siphon piping groups.

If only one group meets the 95 percent and 5 percent performance standard, then limit use of the method to the group of single tanks or tanks connected by siphon piping, which meet the performance standards. Report the P(d) and P(fa) for the group that meets the criteria. The evaluator should not report the overall P(d) and P(fa) for the combined data because the method is limited to one group of tanks.

6.2.1 SIR Method System Size Limitations

To justify extrapolation to larger tank sizes, the results for small and large tanks must be similar. The distribution of tank sizes in the database should be as nearly uniform as practical. The database should not emphasize small tanks. Test data should represent the population of tanks for which the method is intended to be used. The results of an evaluation can be extended to tanks 50 percent larger than the 80th percentile of the tank sizes used in the evaluation data set, if the method is not affected by increasing volume.

Determination of whether tank size affects the performance of the SIR method can be conducted on the entire database as a whole, if the method is found to perform equivalently on single tanks and tanks connected by siphon piping. In this case, there will be only one maximum volume limitation that is applicable to both single tanks and tanks connected by siphon piping.

However, if the procedures reveal that the method does not perform equivalently on single tanks and tanks connected by siphon piping, then the volume's effect on the SIR method's performance must be determined separately for single tanks and tanks connected by siphon piping. In this case, there will be two maximum volume limitations: one that is applicable to single tanks and the other for tanks connected by siphon piping. In addition, if the calculations reveal that the method meets the 95 percent and 5 percent performance standard for only one group of tanks, for example single tanks, then the procedures for determining the effect of volume on performance is limited to single tanks.

Order the tank records by volume from least to greatest and determine the various percentiles. The volume of tanks connected by siphon piping record is the total volume of the tanks connected by siphon piping. Report the smallest, 25^{th}, 50^{th} (median), 75^{th}, 80^{th} percentile, and the largest tank size on the results form. To find a tank size for a given percentile, take the percentile as a percentage of the sample size and count up from the smallest tank until that number of tank records is reached. For example, for the 25^{th} percentile, with n=40 records, take 25 percent of 40 to get 10. Fractions are moved up to the next integer. The 25^{th} percentile is the 10^{th} tank size in the set of ordered tank sizes, counting from smallest to largest. If the result of taking a percent of the sample size is not an integer, use the next larger integer.

In particular, the 80^{th} percentile determines a limitation on tank size. If there are 40 conclusive records, the 80^{th} percentile is the 33^{rd} tank size counting from the smallest to the largest. If a different number of records is used, the 80^{th} percentile is the tank size corresponding to the integer greater than or equal to 0.8n, where n is the number of records, again counting from the smallest tank size to the largest.

If the method is not adversely affected by increasing tank volume, then the maximum tank size limitation is 1.5 times the 80^{th} percentile of tank sizes used in the evaluation. If the method is adversely affected by increasing tank volume, then the maximum tank size limitation is reduced to the smaller of the largest tank in the evaluation, or 1.25 times the 80^{th} percentile.

To justify extrapolation to larger tank sizes, the results for small and large tanks must be similar. To make this comparison, divide the data records into two groups based on volume. The two groups should be of nearly equal size, but if there are many records at one tank size, for example 10,000 gallons, it may be impossible to make the two groups exactly equal.

For example, in a database consisting of 40 conclusive records, suppose 28 are single tank records and 12 are from tanks connected by siphon piping. Suppose it was determined that the method does not perform equivalently on single tanks and tanks connected by siphon piping, but it meets the 95 percent and 5 percent performance standard for both types of tanks. Therefore, the effect that increasing volume has on the performance of the method must be determined separately for single tanks and tanks connected by siphon piping. Divide the 28 single tanks into two groups based on small and large volume as close to the median as possible. Also, divide the 12 records of tanks connected by siphon piping into two groups based on volume as close as possible to the median. The volume of a system with tanks connected by siphon piping is the total volume of the tanks that are connected by siphon piping.

Compare the SDs and Bs of the large and small tanks separately following the same procedures as in Sections 5.1 and 5.2. If the calculated value of F is less than the tabled value, there is no significant evidence that the two population variances are different. In this case, there is justification that the method is not affected by increasing volume. If the absolute value of t_{bp} does not exceed the critical value, then there is no evidence that the B is different for small tanks compared to large tanks. In this case, there is justification that the method is not affected by increasing volume.

If the SDs and Bs of large versus small volume groups are not significantly different, then the SIR method is not affected by increasing volume. In this case, extrapolation to 1.5 times the 80th percentile of tank sizes is justified.

If either the SDs or Bs of large versus small volume tanks are significantly different, that is the calculated value of F exceeds the tabled value or the absolute value of t_{bp} exceeds the percentile from the t-table, then there is evidence that the performance of the method is affected by volume. In this case, determine whether the method is adversely affected by increasing volume. Compare the SDs calculated for the large and small volume groups.

If the SD of the small volume group is greater than the SD of the large volume group, then the method is not adversely affected by increasing volume. In this case, the maximum size limitation is 1.5 times the 80th percentile. On the other hand, if the SD of the large volume group is greater than the SD of the small volume group, then the method is adversely affected by increasing volume. In this case, the maximum tank size limitation is reduced to the smaller of the largest tank in the evaluation or 1.25 times the 80th percentile.

6.2.2 SIR Method Throughput Limitations

The volume of product dispensed from the tank in a month is referred to as the monthly throughput. This is an important factor because the higher the monthly throughput, the fewer and shorter the periods of quiescence for a tank. This would affect the time needed to get a valid test, the relative noise levels of the test, and the amount of data available for the test. The evaluator should design the database to encompass the throughput variation discussed with the vendor. To the extent practical, the test database should represent the distribution of monthly throughputs for the population of tanks for which the system is used. The distribution of throughputs should be approximately uniform.

Determine the monthly throughputs for the tank records in the database. If a test is for less than a month, determine the throughput for the duration of the test from the record and scale up to one month.

Calculate the maximum allowable monthly throughput as 1.5 times the 80th percentile of the throughputs in the evaluation data. Calculate the monthly throughput for each record in the evaluation. For records that are less than one month, determine the recorded throughput for that record. Divide the throughput by the number of days in the record and use fractions if appropriate, then multiply by 31 to get the equivalent monthly throughput. Order these monthly throughputs from least to greatest and compute the 80th percentile. Multiply this by 1.5 to determine the throughput limit for the system.

To justify the extrapolation to the larger throughputs, the results for smaller throughputs and larger throughputs must be similar. To make this comparison, divide the data records into two groups based on monthly throughput. The two groups should be of nearly equal size.

Calculate the mean and SD separately for the two throughput groups. Do this by using the formulas in Section 5 separately on the two throughput groups. Use a two-sample F test to test whether the variances of the two groups are equal using the equation in Section 5.4.

In forming the F ratio, use the SD with the larger calculated value in the numerator. Compare the calculated value of F to the 95th percentile of an F-distribution with (n_1-1) df in the numerator (corresponding to SD_1) and (n_2-1) df in the denominator (corresponding to SD_2). The sample sizes are n_1 and n_2, respectively. If the calculated value of F is less than the tabled value, there is no significant evidence that the two population variances are different. In this case, there is justification for extrapolating to throughputs larger than those in the data base.

If the calculated value of F exceeds the tabled value, the two variances are significantly different at the 5 percent significance level. This is evidence that throughput affects performance of the system. Assuming that the SD for the larger throughputs is the larger, this indicates that the performance of the system is worse for higher throughput tanks. The throughput limit should be reduced to the smaller of the largest throughput in the data or 1.25 times the 80th percentile.

If the SD are not significantly different, test to see if the B is different for the two groups of throughputs. Use a two-sample t-test to test whether there is any significant difference in the B using the equation for t in Section 5.4. Note that Sp is the pooled SD of the two groups and is calculated as in Section 5.4.

Compare t_b to a two-sided 5 percent critical value from a t-distribution with (n_1+n_2 - 2) df. If the absolute value of t_b does not exceed the critical value, then there is no evidence that the B is different for different throughputs. In this case, extrapolation to 1.5 times the 80th percentile of throughputs is justified.

If the absolute value of t_b exceeds the percentile from the t-table, then the system has a significantly different B for the different throughputs.

If you find a significant difference in the performance for different throughputs, note this fact and the reduced throughput limit in the other limitations section of the results form.

Section 7: Reporting Of Results

Appendix B contains templates for an evaluator to develop a standard report for SIR methods. There are three parts to the standard report and each template includes instructions for how to complete it.

- The first part is *Results Of EPA Standard Evaluation*. This is an executive summary of the findings. It provides each tank owner or operator who uses this method of release detection with documentation that the method meets EPA's standards. Structure the report so an owner or operator can easily reproduce it. Report the limitations of the evaluation results on the *Results Of EPA Standard Evaluation* form. This documents that the results are valid under conditions represented by the test conditions. Section 4.2 describes the summary of the test conditions that should be reported as limitations on the results form.

- The second part is the *Description Of Statistical Inventory Reconciliation Method*. The evaluator, with the assistance of the vendor, completes this description form, which provides supporting information on data requirements and approach of the statistical inventory method.

- The third part is a *Reporting Form For Test Results*. The evaluator completes these tables, which summarize the test results obtained from the method and indicate the induced leak rates added to each inventory record.

If the evaluator performs the optional calculation of the minimum threshold, report it to the vendor in a separate section of the report, distinct from the standard report. This will allow a user to identify the parts of the standard report quickly, while still having the supplemental information available if needed.

Appendix A
Definitions And Student's t Distribution

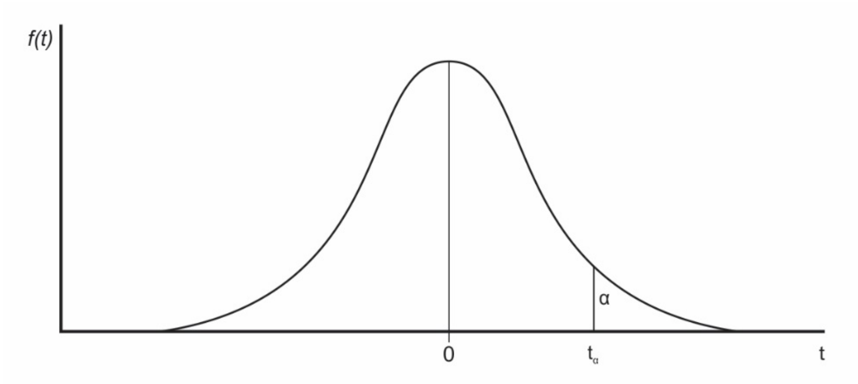
Definitions of terms used throughout the test procedures and the Student's t distribution table in Table A-1 are presented here. For more information on the statistical approach and relationships between the statistics calculated in these test procedures see *General Guidance For Using EPA's Standard Test Procedures For Evaluating Release Detection Methods*.

Accuracy:	The degree to which the calculated leak rate agrees with the induced leak rate on the average. If a method is accurate, it has a very small or zero bias.
Calculated Leak Rate, R:	A positive number, in gallon per hour (gal/hr), estimated by the SIR method and indicating the amount of product leaking out of the tank. A negative leak rate could result from water leaking into the tank, miscalibration, or other causes.
False Alarm:	Declaring that a tank is leaking, when in fact it is tight.
Induced Leak Rate, S:	The actual leak rate, in gal/hr, introduced in the evaluation data sets, against which the results from a given method will be compared.
Mean Squared Error, MSE:	An estimate of the overall performance of a test method.
Method Bias, B:	The average difference between calculated and induced leak rates. It is an indication of whether the SIR method consistently overestimates with a positive bias or underestimates with a negative bias the actual leak rate.
Minimum Detectible Leak Rate, MDL:	A calculated value based on the distribution of the tank data. The MDL is the leak rate that has a Pd of 95 percent for a threshold, $Th_{5\%}$, that has a specified false alarm rate of 5 percent. The minimum detectible leak rate is equivalent to twice the difference between the threshold determined to give a false alarm of 5 percent and the bias (if any); i.e., $MDL = 2(Th_{5\%} - B)$
Precision:	A measure of the test method's ability in producing similar results, that is in close agreement, under identical conditions. Statistically, the precision is expressed as the standard deviation of these measurements.
Probability Of Detection, P(d):	The probability of detecting a leak rate of a given size, R gal/hr. In statistical terms, it is the power of the test method and is calculated as one minus beta (β), where beta is the probability of not detecting or missing a leak rate R. Commonly the power of a test is expressed in percent, as 95 percent.
Probability Of False Alarm, P(fa):	The probability of declaring a tank leaking when it is tight. In statistical terms, this is also called the Type I error and is denoted by alpha (α). It is usually expressed in percent as 5 percent.
Root Mean-Squared Error, RMSE:	An estimate of the overall performance of a test method.

Threshold, Th: The leak rate above which a method declares a leak. It is also called the threshold of the method.

Variance: A measure of the variability of measurements. It is the square of the standard deviation.

Table A-1. Percentage Points Of Student's t Distribution



df	α = .10	α = .05	α = .025	α = .010	α = .005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.333	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704

df	α = .10	α = .05	α = .025	α = .010	α = .005
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
inf.	1.282	1.645	1.960	2.326	2.576

Appendix B
Reporting Forms

Appendix B contains three blank forms. After you complete them, these forms constitute a standard report. Each blank form includes instructions on how to and who should complete it. Listed below are the three forms and who is responsible for completing each.

1. Results Of EPA Standard Evaluation – Statistical Inventory Reconciliation (SIR) Method. The evaluator is responsible for completing this form at the end of the evaluation.

2. Description Of Statistical Inventory Reconciliation Method. The evaluator, assisted by the vendor, is responsible for completing this form by the end of the evaluation.

3. Reporting Form For Test Results – Statistical Inventory Reconciliation Method. The evaluator is responsible for completing this form. The statistician analyzing the data may, however, complete this form. A blank form can be developed on a personal computer, the database for a given evaluation generated, and the two merged on the computer. The form can also be filled out manually. The evaluator and the vendor's test results provide input for the form.

After completing the evaluation, the evaluator collates all the forms in the order listed above into a single standard report. In those cases where the evaluator performed optional calculations, attach the results to the standard report. There is no reporting requirement for these calculations.

Distributing The Evaluation Test Results

The organization performing the evaluation prepares a report describing the results of the evaluation and provides the report to the vendor. The report consists of the forms in Appendix B. The first form is three pages, reports the results of the evaluation, and will be distributed widely. Each tank owner or operator who uses this method of release detection will receive a copy of this form. Owners or operators must retain a copy of this form as part of record keeping requirements. Regulators who must approve release detection methods for use in their jurisdictions will also receive a copy of this form.

The evaluator submits the report, comprised of all forms in Appendix B, to the SIR method vendor. The vendor may distribute the report to regulators who want to see the evaluation data and results. The vendor may also distribute the report to SIR method clients who want additional information before deciding to select a particular release detection method.

The evaluator reports the optional part of the calculations, if done, to the SIR method vendor. This is primarily for the vendor's use in understanding the details of the performance and suggesting how to improve the method. The vendor decides whether to distribute the optional part of the calculations and to whom.

The SIR method evaluator provides the report to the vendor. The vendor is responsible for distributing the report to tank owners or operators and to regulators.

Results Of EPA Standard Evaluation
Statistical Inventory Reconciliation Method

Instructions

The evaluator completes this form after evaluating the method. This form contains the most important information relative to the method evaluation. Complete all items and check the appropriate boxes. If a question is not applicable to the method, write *NA* in the appropriate space.

This form consists of five main parts.

1. Method Description
2. Evaluation Description And Results
3. Test Conditions During Evaluation
4. Limitations On The Results
5. Certification Of Results

Method Description

Indicate the commercial name of the statistical inventory reconciliation method and the version, as well as the vendor's name, address, and telephone number. Since the method is based on software programs that may be updated, the date reported on the last page of this form is considered the date of the version. If the vendor is not the party who developed and uses the method, then indicate the home office name and address to contact for updates.

Evaluation Description And Results

The vendor supplies the criterion for declaring a tank to be leaking. Indicate the leak rate or other criterion in the space provided.

The SIR method may not be able to make a determination of the leak status on some of the inventory records. This may be due to inadequacies in the data or to marginal results that are difficult to interpret. Summarize the reported results by filling in each box of the table on page 1 with the number of inventory records in each category. Calculate and report the totals. The category *inconclusive* is for records that were analyzed but did not give a conclusive result. Vendors may refer to these cases under a variety of terms. The category *not analyzed* is for records where the method identified a data problem and were consequently judged unacceptable for analysis. These are removed from the evaluation database.

The percentages of records that were inconclusive, that is could not be determined to be tight or leaking by the method, are to be calculated and reported separately among the records from tight tanks, among those with induced leaks, and among all tanks.

P(fa) is the probability of false alarm as calculated in Section 5.2.

P(d) is the probability of detecting a leak of specified size as calculated in Section 5.3.

If the P(fa) is 5 percent or less and if the P(d) is 95 percent or greater, then check the *Yes* box. Otherwise, check the *No* box. Cross out the leak rate for which the performance estimates do not apply.

The minimum detectable leak rate is obtained from the calculations in Section 5.5.

Test Conditions During Evaluation

Summarize the conditions of the database in this part. Report the distribution of tank sizes in the categories indicated by inserting the number of records for each size class of tank.

Report the distribution of throughputs for the tank records in the database. Calculate the 25^{th}, 50^{th}, and 75^{th} percentiles of the monthly throughputs in the evaluation database and enter the results on the form.

Report the distribution of the data records by season of year.

Limitations On The Results

The size in gallons of the largest tank to which these results can be applied is calculated as 1.5 times the 80^{th} percentile of the tank sizes used in the data for the evaluation.

The minimum record length needed by the method to achieve the performance results reported here is reported as a limitation on the minimum amount of data. This is the average number of usable days of inventory records in the evaluation database.

Certification Of Results

The person who directed the evaluation work provides his or her name and signature, as well as the name, address, and telephone number of the organization performing the evaluation.

Results Of EPA Standard Evaluation
Statistical Inventory Reconciliation Method

This form tells whether the statistical inventory reconciliation (SIR) method described below complies with requirements of the federal underground storage tank (UST) regulation. The evaluation was conducted by the SIR method vendor or a consultant to the vendor according to EPA's *Standard Test Procedures For Evaluating Release Detection Methods: Statistical Inventory Reconciliation*. The full evaluation report also includes a form describing the method and a form summarizing the test data.

UST owners or operators using this release detection method should keep this form to prove compliance with the federal UST regulation. Owners or operators should check with regulatory authorities to make sure this form satisfies their requirements.

1. Method Description

Name ____________________

Version ____________________

Vendor Information

Vendor ____________________

Street address ____________________

City ____________________

State ____________________

Zip ____________________

Phone number ____________________

2. Evaluation Description And Results

a. Vendor's threshold ____________________ gal/hr

or vendor's criterion ____________________

b. Based on the test results, does the method meet the federal performance standards established by EPA of 0.1 gal/hr (or 0.2 gal/hr) at P(d) of 95 percent and P(fa) of 5 percent?

☐ Yes ☐ No

Probability of false alarms, P(fa), based on the vendor's threshold is__________%

Probability of detection, P(d), is ________%

The minimum detectable leak rate is ______gal/hr

This is valid for a leak rate of (check one):

☐0.1 gal/hr ☐0.2 gal/hr

		Reported Results				
		Tight	Leak	Inconclusive	Total Analyzed	Not Analyzed
Actual	Tight					
	Induced Leak					
	Total					

c. The proportions of inventory records reported inconclusive are:

__________ % among tight tanks

__________ % among leaking tanks

__________ % among all tanks

3. Test Conditions During Evaluation

a. The data evaluation set includes data from tanks of the following sizes.

Tank Size (Gallons)	<5,000	5,000-10,000	10,000-15,000	>15,000	**Total Number Of Records**
Number Of Records					

b. The tanks had various monthly throughputs.

Percentile Of Records	25	50 (Median)	75
Monthly Throughput (Gallons)			

c. The data included:

_______ Records during hot weather months.

_______ Records during mild weather months.

_______ Records during cold weather months.

---

4. Limitations On The Results

The performance estimates above are only valid when:

- The method has not been substantially changed.
- The vendor's instructions for using the method are followed.
- The tank is no larger than _______ gallons.
- Any tanks connected by siphon piping used in the evaluation is no larger than ________ gallons.
- The monthly throughput is no more than _________ gallons per month.
- The data records cover ________days or more.
- Other limitations specified by the vendor or determined during testing ____________________ ____ ____

> **Safety disclaimer: The test procedure only addresses the issue of the method's ability to detect leaks. It does not test data recording equipment for safety hazards.**

Additional explanations or comments

____
____
____
____
____
____
____
____

5. Certification Of Results

I certify that the statistical inventory reconciliation method was applied according to the vendor's instructions. I also certify that the evaluation was performed according to the standard EPA test procedures for statistical inventory reconciliation and that the results presented above are those obtained during the evaluation.

Printed name	Organization performing evaluation
Signature	Street address
Date	City, state, zip
Telephone number	

Description Of Statistical Inventory Reconciliation Method

Instructions

The evaluator, with assistance from the vendor, completes this form after evaluating the method. This form provides supporting information on the data requirements and approach of the statistical inventory method.

To minimize the time needed to complete this form, we provide the most frequently expected answers to the questions. For questions that require additional information, provide explanations in the area adjacent to the question. Use the answer that applies most often or in typical cases.

The form consists of five parts.

1. General Information
2. Data Requirements
3. Reporting Of Leak Status
4. Identification Of Causes For Discrepancies
5. Exceptions

In the first part, provide the commercial name and other identifying information. Since software is often updated, give the software's version and date applied to the method in the evaluation.

For the four remaining parts, check appropriate boxes. Check more than one box per question if it applies. If you check a box *Other*, explain or specify. Use additional white space for additional explanation.

Description Of
Statistical Inventory Reconciliation Method

This section describes important aspects of the statistical inventory reconciliation method. It does not provide a complete description of the principles behind the SIR method and associated computer software.

1. General Information

Method Description

Name ______________________

Version ______________________

Revision number ______________________

Date ______________________

Vendor Information

Vendor ______________________

Street address ______________________

City ______________________

State ______________________

Zip ______________________

Phone number ______________________

2. Data Requirements

a. Vendor Specified Data Form

☐ Yes ☐ No

b. Method Of Recording Inventory Data

☐ Manually, on provided forms
☐ Manually, no forms provided
☐ Hand entered into a computer
☐ Direct entry from ATGS
☐ Other:

____

c. Vendor Recommended Number Of Daily Records

☐ 30 daily records
☐ 60 daily records
☐ Other:

____

d. What is the required number of usable daily inventory records necessary to detect the indicated leak rate (gal/hr) with 95 percent confidence?

If leak rate is 0.1 gal/hr: ______________

If leak rate is 0.2 gal/hr: ______________

e. Does the method allow for closure of the facility on one or more consecutive days of the week?

☐ Yes ☐ No

f. Does the method require meter calibration?

☐ Yes, specify frequency: ☐ No

____

3. Reporting Of Leak Status

a. Is the leak status reported in terms of a leak rate (gal/hr or gal/day)?

☐ Yes ☐ No If no, how are the results reported?

____

b. What criterion does the method use to declare that a tank is leaking?

☐ Average daily discrepancy exceeds threshold of ___________ gal/hr

☐ Daily discrepancy relative to variability exceeds threshold of ___________ gal/hr

☐ Water level change exceeds threshold of ____________ inch

☐ Statistically significant continuous loss at the ____________ level of significance

☐ Other:

4. Identification Of Causes For Discrepancies

Which of the following factors does the method consider? Check the appropriate categories.

Factor	Identify Only	Compensate	Not Considered
Dispensing meter errors			
Calibration errors			
Conversion chart miscalibration			
Vapor loss			
Thermal effects			
List others			

Which of the following effects does the method identify and quantify? Check the appropriate categories.

Effects	Identify Only	Quantify	Not Considered
Leak rate			
Delivery errors			
Unexplained losses or gains			
Water inflow			
Water outflow			
Product level measurement errors			
List others			

5. Exceptions

a. Are there any conditions under which the statistical inventory method is inadequate?

- ☐ Insufficient number of usable records
- ☐ Irregular time intervals between product level readings
- ☐ Unacceptable daily variability of inventory records
- ☐ Other: ____

b. What elements in the record keeping are left to the discretion of the personnel on site?

- ☐ Length of record keeping, if beyond minimum requested
- ☐ Other: ____
- ☐ None

c. If applicable, attach a copy of the inventory data collection forms the vendor provided to the user.

Additional explanations or comments

____

Reporting Form For Test Results
Statistical Inventory Reconciliation Testing Method

Instructions

The evaluator completes this form after evaluating the method. One page contains space for 40 test results. Use as many pages as necessary to summarize all attempted tests. Indicate the commercial name and version of the SIR method.

In general, the evaluator analyzing the data will complete this form. A blank form can be developed on a personal computer, the database for a given evaluation generated, and the two merged on the computer. The form can also be filled out manually. The evaluator and the vendor's test results provide input for the form.

The table consists of 6 columns. One line is provided for each inventory record used to evaluate the method. If a test was inconclusive, please note and explain this.

The Record Code No. in the first column refers to the code the evaluator assigns to each record for decoding purposes during the evaluation process.

See below for information required on the form and the source of that information.

Under This Column Heading	Information Provided By
Record Code No.	Evaluator
Induced Leak Rate (gal/hr)	Evaluator
Estimated Leak Rate (gal/hr)	Vendor's reporting form
Estimated – Measured Leak Rate (gal/hr)	By subtraction
Interpretation – Tank Tight? (Yes, No, or Inconclusive)	Vendor's reporting form
Vendor's Comments	Vendor's reporting form

Method name and version ______________________________

Date ______________________

Record Code No.	Submitted: Induced Leak Rate (gal/hr)	Results Reported by Vendor: Estimated Leak Rate (gal/hr)	Estimated - Measured Leak Rate (gal/hr)	Interpretation Tank Tight? (Yes, No, or Inconclusive)	Vendor's Comments
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					

Method name and version ____

Date ______________________

Record Code No.	Submitted: Induced Leak Rate (gal/hr)	Results Reported By Vendor: Estimated Leak Rate (gal/hr)	Estimated - Measured Leak Rate (gal/hr)	Interpretation Tank Tight? (Yes, No, or Inconclusive)	Vendor's Comments
41					
42					
43					
44					
45					
46					
47					
48					
49					
50					
51					
52					
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					
66					
67					
68					
69					
70					
71					
72					
73					
74					
75					
76					
77					
78					
79					
80					

Appendix C
SIR Reliability Comparison Protocol

Reliability Comparison Protocol

SIR Comparison Limited Application Protocol For Determining Reliability When SIR Program Language Is Updated

Disclaimer

This protocol may not be suitable in all applications when modifications are made to SIR methods. To determine if this protocol is applicable for the intended use, the user must consult the National Work Group on Leak Detection Evaluations (NWGLDE) prior to using this SIR protocol and receive guidance from NWGLDE concerning its applicability.

Purpose

The purpose of this protocol is to compare the reliability of two versions of the same method; that is, to determine whether or not the two versions of the same SIR program written in different computer languages produce the same results. Thus it is important to establish that this is a reliability protocol and not a validity protocol.

The two terms, reliability and validity, are often used interchangeably when they are not related to statistics. When critical readers of statistics use these terms, however, they refer to different properties of the statistical or experimental method.

Reliability is another term for consistency. If one person takes the same personality test several times and always receives the same results, the test is said to be reliable.

A test is valid if it measures what it is supposed to measure. If the results of a personality test showed that a very shy person was in fact very outgoing, the test would not be valid.

Reliability and validity are independent of each other. A measurement may be valid, but not be reliable, or be reliable but not valid. If a bathroom scale was reset to read 10 pounds lighter than a person's true weight, then the scale would be reliable if the weights were the same each time a person steps on it, but not valid because the weights would always be wrong.

An independent third party evaluation according to an NWGLDE acceptable protocol is used to certify the validity of a statistical method; that is, it certifies that a particular method can find a product release or leak with a probability of a false alarm of 0.05 or less and a probability of detection of 0.95 or greater. This reliability protocol does not certify the validity of a SIR method, and thus can only be used on methods which have already been evaluated with an approved SIR protocol and listed on the NWGLDE list.

The Need

The basic reason that a method might have two or more versions is so the method can be run using different operating systems. This can be accomplished in one of two ways: the software source code can be written for different languages (for example, C++, Visual Basic, etc.) or a cross compiler can be used which reads the same software source code but produces different executable binary code for different operating systems (for example, Windows, Apple OS, Linux, etc.).

Methodology

This protocol must be performed by a party independent of the SIR vendor with no financial interest in the SIR vendor's company whose method is being examined.

The person performing the comparison must be familiar with statistical analytical methods and have access to a computer capable of running the versions of the SIR program being evaluated. In this document, we refer to this person as the evaluator. The evaluator may be asked to provide information to NWGLDE which supports his suitability to perform the reliability comparison.

Steps

1. The evaluator must have at least 100 monthly inventory data, or SIR datasets, from tanks. Datasets do not have to be from 100 different tanks, but must be discrete data periods.

2. The evaluator obtains a copy of the SIR program from the vendor with instructions on how to load and operate the SIR program on the evaluator's computer. This version must be the originally listed NWGLDE version with no modifications from the full evaluation performed when the method was evaluated. This is referred to in this protocol as the standard version and may be identified in a subsequent report by a version number, name, or other identifier.

3. The vendor also supplies the evaluator with the modified SIR program and must provide a detailed description of changes made in this version from the standard version. This version is referred to in this comparison protocol as the revised version. At this juncture in the process, the vendor or evaluator, or both, should have made contact with NWGLDE to get guidance on the suitability of using this protocol on the revised version. It may be that a full SIR evaluation, or validity evaluation, on the revised version will be necessary due to the nature of the modifications made to the SIR program or to take into account the wishes of the SIR vendor. Note: a full SIR evaluation is always an option available to a SIR vendor when revisions to SIR programs are made and most likely will be mandatory when changes affecting the analytical programming which affect the results are made. If there are doubts or questions regarding a full evaluation, NWGLDE should always be consulted very early in the process.

4. The evaluator should run the 100 monthly SIR datasets on each version and tabulate the statistics produced by each version; see table below. If for any reason an analysis fails to run to completion on any dataset, the dataset may be discarded and another substituted until there are 100 total results. Do not use the discarded dataset in step 5 below; use the substituted dataset in step 5.

5. The evaluator must use the same 100 monthly SIR datasets successfully used on the standard version, and run them on the revised SIR version and tabulate the same parameters collected in step 4.

6. The evaluator will prepare a table comparing the results of step 4 and step 5. You may use the following table as a model. The italicized row presents an example of SIR data to be included in the comparison table.

Variables To Be Compared

Standard Method Or Version						**Revised Method Or Version**				
SITE	TANK #	TH	MDL	CLR	P/F/I	TANK #	TH	MDL	CLR	P/F/I
1	*001*	*0.086*	*0.173*	*0.065*	*P*	*001*	*0.086*	*0.173*	*0.065*	*P*

7. The evaluator compares the results of the two versions. The versions are considered comparable if the results conform to the following:

   Quantitative Comparisons: All quantitative comparisons (TH, MDL, and CLR) must agree to at least two places beyond the decimal. A discrepancy greater than 0.01 between any of the three quantities will result in the evaluator having to declare the two versions to be different.

   Qualitative Comparisons: The two versions must be in 100 percent agreement on the qualitative variable of pass, loss, gain, or inconclusive. Any disagreement results in the evaluator having to declare the two versions to be different.

8. The evaluator may prepare a report confirming the results of his comparison. The report must contain at a minimum, a summary of the work done, conclusions reached, and the table of the results obtained from his comparison as shown in the table above. The evaluator must sign and date the comparison and print it on the evaluator's letterhead. The evaluator may include additional information at the request of the SIR vendor or NWGLDE.

As an alternative to steps 2 and 3 above, if an evaluator lacks the appropriate computer infrastructure to do the reliability protocol at his place of business or home, it is acceptable for the evaluator to perform the evaluation at a different location. If a different location is needed, the evaluator should discuss this with NWGLDE prior to attempting the reliability protocol. The evaluator’s report, as noted in step 8, must state where the reliability protocol was performed and whether multiple computers or what equipment was used, as well as describe any other deviation allowed by NWGLDE.

United States Environmental Protection Agency | Land And Emergency Management 5401R | EPA 510-B-19-004 May 2019 www.epa.gov/ust