SECTION 10

INVENTORY RECONCILIATION TECHNIQUES

I. INTRODUCTION

Analysis of properly collected motor fuel inventory data offers the potential for an inexpensive, readily available approach to detecting tank-product losses, including those resulting from leaking tanks and piping. Inventory data is collected typically at the close of each day of operation by the tank establishment operator, records (1) volume of fuel in each tank as measured by metering stick or gauge, (2) delivery volumes, and (3) dispensing meter readings. From these measurements, the volume of fuel metered through a tank system's dispensers is reconciled with the physical measurement (based on stick readings and delivery volumes) of product gone from the tank.

Rarely are these two measurements of volume of daily through-put (physical versus dispensing meter) numerically equal, even when there has been no loss of product. Rather, they will show some variance either as an "overage" (a numerical excess of product in the tank) or an "underage" (a numerical loss of product). Some part of the daily variance will be due to random errors of measurement; however, inaccurate gauging and metering devices add to the variance, as do temperature-induced product shrinkage and expansion, vapor loss, theft, and leakage of product from (or of water into) the system. To further complicate interpretation of inventory reconciliation data, several of the above factors may contribute both to overages and underages.

There are currently available several commercially developed computerized models which can be used to identify and quantify factors contributing to daily inventory variances. These models are proprietary and, in any case, are much too sophisticated to be implemented and interpreted by a typical tank owner or operator. To monitor inventory using these models, a tank owner must contract for the services of one of the firms which has developed such a model. To provide tank owners and operators with a simple, inexpensive procedure for monitoring inventory, EPA has developed a procedure based on counts of the number of negative daily variances (underages) in successive monthly periods.[1] Because of its simplicity, the EPA method would not be expected to be as accurate as the more sophisticated modeling techniques. Tank tightness test data and inventory reconciliation data collected during two phases of the National Survey of Underground Motor Fuel Storage Tanks make possible a determination of the extent of agreement between tank tightness tests and the various methods of inventory reconciliation analysis, as well as the extent of agreement among the latter.

At the time the inventory analysis was conducted, the survey had provided complete, properly-collected inventory data on 855 tank systems. While this represents only the first portion of the inventory returns, the inventory data collection and editing effort resulted in 41 percent of the attempted cases providing usable data. Of these 855 tank systems, 511 were analyzed using the inventory reconciliation model developed by Warren Rogers

---

[1] USEPA, Office of Toxic Substances, "More about Leaking Underground Storage Tanks: a Background Booklet for the Chemical Advisory," (October 1984).

Associates.[2] In addition, in a smaller study, 18 tank systems were analyzed by Entropy Limited.[3] Tightness test results were available for 189 of the 855 tanks for which usable inventory data was available. The EPA inventory analysis method was applicable, in modified form, to all 855 tanks; modification was required since the EPA method was intended for application to on-going monitoring programs and not to a single set of one-time inventory data.[4]

The present section provides an analysis of the extent of agreement of the above inventory reconciliation methods with one another and with tightness-test results, based on data collected in the survey. In addition, we report the results of a small quality control study in which the various inventory approaches were applied to a simulated set of inventory data for five tanks. Mathematical techniques were used to simulate various combinations of stick error, leakage and theft for the five tanks. The results of the inventory analyses were then compared with the true condition of the hypothetical tanks which in this case, of course, was known exactly.

II. METHODS AND DATA

The survey protocol called for the collection of 30 days of inventory data on each tank or manifolded tank system at the

---

[2] Warren Rogers Associates, Inc., "Inventory Reconciliation System," (undated)

[3] Entropy Limited, "Precision Tank Inventory Control," (1984).

[4] USEPA, Office of Toxic Substances, "More about Leaking Underground Storage Tanks: a Background Booklet for the Chemical Advisory," (October 1984).

sampled establishments. Many respondents were unable to supply proper inventory data (see Appendix B for details), with the result that complete, usable data was obtained for only 855 tanks or manifolded tank systems. Of these tanks, 511 were analyzed by Warren Rogers Associates (WRA). Eighteen of the 511 tanks were also analyzed by Entropy Limited (EL). Of the 439 tank systems in the tightness test (TT) sample, 189 had usable inventory data. Table 10-1 shows the available sample sizes for all possible pairwise comparisons between methods.

Table 10-1. Sample sizes for pairwise comparisons between methods

	EPA	WRA	EL	TT
EPA	855	511	18	189
WRA		511	17	106
EL			18	17
TT				439

The EPA-developed inventory analysis method is based on a simple count of the number of days for which the inventory reconciliation shows a negative variance, i.e., a numerical loss of product. An excess of days with negative variance over those with positive (or zero) variance may, under certain circumstances, be interpreted as a loss of product due to leakage or some other cause. The EPA method was developed and calibrated for application to an on-going monitoring program, in which cumulative month-by-month counts of negative variances would be compared with statistically-derived "action numbers" to determine whether there was evidence of a systematic deficit in inventory. Calculations indicated that the method would be effective in detecting even relatively small leaks over a sufficiently long period of monitoring, at least when reasonably good inventory records were kept.[5] The method was not intended for application to a one-time collection of 30 days inventory and, indeed, would not be expected to detect smaller leaks based on such a small data set. For the present comparisons, the method was modified and the following decision rule adopted: A tank system is declared leaking if the 30-day record exhibits 18 or more negative daily variances. Calculations detailed in Appendix E, indicate that this rule has approximately a five percent false-positive (or "false alarm") rate. That is, there is a five percent chance that a tank system which is not leaking and whose inventory record is subject only to random stick measurement error, would be declared leaking. This is consistent with the definition of a leak adopted for the tightness test procedure, see Section 8. However, the modified EPA inventory analysis method has significantly poorer detection capability than tightness testing, even in optimistic scenarios where the inventory record is subject only to stick measurement error and not to other sources of discrepancy, such as delivery errors. For example, the chance of detecting a leak of 0.1 gallons per hour (2.4 gallons per day) in a typical tank is approximately 17 percent, as opposed to 95 percent for tightness testing. A detection capability of 80 percent or greater was obtained using the EPA method for leaks in excess of 0.37 gallons per hour (8.9 gallons per day). By substantially increasing the number of days of inventory data, an 80 percent detection capability would be possible for smaller leaks.

---

[5] David C. Cox, "Performance of the Chemical Advisory Inventory Analysis Method Under Various Scenarios," Report from Battelle Columbus Laboratories to EPA under Contract No. 68-01-6721 (April 1984)

Both the WRA and EL procedures are proprietary, so that details of the methodology and decision rules are available only in sketchy form. A description of the methods, based on literature provided by their developers, is given in Appendix E. The literature mentioned is partly promotional in nature. Claims made therein have not been investigated or verified by EPA, except to the extent reported here, so that no endorsement of the methods should be inferred. In order to place the WRA and EL methods on the same footing as the EPA and tightness test approaches, it is important that the false positive rate of the WRA and EL methods also be five percent. According to the developers of these methods, the inventory analysis results they have provided meet this requirement.

The WRA, EL and tightness test methods occasionally fail to produce a definite conclusion as to whether a tank system is leaking or not leaking. This occurs for the inventory methods whenever the data is excessively noisy due, for example, to large stick errors or very frequent deliveries. For tightness tests, various physical problems may lead to an indeterminate test result (see Section 8.) Table 10-2 presents a breakdown of the results for each method. It is reasonable that the tightness test procedure reports the largest percent leaking; this method should have the best detection capability. Likewise, the EPA procedure should have the lowest percent leaking, as it does. In the next section, we examine agreement between the methods, i.e., the extent to which they agree, not just on percent leaking but on <u>which</u> tank systems are leaking and which are not leaking.

Table 10-2. Number and percent of tank systems judged to be leaking, judged not to be leaking, and providing inconclusive results for inventory methods and tightness testing

		Tank Systems Judged to be Leaking		Tank Systems Judged to be Tight		Tank Systems with Inconclusive Results	
Method	Sample Size	Number	Percent[1]	Number	Percent[1]	Number	Percent[1]
TT	439	152	35%	259	59%	28	6%
EPA	855	149	17%	706	83%	0	0%
WRA	511	160	31%	294	58%	57	11%
EL	18	4	22%	11	61%	3	17%

---

[1] May differ from national estimates because survey sampling weights are not considered here.

III. COMPARISON OF METHODS

In comparing two methods of deciding whether or not a tank system is leaking, one cannot focus simply on the degree of agreement between the predictions of the methods. To see why, consider two methods which give a correct prediction in 70 percent of cases. If the methods were completely independent, one would expect them to agree in 49 percent of cases, according to the rules of probability. This 49 percent represents the degree of agreement expected purely by chance and not due to any tendency for the methods to act in the same direction. In this example, agreement in significantly more than 49 percent of cases is required before one can conclude that the methods really are in substantial agreement. In this section, a statistic, K, is used to measure the agreement between methods above and beyond what is expected by chance.[6] The value K = 0 corresponds to purely chance agreement, while K = 1 means perfect agreement. Values of K between 0 and 1 may be interpreted on an ordinal scale, i.e., the larger K is, the better the agreement. Quantitative interpretation of K is more elusive, e.g., it is difficult to determine just how much agreement is represented by a value of, say, K = 0.5.

For each pairwise comparison between methods considered, we carry out a statistical test to determine whether there is sufficient evidence to conclude that K is positive, i.e., that there is more than chance agreement between the methods. In performing this test we have ignored inconclusive results for the various methods. We have also treated the data as if it were generated by a simple random sample of tank system, ignoring the

---

[6] Yvonne M. M. Bishop, Stephen E. Fienberg and Paul W. Holland, "Discrete Multivariate Analysis: Theory and Practice," MIT Press, Cambridge, MA (1975)

survey sample weights. Tables 10-3 through 10-5 present comparisons between the tightness test, WRA and EPA.

The two inventory methods show agreement with each other beyond what one would expect by chance alone. However, each inventory method exhibits only chance agreement with tightness test results. This conclusion should be regarded as tentative since the sample sizes for even the overall inventory -- tightness testing comparisons reported here are not very large. More detailed analyses of the agreement between the methods would not be statistically meaningful. It is, however, worth pointing out that the agreement does not appear to be improved even if we restrict attention to large leaks (as measured by tightness testing). For example, consider quantifiable leaks exceeding 4 gallons per day. The EPA procedure found 6 out of 23 (26%), for which a comparison was possible, to be leakers; WRA found 4 out of 7 (57%) leaking.

Statistically meaningful comparisons with the EL results are not possible because of the very small number of tanks evaluated by Entropy. However, the data confirm the above two findings in a general way. Inventory methods agree with one another but not with tightness test results. The extent to which inventory and tightness testing may be measuring differing phenomena, as is suggested by these results, is not clear. It is possible that certain measured leaks may not represent operational leaks. For example, leaks at the very top of the tank would occur in a tightness test, but might not occur in practice if the tank were never filled to the top. Likewise, a phenomenon such as theft may be reported as a leak by the inventory methods while the tank system tests tight. The resolution of this question will require more detailed analyses of the survey data and, possibly, collection of longer series (more than 30 days) of inventory data.

Table 10-3. Comparison of EPA Inventory Reconciliation Method with Warren Rogers Associates Inventory Reconciliation Method

		WRA		
		Number of tank systems judged to be leaking	Number of tank systems judged to be tight	Inconclusive
EPA	Number of tank systems judged to be leaking	61	19	10
	Number of tank systems judged to be tight	99	275	47

Percent agreement = 74%

K = 0.36 (STATISTICALLY SIGNIFICANT)

Table 10-4. Comparison of EPA Inventory Reconciliation Method with Tightness Testing

		TT		
		Number of tank systems judged to be leaking	Number of tank systems judged to be tight	Inconclusive
EPA	Number of tank systems judged to be leaking	13	22	1
	Number of tank systems judged to be tight	37	102	14

Percent agreement = 66%

K = 0.09 (NOT STATISTICALLY SIGNIFICANT)

Table 10-5. Comparison of Warren Rogers Associates Inventory Reconciliation Method with Tightness Testing

		TT		
		Number of tank systems judged to be leaking	Number of tank systems judged to be tight	Inconclusive
WRA	Number of tank systems judged to be leaking	9	28	0
	Number of tank systems judged to be tight	11	38	5
	Inconclusive	5	8	2

Percent agreement = 55%

K = 0.02 (NOT STATISTICALLY SIGNIFICANT)

IV. QUALITY CONTROL SAMPLES

A limited quality-control study of the 3 inventory reconciliation methods was conducted in order to compare performance on a data set for which the true leak status could be unequivocally determined. Inventory data for a total of 5 tanks at 2 sites was generated using mathematical techniques to simulate various combinations of stick error, leakage and theft. To maximize realism, actual tank conversion charts were used to simulate the effect of random measurement error due to sticking the tank. The simulated data was provided blind to WRA and EL. Table 10-6 shows the scenarios simulated. Table 10-7 shows the results of the inventory analyses. In addition to EPA, WRA and EL, we have added a simple t-test. The test is a standard t-test with 29 degrees of freedom based on the 30 daily variances in the inventory record. For consistency with the other inventory methods, the false positive rate of the test is set at 5 percent.

Site 1 represents clean inventory data. There is no noise in the record other than random measurement error due to sticking the tank. The EPA method and the t-test correctly predicted leak status for both tanks. WRA detected the leak but also classified the non-leaker as a leaker. EL correctely classified the non-leaker but reported the leaker as inconclusive (accurately estimating the true leak rate, however).

Site 2 represent a more difficult test of the inventory methods. Stick measurement error is unusually large. Moreover, both the random pattern of theft in tank 2 and the relatively small leak (3 gallons per day) in tank 3 would be expected to be difficult to detect by any inventory method based on only 30

Table 10-6. Simulated quality control inventory data

Site 1

Tank	Size	Product	Description
1	10,000	Reg. unleaded	5 gals/day leak
2	10,000	Reg. leaded	No leak, stick error only

Readings to nearest 1/2" on dipstick -- typical random measurement error of 14-19 gallons

Site 2

Tank	Size	Product	Description
1	6,000	Reg. leaded	Stick error only
2	6,000	Prem. unleaded	Theft 15-20 gallons on 9 days
3	10,000	Reg. unleaded	Leak, 3 gals/day

Readings to nearest 1" on dipstick -- typical random measurement error is 20-25 gallons for 6,000 gallon tanks, 35-40 gals for the 10,000 gallon tank

Table 10-7. Inventory analysis of quality control samples

			EPA		WRA			EL			t-TEST	
Site	Tank	True Status	LS	NEG	LS	LR	Comments	LS	LR	Comments	LS	LR
1	1	L - 5 gals/day σ = 17 gals	L	19	L	8.4	Delivery discrepancies (2), Unexplained gains or losses (4), Large stick errors (1).	?	4.6	Thermal shrinkage, vapor loss reported	L	6.6
1	2	NL - Stick error only σ = 17 gals	NL	15	L	2.8	Delivery discrepancies (2), Large stick errors (2)	NL	-	Thermal shrinkage, vapor loss reported	NL	-
2	1	NL - Stick error only σ = 22 gals	L	18	L	11.1	Unexplained gains and losses (3), Large stick errors (1). Average stick error low.				NL	-
2	2	NL - Theft on 9 days/30, approx. 17 gals σ = 22 gals	L	19	L	10.3	Unexplained gains and losses (6). Average stick error low.				L	10.0
2	3	L - 3 gals/day σ = 40 gals	NL	13	NL	-	Unexplained gains and losses (1). Data quality code = 2, meaning results explained by frequent deliveries??				NL	-

Legend: LS = Leak status
LR = Leak rate
L = Leak
NL = No leak
NEG = Number of negative daily variances
? = Inconclusive
σ = Standard deviation of stick measurement error

days' data. (The change of detecting the 3 gallons per day leak using the EPA method is only about 10%.) Both EPA and WRA predicted that the tight tank was leaking and that the leaker was not; both found the theft case to be a leak, thus successfully detecting the negative trend in the inventory, although attributing it to the wrong cause. The t-test correctly evaluated the non-leaker but also got the other two tanks wrong. A surprising feature of both sophisticated methods (WRA and EL) was a tendency to find effects in the data which were not in fact present. Thus WRA found numerous unexplained gains and losses, delivery discrepancies and large stick errors, while EL found theoretical shrinkage and vapor loss. These spurious findings apparently tended to obscure the true status of the tank systems. It should be pointed out, of course, that a method of analysis tailored to perform well on real-world, noisy data will of necessity be less than optimal for unusually clean data such as we have here. Moreover, the simulated data does not reflect the effects of factors such as location and time-of-year that may be very important to account for in real data analysis. The results reported here must be interpreted in this light.

V. CONCLUSIONS

We have compared a number of inventory reconciliation techniques, with each other and with the results of tank tightness tests, using data from the survey, as well as a small set of simulated inventory records. The sample sizes, especially for inventory vs. tightness test comparisons, were somewhat small. Finally, the data were analyzed as if they were generated by a simple random sampling technique, rather than the sampling procedures actually used in the survey. Thus, the conclusions reported here sophisticated methods may have some tendency to "detect" noise in the data from effects that are not, in fact, present.