Correlating Tube Failures with Operating Severity

Detailed report on the results of an extensive survey of reformer catalyst tube failure experience in side- and bottom-fired reformer furnaces.

R. M. Osman and P. A. Ruziska,
Exxon Chemical Co.,
Florham Park, N. J.

The AICLE Symposia on Ammonia Plants and Related Facilities have included a number of papers on failure statistics for reformer catalyst tubes (*1*, *2*, *3*, *4*).

The data was primarily based on experiences of plants having similar *top-fired* furnace designs. Little quantitative comparison of failure data with operating severity was included. This is a report on results of a survey of reformer catalyst tube failure experience in *side-* and *bottom-fired* reformer furnaces. To obtain this failure experience, a questionnaire was mailed to operators of such furnaces in November, 1974. To permit correlation of tube failure frequency with operating severity, furnace operating data was requested along with tube failure information.

In response, 30 completed questionnaires were received. Responses represented ammonia, hydrogen, and methanol plants, Selas, Foster-Wheeler, Chemico, and Exxon furnace designs, and both staggered and straight tube row arrangements. All furnaces use HK-40 for their tube material, although one had recently installed a few Manurite 36X tubes (Manoir-Pompey proprietary alloy). A small number of internally machined HK-40 tubes were also reported. The furnaces reporting contain a total of more than 5,500 catalyst tubes, and at the time of questionnaire completion, had amassed more than 40 thousand tube-years of operating experience.

The primary conclusions reached from an analysis of the questionnaire data are as follows:

1. Most reported reformer catalyst tube failures and preventive replacements were due to creep rupture (primarily from long-term aging but with significant numbers attributed to localized overheating or plugging/coking).
2. Other major causes of failures and preventive replacements were outlet pigtail/coupling problems; severe bowing, carburization (indicated by inspection), and tube weld cracks. Furnaces with staggered tube rows accounted for most of the bowing and outlet pigtail/coupling problems.
3. The Weibull probability distribution gives generally excellent fits to individual plant creep rupture data, and thus can be used to predict tube failure frequency vs. time.
4. Median creep rupture failure times predicted from Weibull analysis can be correlated atainst operating severity, with severity defined in terms of a Larson-Miller analysis using tube metal temperature (TMT) and tube stress level.
5. Tube life predictions based on Larson-Miller analysis tended to be optimistic.
6. Severity calculations based on computer predicted TMT's gave better correlations (less scatter) than those based on measured TMT's.
7. No significant difference in severity vs. median failure time correlation among the various furnace designs was noted.

Creep rupture predominant failure mode

Of the 30 plants responding, 24 reported experiencing one or more catalyst tube failures. A total of 818 tube failures, or incipient failures requiring preventive replacement, were reported (excluding failures due to outlet header problems—reported by two plants). The breakdown of the reported failures and preventive replacements among the various reported causes is given in Table 1.

The table shows that 71% of all failures or preventive replacements reported were due to a creep rupture mechanism. Even excluding failures due to local overheating, plugging or coking, more than 2/3 of the failures were reported as being from creep rupture.

It can therefore be concluded that creep rupture (from now on referring only to creep ruptures due to aging) was the predominant mode of reformer catalyst tube failure. This statement must be tempered somewhat by the fact that some reported creep rupture failures could have been due to an unrecorded local overheating situation. Also, 343 of the 555 reported creep rupture failures occurred at 3 of the 21 plants which experienced such failures.

The outlet pigtail/coupling failure (or preventive replacement) category represented almost 15% of the total reported. This category included pigtail weld cracks, outlet coupling cracks, and "weldolet" bulging and ruptures. Note that none of these can actually be considered as *tube* failures. Most (> 80%) of the outlet pigtail/coupling problems were in a furnace having staggered tube rows, and where the thermal cycling of long vs. short pigtails was given primary blame for the failures.

Bowing problems were blamed directly for more than 8% of the reported failures/replacements, and were also cited as secondary problems contributing to creep rupture failures at three plants. Two-thirds of the plants reporting bowing problems had staggered tube rows. More significantly, these staggered tube row plants accounted for 97%

Table 1. Summary tabulation of reported primary failures

Reported primary failure or preventive replacement cause		No. of failures	Percent of total	No. of plants reporting
Creep Rupture	Aging	555	67.9	21
	Local Overheating	19	2.3	7
	Plugging/ Coking	8	1.0	3
Outlet pigtail/coupling problems		120	14.7	3
Severe bowing		69	8.4	3
Carburization (per inspection)		28	3.4	4
Tube weld cracks		17	2.1	3
Other problems		2	0.2	2
Totals		818	100.0	24

of the tubes removed due to bowing as the primary reported mechanism.

Carburization, as indicated by magnetic inspection, was cited as the reason for preventive tube replacement at four plants. No actual failures were reported, however.

Tube weld cracks accounted for about 2% of the reported failures. Of these, about half were reported as being at the top flange weld. For the rest, no location was given, so it is not known whether top flange welds or firebox welds were at fault.

In the ''other'' category, two problem types were recorded. One tube failed due to thermal shock from steam carryover, and another was preventively removed due to heavy pitting (cause not recorded).

Type of data requested in questionnaire

Completed questionaires contained a summary of the furnace tube replacement failure history. Specific data requested for failed for replaced tubes were: time of failure or replacement; whether replaced before failure or actually failed; type of failure or reason for replacement; and any changes in specification for replacement tube.

Also requested in the failure history section of the questionnaire was a summary of the present ages of all tubes currently in the furnace. Usually (but unfortunately not always) the questionnaire data was sufficient to establish a complete picture of what had happened to the various furnace tubes.

Since a primary interest was creep rupture failures, special attention was given to data describing failures or replacements due to that mechanism. One problem was how to handle data on tubes that were replaced due to incipient creep rupture (usually indicated by severe bulging) before they had actually failed. To consider them as failures at the time of replacement would bias the analysis toward earlier failure times, whereas to ignore these tubes completely would give a very optimistic picture of tube life. As a reasonable compromise, it was assumed that all tubes preventively replaced due to incipient creep rupture would actually have failed within 12 months of the reported replacement time, had they been left in service.

Inasmuch as a probablistic analysis of creep failure data requires cumulative *fractions* of tubes failed as a function of time, special attention had to be paid to properly defining the total population of tubes that the failures should be referenced to. The situation is complicated by the fact that, in most cases, failed tubes are replaced by new tubes, which usually remain in the furnace for significantly long periods, and which sometimes fail themselves. Also, tubes often fail (or are replaced) for reasons other than creep rupture of incipient creep rupture.

To handle these difficulties, the following approaches were used:

1. For plants having new tubes installed as replacements, several tube populations were considered, each having different total service times. When computing cumulative failure fraction for time t, the numerator is the cumulative number of failures at time t for all tube populations having service times equal to or greater than t, and the denominator is the total number of tubes in all such populations.

2. Tubes which, at time t, fail and/or are removed for reasons other than creep rupture, were considered to be part of the total tube population up until time t, but were eliminated from the population afterwards. Note that tubes whose failures and/or removals were attributed to creep rupture were considered to be part of their original population for the full life of that population.

For some responding plants, the data indicates very high failure rates during the early life of the plants, with considerably lower failure rates later on. Since the startup and early post-startup phases of plant operation often result in severe upsets in furnace operation, it was assumed that the failure vs. time data on the initial populations of tubes from such plants were not representative. In these cases, only the failure data on subsequent tube populations were used in the general correlation work.

Use Of Weibull probability distribution

In correlating reformer tube failure rates against operating severity, it was decided to use the median time to creep rupture failure as the correlation basis. This was defined as the time required to fail 50% of the tubes in a given population operating at a given severity.

This basis has the advantage of being independent of any maldistribution of firing, process gas flow, coking, metal defects, etc., among the various tubes in a given furnace. Theoretically at least, two furnaces operating at the same severity but with different degrees of tube-to-tube maldistribution, would have the same median failure time. Data from two such hypothetical furnaces would show more early failures, but a lower cumulative failure vs. time slope, for the furnace with more maldistribution than for the more evenly balanced furnace. However, the two cumulative failure vs. time curves would cross each other at the 50% failure point.

Of the 30 plants responding to the reformer tube failure questionnaire, 12 reported having either no creep rupture failures, or only one. Therefore, the data for these plants could not be used in formulating a correlation between median failure time and operating severity, because at least two data points are needed to make a projection to the 50% failure time.

Of the remaining 18 plants responding, all of which reported two or more creep rupture failures or replacements, 16 provided failure data that were sufficiently descriptive to yield complete pictures of their failure histories. These 16 failure data sets formed the basis for an attempt to correlate median failure time against operating severity.

The Weibull probability distribution has been widely used to describe failure rate data (*5*). The Weibull cumulative distribution function can be expressed mathematically as:

$$P(t) = 1 - \exp\left\langle -\left(\frac{t}{\theta}\right)^c \right\rangle \qquad (1)$$

where: $P(t)$ is the cumulative failure probability
t is time
θ is the Weibull scale parameter
c is the Weibull shape parameter.

Fortunately, the expression given in equation 1 can be rearranged into the following more convenient form:

$$\log \ln\left[\left\langle \frac{1}{1 - P(t)} \right\rangle\right] = c \log t - c \log\theta \qquad (2)$$

The form of equation 2 means that when log $[\ln \langle 1/(1 - P(t)) \rangle]$ is plotted against log t, a straight line will result if a Weibull probability distribution is followed. Such a straight line can be extrapolated to predict a median failure time by setting $P(t)$ equal to 0.5 and solving as follows:

$$\log t_{50} = \log\theta + \frac{\log[\ln\langle 2\rangle]}{c} \qquad (3)$$

or:

$$\log t_{50} = \log\theta + \frac{0.1592}{c} \qquad (4)$$

where t_{50} is the median time to failure in months.

Table 2 summarizes the results of Weibull probability distribution least squares fits for the 16 failure data sets having two or more discrete tube failure times. Included in the table are values for the product-moment correlation coefficient, which is a statistical measure of how well a straight line fits the data. Also included, but only for plants with three or more failure points, are values of the 95% confidence intervals (I) on the log t_{50} projections. Based on the Weibull model, this means that it can be said, with 95% certainty, that:

$$(\log t_{50})_{\mathrm{actual}} = (\log t_{50})_{\mathrm{predicted}} + I \qquad (5)$$

The Weibull failure time model was also used to determine if certain early failures could be discounted in projecting median failure times. Figure 1 gives a typical Weibull cumulative failure fraction vs. time plot (for plant No. 2).

By comparison, Figure 2 shows a Weibull plot for plant No. 4, which had two early failures at 4 and 19 months tube life, with the rest of the failures at 78+ months life. Note the skewed distribution indicating that the early failures do not correlate with the rest of the failures, and should not be used for projecting median failure time. Figure 3 shows a revised Weibull plot for plant No. 4, with the first two failure points omitted. Note that the data now correlates well using a Weibull distribution.

Predicting tube metal temperatures

The reformer tube failure questionnaire requests data on plant operating conditions such as feed rate and composition, steam rate, catalyst tube inlet and outlet temperature and pressure, and catalyst tube outlet gas analysis. Physical data such as tube and catalyst dimensions were also requested.

With a proprietary Exxon computer model for reformer furnaces, the operating and physical data supplied for each plant were used to generate tube metal temperature (TMT) and pressure stress profiles for the catalyst tubes. No attempt was made to add a maldistribution safety margin because a correlation of failure times predicted from *average* furnace conditions against Weibull *median* failure time projections was planned.

Significant inaccuracies in the predicted TMT profiles may be present, due to the uncertainties in the base operating data. The outlet conditions of the reformer furnace must be defined in order to perform the TMT and stress calculations. These conditions were determined for each case by fixing the feed composition, outlet temperature and outlet pressure, and then varying the temperature approach to reforming equilibrium until the calculated ratio of outlet methane to total carbon matched that reported in the outlet analysis. Unfortunately, the approach to reforming equilibrium was often significantly negative, indicating better than equilibrium catalyst performance. This would suggest errors in the outlet temperature, outlet anlysis, or both. Fortunately, these factors tend to be compensating to some extent, which would reduce the expected errors in the predicted TMT profiles.

Table 3 summarizes the calculated outside maximum TMT's and the corresponding calculated approaches to reforming equilibrium. Since these TMT's represent average furnace conditions (no maldistribution factor), they will be abbreviated as "AMT" (average maximum temperature). Also tabulated for comparison are the reported

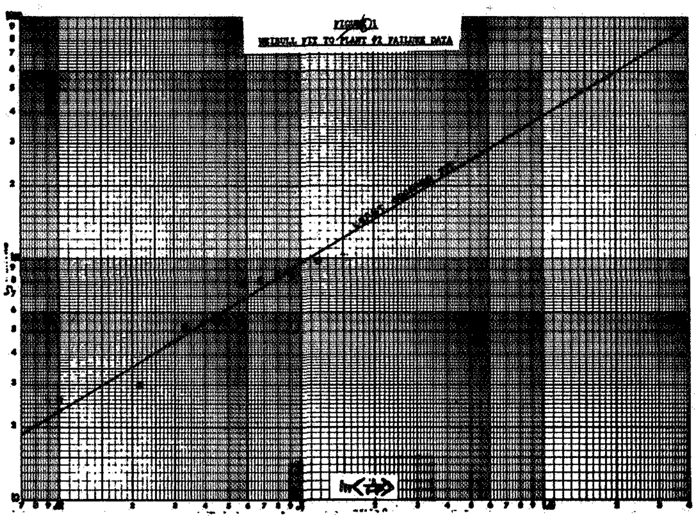


Figure 1. Weibull fit to Plant No. 2 failure data.

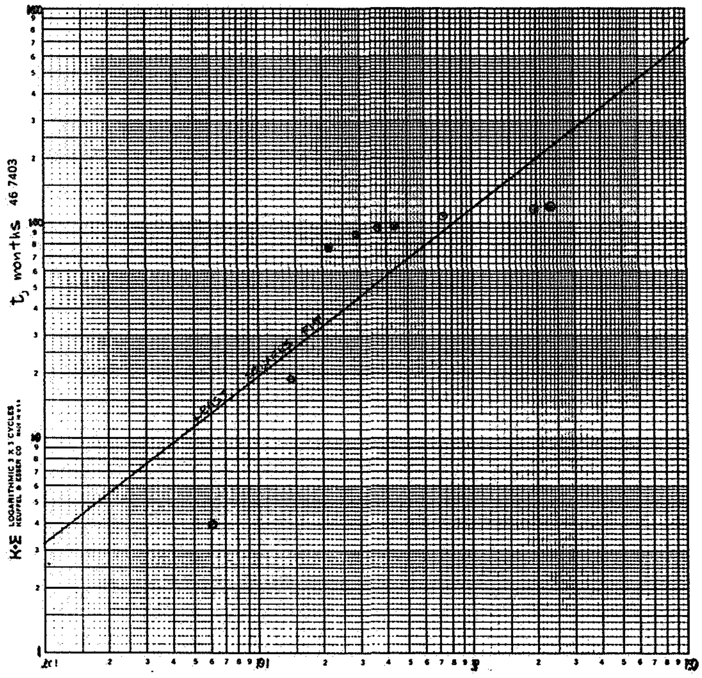


Figure 2. Weibull fit to Plant No. 4 failure data (all points included).

Table 2. Summary of results of Weibull analysis of plant creep rupture tube failure data

Plant no.	Furnace type	Rows straight Or staggered	Weibull Constants C	θ Months	Product moment correlation coefficient, r	Median failure time, months	Log (median failure hours), W	95% Log confidence limit I	Number of discrete failure times	Correlation weight factor
2	Selas	Straight	1.635	388.5	0.975	310.5	5.36	0.146	9	1
3	Selas	Straight	0.936	2,077.1	1.000	1,404.5	6.01	—	2	2
4	Selas	Staggered	8.416	146.0	0.983	139.8	5.01	0.037	7	1
6	Selas	Straight	3.636	201.9	0.805	182.5	5.12	0.485	4	2
7	F-W	Staggered	7.928	56.8	0.996	54.2	4.60	0.016	10	1
8	F-W	Staggered	8.389	123.3	0.816	118.0	4.94	0.257	4	1
11	Selas	Staggered	10.820	146.0	0.958	141.2	5.01	0.033	6	1
13	F-W	Straight	8.367	116.1	1.000	111.1	4.91	—	2	2
16	F-W	Staggered	6.351	51.9	1.000	49.0	4.55	—	2	2
18	F-W	Staggered	6.649	67.8	1.000	64.2	4.67	—	2	2
19	F-W	Straight	19.246	178.2	0.993	174.8	5.11	0.017	3	1
20	Selas	Staggered	2.658	228.0	0.929	198.6	5.16	0.169	10	1
22	F-W	Straight	3.194	177.2	0.987	158.0	5.06	0.041	11	1
23	F-W	Staggered	4.405	190.6	0.970	175.4	5.11	0.119	6	1
24	Selas	Staggered	7.761	94.4	0.997	90.0	4.82	0.057	3	1
27	Selas	Straight	2.804	127.3	0.996	111.7	4.91	0.107	3	1

AMT's as actually measured (five plants reported using infrared pyrometers—the rest optical pyrometers), plus the measured hottest tube single maximum TMT's (SMT). There does not appear to be any significant correlation between calculated approaches to reforming equilibrium, and the differences between calculated and measured average maximum TMT's (AMT's). Because of this, no data points were rejected, even if large negative reforming equilibrium approaches were indicated.

Larson-Miller stress-time-temperature parameter

Laboratory data on the creep rupture phenomenon have indicated that there is an "equivalency" between operating temperature and time to rupture. This "equivalency" permits short-time high temperature tests to be used to predict what would happen after long times at lower temperatures. The most widely used temperature-time equivalency method is by a Larson-Miller diagram.

A typical Larson-Miller diagram for HK-40 is shown in Figure 4. (*6*) The ordinate is simply the stress level in lb./sq. in., and the abscissa is the Larson-Miller parameter P. Once P has been determined from the stress level, the expected creep rupture failure time can be determined from:

$$P = \frac{T(15.0 + \log t)}{1000} \quad (6)$$

or

$$\log t = 1000 \frac{P}{(T)} - 15.0 \quad (7)$$

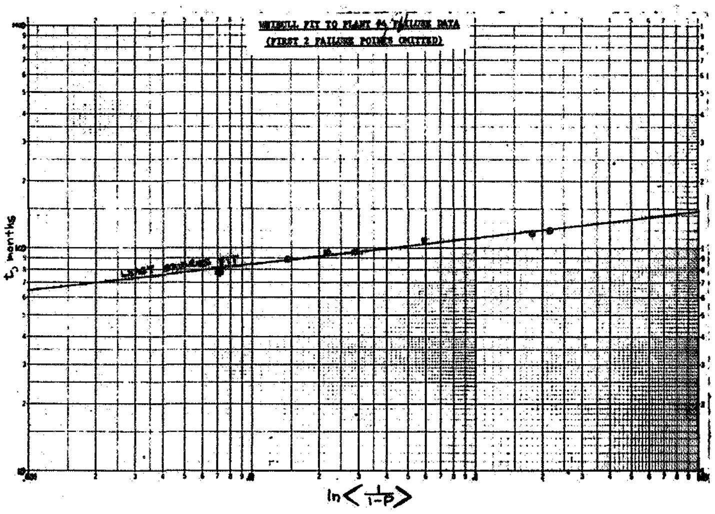


Figure 3. Weibull fit to Plant No. 4 failure data (first two failure points omitted).

where t is the time to failure in hours
T is the temperature in degrees Kelvin.

For each plant, the TMT and pressure stress level profiles generated by the Exxon reformer tube computer program, plus a Larson-Miller diagram for HK-40, could be used to determine the tube cross-section seeing the most severe operating condition. This most severe point is defined as that location on a tube where the combination of TMT and pressure stress yield the shortest Larson-Miller life projection. These minimum life projections are sum-

Table 3. Comparison of measured and calculated tube metal temperatures

Plant no.	Furnace type	Rows straight or staggered	Calculated AMT, °F	Measured AMT, °F	Measured SMT, °F	Measured AMT minus calculated AMT, °F	Approach to reforming equilibrium, °F
1	F-W	Staggered	1,651	1,702	1,770	51	(60)
2	Selas	Straight	1,543	—	1,640	—	(6)
3	Selas	Straight	1,550	—	1,620	—	(42)
4	Selas	Staggered	1,589	1,625	1,670	36	5
5	F-W	Straight	1,552	1,580	1,616	28	20
6	Selas	Straight	1,666	1,690	1,720	24	(66)
7	F-W	Staggered	1,564	1,590	1,775	26	(33)
8	F-W	Staggered	1,571	1,600	1,700	29	(60)
9	F-W	Straight	1,573	1,610	1,650	37	(20)
10	Selas	Straight	1,554	1,680	1,750	126	115
11	Selas	Staggered	1,712	1,660	1,700	(52)	(58)
12	Selas	Staggered	1,674	1,630	1,650	(44)	11
13	F-W	Straight	1,661	1,675	1,800	14	25
14	Chemico	Straight	1,663	1,630	1,700	(33)	20
15	Chemico	Straight	1,760	1,650	1,750	(110)	45
16	F-W	Staggered	1,647	1,590	1,650	(57)	(29)
17	Selas	Staggered	1,681	1,697	1,742	16	0
18	F-W	Staggered	1,649	—	—	—	25
19	F-W	Straight	1,617	1,460	1,770	(157)	(180)
20	Selas	Staggered	1,566	1,534	1,590	(32)	(17)
21	F-W	Straight	1,680	1,625	—	(55)	(85)
22	F-W	Straight	1,735	1,680	1,710	(55)	75
23	F-W	Staggered	1,656	—	1,760	—	29
24	Selas	Staggered	1,739	1,730	1,800	(9)	105
25	F-W	Staggered	1,686	1,760	1,800	74	(77)
26	F-W	Straight	1,594	1,710	1,725	116	(8)
27	Selas	Straight	1,599	—	1,616	—	(15)
28	Selas	Staggered	1,711	—	1,745	—	(72)
29	Exxon	Straight	1,657	—	1,655	—	75
30	Chemico	Straight	1,584	1,590	1,650	6	75

Notes: AMT is average maximum temperature; SMT is hottest tube single TMT.

marized in Table 4 on a log ⟨hours to failure⟩ basis. Assuming the same pressure stress profiles, Larson-Miller tube life expectancies were also calculated using *measured* AMT's and SMT's as bases. Table 4 also summarizes these life projections.

Two failure times are correlated

To find the dependence of tube life upon operating severity, a correlation of median failure times determined from Weibull projections of actual failure data was made against Larson-Miller life projections. The Weibull log ⟨median failure hours⟩ was used as the dependent variable, henceforth abbreviated as W. The Larson-Miller projected log ⟨median failure hours⟩ values were used as independent variables. Specifically the log ⟨median failure hours⟩ determined via the Exxon reformer simulation computer program was independent variable $M1$, the corresponding term based on measured AMT was $M2$, and the term based on measured SMT was $M3$. In other words, a correlation in the form of equation 8 was sought:

$$W = f(M1, M2, M3) \qquad (8)$$

There were additional complicating factors in the analysis, however. The questionnaire requested an estimate of the fraction of time the responding plant actually operated at the reported most severe condition. The responses indicated that this fraction ranged from 20 to 95% depending on the plant. It did not seem reasonable to compare, on an equal basis, a plant operating 95% of the time at peak severity, with another operating only 20% of the time of peak severity. More reasonably, a longer life relative to severity would be expected for the plant operating only a small fraction of the time at the reported most severe condition.

All plants responding to the questionnaire indicated fractional time at peak severity falling into one of the following three groups: I, 20-30%; II, 50-68%; and III, 80-95%.

Since no plants are expected to run at their peak (most severe) condition 100% of the time, group III (80-95%) was taken as the base point for the correlation. Plants in Group I, having only a small fraction of time at their most severe conditions, were assumed to have lives K times longer than predicted by the correlation, where K would be determined by the best least squares fit to the data. Thus, a correlation in the form of

$$W = f([M1 + E\log K], [M2 + E\log K], [M3 + E\log K]) \qquad (9)$$

was called for. E would be 0.0 for group III plants (80-95% at peak severity) and 1.0 for group I plants (20-30% at peak severity). Group II plants, being intermediate between groups I and III, would be assigned an E value of 0.5. By this procedure, the fractional severe operation factor, E, became a fourth independent variable in the correlation.

An additional complicating factor was the fact that not all Weibull median life projections could be made with equal confidence. Plants that had only two failures did not even permit a confidence limit to be estimated (confidence limit calculation formula becomes indeterminate for two points), so median failure times for those plants were considered to carry less weight.

For the plants were confidence limits could be determined (i.e., three or more failures), it was arbitrarily assumed that a median life projection whose 95% confidence limit exceeded a factor of two (i.e., $I > 0.301$) would count for less than life projections with narrower confidence limits. Thus, the Weibull data was split into two groups: one group getting full weight in the correlation, with a new parameter, "*WT*", equal to 1, and another group and lower weight data having "*WT*" equal to 2.

Table 4. Larson-Miller failure time projections

Plant no.	Furnace type	Rows straight or staggered	Larson-Miller log ⟨hours to failure⟩ based on calculated AMT (M1)	Based on measured AMT (M2)	Based on measured SMT (M3)
1	F-W	Staggered	5.22	4.75	4.14
2	Selas	Straight	6.18	—	5.20
3	Selas	Straight	6.06	—	5.35
4	Selas	Staggered	4.61	4.27	3.87
5	F-W	Straight	5.86	5.57	5.22
6	Selas	Straight	5.48	5.25	5.00
7	F-W	Staggered	6.22	6.13	4.38
8	F-W	Staggered	5.66	5.36	4.42
9	F-W	Straight	6.21	5.84	5.44
10	Selas	Straight	7.30	5.99	5.32
11	Selas	Staggered	5.81	6.31	5.92
12	Selas	Staggered	5.99	6.43	6.23
13	F-W	Straight	6.25	6.11	4.94
14	Chemico	Straight	4.99	5.31	4.65
15	Chemico	Straight	5.18	6.23	5.27
16	F-W	Staggered	4.80	5.94	5.34
17	Selas	Staggered	5.52	5.37	4.95
18	F-W	Staggered	5.14	—	—
19	F-W	Straight	6.43	8.18	4.96
20	Selas	Staggered	6.18	6.52	5.93
21	F-W	Straight	5.79	6.34	—
22	F-W	Straight	5.26	5.79	5.50
23	F-W	Staggered	5.28	—	4.33
24	Selas	Staggered	5.14	5.22	4.59
25	F-W	Staggered	5.67	4.98	4.63
26	F-W	Straight	5.64	4.54	4.40
27	Selas	Straight	5.79	—	5.61
28	Selas	Staggered	5.36	—	5.05
29	Exxon	Straight	5.37	—	5.39
30	Chemico	Straight	6.19	6.13	5.53

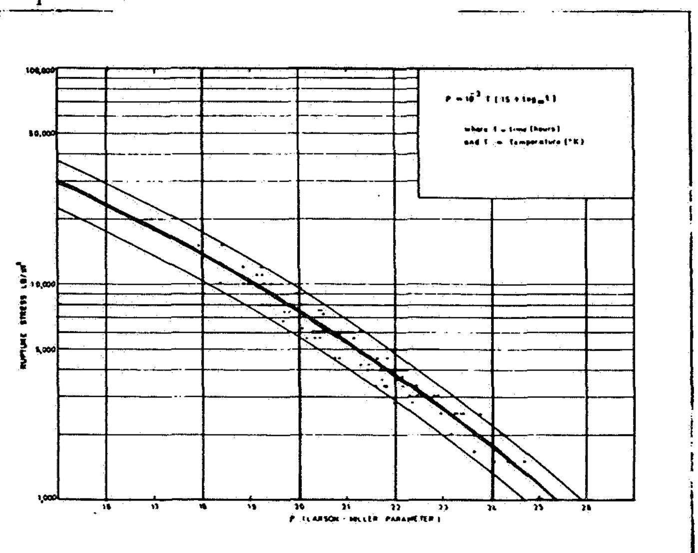


Figure 4. Typical Larson-Miller plot for HK-40 alloy (6). The middle line represents average values.

Table 5 gives a final summary of the data base used in trying out various correlations of Weibull median failure time against Larson-Miller predictions. To keep the resulting models simple, only linear correlations were considered. Thus, all correlations explored fit the general form:

$$W = Z(0) + Z(1)[M1 + E\log K] + Z(2)[M2 + E\log K] + Z(3)[M3 + E\log K] \quad (10)$$

Models considered did not utilize all the terms in equation 10 simultaneously, but set one or more of the Z correlation constants to zero. For each model considered, the non-zero Z terms were determined by a least squares multiple regression of the data. Data considered as carrying less weight ($WT=2$) was handled by dividing the correlation deviations (actual W − predicted W) by a poor confidence weight factor before the squaring and summing step in the least squares determination of best correlation constants. Poor confidence weight factors of 1 (i.e., counting deviations at full value), $\sqrt{2}$ (i.e., counting *squared* deviations at half value), and 2 (i.e., counting deviations at half value) were all tried in the various correlations.

Most of the correlation work was aimed toward estimating W from simple multiplicative functions of $M1$, $M2$ and $M3$. Thus $Z(0)$, the constant term, was set equal to zero. In various trials, W was correlated against $M1$ alone ($Z(2) = 0$; $Z(3) = 0$), $M2$ alone ($Z(1) = 0$; $Z(3) = 0$), $M3$ alone ($Z(1) = 0$; $Z(2) = 0$), $M1$ and $M2$ ($Z(3) = 0$), $M1$ and $M3$ ($Z(2) = 0$), $M2$ and $M3$ ($Z(1) = 0$), and finally $M1$, $M2$, and $M3$ (only $Z(0) = 0$).

Finding the best fit to the data

Comparing the standard deviations (s) determined for the various models, we found that if a single Larson-Miller life projection must be used, the values based on the Exxon reformer computer program results ($M1$) gave the best match to projections based on actual failure data. The $M1$ correlation also yielded lower errors than the correlation based on both measured AMT ($M2$) and measured SMT ($M3$).

Keeping the $M1$ dependency (i.e., $Z(1) \neq 0$), and adding terms for $M2$ and/or $M3$ dependency gave, as expected, lower standard deviations (s). However, both the $M1/M2$ model and the $M1/M2/M3$ models yielded negative values for $Z(2)$, which is not physically realistic (it would imply higher severity gives longer life). The $M1/M3$ model, although giving lower standard deviations than the $M1$ model, was such a slight improvement ($s = 0.341 - 0.380$ vs. $s = 0.347 - 0.388$), that adding the $M3$ dependency cannot be considered as having had any significant effect. Thus it was concluded that the best model for all the points considered was:

$$W = Z(1)[M1 + E\log K] \quad (11)$$

where, depending on poor confidence weight factor assumed, $Z(1)$ was in the range of $0.847 - 0.849$, with corresponding K's ranging from 5.12-5.00.

Since the choice of weight factor did not significantly affect the values of $Z(1)$ and K, it was decided to use the middle values associated with a $\sqrt{2}$ factor. This gave a final model (henceforth referred to as Model I) as follows:

$$W = 0.848[M1 + E\log(5.08)] \quad (12)$$

For Model I, s, the standard deviation, was 0.361.

Note that this model indicates that actual tube life is shorter than predicted from a Larson-Miller analysis ($0.848 < 1$). This could be due to actual TMT's being higher than calculated (note the many points in Table 3 indicating large negative approaches to reforming equilibrium), or due to the effects of cycling, thermal stresses, and operating upsets. The model also indicates that, on the average, a plant operating only 20-30% of the time at peak severity, will have a median tube life of about 5 times as long as one operating 80-95% of the time at peak severity.

No significant differences among furnace designs

To see if there were any significant differences among the various furnace designs with respect to tube life, the same seven correlation models (W vs. $M1$, $M2$, $M3$, $M1/M2$, $M1/M3$, $M2/M3$, $M1/M2/M3$) were checked against data from plants in each of the following furnace design categories (not mutually exclusive): Selas, Foster-Wheeler, straight tube rows, and staggered tube rows.

When the Z values for the various design groups were

Table 5. Summary of data used in developing correlations

Plant	Furnace type	Rows straight or staggered	M 1	M 2	M 3	E	W	WT
2	Selas	Str.	6.18	—	5.20	1	5.36	1
3	Selas	Str.	6.06	—	5.35	5	6.01	2
27	Selas	Str.	5.79	—	5.61	0	4.91	1
6	Selas	Str.	5.48	5.25	5.00	1	5.12	2
4	Selas	Stag.	4.61	4.27	3.87	1	5.01	1
11	Selas	Stag.	5.81	6.31	5.92	0	5.01	1
20	Selas	Stag.	6.18	6.52	5.93	0	5.16	1
24	Selas	Stag.	5.14	5.22	4.59	0	4.82	1
13	F-W	Str.	6.25	6.11	4.94	0	4.91	2
19	F-W	Str.	6.43	8.18	4.96	0	5.11	1
22	F-W	Str.	5.26	5.79	5.50	1	5.06	1
7	F-W	Stag.	6.22	6.13	4.38	0	4.60	1
8	F-W	Stag.	5.66	5.36	4.42	0	4.94	1
16	F-W	Stag.	4.80	5.94	5.34	1	4.55	2
23	F-W	Stag.	5.28	—	4.33	0	5.11	1
18	F-W	Stag.	5.14	—	—	0	4.67	2

compared, it was found that the differences were generally small compared to the standard deviations of the predictions from the various models. Thus, no statistically significant differences can be detected among the various furnace design types with respect to the correlation of tube life against operating severity. This means that the model given by equation 12 is a general one, and can be used for any of the furnace design types included in this survey.

No constant term was included in any of the previously discussed correlation work (i.e., $Z(0) = 0$). As a comparison, a correlation based on weighted average values of W alone, without regard to operating severity, was also made. This correlation sets $Z(1)$, $Z(2)$, and $Z(3)$ equal to zero, thus eliminating any dependency on severity, and just determines Z (0). This approach was tried for all data points together, and separately for the various furnace design data point subgroups.

Comparing the results to those based on assuming $Z(0) = 0$, it was distrubing to note that, with few exceptions, the life predictions based simply on the weighted mean values of W were better (as indicated by lower standard deviations) than those based on operating severity. However, since only plants which had had two or more creep rupture failures permitted a Weibull median failure time estimation, and since the operating times for various plants were all on the same order of magnitude, it should not be too surprising that there is little variation in the *logarithms* of the projected median failure times.

Values of $Z(0)$ were determined for each of the various furnace design types, as were the corresponding standard deviations. These were compared to overall values for the entire population, and it was again concluded that a single common model was adequate. This model, henceforth referred to as Model II, is as follows:

$$W = 5.01 \tag{13}$$

For Model II, s was 0.266 (vs. 0.361 for Model I).

Since models based only on an average value and only on $M1$ were good, one based on both should be better. On this theory, correlation constants to fit the form:

$$W = Z(0) + Z(1)[M1 + E \log K] \tag{14}$$

were developed. As expected, the weighted standard deviations were lower than those for either Model I or Model II. Again using the $\sqrt{2}$ poor confidence weight factor values as a reasonable middle ground, the following model, henceforth referred to as Model III, was developed:

$$W = 3.485 + 0.255[M1 + E \log(11.73)] \tag{15}$$

For Model III, s was 0.229.

Successful prediction of no-failure data

As previously noted, all plants included in the correlation work could be expected to yield log ⟨median failure time⟩ values that were not very much different from each other, because only plants reporting two or more failures were in the sample. A good test of the models resulting from such a sample would be to see what life they would predict for plants which have not had any failures, and to compare those predictions to the actual times the plants have operated without failures.

A difficulty arises here, in that the correlation models developed predict median failure times, but what are really needed for comparison to no-failure data are predictions of *first* failure times. For each of the 16 plants included in the median failure time correlation work, the difference (Δ) between log ⟨first failure time⟩ and log ⟨projected median failure time⟩ was determined. (Δ is, in fact, the log of the ratio of first to median failure times.)

Note that in two cases, very early failures which had been discounted in the Weibull analysis work due to poor correlation with other failure points, were also discounted here, and not considered as first failures. The justification for this is the assumption that these very early failures were the result of a tube structural defect, or some unreported local overheating phenomenon, rather than creep rupture from metal aging.

The first failure vs. median failure time Δ's were placed in ascending order, and used to generate a cumulative frequency diagram intended to approximate the probabilities of Δ's exceeding various values. This cumulative frequency diagram is shown in Figure 5. With this figure, and a median failure time prediction from one of the correlation models, a probability (P_1) of experiencing the first creep rupture failure (due to aging) before a given time can be determined. Note that according to Figure 5, on the average, the median failure time will be about double the first failure time ($\Delta_{50} = 0.295$; log 2 = 0.301), and in 80% of the cases, the median failure time will be at least one and a half times the first failure time ($\Delta_{80} = 0.175$; log 1.5 = 0.176).

Eight of the plants responding to the questionnaire reported no creep rupture failures. For each of these plants the $M1$, $M2$, $M3$, and E values were determined from the reported operating data. These values, well as L, the log ⟨actual operating hours⟩, are summarized in Table 6. From these operating severity values, values of W, the predicted log ⟨median failure hours⟩ were computed using each of Models I, II, and III. From W and L, values of Δ were computed as follows:

$$\Delta = W - L \tag{16}$$

Given the calculated Δ for each case, a corresponding predicted probability (P_1) of having had the first failure on or before the reported failure-free operating time can

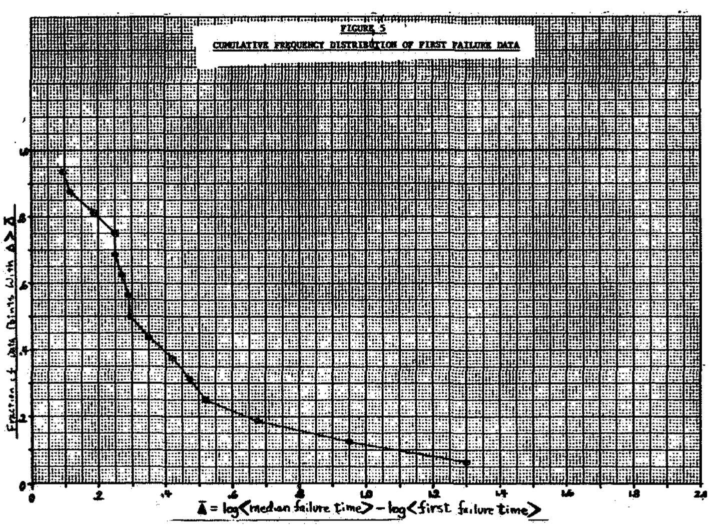


Figure 5. Cumulative frequency distribution of first failure data.

be estimated from Figure 5. Table 6 summarizes the calculated values of Δ and P_1 for each of Models I, II, and III, with cases having $P_1 > 50\%$ indicated by a box. Note that Model I, using only a multiplier on $M1$ and no constant term, would predict probable first failure only for one case. Models II and III, each containing a constant term, would predict probable first failure for five and three cases respectively. Thus, we can conclude that Model I best simulates the actual physical situation.

The Model I prediction for Plant 15 was puzzling since it not only predicted that the first failure time had been exceeded, but that the time to 50% failures had been exceeded as well. The questionnaire data for Plant 15 was reviewed, and an inconsistency was noted. Although no failures or replacements were reported, the ages of the tubes currently in the furnace indicated that 52 (of 234) of them were not from the original batch, and that 12 of these newer tubes had been installed as early as six months after startup.

Thus, it seems apparent that although Plant 15 may not have had any actual failures, it must certainly have had some preventive replacements. Therefore, it is not too surprising that Model I would not predict it to be in the failure-free category. Note that no similar data inconsistencies were found for any of the other plants reporting no creep rupture failures.

Good job of predicting no-failure data

So far, all the models tested against no-failure data have not depended at all on measured TMT's. Since the best model tested (Model I) requires the use of Exxon's proprietary reformer computer simulation program to determine $M1$, it is of limited usefulness to plants not having access to that program. A model that would be more generally useful would be one that depended only on $M2$ and/or $M3$, which can be derived from the AMT and SMT.

If $M1$ cannot be considered, the correlations against $M2$ and $M3$ together were found to be significantly better than those against either $M2$ or $M3$ alone. However, depending on the poor confidence weight factor assumed, significantly different predictions of median tube life for the no-failure plants were reached. These predictions were compared and it was determined that the constants corresponding to a weight factor of 2 gave the best prediction of the no-failure data. The model using this weight factor, henceforth referred to as Model IV, is as follows:

$$W = 0.325[M2 + E\log(5.10)] + 0.560[M3 + E\log(5.10)] \tag{17}$$

The standard deviation for Model IV predictions of actual failure data was 0.480.

Note that the low time at peak severity life increase factor, K, is 5.10, which compares very well to the 5.08 value determined for Model I. Also note that Model IV, like Model I, predicts median failure times that are shorter than those predicted by Larson-Miller theory (i.e., $0.325 + 0.560 < 1.0$).

Table 6 indicates how well Model IV predicts no-failure data. With the exception of the questionable Plant 15 data, Model IV predicts less than 50% probability of first creep rupture failure for all of the plants that actually did not experience any such failures. Thus Model IV can be considered as a useful tool for predicting reformer furnace catalyst tube life. The correlation standard deviations are significantly higher than those for Model I however (0.480 vs. 0.361), so Model I is still preferred where its use is possible.

Assuming that a given plant actually operates at its design severity a high percentage (≥ 80%) of the time, Model I reduces to the simple form:

$$W = 0.848M1 \tag{18}$$

Remembering that W and $M1$ are both logarithms of time,

Table 6. Test of various correlation models against plants with no creep rupture failures

Plant no.	Log ⟨life⟩, L	Operating data M1	M2	M3	E	Model I W	Δ	P_1
5	4.77	5.36	5.57	5.22	1	5.14	0.37	42
10	4.73	7.30	5.99	5.32	0.5	6.49	1.76	>6
12	3.91	5.99	6.43	6.23	0	5.08	1.17	9
15	4.79	5.18	6.23	5.27	0	4.39	(.40)	$\boxed{100}$
21	4.83	5.79	6.34	—	0.5	5.21	0.38	41
25	4.52	5.67	4.98	4.63	1	5.41	0.89	14
29	4.71	5.37	—	5.39	1	5.15	0.44	35
30	4.97	6.19	6.13	5.53	0.5	5.55	0.58	23

Plant no.	Model II W	Δ	P_1	Model III W	Δ	P_1	Model IV W	Δ	P_1
5	5.01	0.24	$\boxed{76}$	5.12	0.35	44	5.27	0.59	22
10	5.01	0.28	$\boxed{59}$	5.48	0.75	17	5.20	0.51	26
12	5.01	1.10	10	5.01	1.10	10	5.58	1.68	<6
15	5.01	0.22	$\boxed{78}$	4.81	0.02	$\boxed{>94}$	4.98	0.19	$\boxed{81}$
21	5.01	0.18	$\boxed{82}$	5.10	0.27	$\boxed{63}$	—	—	—
25	5.01	0.49	29	5.20	0.68	19	4.75	0.32	47
29	5.01	0.30	49	5.13	0.42	38	—	—	—
30	5.01	0.04	$\boxed{>94}$	5.20	0.23	$\boxed{77}$	5.36	0.43	36

the 0.848 factor represents quite a significant reduction in life. Some quantitative examples of the effect of this factor follow:

1. If a set of furnace tubes are designed for 100,000 hr. life ($M1 = 5.00$) based on a Larson-Miller analysis, and no TMT or stress safety margin is used, the actual median tube life would only be 17,400 hr. ($W = 4.24$).

2. If an actual median tube life of 100,000 hr. is desired, a TMT safety factor of 4.48% (based on degrees *Rankine*) must be used. For outside TMT's in the typical range of 1500-1700°F, this corresponds to safety factors of 88-97°F.

It is interesting to note that W. J. Salot came to a very similar conclusion. (*4*) He noted that, on the average, the failure behavior of the tubes in the top-fired reformers participating in the "Reformer Information Network" was the same as that for a creep-rupture test specimen overheated by 120°F above the design calculated mean diameter temperature of the tubes.

For the side- and bottom-fired reformers participating in the current survey, and average 35°F difference between mean diameter temperature and outside temperature was calculated. If this 35°F is added to the required safety margins predicted by Model I (based on outside TMT), a mean diameter temperature safety margin in the range of 123-132°F is indicated. This compares very well to the 120°F margin reported by Salot.

Principal conclusions

Summing up the conclusions of the work to correlate reformer catalyst tube creep rupture failure data against operating severity and furnace type:

1. The Weibull probability distribution is a useful tool for extrapolating cumulative failure vs. time data.

2. The most successful correlation of median creep rupture time against operating severity (based on outside TMT's calculated using the Exxon reformer computer model) can be expressed as in equation 12:

$$W = 0.848[M1 + E\log(5.08)]$$

where W is the log $\langle$median failure hours$\rangle$ from Weibull analysis, $M1$ is the log $\langle$failure hours$\rangle$ from Larson-Miller analysis based on computer calculated outside TMT, and E is the short time at peak severity correction factor.

3. If only measured TMT data are available, the creep rupture time vs. severity correlation yielding the best results is in equation 17:

$$W = 0.325[M2 + E\log(5.10)] + 0.560[M3 + E\log(5.10)]$$

where $M2$ is the log $\langle$failure hours$\rangle$ from Larson-Miller analysis based on measured average maximum TMT (AMT), and $M3$ is the log $\langle$failure hours$\rangle$ from Larson-Miller analysis based on single maximum measured TMT (SMT).

4. Both correlation models indicate that actual tube life is normally significantly less than that predicted from a simple Larson-Miller analysis (assuming continuous operation at the severity level assumed in that analysis).

5. No statistically significant differences among the various furnace design types (Selas vs. Foster-Wheeler and straight tube rows vs. staggered tube rows) were found in the correlation of median creep rupture failure time against operating severity.

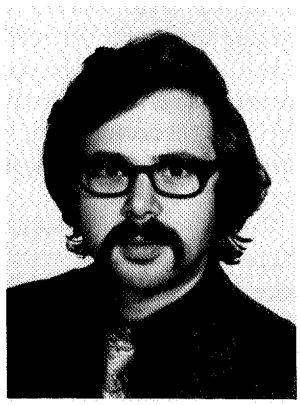
Acknowledgments

The authors wish to thank the 30 ammonia, hydrogen and methanol plants which submitted data to this survey. We will be sending out individual correlation results, including pertinent Weibull cumulative failure vs. time projections, to each of these responding plants.

Literature cited

1. Zeis, L. A., and Heinz, E. "Catalyst Tubes in Primary Reformer Furnaces," 66th National AIChE Meeting, August 25-27, 1969.
2. Salot, W. J., "Anticipating Reformer Tube Life in Normal Exposure Service," 70th National AIChE Meeting, September 1, 1971.
3. Salot, W. J., "Primary Reformer Catalyst Tube Performance Updated," 73rd National AIChE Meeting, August 27-30, 1972.
4. Salot, W. J., "The Trouble With HK-40 High-Pressure Reformer Tube Operation," 78th National AIChE Meeting, August 18-21, 1974.
5. Kao, J. H. K., "A Summary Of Some New Techniques On Failure Analysis," Proceedings 6th National Symposium On Reliability Quality Control In Electronics, 1960.
6. Estruch, B., and Lyth, C., "High-Temperature Alloys For Use In Reformer Furnaces," a chapter in "Materials Technology In Steam Reforming Processes," C. Edeleanu (Ed.), Pergamon Press (1966).

P.A. Ruziska



R.M. Osman

DISCUSSION

BILL SALOT, Allied Chemical Co.: I was pleased to see several correlations between your data and the previously published data on top fired reformers. One might say that your model probably applies to top fired-reformers also. The correlations that I had in mind were:

1. The relation between tube performance and creep rupture test specimen performance was similar for both groups of reformers.
2. The predominant mechanism of failure in both groups of reformers is the same, namely longitudinal creep rupture.
3. In both groups, many reformers have premature failures that cannot be plotted on a single straight line on any set or coordinates.
4. The value of the weibull factor W (the log of the 50% failure time) was in the vicinity of five in most of your cases. In the case of the top fired reformers, it was only slightly more than five. Would you care to comment on the significance of the shape parameter C, which was listed in one of your tables. It's my understanding that when this shape parameter is small, the reformer is doing well and when it is very high, the failures are excessive. If the number exceeds 2.3 for instance, the failures in that reformer are failing more frequently than normal probability would predict.

In the case of the top fired reformers, the average for a large number was 3.2

OSMAN: Bill, with regard to your question about the Weibull shape parameter, C, I'd first like to say that, depending on the value of the shape parameter, you can describe the various portions of the classic bathtub failure curve. If C is less than one, you have the early, "burn-in" period where the failure rate is declining. If C is equal to one, you have the constant rate, random failure period, and if C is greater than one, you have the wear-out period, where failure rate increases. In almost all cases, without throwing out any early data points, we came up with values of C which were significantly greater than one, indicating that the failures were occurring during the wear-out period. In some of the plants, we did have to throw out a few early failures which did not fit the correlation, and which we felt were due to some burn-in problems with tube defects or perhaps unreported upsets.

Considering the significance of a high value of the shape parameter, I would not say that this represents a poor furnace operation. In fact, directionally, I would say that the higher the value of C, the better the furnace operation. This is because the C shape parameter effectively describes the steepness of the failure versus time curve. This means that a plant with perfect firing distribution, identical metallurgy and dimensions for each tube, and perfect flow distribution, would theoretically have every tube in the furnace fail at the same moment. This would give an infinite value of the C parameter. Thus, the higher the value of C, the closer the furnace approximates this ideal situation. Low values of C indicate a relatively sloppy furnace operation, with a relatively large number of early failures due to, for example, poor firing distribution.

Note that the factor which characterizes the average tube life, is really the scale parameter, 0, which relates to the 50% failure time.

JOHN LANCASTER, Kellogg International, London: I'd like to comment that, quite independently of Bill Salot and the Exxon people, we carried out a survey of top fired reformer furnaces in Europe and came up with results which really correlate extremely well with these other two surveys, except that the 50% mean failure time came out somewhat higher; in the range 10^6 hours, rather than 10^5 hours.

But I would like to comment on one particular point and that is the nature of the failures that we are observing. You can, of course, get rupture failures in reformer tubes for a number of reasons. One possibility is that the tube has been operated at its steady design temperature and it's come to the end of its natural life. Other possibilities are that the tube has been locally overheated due to flame impingement, due to catalyst blocking, or due to one of a great number of things.

We are particularly interested in this aspect of the problem because, clearly, if we are dealing entirely with aging failures, then as soon as there are two or three failures in the furnace, pretty soon there is going to be an avalanche of failures and you really ought to change all your tubes. If, on the other hand, the failures are of an accidental character due to flame impingement, etc., then it may be possible to continue operating provided, of course, you replace the defective tube on the next shutdown.

Now there are two or three possible diagnostic methods of determining which type of failure you've got. One of them is by taking microsections of the failure. We found by experiment that long term failures are characterised by the presence of creep cavitation in the microsection, whereas short term failures don't have such creep cavities in them.

Another characteristic is that in long term failures the weld tends to be the weaker part, whereas with short term failures the weld is stronger, so that any swelling that takes place is below and above a weld, and the weld appears to form a constriction in the tube. But the third indication is that, if the slope of the Weibull curve is one or near to one, this is an indication that the distribution of failures is of a random type. This type of distribution is one where you get the same number of failures occurring in a given time, this year, next year and ten years hence. Now we found, and I think everybody else has found here, that, in fact, the slope of the Weibull curve is not one, but is slightly greater than one. However, it's a very long way from the 3.2 slope which would indicate the type of distribution you expect from a pure aging type of failure. And I would interpret the results that we got as indicating that the failures we are seeing are essentially random failures due to local overheating with a certain amount of aging built into it.

OSMAN: John, let me comment on a few of your points. One, you noted that for the European top-fired reformers, the average life is more on the order of 10^6 hours, rather than 10^5 hours reported in this paper. You have to remember that we were using a biased sample here which tended to give shorter failure times. If you recall, the typical 10^5 hours life figure applies only to those 16 plants having two or more creep rupture failures. I'm sure that if you actually operated the rest of the plants responding to our survey, those which currently have no or only one creep rupture failure, out to the point where they had significant numbers of creep rupture failures, you would find that the average time to 50% failures would have been much longer.

Now considering your point about the local overheating phenomenon, we did attempt to split out local overheating and upsets from the correlation work where the failures were reported as such. Also we found for most of the plants, that the Weibull shape parameter C was in fact greater than 3.2, in some cases 8 or 10, or even 19, indicating that the aging phenomenon was very much the dominant reason for failure.

Q. Do you believe there is any correlation between the frequency of thermally cycling the furnace and tube life?

OSMAN: I think it's fairly certain that there is such a correlation, and in fact Battelle has a computer program that enables you to include the effects of thermal stresses and also cycling in your analysis of the operating severity. Now unfortunately, timing and problems with our particular version of Battelle's program did not permit us to explore this, but we can almost definitely say that cycling would have an adverse effect on life. Shortly after startup there are very high thermal stresses across the tube wall, sometimes as high as 10 times the pressure stresses. However, these thermal stresses tend to relieve themselves quite quickly, so that if you only had a single cycle for example, their effect would be close to negligible. However, if you have enough cycles, and each time this thermal stress effect is repeated, we would expect to see a shortening of the life of the tube.

PHIL RUZISKA, Exxon Chemical: I would like to elaborate a little further regarding the practical significance of this correlation method. I think that the quantitative correlation, which includes the effect of severity, can be used for several meaningful purposes by plant personnel. For instance, we can look at the effects of increasing furnace duty, such as operating at higher throughput or higher coil outlet temperature, to determine what effects this would have on tube life and time to first failure.

Secondly we can look at increasing the tube wall thicknesses to determine what effect this might have on extending tube life. Finally, for plants which have not experienced their first failure, we can look at when initial failures are expected, and so can plan an appropriate program for turnaround tube inspections.

Q. In this survey you say you have 30 response, that is a mixture of ammonia, methanol, and hydrogen reformers. How many of the 30 are actually high pressure reformers, and how many are not, which I assume to be in the lower pressure range of operation for methanol and hydrogen? Also, did you make any attempt to distinguish between the data in the ammonia area and the methanol/hydrogen area?

OSMAN: Off the top of my head, I can't tell you what the split was. I believe the majority were high pressure ammonia reformers. We did not in fact make any attempt to split out the high pressure reformers from low pressure reformers. We felt that the severity characterization factor would account for the difference between the types, in the parameter itself.

Q. But you said, in effect, that the severity factor was not a factor in correlating the data. In other words, the fact that it was low pressure/high temperature, or high pressure/low temperature had no factor in the way the data was put together for life predictions.

OSMAN: That's correct. We made no attempt to split out the high pressure reformers from the low pressure reformers in the data workup. This could have been done but we didn't do it. The point is that the severity term which includes temperature and pressure in effect correlates both the high pressure and low pressure furnace experience.