Test Monitoring Center
6555 Penn Avenue
Pittsburgh, PA 15206-4489
(412) 365-1000

MEMORANDUM: 02-102

DATE: November 8, 2002
(Correction Issued December 3, 2002)

TO: Dr. Clifford Venier
Chair D02.B07 Engine Oil Volatility Test Surveillance Panel

FROM: Tom Schofield

SUBJECT: D5800 A & B 2002 Post-Workshop Round-Robin Statistical Summary

A D5800 Procedures A & B round-robin matrix was completed by labs participating in the workshop held on March 13 -14 at PAC's facilities in Pasadena, TX, under the auspices of ASTM D02.B07 Engine Oil Volatility Test Surveillance Panel (EOVTSP). This report summarizes the TMC's statistical analysis of the round-robin data.

Calibration Monitoring Trends and Purpose of the Round-Robin

Attached Figures 1A and 1B are CUSUM severity plots of TMC calibration data since the TMC began monitoring the test in May 1996. The slope of the plots indicates a persistent overall severe trend with the calibration data. The plots show the severity trend is fairly consistent over the entire time the test has been monitored by the TMC. Figure 1B shows that the severe trend continued even after new test targets were introduced and reference oil changes were made effective September 26, 2000 (a round-robin was run at that time to establish the new targets).

Specifically, over the recent one-year period from October 1, 2001 through September 30 2002, the overall severity of the test is 0.9 standard deviations severe of targets. That is, each and every operationally valid reference test reported to the TMC over that one-year period was, on average, 0.9 standard deviations severe of target. (It should be noted that this analysis includes both Procedure A and Procedure B tests, with mostly Procedure B test being reported).

Table 1 shows a breakdown of the one-year calibration results, overall and by oil:

Table 1
D5800 TMC Calibration Data 20011001 – 20020930
Operationally Valid Tests Only

Oil	n	Target Mean	Mean for Period	Target s_R	s_R for Period	Mean Δ/s_R	% Fail Rate
All	68	-----	-----	0.56	0.75	0.90	20.59
52	21	13.61	13.76	0.49	0.66	0.30	9.52
55	23	16.39	17.05	0.66	0.82	1.00	21.74
58	24	14.46	15.15	0.52	0.74	1.33	29.17

Due to the persistent overall severity, the poor precision and the increasingly high fail rate of the calibration tests reported to the TMC as operationally valid, ASTM D02.B07 directed the EOVTSP to conduct a workshop to try to address any procedural differences between labs, and to develop a list of recommended practices to further improve conformance between labs. At the workshop, a consensus for a follow-up round-robin was reached. The purpose of the round-robin is to see if improvements in severity and/or precision could be achieved after the workshop and by following the new recommended practices, and possibly to evaluate if new reference oil targets can be developed from the data.

The Round-Robin Design

The workshop and follow-up round-robin addressed only D5800 Procedure A & B tests (the TMC receives very few Procedure C calibrations). A separate Procedure C workshop was conducted by Tannas Co., with a follow-up round-robin pending under the auspices of D02.06.0B, Task Group on Volatility, Chaired By Mr. Greg Miiller. The Procedure A & B round-robin matrix of this report consisted of:

- 4 oils plus a check oil (check oil data was not included in the TMC's analysis)
- 10 labs, of which 8 participate in the TMC's calibration monitoring system
- 15 separate instruments (with five labs contributing data from two instruments)
- 12 Procedure B instruments and 3 Procedure A instruments
- The 3 labs reporting Procedure A runs also contributed Procedure B data

Each of the four test oils were run in duplicate by each lab on each instrument providing a data set of 120 runs (4 oils x 15 instruments x duplicate runs), with 24 Procedure A results and 96 Procedure B results. The labs were asked to run a CRM check sample to ensure their instruments were properly calibrated before running the other four test oils under repeatable conditions (same operator, same instrument, as close to consecutively as possible). The labs were asked to finish by running the CRM check again, without adjusting the instrument, to see if there was any significant drift during testing.

The Data and Reported Operational Anomalies

The test results were reported to the TMC by spreadsheet. A master spreadsheet of the data is available for download from the TMC's web site:

ftp://ftp.astmtmc.cmu.edu/refdata/bench/d5800/data/D5800_2002_WS_Matrix1.xls

Lab A ran the CRM in duplicate at the start of the test rather than running it first and last, as requested. Lab E is the only lab with a CRM result outside of the nominal range at the end of testing (3.8s mild of target while the suggested acceptance bands are approximately 2s). These deviations did not seem significant enough to exclude any data from the TMC's analysis. All reported data is included in the analysis in this report.

Overall Analysis and New Reference Oil Targets

Table 2 shows the overall statistical summary of the round-robin data. Note there are no targets or prior calibration data for SNA 150 (SNA 150 is not a current TMC reference oil).

Table 2

Overall Statistical Summary of ASTM D5800 2002 Post-Workshop Round-Robin Matrix
Parameter: Sample Evaporation Loss, Mass%
(10 labs; 15 Total Instruments; 12 Procedure B Instruments; 3 Procedure A Instruments)

Data Set	Oil	n	*Mean Δ/s	Mean	s_R	s_r	Acceptance Bands (95%)
Round-Robin 2002	Pooled (All Oils)	90 / 120	0.5 (n=90)	-----	**0.69 (n=120)	**0.37 (n=120)	----------
Current Targets	Pooled (All Oils)	178	-----	-----	**0.56	**0.32	----------
^TMC Calibration	Pooled (All Oils)	68	0.90	-----	**0.75	**0.73	----------
Round-Robin 2002	52	30	0.01	13.62	0.76	-----	12.1 – 15.1
Current Targets	52	59	-----	13.61	0.49	-----	12.6 – 14.6
^TMC Calibration	52	21	0.30	13.76	0.66	-----	----------
CEC 2002 Round-Robin	52	88	-----	13.61	0.58	-----	----------
Round-Robin 2002	55	30	0.49	16.71	0.71	-----	15.3 – 18.1
Current Targets	55	60	-----	16.39	0.66	-----	15.1 – 17.7
^TMC Calibration	55	23	1.00	17.05	0.82	-----	----------
CEC 2002 Round-Robin	55	88	-----	16.57	0.69	-----	----------
Round-Robin 2002	58	30	1.11	15.04	0.73	-----	13.6 – 16.5
Current Targets	58	59	-----	14.46	0.52	-----	13.4 – 15.5
^TMC Calibration	58	24	1.33	15.15	0.74	-----	----------
CEC 2002 Round-Robin	58	88	-----	14.90	0.63	-----	----------
Round-Robin 2002	SNA 150	30	-----	11.76	0.72	-----	10.3 – 13.2

*Mean Δ/s (n=90) compared to current targets (effective 20000926); excludes SNA 150 (which has no targets)
**Pooled precision across all four oils (n=120). ^One-year period 20011001 – 20020930.

D5800 –00a suggests a s_R of 0.57 and s_r of 0.29

15 of 90 (16.7%) round-robin results on Oils 52, 55 & 58 would have failed the current calibration acceptance bands.
5 of 30 (16.7%) round-robin results on Oil 52 would have failed the current calibration acceptance bands.
2 of 30 (6.7%) round-robin results on Oil 55 would have failed the current calibration acceptance bands.
8 of 30 (26.7%) round-robin results on Oil 58 would have failed the current calibration acceptance bands.
Only one Procedure A run would have failed the current calibration acceptance bands.

Comparing the overall performance (Table 2) found in the round-robin to current reference oil targets for Oils 52, 55 & 58 shows the three oils in the round-robin running, overall, 0.5 s severe of the current reference oil targets while reference testing over the past year is running 0.9 s severe of targets. Overall round-robin precision (including the SNA 150 oil) is somewhat worse (at s_R = 0.69) than target precision (s_R = 0.56), but better than the calibration testing (s_R = 0.75). Comparison of the intermediate precision (s_r; within labs) between the current round-robin and the previous round-robin (used to establish the Current Targets in Table 2) shows they are comparable. However, it is interesting to note that the intermediate precision (s_r; within labs) of the TMC calibration data is much worse, and even comparable to the reproducibility precision (s_R; across labs). This finding implies that within lab precision is fairly good for consecutive runs such as found in a round-robin, but over time (as calibration testing is run), the overall repeatability (within labs) is not much better than the overall reproducibility (across labs).

It would not appear, from the round-robin data, or from reviewing the TMC reference testing results from before and after the workshop, that the workshop conducted in March 2002 has had much of an effect on improving the overall severity or precision of D5800 testing.

Looking at the oils individually, the Oil 52 round-robin mean is almost identical to the current target mean, but the round-robin precision is considerably worse than the target precision. Reference testing on Oil 52 is trending slightly severe for the year (Mean Δ/s = 0.3 s). Adjusting the targets and acceptance bands based on the current round-robin will serve only to broaden the bands (allowing more passing calibrations), but would not be expected to significantly change the severity of reference testing on Oil 52.

Oil 55 is showing the round-robin performance 0.49 s severe of current targets, while recent reference testing is averaging 1.00 s severe. The round-robin precision is worse than target precision but better than reference testing precision. Changing targets based on the latest round-robin would serve to shift the calibration testing 0.5 s milder (by moving the target mean 0.5 s more severe compared to the current target mean), and would broaden the acceptance bands somewhat. Calibration tests on Oil 55 would be expected to run about 0.5 s severe of target (rather than 1 s), and more calibration tests would be expected to pass because of the shift in the target mean and because of the broader acceptance bands based on the poorer precision of the round-robin.

Oil 58 round-robin severity is substantially severe (1.11 s) of the current target mean, as is current calibration testing (1.33 s severe). Precision of the round-robin is substantially worse than target, but is comparable to recent calibration testing. Adjusting targets based on the round-robin would move both severity and precision much closer to that seen in recent calibration testing, but would also serve to substantially broaden the acceptance bands (due to the poorer precision).

TMC Oil 58 has a higher “fail rate” than the other two oils, both in reference testing and in the round-robin. However, the precision on Oil 58 is comparable to Oils 52 & 55 (in fact, Oil 55 has the worse precision in reference testing). The data shows that Oil 58 is not an exceptionally variable oil in D5800 testing (as some have suggested), but that the current target mean on reference Oil 58 is, likely, not accurate.

The precision on Oil SNA 150 in the round-robin is no better or worse than on the three TMC reference oils. The performance mean of SNA 150 is the mildest of four round-robin test oils.

Figure 2 (attached) shows how the calibration testing CUSUM severity plot of Figure 1A would look if we substitute the targets suggested by the round-robin (Table 2) for the reference oils since October 1, 2001. Note the slope of the severity trend changes abruptly to less severe at the 01OCT01date line on the plot. This change would suggest what might be expected with future calibration testing should new targets be adopted based on the round-robin data (less severe overall, but still trending somewhat severe).

Moving oil targets closer to the mean results that we are actually seeing in calibration testing may have technical merit (particularly for Oil 58), but broadening acceptance bands because of the poorer round-robin precision is not desirable. Unfortunately, using new means derived from the round-robin data would also suggest using the poorer precision of the round-robin. However, it is incautious to change targets without really understanding the reason that caused the change. (Why is the performance in this round-robin different from the previous round-robin?)

Note that changing targets will not affect reference testing precision (the precision of a data set is absolutely calculated, and not influenced by targets), although using a poorer target precision might make poorer precision in calibration testing look relatively better by comparison.

The Coordinating European Council (CEC) ran a round-robin in 2002 using TMC reference Oils 52, 55 & 58. The CEC's 2002 results on the individual oils are included in Table 2 for comparison. The CEC data is somewhat more precise than the ASTM round-robin, with somewhat milder results for Oils 55 and 58.

<u>Round-Robin Comparison of Procedures A & B</u>

The round-robin data reported to the TMC consisted of mostly Procedure B data (96 Procedure B observations compared to 24 by Procedure A). All three labs reporting Procedure A data also reported Procedure B data. Table 3 summarizes and compares the round-robin results by procedure. For comparison, the CEC's 2002 round-robin results are summarized in Table 4.

Table 3
Statistical Summary of ASTM D5800 2002 Post-Workshop Round-Robin Matrix by Procedure
Parameter: Sample Evaporation Loss, Mass%
(10 labs reporting)

Oil	n A/B	Procedure A Mean	Procedure B Mean	Δ Mean	Procedure A s_R	Procedure B s_R
Pooled (All Oils)	24/96	14.0	14.3	0.3	0.56	0.71
52	6/24	13.3	13.7	0.4	0.66	0.78
55	6/24	16.5	16.8	0.3	0.55	0.75
58	6/24	14.6	15.2	0.6	0.53	0.73
SNA 150	6/24	11.7	11.8	0.1	0.47	0.54

Table 4
Statistical Summary of CEC SG-L-040 NOACK 2002 Round-Robin Matrix by Procedure
Parameter: Sample Evaporation Loss, Mass%

Oil	n A/B	Procedure A Mean	Procedure B Mean	Δ Mean	Procedure A s_R	Procedure B s_R
52	35/53	13.3	13.8	0.5	0.59	0.62
55	35/53	16.3	16.7	0.4	0.75	0.64
58	35/53	14.6	15.1	0.5	0.64	0.62

Overall, the round-robin suggests Procedure B is somewhat more severe than Procedure A, with Oil 58 showing the largest absolute difference, and SNA 150 the smallest. The observed overall difference between the procedures is of approximately the same magnitude as CEC observed. However, the data also provides significant interactions that make defining the magnitude of the procedure differences very difficult to assess in this round-robin.

Interactions

The TMC's analysis of the ASTM round-robin data showed statistically significant (95% confidence) interactions in the data. In the overall analysis (Table 2) not all labs showed the oils performing the same. However, as Figure 3 (attached) indicates, all the labs ranked the oils in the same order, therefore, the interactions are probably not of practical significance in the overall analysis.

However, statistically significant interactions were also observed in the comparison of Procedure A & B analysis (Table 3). Because we are trying to compare the methods directly, the TMC isolated for study the data that was supplied by the three labs that contributed both Procedure A & B data with the premise that this may help to limit lab effects on the analysis. In doing so, we found that two of the three labs ranked Procedure B as slightly more severe than Procedure A, while the third lab reported the opposite observation. This interaction (labs ranking the procedures' performance differently) is illustrated in Figure 4 (attached). Figures 5 and 6 (attached) show that the three labs rank the four oils the same in both procedures.

Because of the uneven distribution of the data (with Procedure B results weighted more heavily than Procedure A), and because at least one of the three labs running both procedures shows conflicting results, the data raises doubts about accurately quantifying the severity differences between procedures.

Lab Hood Effects on Procedure B Testing

Seven laboratories (56 test results) indicated they ran their Procedure B tests in an operating fume hood while five laboratories (40 test results) reported not using a fume hood during testing. Since all labs ran all oils on each instrument, a comparison of the overall mean and precision will give a good comparison estimate of severity differences. Table 5 provides this comparison.

Table 5
Statistical Summary of ASTM D5800 2002 Post-Workshop Round-Robin Matrix by Hood Use
Parameter: Sample Evaporation Loss, Mass%
(10 labs reporting)

Hood	n	LS Mean (all oils)	Pooled s_R	Pooled s_r
Yes	40	14.64	0.54	0.31
No	56	14.14	0.76	0.31
	Delta	0.50		

The round-robin data suggests that Procedure B tests run in a laboratory fume hood appear to be, overall, more severe, and more precise than those run outside of a hood.

Conclusions

1. Overall round-robin precision (including the SNA 150 oil) is somewhat worse (at $s_R = 0.69$) than target precision ($s_R = 0.56$), but better than the calibration testing ($s_R = 0.75$). Comparison of the intermediate precision (s_r; within labs) between the current round-robin and the previous round-robin (used to establish the Current Targets in Table 2) shows they are comparable.

2. Overall round-robin severity is 0.5 standard deviations more severe than the current reference oil targets would suggest (while recent calibration testing is running 0.9 s severe).

3. The round-robin data, and TMC blind calibration testing since the workshop, give no indication that the March 2002 D5800 Procedure A & B Workshop has helped to improve overall test precision. (Though a letter of recommended practices was generated from the workshop, which ultimately should prove beneficial.)

4. Overall analysis by oil shows worse precision in the 2002 round-robin for all three oils (compared to the target precision for each oil from an earlier round-robin), and severe performance compared to current targets on Oils 55 & 58.

5. Changing oil targets to reflect the current (2002) round-robin statistics would serve to broaden acceptance bands (due to the poorer precision), and shift Oils 55 & 58 targets more severe. (The TMC offers no recommendation on whether or not this would be appropriate, other than to say both reference data and the round-robin suggest that the targets mean for Oil 58 should probably be changed). New reference oil targets and acceptance bands based on the round-robin are proposed in Table 2 if the panel wishes to pursue this action.

6. The round-robin data cannot, with reasonable confidence, be used to provide a reliable determination of the procedure severity differences. Comparison of Procedures A & B is inconclusive due to the small sample size of Procedure A runs and the interaction observed in the data trying to rank Procedure A with Procedure B. The overall procedure comparison analysis (Table 3) shows Procedure B to be more severe than Procedure A, and shows reasonably good overall correlation with the proposed "translation" equation of Procedure B Result = 0.027 x Procedure A Result. However, in applying this equation to individual test results in this round-robin, the applicability of the equation to specific test results becomes questionable. Using the translation equation in the instances where Lab U finds Procedure A more severe than Procedure B would be especially ill advised.

7. The round-robin data suggests that Procedure B tests run in a laboratory fume hood appear to be, overall, more severe, and more precise than those run outside of a hood.

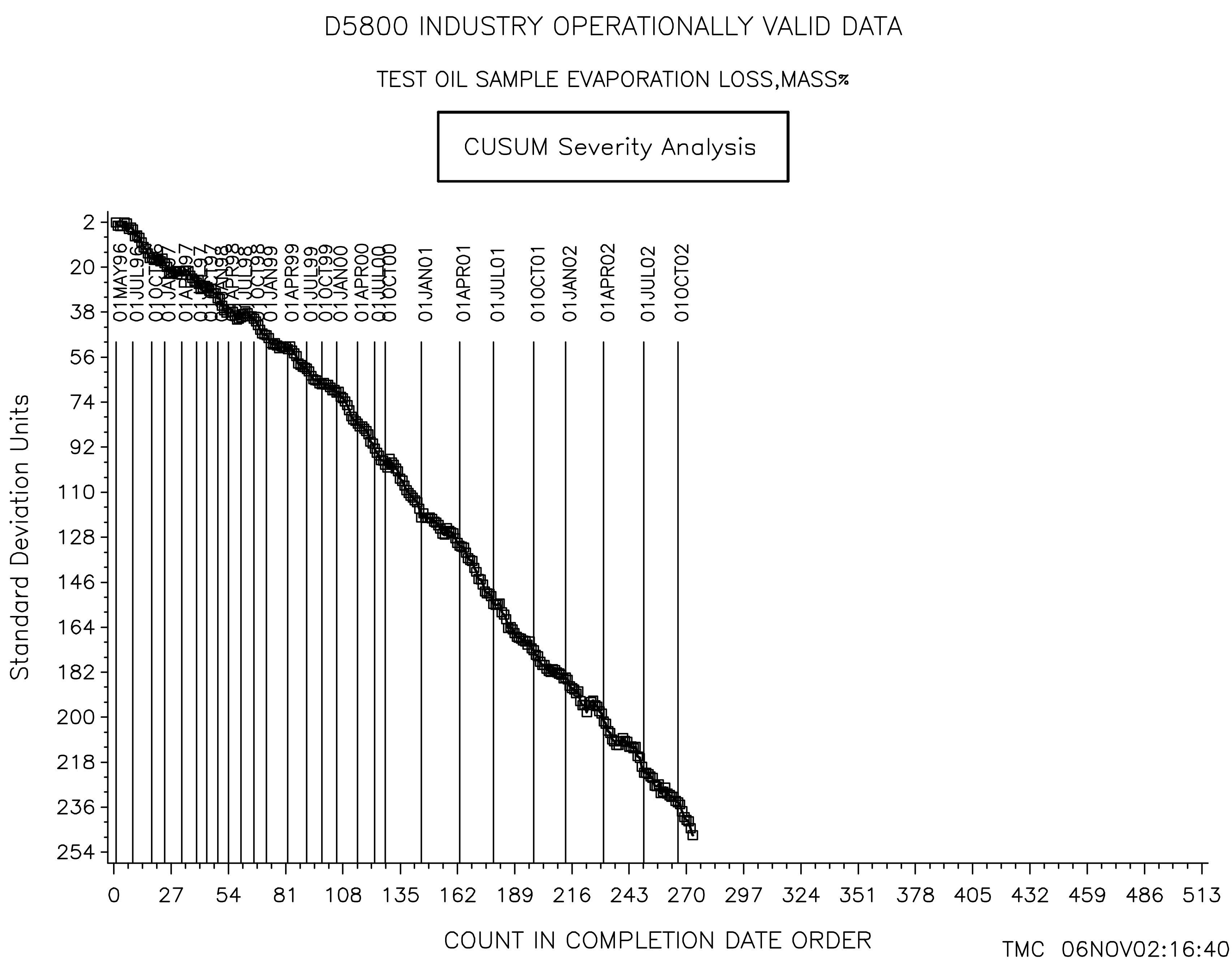
TMS/tms

Attachments

c: D02.B07 EOVTSP and CONTACTS Mailing List
D02.B07 EOV D5800 Mailing List
D02.B07 EOV D5800 Participants
ftp://www.astmtmc.cmu.edu/docs/bench/d5800/mem02-102.pdf

Distribution: Email

Figure 1A



D5800 INDUSTRY OPERATIONALLY VALID DATA
TEST OIL SAMPLE EVAPORATION LOSS,MASS%
CUSUM Severity Analysis
Standard Deviation Units
2
20
38
56
74
92
110
128
146
164
182
200
218
236
254
01MAY96
01APR99
01JUL99
01OCT99
01JAN00
01APR00
01OCT00
01JAN01
01APR01
01JUL01
01OCT01
01JAN02
01APR02
01JUL02
01OCT02
0
27
54
81
108
135
162
189
216
243
270
297
324
351
378
405
432
459
486
513
COUNT IN COMPLETION DATE ORDER
TMC 06NOV02:16:40

Figure 1B

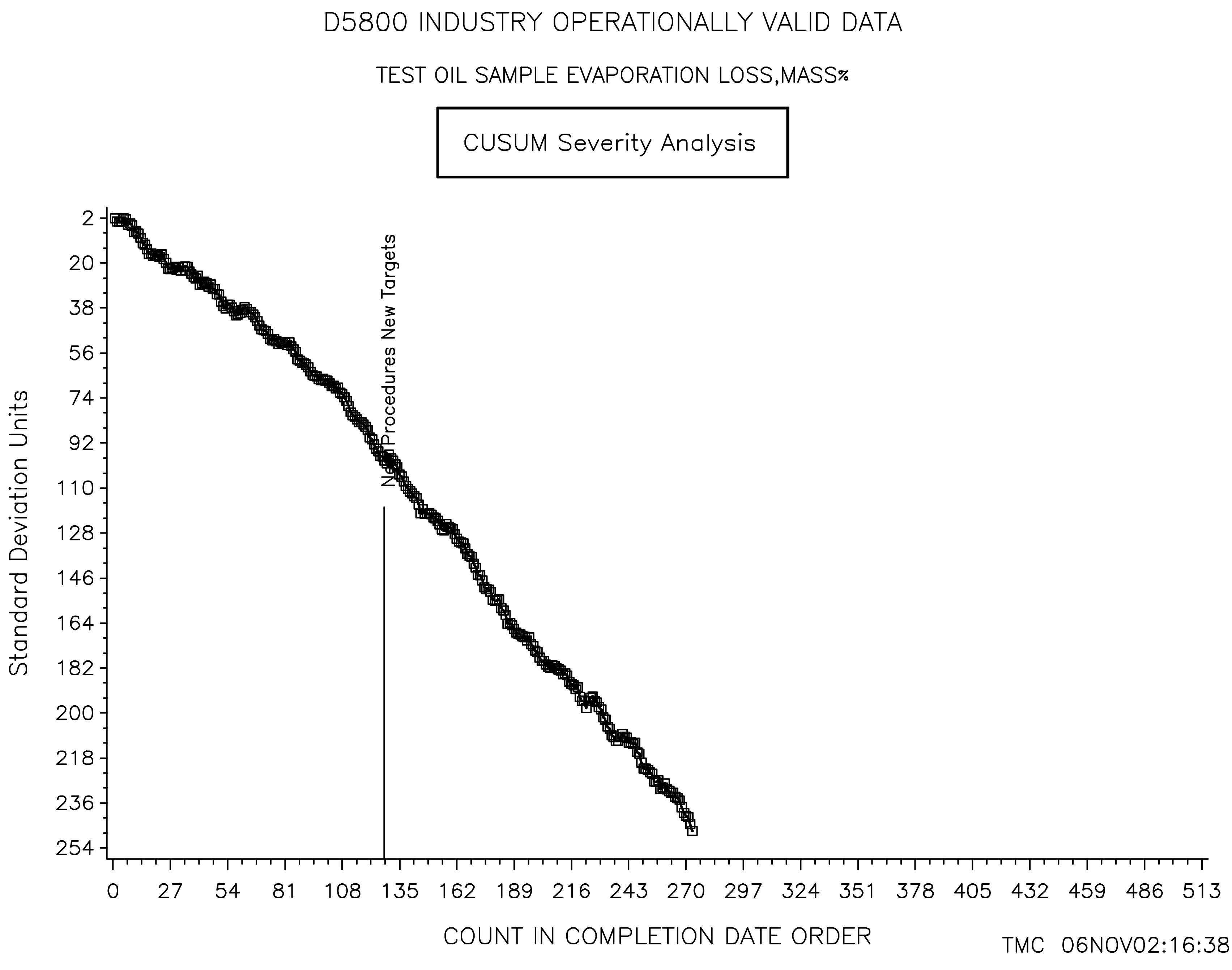

D5800 INDUSTRY OPERATIONALLY VALID DATA
TEST OIL SAMPLE EVAPORATION LOSS,MASS%
CUSUM Severity Analysis
New Procedures New Targets
Standard Deviation Units
2
20
38
56
74
92
110
128
146
164
182
200
218
236
254
0
27
54
81
108
135
162
189
216
243
270
297
324
351
378
405
432
459
486
513
COUNT IN COMPLETION DATE ORDER
TMC 06NOV02:16:38

Figure 2

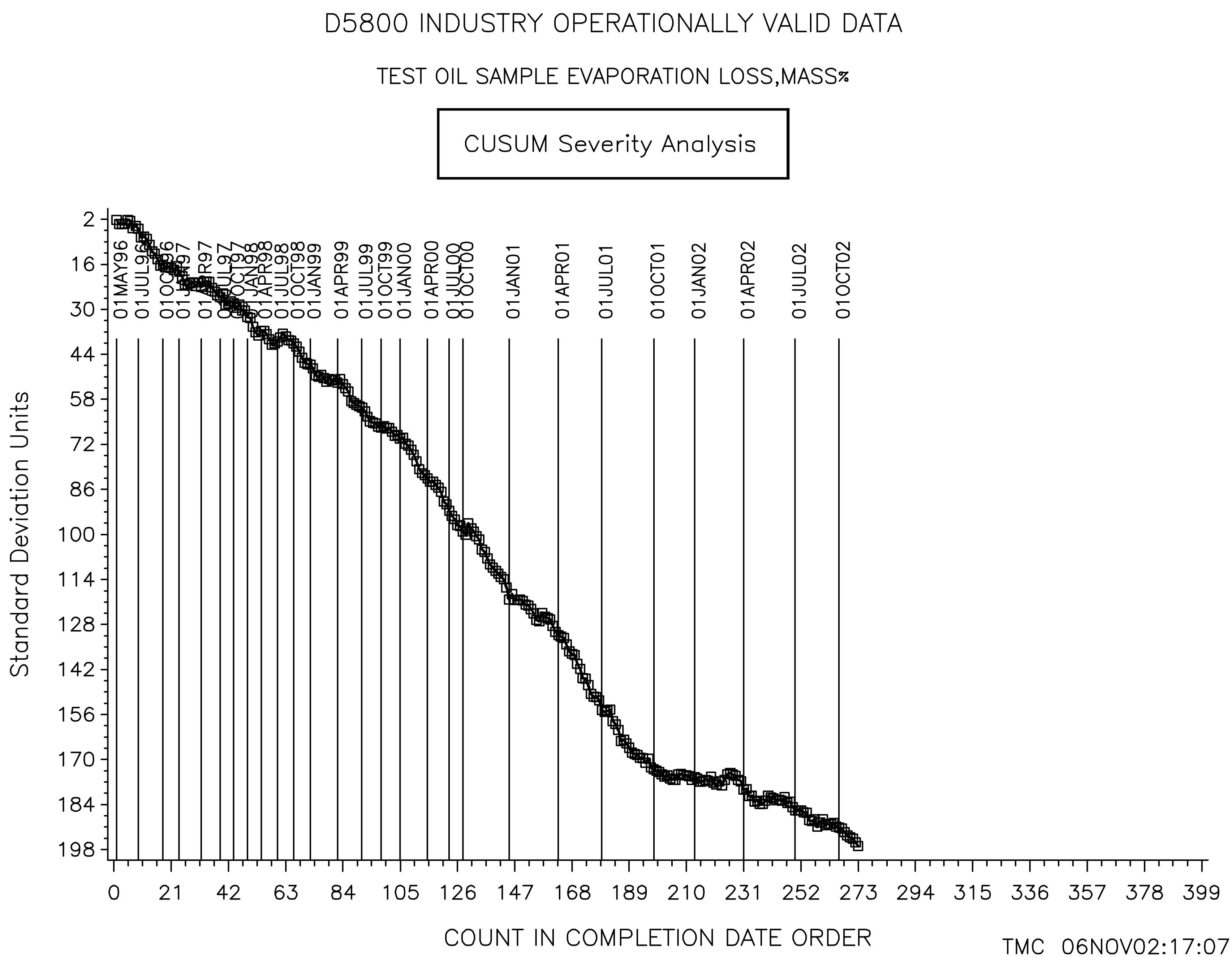

D5800 INDUSTRY OPERATIONALLY VALID DATA
TEST OIL SAMPLE EVAPORATION LOSS,MASS%
CUSUM Severity Analysis
Standard Deviation Units
2
16
30
44
58
72
86
100
114
128
142
156
170
184
198
01MAY96
01APR99
01JUL99
01OCT99
01JAN00
01APR00
01JAN01
01APR01
01JUL01
01OCT01
01JAN02
01APR02
01JUL02
01OCT02
0
21
42
63
84
105
126
147
168
189
210
231
252
273
294
315
336
357
378
399
COUNT IN COMPLETION DATE ORDER
TMC 06NOV02:17:07

Figure 3

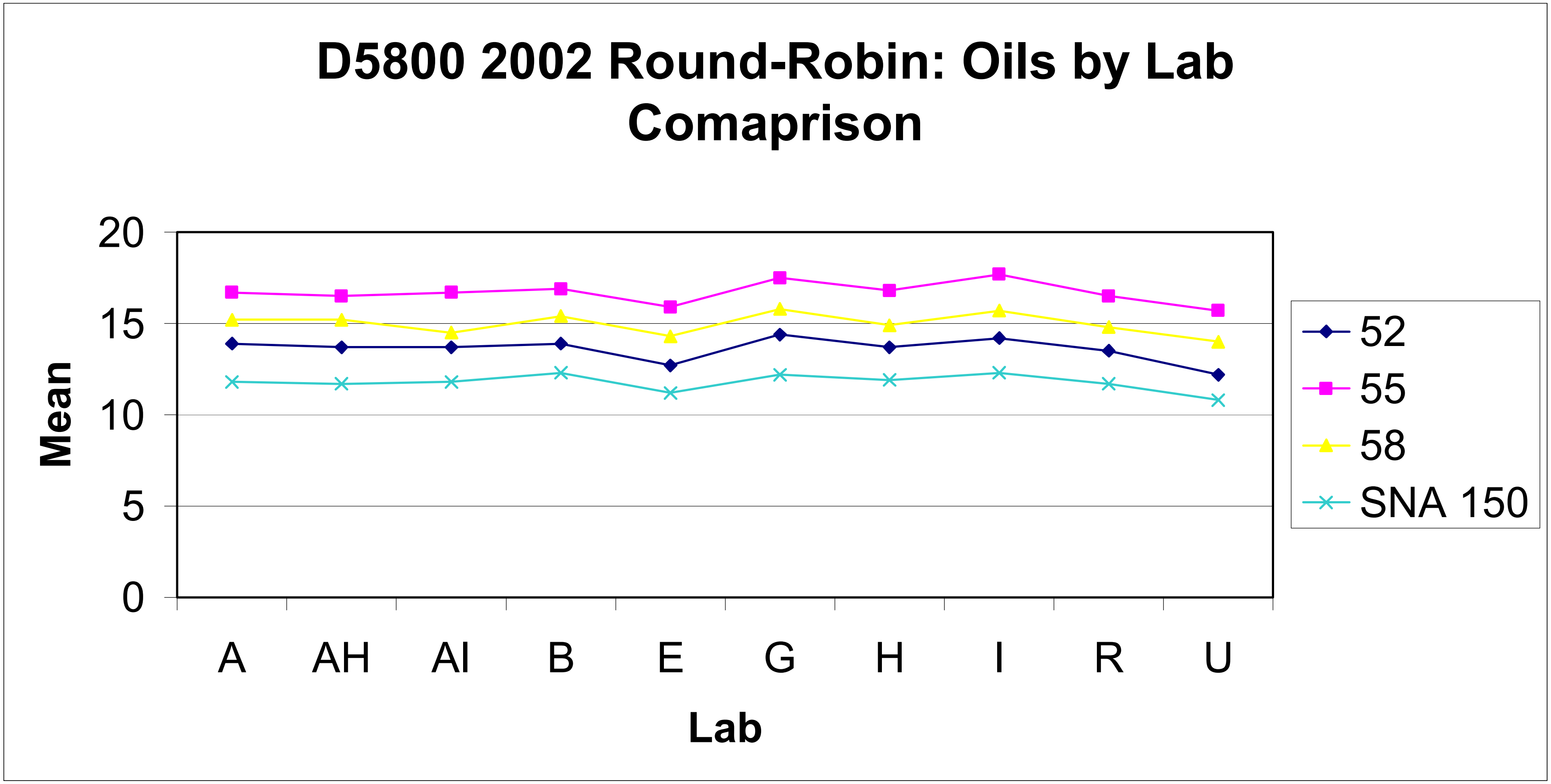

D5800 2002 Round-Robin: Oils by Lab Comaprison
Mean
20
15
10
5
0
A
AH
AI
B
E
G
H
I
R
U
Lab
52
55
58
SNA 150

Figure 4

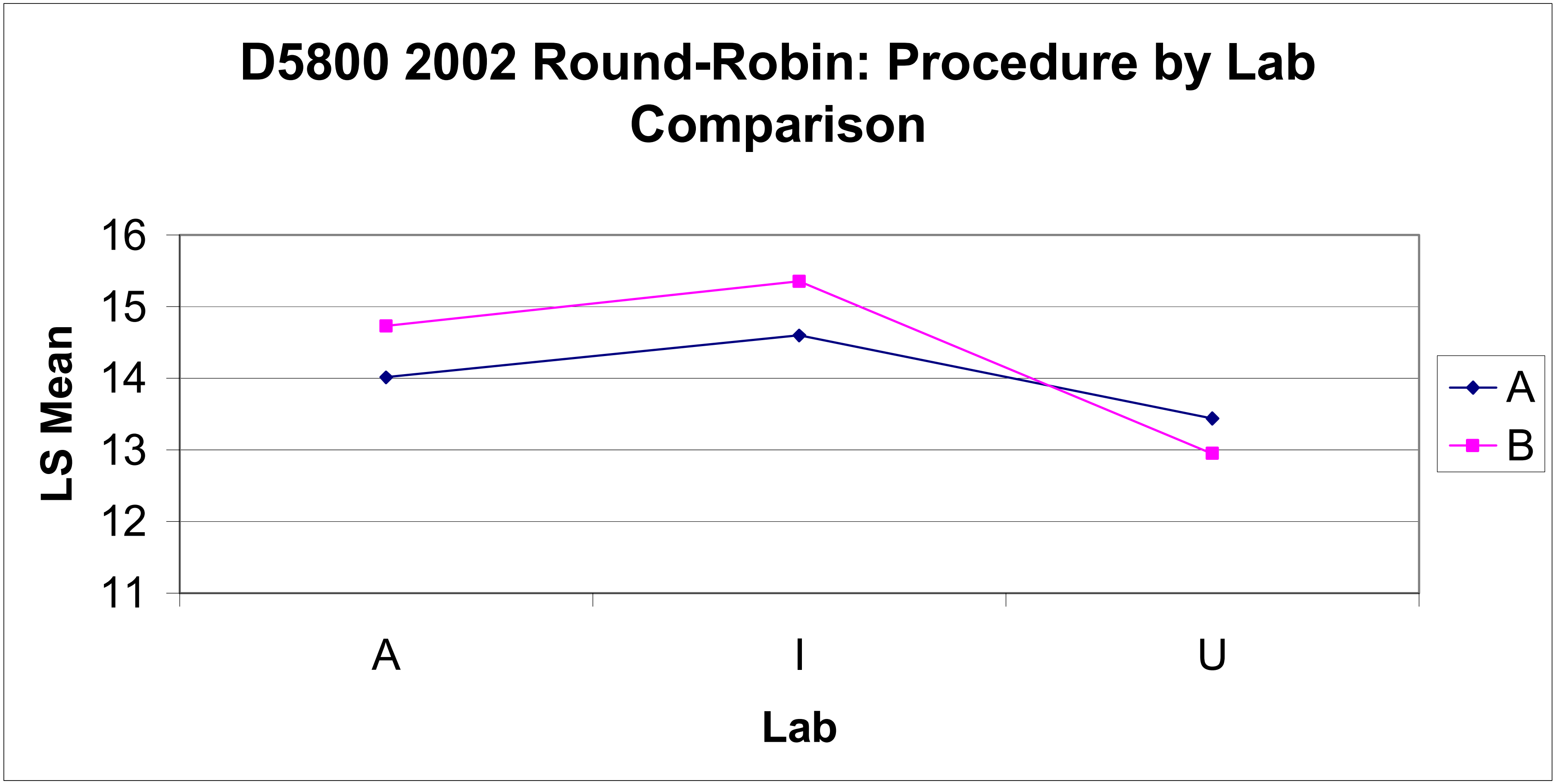
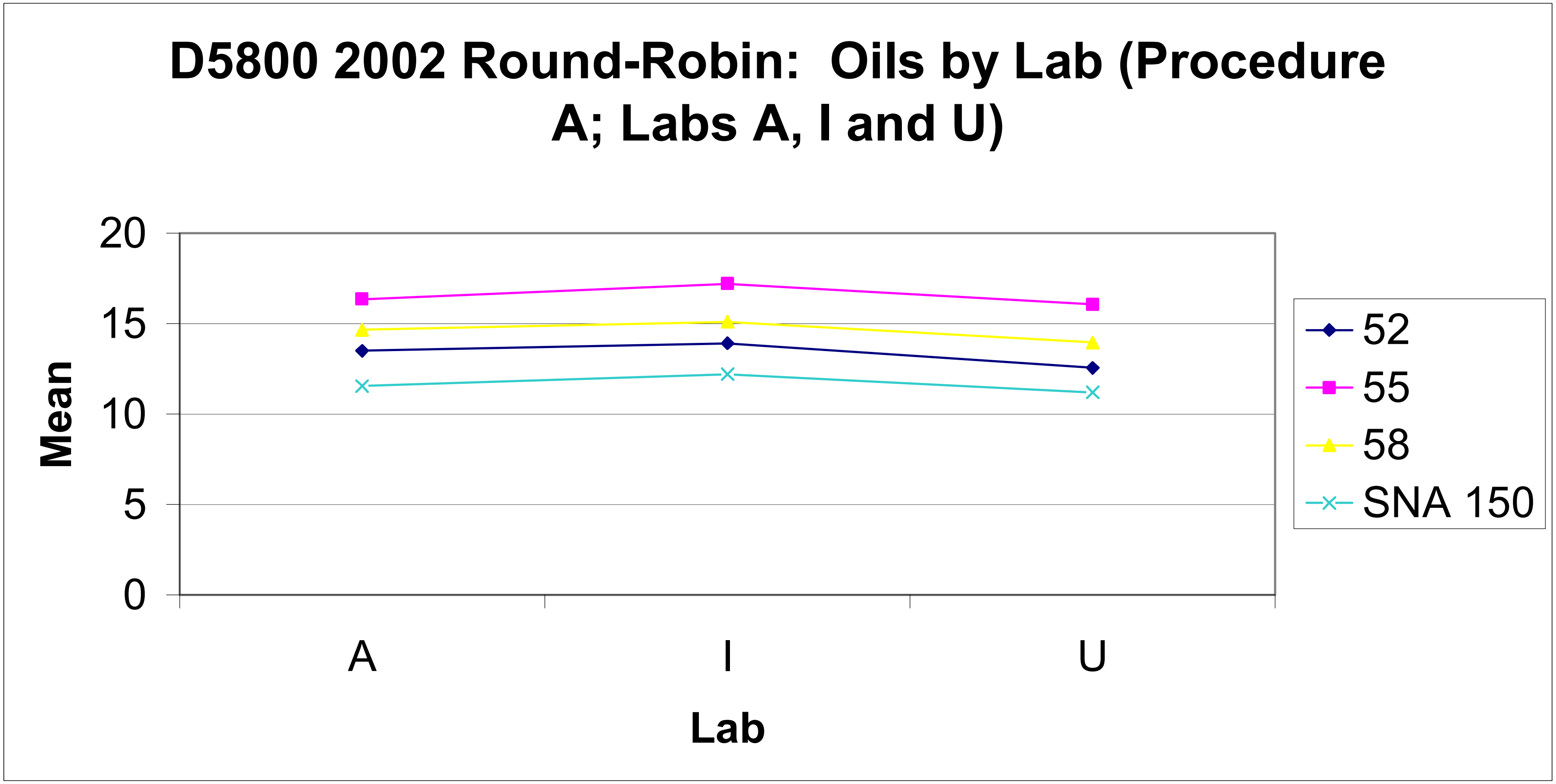
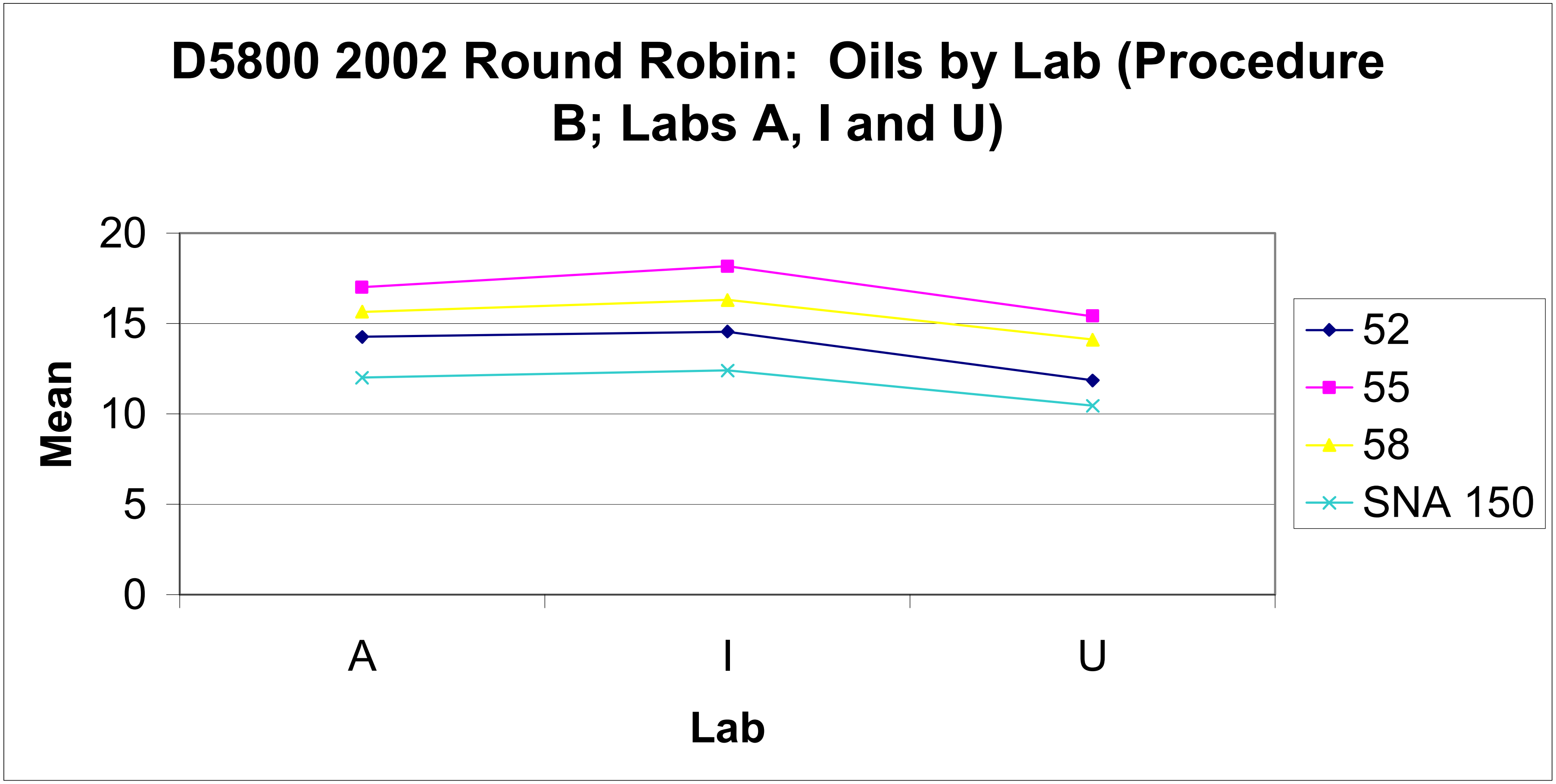

D5800 2002 Round-Robin: Procedure by Lab Comparison
LS Mean
16
15
14
13
12
11
A
I
U
Lab
A
B

Figure 5



D5800 2002 Round-Robin: Oils by Lab (Procedure A; Labs A, I and U)
Mean
20
15
10
5
0
A
I
U
Lab
52
55
58
SNA 150

Figure 6


D5800 2002 Round Robin: Oils by Lab (Procedure B; Labs A, I and U)
Mean
20
15
10
5
0
A
I
U
Lab
52
55
58
SNA 150