

NPS55-83-013

NAVAL POSTGRADUATE SCHOOL

Monterey, California



THE CEMS IV OAP ALGORITHM

by

H. J. Larson

T. Jayachandran

May 1983

Approved for public release; distribution unlimited.

Prepared for:
Kelly Air Force Base
San Antonio, TX 78241

FedDocs
D 208.14/2
NPS-55-83-013

2.10^{ns}
02 1412 NPC-55-83-013

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral J. J. Ekelund
Superintendent

D. A. Schradly
Provost

The work reported herein was supported with funds provided by the
Directorate of Material Management, Kelly Air Force Base.

Reproduction of all or part of this report is authorized.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55-83-013	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE CEMS IV OAP ALGORITHM		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) H. J. Larson T. Jayachandran		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, CA 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS MME-83--004
11. CONTROLLING OFFICE NAME AND ADDRESS Directorate of Material Management Kelly Air Force Base ATTN: Henry Solar, San Antonio, TX 78241 SA-ALC/MMEI		12. REPORT DATE May 1983
		13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Oil analysis program Comprehensive Engine Management System CEMS IV		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The Comprehensive Engine Management System (CEMS) Phase IV, will provide real time data analysis capability for all Air Force oil analysis laboratories. This paper describes the statistical algorithm used by this system to aid the oil analysis technician in making his recommendations. The algorithm incorporates usage and oil consumption variables, and employs least squares to minimize the effects of the random errors in the spectrometer readings.		

THE CEMS IV OAP ALGORITHM

by

H. J. Larson
T. Jayachandran

The Air Force Oil Analysis Program (OAP) uses spectrometric analyses of used oil to monitor the levels of metallic contaminants in the lubricating systems of aircraft engines and other types of equipment. Oil samples are analyzed at more or less regular intervals (every 10 operating hours, after every flight), allowing observation of the temporal evolution of the contaminants being monitored. These temporal observations are used, in turn, to recommend special maintenance actions when required, generally triggered by "high" levels of one or more contaminants.

During fiscal year 1982, the Air Force contracted for the development of the Comprehensive Engine Management System (CEMS), phase IV. This system is to operate at two different levels:

- (a) it is to gather maintenance and other data at the level of the individual base, making this data available at the base level to aid in maintenance decisions.
- (b) a central data bank is to be located at Tinker Air Force Base, bringing together data from all the individual bases.

The OAP data generated by an Air Force base is to be one of the data elements in the base level CEMS IV system. Because of this, upon the adoption of the CEMS IV system it is planned that every Air Force oil analysis laboratory will be equipped with a computer terminal, linking it to the base's CEMS IV computer, allowing real time processing of the oil analysis data.

The use of regression methodology to aid the OAP decision making process has been suggested several times (see references [1] through [6]). The main drawback to implementing this type of approach has been the data-analytic requirements of such a system. Many of the OAP laboratories maintain a heavy workload of sample analyses and are not equipped to handle

extensive number crunching prior to giving a laboratory recommendation, based on the numbers produced by the spectrometric analysis of a used oil sample. Availability of real time data processing, and a direct link between the spectrometer and the CEMS IV computer, will remove this drawback.

For any aircraft participating in OAP, used engine oil samples are removed on a regular basis. The frequency of such samples depends on the aircraft type involved. Some aircraft are sampled after every flight, some every flying day, some every 10 flight hours, etc. In every case the used oil sample is delivered to an oil analysis laboratory where it is analyzed on a spectrometer. The spectrometer produces a digital readout of the part per million content of any of 20 different contaminants; generally the laboratory only analyzes for those metals that are in contact with the lubricating oil in the engine sampled. In current practice, the reading for each metal of interest is subjected to two comparisons defined by the Technical Order (T.O.) table for the specific engine sampled. The T.O. table gives a trend value and a number of range values, for each contaminant of interest. The range values provide absolute limits for the amount of the contaminant, regardless of how long it has been since the last sample was taken or since the oil was changed. Thus for iron, for example, the table may say the normal range is 0-12 ppm, the marginal range is 13 to 15 ppm, the high range is 16 to 18 ppm and the abnormal range is 19 ppm and over. The spectrometer operator then observes where the current iron reading falls, relative to these values, and notes the appropriate recommendation.

The trend value is used to judge the change in contaminant level from the preceding sample to the current reading. For iron, say the T.O. trend value is 4 ppm. The spectrometer operator must then take the iron reading for the current sample, subtract the iron reading for the previous sample,

multiply by 10 and then divide by the change in operating hours. This computation gives an estimate of the rate of change of the iron contamination, standardized to a 10 hour operating period. If the computer value is no more than 4 (for this example) the sample has a normal trend; otherwise the T.O. recommendation is to declare the trend abnormal and appropriate recommendations are made, based on the range and trend results.

The oil sample received by the spectrometric laboratory is removed from the aircraft's oil sump while the engine is still hot, hopefully ensuring that the fluid in the sump is homogeneous and that the contamination in the sample removed is representative of that in the sump. The sample is placed in a small bottle, labelled to identify the aircraft, date and cumulative hours since the oil was changed, and then is sent to the laboratory. At the laboratory the sample is well mixed and then a small part of the oil in the bottle is burned by the spectrometer to produce the ppm contaminant counts.

The counts produced by the spectrometer are in fact observed values of random variables. This apparent randomness in the readings is caused by a number of factors, including the following. If the oil in the same sample bottle is analyzed two or more times, the counts produced are not the same. This variation in readings is caused by voltage fluctuations, temperature variations, actual length of time the spectrometer uses to burn the oil, variations in the actual contamination contents in the small amounts burned, etc. From a broader view, if a second sample were removed from the aircraft sump it is quite likely that the actual contamination levels in the two bottles are not in fact identical and, of course, both may differ from the actual contamination levels of the sump itself, the quantity of interest. Earlier studies ([3], [4]) have shown that the readings produced by the

spectrometer appear to be well described by a normal (Gaussian) distribution. Additionally, it has been shown ([2], [3], [4]) that different serial numbers of the same type of equipment appear to present individual signatures for contaminant buildup in the oil sump, even though they are presumably "identical" in construction.

Previously the only useful environmental variables readily available to the technician in making oil analysis recommendations were the number of flight hours since the oil was changed and the number of hours since overhaul. With the advent of CEMS IV a new environmental variable is now available from other maintenance sources: whether or not fresh oil was added to the sump. Our major task for this year was to provide an implementable statistical algorithm to aid the operator in making recommendations. This algorithm was to take into account (minimize the effects of) the random noise of the spectrometer, the number of flight hours since oil change and the oil addition records.

For the A-10 aircraft, maintenance procedures call for oil to be added (if needed) after the oil sample has been removed for the oil analysis program. The maintenance form on which additions are recorded allows entry of the number of whole units (pints, quarts or gallons) added to the sump; for the A-10 aircraft the unit used is pints. No provision is made for the entry of fractional amounts of units being added to the sump. Thus the oil added values consist of 0's and 1's, indicating whether or not a one pint can of oil was opened and used to top off the A-10 sump. In actual practice, of course, the amount added is generally a fraction of a pint, but this is not reflected in the records available to CEMS IV.

Among other topics, reference [6] discusses the use of oil addition records to estimate wear metal production rates for aircraft engines. This

approach is based on several assumptions which may or may not be universally acceptable for all aircraft types. A tacit assumption apparently made is that oil is lost through a leaking or burning phenomenon and that the metallic contaminants are also lost in direct proportion to the oil lost. That is, if the iron contamination level (as measured by the spectrometer) is, say, 10 ppm and one pint of oil has been added to an aircraft sump which holds 10 pints, reference [6] suggests that the iron contamination level should be "corrected" to read 11 ppm (a 10% upward adjustment to account for the 10% addition of fresh oil). The procedure suggested in [6] also tacitly assumes a relatively accurate record made of the amount of oil added on each occurrence, rather than a simple 0-1 variable indicating whether a one pint can was opened and partially poured into the sump. This suggested procedure then goes on to suggest fitting least squares regression curves with the corrected concentration as the dependent variable and number of flight hours since oil change as the independent variable. The least squares regression approach certainly seems justified, to minimize the effects of the spectrometer errors of measurement of the ppm concentration. The use of the "corrected" concentration, though, does not seem wise for the CEMS IV algorithm, in part because of the crude indication of how much oil was added (and thus how much the spectrometer reading should be corrected). Of equal importance, it may not be true that the iron contamination is lost at the same rate as the oil itself. If an evaporative mechanism were causing the oil loss, it seems possible the iron contamination may not evaporate at the same rate as the oil, if at all; if this were true the "corrected" concentrations would then be too high.

The algorithm employed in the CEMS IV prototype uses least squares regression methodology to minimize the effects of the random spectrometer

errors. It utilizes the number of hours since oil change as an independent variable and, if oil addition records are available, it lets least squares itself determine the "corrections" to be applied to the spectrometer contamination readings. This algorithm will now be described.

The contaminants monitored for the TF34 engine in the A-10 aircraft are Fe, Ag, Al, Cr, Cu, Ni, Ti . These 7 different contaminants are treated separately and in the same way. The following discussion refers to only one contaminant; it is understood that the same procedures are applied to each and that the data from each different serial number are treated separately.

Define

Y_i = Spectrometer contaminant reading for the i^{th} sample.

T_i = Number of hours since oil change when i^{th} sample is taken.

a_i = Amount of oil added to the sump, after the i^{th} sample is sent to the spectrometer.

$X_1 = 0$

$X_i = \sum_{j=2}^i a_{j-1} Y_j / 11$, the accumulated "correction" to the spectrometer reading in the spirit of reference [6]; the TF34 sump is assumed to hold 11 pints, the reason for the divisor of 11.

e_i = random spectrometer measurement error on the i^{th} sample.

Formally, the computations in the algorithm then are consistent with the assumption that

$$(1) \quad Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + e_i ,$$

where the e_i 's are independent normal random variables with mean 0 and variance σ^2 . Standard formulas for unweighted least squares, with two

independent variables (T_i and X_i) are employed to estimate the unknown parameters β_0 , β_1 , β_2 and σ^2 . Detailed definition of these formulas is provided in Appendix I.

The synthetic variable X_i in equation (1) is created from the oil added values (a_i) and the spectrometer contaminant readings; indeed X_i is the "correction" to be applied to the i^{th} contaminant reading by one of the procedures described in [6], granted the full pint was added to the sump. If we were to assume that $\beta_2 = -1$ then equation (1) is equivalent to the wear metal production rate estimation procedure given in [6]. Use of X_i in this way gives (1) an autoregressive flavor. It is interesting to note that the use of regular unweighted least squares on (1) does in fact yield true least squares estimates of β_0 , β_1 and β_2 , as is proved in Appendix II. With the added assumption of normality of the e_i 's, the estimates used are also maximum likelihood.

The CEMS IV algorithm applies two statistical tests to the spectrometer reading, in addition to the T.O. limits mentioned earlier. These two tests are meant to be similar in spirit to the T.O. range and T.O. trend comparisons; they differ from these T.O. comparisons in that they are determined by the historical data base for the serial number being analyzed. These two statistical tests are called the Primary test (similar to the T.O. range comparison) and the Secondary test (similar to the T.O. trend comparison).

After the spectrometer analysis has been made for an incoming sample (say from serial number 1111), the computer calls up the prior data base for serial number 1111, and uses this data to estimate β_0 , β_1 , β_2 and σ^2 from equation (1). This estimated equation then is used to extrapolate forward to the hours since oil change value (T_i) and the X_i value for the new sample

FIGURE 1

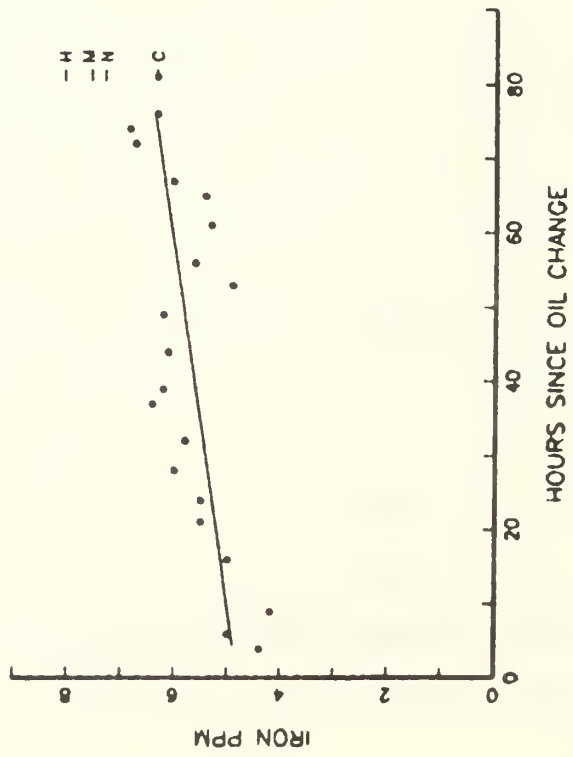
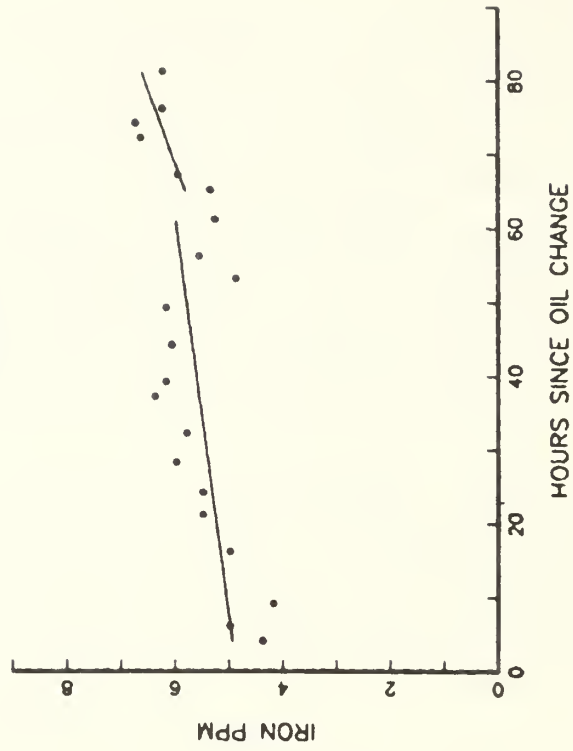


FIGURE 2



just analyzed. This extrapolation produces three numbers, labelled N, M, H in Figure 1 (plotted for a case in which no oil has been added); these numbers are 90%, 95% and 99% prediction limits for what contaminant reading one would expect at this time, based on the data base. The current contaminant reading (just produced by the spectrometer) is labelled C. Table 1 gives the result of the primary test, based on the relationship of C to N, M, H.

Table 1

Current value	Primary test result
$C \leq N$	Normal
$N < C \leq M$	Marginal
$M < C \leq H$	High
$H < C$	Abnormal

As pictured in Figure 1 the primary test result would be normal.

The Secondary test adds the current reading C to the data base and then splits the data base into two parts: the earliest 75% of the data records versus the latest 25% of the data records. Equation (1) is then fit separately to each of these two pieces, giving two estimates of β_1 , the rate of change of the contaminant. These two estimates of β_1 are then compared (using a Student's t statistic) to see if it appears likely that the rate of change in the later data is larger than that in the earlier part. This test is pictured in Figure 2, again for a case with no oil additions. If the computed t statistic exceeds the 95th quantile of the appropriate t distribution, the secondary test produces an abnormal recommendation; otherwise the recommendation returned is normal. For the two slopes pictured in Figure 2, the Secondary test result is normal.

At this point the computer has evaluated four recommendations for each element: the T.O. range and trend values, as well as the Primary and Secondary statistical test results. It then takes the worst case of the T.O. range and Primary tests, and the worst case of the T.O. trend and Secondary tests and uses these as entries in the T.O. decision making guidance table. The result of this is the computer's recommendation for each element for the current sample (standard A, B, C, E, F, H, J, P, S, T, or U as used in the JOAP laboratory manual). This computer recommendation may be accepted (used) by the OAP technician or may be overridden and changed by him if he feels that to be appropriate.

This discussion of the Primary and Secondary tests has referred to a historical data base for each serial number. For many aircraft types, oil changes are widely separated in time and for some types, the oil is never changed. Thus, if all historical data were maintained, the data base could become quite large for each serial number, requiring a very large, accessible data storage facility for each engine at each base. Of equal importance, it seems intuitively reasonable that older data gets "stale", that ancient history has little bearing on the judgment of the current state of an engine's health. Because of this the statistical data base for each engine consists only of the 20 latest historical records which were accepted as being normal for the given engine. The current spectrometer reading makes the 21st record, the largest number used in the statistical algorithm; as each new record is accepted as being normal, it replaces the oldest record in the data base.

When a new engine enters the program, or an old engine has an oil change, for the first 7 records only the T.O. range and trend computations

are used for the computer recommendation. When the 8th record becomes available, the Primary test is also applied (but the Secondary is not); this is also the case for records 9, 10, 11 and 12. The 13th and all subsequent records are subjected to both the Primary and Secondary tests. The statistical algorithm has been programmed in APL at the Naval Postgraduate School and in Fortran by the CEMS IV contractor. Several data sets have been used, giving identical results from both programs.

Formulas for CEMS IV OAP

H. Larson, T. Jayachandran

I Primary Test Model 1

Element values Y_1, Y_2, \dots, Y_{N-1}

Time values T_1, T_2, \dots, T_{N-1}

Computed "oil addition" values X_1, X_2, \dots, X_{N-1}

Compute sums of squares and cross products

$$SS_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N-1}$$

$$SS_T = \sum T_i^2 - \frac{(\sum T_i)^2}{N-1}$$

$$SS_X = \sum X_i^2 - \frac{(\sum X_i)^2}{N-1}$$

$$SP_{YT} = \sum Y_i T_i - \frac{(\sum Y_i)(\sum T_i)}{N-1}$$

$$SP_{YX} = \sum Y_i X_i - \frac{(\sum Y_i)(\sum X_i)}{N-1}$$

$$SP_{TX} = \sum T_i X_i - \frac{(\sum T_i)(\sum X_i)}{N-1}$$

Compute denominator $D = (SS_T)(SS_X) - (SP_{TX})^2$

Compute coefficients

$$\hat{\beta}_1 = [(SS_X)(SP_{YT}) - (SP_{TX})(SP_{YX})] \div D$$

$$\hat{\beta}_2 = [(SS_T)(SP_{YX}) - (SP_{TX})(SP_{YT})] \div D$$

and means $\bar{Y} = \frac{\sum Y_i}{N-1}$, $\bar{T} = \frac{\sum T_i}{N-1}$, $\bar{X} = \frac{\sum X_i}{N-1}$

Compute residual

$$RES = SS_Y - \hat{\beta}_1(SP_{YT}) - \hat{\beta}_2(SP_{YX})$$

Current values for time, "oil addition" - T_N , X_N

Current value for element - Y_N

Compute increment

$$INC = \left\{ \left(\frac{N}{N-1} + [(T_N - \bar{T})^2 SS_X + (X_N - \bar{X})^2 SS_T - 2(T_N - \bar{T})(X_N - \bar{X})SP_{TX}] \div D \right) \frac{RES}{N-4} \right\}^{1/2}$$

Compute predicted element value

$$P = \bar{Y} + \hat{\beta}_1(T_N - \bar{T}) + \hat{\beta}_2(X_N - \bar{X})$$

Compute limits

$$L_1 = P + t_{.9}(N-4)INC$$

$$L_2 = P + t_{.95}(N-4)INC$$

$$L_3 = P + t_{.99}(N-4)INC$$

where $t_{.9}(N-4)$, $t_{.95}(N-4)$, $t_{.99}(N-4)$ are quantiles of the t-distribution, N-4 degrees of freedom

Sample reading is normal if $Y_N \leq L_1$

Sample reading is marginal if $L_1 < Y_N \leq L_2$

Sample reading is high if $L_2 < Y_N \leq L_3$

Sample reading is abnormal if $L_3 < Y_N$

II Primary test Model 2

Element values Y_1, Y_2, \dots, Y_{N-1}

Time values T_1, T_2, \dots, T_{N-1}

Compute sums of squares and cross product

$$SS_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N-1}$$

$$SS_T = \sum T_i^2 - \frac{(\sum T_i)^2}{N-1}$$

$$SP_{YT} = \sum Y_i T_i - \frac{(\sum Y_i)(\sum T_i)}{N-1}$$

Compute coefficient

$$\hat{\beta} = SP_{YT}/SS_T$$

and means $\bar{Y} = \frac{\sum Y_i}{N-1}$, $\bar{T} = \frac{\sum T_i}{N-1}$

Compute residual

$$RES = SS_Y - \frac{(SP_{YT})^2}{SS_T}$$

Current values for element, time Y_N, T_N

Compute increment

$$INC = \left\{ \left(\frac{N}{N-1} + \frac{(T_N - \bar{T})^2}{SS_T} \right) \frac{RES}{N-3} \right\}^{1/2}$$

Compute predicted element value

$$P = \bar{Y} + \hat{\beta}(T_N - \bar{T})$$

Compute limits

$$L_1 = P + t_{.9}(N-3)INC$$

$$L_2 = P + t_{.95}(N-3)INC$$

$$L_3 = P + t_{.99}(N-3)INC$$

where $t_{.9}(N-3)$, $t_{.95}(N-3)$ $t_{.99}(N-3)$ are quantiles of the t-distribution, N-3 degrees of freedom.

Sample reading is normal if $Y_N \leq L_1$

Sample reading is marginal if $L_1 < Y_N \leq L_2$

Sample reading is high if $L_2 < Y_N \leq L_3$

Sample reading is abnormal if $L_3 < Y_N$

III Secondary test model 1

Element values Y_1, Y_2, \dots, Y_N

Time values T_1, T_2, \dots, T_N

Computed "oil addition" values X_1, X_2, \dots, X_N

Data is split into 2 segments

Segment 1 Earliest $N_1 = [.75N]$ values

Segment 2 Remaining $N_2 = N - N_1$ values

For each segment separately, $j = 1, 2$

Compute sums of squares and cross products

$$SS_{Yj} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N_j}$$

$$SS_{Tj} = \sum T_i^2 - \frac{(\sum T_i)^2}{N_j}$$

$$SS_{Xj} = \sum X_i^2 - \frac{(\sum X_i)^2}{N_j}$$

$$SP_{YTj} = \sum Y_i T_i - \frac{(\sum Y_i)(\sum T_i)}{N_j}$$

$$SP_{YXj} = \sum Y_i X_i - \frac{(\sum Y_i)(\sum X_i)}{N_j}$$

$$SP_{TXj} = \sum T_i X_i - \frac{(\sum T_i)(\sum X_i)}{N_j}$$

Compute denominators $D_j = (SS_{Tj})(SS_{Xj}) - (SP_{TXj})^2$

Compute coefficients

$$\hat{\beta}_{1j} = [(SS_{Xj})(SP_{YTj}) - (SP_{TXj})(SP_{YXj})] \div D_j$$

$$\hat{\beta}_{2j} = [(SS_{Tj})(SP_{YXj}) - (SP_{TXj})(SP_{YTj})] \div D_j$$

and means $\bar{Y}_j = \frac{\sum Y_i}{N_j}$, $\bar{T}_j = \frac{\sum T_i}{N_j}$, $\bar{X}_j = \frac{\sum X_i}{N_j}$

Compute residuals

$$RES_j = SS_{Yj} - \hat{\beta}_{1j}(SP_{YTj}) - \hat{\beta}_{2j}(SP_{YXj})$$

Compute test value

$$\text{TEST} = (\hat{\beta}_{12} - \hat{\beta}_{11}) \div \left[\frac{(\text{RES}_1 + \text{RES}_2)}{N-6} \left(\frac{\text{SS}_{X1}}{D_1} + \frac{\text{SS}_{X2}}{D_2} \right) \right]^{1/2}$$

Sample trend is normal if $\text{TEST} \leq t_{.99}(N-6)$

Sample trend is abnormal if $t_{.99}(N-6) < \text{TEST}$

where $t_{.99}(N-6)$ is a t-distribution quantile, N-6 degrees of freedom.

IV Secondary test Model 2

Element values Y_1, Y_2, \dots, Y_N

Time values T_1, T_2, \dots, T_N

Data is split into 2 segments

Segment 1 Earliest $N_1 = [.75N]$ values

Segment 2 Remaining $N_2 = N - N_1$ values

For each segment separately, $j = 1, 2$

Compute sums of squares and cross products

$$\text{SS}_{Yj} = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N_j}$$

$$\text{SS}_{Tj} = \sum T_i^2 - \frac{(\sum T_i)^2}{N_j}$$

$$\text{SP}_{YTj} = \sum Y_i T_i - \frac{(\sum Y_i)(\sum T_i)}{N_j}$$

Compute coefficients $\hat{\beta}_j = \frac{SP_{YTj}}{SS_{Tj}}$ $j = 1, 2$

and means $\bar{Y}_j = \frac{\sum Y_i}{N_j}$, $\bar{T}_j = \frac{\sum T_i}{N_j}$, $j = 1, 2$.

Compute residuals

$$RES_j = SS_{Yj} - \hat{\beta}_j SP_{YTj} \quad j = 1, 2$$

Compute test value

$$TEST = (\hat{\beta}_2 - \hat{\beta}_1) \div \left[\frac{RES_1 + RES_2}{N-4} \left(\frac{1}{SS_{T1}} + \frac{1}{SS_{T2}} \right) \right]^{1/2}$$

Sample trend is normal if $TEST \leq t_{.99(N-4)}$

Sample trend is abnormal if $t_{.99(N-4)} < TEST$

V. Residual SS = 0? test

Residual SS is RES in II

VI. Residual SS = 0? test

Residual SS is RES in I

VII. Residual₁ = Residual₂ = 0? test

Values are RES₁ and RES₂ in IV

VIII. Residual₁ = Residual₂ = 0? test

Values are RES₁ and RES₂ in III

IX. T_i, X_i correlation = 1? test

With definitions in I the correlation is

$$r = \frac{SP_{TX}}{\sqrt{SS_X} \sqrt{SS_T}}, \text{ compare with 1.}$$

When the data is split into 2 segments, compute the correlation separately for each

$$r_j = \frac{SP_{TXj}}{\sqrt{SS_{Xj}} \sqrt{SS_{Tj}}} \quad j = 1, 2.$$

X Oil addition computation

The values for X_1, X_2, \dots, X_N are computed using only the data in the current window and do not depend on earlier values.

Formula used: $X_1 = 0$

$$X_{i+1} = X_i + a_i Y_{i+1} / D$$

where Y_i is the i th element reading, a_i is the i th "oil" value (generally 0 or 1), D is the sump capacity (unrelated to D used in I).

Example

i	OIL = a_i	FE = y_i	X_i	$D = 10$
1	0	4	0	
2	0	3	0	
3	1	3	.3	
4	0	2	.3	
5	1	4	.7	
6	0	5	.7	
7	0	4	.7	
8	0	3	.7	
9	1	3	1.0	
10	1	4	1.4	

Appendix II

For any fixed number, n , of data points the CEMS IV algorithm is consistent with

$$(1) \quad Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + e_i, \quad i = 1, 2, \dots, n$$

where Y_i = spectrometer contaminant reading, sample i

T_i = time since oil change, sample i

$$X_1 = 0$$

$$X_i = \sum_{j=2}^i a_{j-1} Y_j / 11, \quad \text{where } a_j = \text{amount of oil added, sample's}$$

e_i = random spectrometer additive noise, sample i .

In standard least squares notation, define Y to be the $n \times 1$ vector of contaminant readings, define X to be the $n \times 3$ matrix whose first column is all 1's, whose second column contains the $n - T_i$ values and whose third column contains the synthetic X_i values, and define β to be the 3×1 vector with components $\beta_0, \beta_1, \beta_2$. Then it is well known that the simple unweighted least squares estimate for β is given by

$$\hat{\beta} = (X'X)^{-1} X'Y$$

where X' denotes the transpose of X .

To see the form of the least squares estimates of β_0, β_1 and β_2 for equation (1), let A represent the lower triangular $n \times n$ matrix whose first row is all 0's, whose i^{th} diagonal and all elements below it in that column are a_{i-1} if $a_{i-1} \neq 0$, and whose remaining elements are all 0. A

little reflection shows that the elements of the $n \times 1$ vector $V = AY$ then are the values for X_1, X_2, \dots, X_n , the synthetic oil added variable. Additionally, let U represent the $n \times 2$ matrix consisting of the first 2 columns of the matrix X , let γ be the 2×1 vector with components β_0 and β_1 (the first two components of β) and let e be the $n \times 1$ vector with components e_1, e_2, \dots, e_n . Then (1) can be written

$$\begin{aligned} Y &= U\gamma + \beta_2 V + e \\ &= U\gamma + \beta_2 AY + e ; \end{aligned}$$

rearranging terms, this also can be written

$$Y - \beta_2 AY = U\gamma + e$$

or

$$(I - \beta_2 A)Y = U\gamma + e .$$

The unweighted least squares estimates are by definition the parameter values which minimize

$$\begin{aligned} Q &= e'e = ((I - \beta_2 A)Y - U\gamma)'((I - \beta_2 A)Y - U\gamma) \\ &= Y'(I - \beta_2 A)'(I - \beta_2 A)Y - 2\gamma'U'(I - \beta_2 A)Y \\ &\quad + \gamma'U'U\gamma . \end{aligned}$$

Letting $\frac{\partial Q}{\partial \gamma}$ represent the vector whose components are $\frac{\partial Q}{\partial \beta_0}$ and $\frac{\partial Q}{\partial \beta_1}$, respectively, it is easy to see that

$$\frac{\partial Q}{\partial \gamma} = -2U'(I - \beta_2 A)Y + 2U'U\gamma ,$$

from which it is evident that

$$\hat{\gamma} = (U'U)^{-1}U'(I - \hat{\beta}_2 A)Y .$$

Remembering that β_2 is a scalar, a little algebra shows that

$$\frac{\partial Q}{\partial \beta_2} = -Y'(A' + A)Y + 2\beta_2 Y'A'AY + 2\gamma'U'AY ;$$

setting this equal to zero, and substituting the above solution for $\hat{\gamma}$ then easily gives

$$\hat{\beta}_2 = \frac{Y'(\frac{1}{2}(A'+A))Y - Y'U(U'U)^{-1}U'AY}{Y'A'(I - U(U'U)^{-1}U')AY}$$

as the least squares estimate for β_2 .

That these solutions for $\hat{\beta}_0$, $\hat{\beta}_1(\hat{\gamma})$ and $\hat{\beta}_2$ are in fact the same as the earlier quoted simple unweighted least squares values is easily seen by realizing that

$$\begin{aligned} e &= Y - U\gamma - \beta_2 V \\ &= Y - (U|AY)\left(\frac{\gamma}{\beta_2}\right) \\ &= Y - X\beta \end{aligned}$$

since the X matrix is $(U|AY)$ and the β vector is $\left(\frac{\gamma}{\beta_2}\right)$. The well known solution $\hat{\beta} = (X'X)^{-1}X'Y$ then must consist of the subvectors $\hat{\gamma}$ and $\hat{\beta}_2$ defined above. If we also assume that e is multivariate normal with mean vector 0 and covariance matrix $\sigma^2 I$, it is immediately apparent that $\hat{\gamma}$

and $\hat{\beta}_2$ are also the maximum likelihood estimates and the maximum likelihood estimate for σ^2 is

$$\hat{\sigma}^2 = \frac{Y'Y - \hat{\beta}'X'Y}{n}$$

The CEMS IV algorithm actually uses

$$s^2 = \frac{n\hat{\sigma}^2}{n - 3}$$

as the estimate for σ^2 .

References

- [1] Barr, D. R. and Larson, H. J. "Objective Identification Procedures for the Naval Oil Analysis Program", Technical Report NPS-Bn-La 9091A, September 1969.
- [2] Barr, D. R. and Larson, H. J. "Identification of Failing Mechanical Systems Through Spectrometric Oil Analysis", Applied Spectroscopy, Vol. 26, Number 1, 1972.
- [3] Carty, J. J. "An Analysis of Oil Sample Data Obtained from Aircraft Engines by Spectrometry", Naval Postgraduate School Thesis, MS in Operations Research, October 1969.
- [4] Riceman, J. P. "A Statistical Study of Spectrometric Oil Analysis Data from the Naval Oil Analysis Program", Naval Postgraduate School Thesis, MS in Operations Research, October 1969.
- [5] Scheller, K. "Examination of Errors in the Joint Oil Analysis Program Trending Procedures - Suggested Improved Techniques for Trend Analysis of Wear Metal Measurements", Technical Report AFML-TR-79-4194, December 1979.
- [6] Scheller, K. "Statistical Trend Analysis of Wear Metal Concentration Measurements - Calculation of Significant Wear Metal Production Rates", Technical Report AFML-TR-79-4195, December 1979.

DISTRIBUTION LIST

	No. of Copies
SA-ALC/MMEI Kelly Air Force Base San Antonio, TX 78241	20
Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	4
Dean of Research Code 012A Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
Professor H. J. Larson Code 55La Naval Postgraduate School Monterey, CA 93940	45
Associate Professor T. Jayachandran Code 53Jy Naval Postgraduate School Monterey, CA 93940	5

DUDLEY KNOX LIBRARY



3 2768 00329690 6