89-e04

CHRISTINE O. GREGOIRE
Director

Segment No. 25-00-01
WA-15-0020

STATE OF WASHINGTON

DEPARTMENT OF ECOLOGY

7171 Cleanwater Lane, Building 8, LH-14 • Olympia, Washington

July 19, 1989

TO: Nigel Blakely

FROM: Jim Cubbage

SUBJECT: Eagle Harbor Remedial Investigation Draft Report

As you requested, I have reviewed portions (Chapters 2-4) of the Eagle Harbor report and my comments follow. I appreciate the work that has gone into this report. It is obviously a large effort. I was unable to review the risk assessment portions.

Generally, I am discouraged at the quality of the research and presentation. Throughout the report in the sections I reviewed, important assertions are made without statistical tests, references, or qualifications. The myriad problems with the data are dismissed. What were originally considered qualitative results because of analytical problems become significant to three decimal places (e.g., HPAH/LPAH ratios) in some parts of the report. Results are poorly summarized so that the reader must refer to the raw data in the data report. Calculations are often made without showing the data, the equations, or statistical measures of confidence.

Regrettably, the report is founded on questionable chemistry data. Most of the PAH data failed several quality assurance tests and are estimates only. In some cases the estimates are termed qualitative in the data report. Yet they are used in the RI report often without any qualification or caution. Different data sets are merged and averaged without regard to the different sources and quality of data. The qualitative attributes of the data must be shown along with all data presentations.

This work will eventually be judged on its ability to guide the cleanup of Eagle Harbor. Now I am judging it on standards of scientific integrity and veracity. I suspect that the latter standards will be applied to the work later when the ROD (record of decision) is released. Basic problems with the work may go unchallenged for a long time, but the silence of potential responsible parties could reflect caution rather than acquiescence.

I hope these comments are helpful.

Report needs the executive summary.

Chapter 2. Sampling and Analysis Procedures

Page 2-18 2.2.6.3
Current meter deployment duration and location is vague. Table information about the current meters. Show site number, when they were deployed, retrieved, and hours of data gathered at each station.

P 2-42 2.5.2.1 first paragraph
I agree that the data include a significant number of results with qualifiers. Most all the results are qualified. Some studies have very few qualified data, or they are qualified because the results are below contract required detection limits. In this study many of the data are qualified for several reasons (problems of calibration, elution time, surrogate recoveries, matrix spike recoveries, and matrix spike duplicate precision).

I do not have EPA CLP procedures for HPLC analysis of PAH. I am applying the procedures for GC-MS QA. While the qualifiers may have been assigned in accordance with EPA CLP procedures, the QA in the analysis was not. The EPA CLP requires re-analysis if the surrogates are outside the recovery limits. There is no evidence in the data report that this was done.

I am less sanguine than the report when it says "In describing the nature and extent of contamination, the data has been used in a way that does not place undue weight on any one analytical result." Some data have multiple problems that are not clearly highlighted. Results are then used to draw boundaries of trouble spots and calculate ratios of contaminants for source identification.

The analytical quality assurance suffers from lack of replicates (split samples from one source submitted to lab blind) to assess laboratory variability. Field duplicates were taken (multiple grabs at the same station) but these do not assess laboratory precision independent of the field variability.

P 2-45 - 2-46 Figures 2-7a,b,c
Report statistics (n, p) in these analyses. Table the data set being analyzed. To show correlations, both axes should have the same scale so the viewer can see if the correlations reveal bias. What units are the scales in Figure 2-7b?

P 2-47 Table 2-4
In comparing HPLC to GCMS with correlations, Raleigh Farlow removed 4 sites from consideration because they didn't agree very well. This table looks like the same data without the description that some sites are not considered (EH716, EH725, EH659, EH120). He wrote "Results of the HPLC analyses for these stations, when compared to the GC/MS RAS and SAS results, suggests that there may be a discrepancy or error in the analyses or reporting. The HPLC results do not appear to be entirely representative of the station sample. Resolving this discrepancy may require resampling and/or re-analysis." Please respond to his concerns.

Were re-analyses performed? Are the correlations as apparently robust if these problem sites are included? What percent of the total are these sites (appears to be about 20%).

P 2-50 2nd Paragraph
"The analytical variabilities demonstrated within a method and between methods were comparable to the measurable differences in blind field duplicate pairs and the differences between adjacent stations." I don't understand. I would expect the field variabilities to always be higher than the lab. Conversely, could the field variability ever be less than lab variability? (the field variability includes the lab variability). Are you testing variance between methods on split samples against variance within a method between field duplicates? Please show the data sets and please compare variances statistically on log transformed data.

Chapter 3. Physical and Ecological Characteristics of the Study Area

Overall Comment: I would appreciate some way to discern what are data in the section and what are model outputs. One way to show this would be to superimpose current meter data vectors on models of tidal ebb and flow.

Page 3-37 Figure 3-7, 3-8
I don't understand how this relates to the RI.

Page 3-39 Figure 3-9
Show the tides in Eagle Harbor derived from the tide gauges.

Page 3-42 3.3.2.6
What size sediment is capable of being transported by the ferry? (sand?, silt?). What are the inputs to the propeller transport model? (depth, prop pitch, engine horsepower, direction of rudder, rpm) (This might have been answered in the appendix I just received). "The movement of sediment is significant." Define significant in proportional terms.

Page 3-43; 3-48-3-49 Figure 3-10,11,12
Include numbers (speeds, angles, depth).

Page 3-58 Figure 3-14
Prediction doesn't match reality in figure 3-16 Page 3-72.

Page 3-73 Figure 3-17
Unexplained gap in shading appears over shoal. No datapoints at gap to predict the gap. Please reconcile.

Chapter 4. Nature and Extent of Contamination

A general comment: The statistical and methodological uncertainties in the data derived from HPLC PAH analysis are not clearly shown. Some tables show no qualifiers. The reasons for qualification are not always clear. Assertions are made without statistical tests. Please preserve the understanding of the limits of the data.

Page 4-25 Table 4-1
Any time LPAH and HPAH values are shown they must carry the qualifier if any of the component compounds were qualified. (i.e., LPAH should be qualified if naphthalene was qualified.)

P 4-21 - 4-23 Figures 4-6, 4-7, 4-8
Caption should note these are estimates because of unacceptable surrogate recoveries, precision, and matrix spike recoveries.

Page 4-30 4.1.3
Underlined assertion about amounts of NCAC compared to PAH needs a citation. Also is this true for PAH and NCAC in the environment after weathering or only fresh product in the tank? Note that NCAC/PAH ratios will be compared later in the report.

P 4-31 Table 4-2
Caption for 'J' states "value is qualified for use" implying that the samples passed some test not that the analysis failed several. More appropriate wording might be:
"J = means value is estimate due to quality control exceedences." or
"J= the associated value is an estimated quantity".
(Later in the report the phrase "qualified for use as an estimate" appears regularly. Again, the phrasing implies something passed a test and is now qualified. I offer the suggested phrases above as being more descriptive and clear about why the 'J's are appended.)

P 4-32 4.1.3.2 Last para
"..NCAC in clam tissues do not reflect the apparent sources of NCAC at intertidal seeps..." What does this sentence mean? That the PAH/NCAC ratios don't match? That the concentrations in tissues are low compared to the seeps? Please elaborate.

Last sentence speculates that high solubility of NCAC may explain low concentrations in tissues. Please verify: Handbook of Chemistry and Physics lists Carbazole as insoluble in water. If NCAC were highly soluble in water, then it probably would not be found in sediment either.

P 4-34 4.1.3.2
Putrefaction as source of indole needs citation.

P 4-35 4.1.4
Chlorophenols are important in the investigation and were one of the data gaps identified by the Preliminary Investigation. "The estimated qualities are possibly low" should be changed to "probably low."

P 4-35 4.1.4.2 "Pentachlorophenol was indicated in trace amounts in all samples." This sounds like it was found above detection limits in all samples. Is that what is meant? Please clarify. The QA report states that matrix spike data are unusable and that the low detection quantifications are subject to random errors and perturbations (noise). The "J's" are assigned because of the low detection limits attempted. How much faith do the authors have that every clam tissue had pentachlorophenol?

P 4-39 4.1.5.2
"Concentrations of PAH detected in clam tissues by SAS agreed very well with concentrations in split samples analyzed by HPLC." All except one value from SAS were flagged UJ meaning the detection limit was an estimate. In other words, no semivolatiles were found in any SAS analyses of tissue except for Benzo(a)pyrene in sample EH-006L. It is not clear to me how this is can be considered good agreement to the HPLC analyses which reported concentrations between .05 and 3 times the SAS detection limits. Please use a statistical test to compare these samples or clarify that the SAS detection limits were inadequate to compare the different methods.

P 4-50 4.2 (3rd paragraph)
I do not agree that the data between the RI and the earlier PI are comparable. From Figure 4-12 I see great variability between studies and in the case of mercury, a clear bias of about 2 orders of "e" (why were these log transformed to base "e" rather than base 10?) so that this study finds higher concentrations of mercury than the Preliminary Investigation. Please clarify or acknowledge the differences.

To follow-up the analysis performed in the RI, I replicated the comparison of RI sites to PI sites. I compared TPAH in sediment taken from the two studies that were within 150 meters of each other (the RI states that they compared adjacent sites). Table 1 reviews the data extracted from the RI reports Table 4-7 (the Table does not show which analyses (SAS or HPLC) produced the results). Figure 1 depicts the correlation in log base 10 units. As you can see, there is no statistically significant ($p>0.05$) correlation between the concentrations found in the Preliminary Investigation (PI) and the concentrations found in this Remedial Investigation. If the outliers are removed, the correlation coefficient decreases.

Recall the PI had good quality assurance. The values are not qualified nor are they considered estimates. They are based on the isotope dilution method and loss in extraction and analysis is corrected for each sample. Thus, the PI had the better and probably more accurate sediment data than the RI.

I am not sure that the RI is justified in merging the two data sets. The fundamental differences in methods and results quality between the investigations must be highlighted. The ARARs will likely be based on the Apparent Effects Threshold model and most of the supporting data were gathered with the isotope dilution method used in the PI. I appreciate the high cost of the isotope dilution method. The RI must appreciate that the data from HPLC had problems and may not be strictly comparable to past work. If AET's were calculated based on the current RI, please present them. As one solution I suggest that the RI and PI data sets be separated for the kriging exercise. Create isoconcentration lines of the same value for each separate data set and compare the results. If they delineate the same areas, my concerns are mollified.

P 4-53 4.2.1.1
"Because pyrene was not included, the kriging results underestimate the average concentrations of HPAH by about 25 to 27 percent." This implies an accuracy in the analysis that may not be warranted. For example, all eight matrix spikes (HPLC) had some PAH recovery out of control limits (even for UV detector) and varied between 0% and 559% recovery. Apparently all samples had surrogate recovery problems (Data report). A more accurate sentence might read: "Because of analytical problems, the kriging results may underestimate HPAH in sediments and are considered estimates only."

P 4-54 Table 4-7. No qualifiers are shown for the data! Reveal the problems with the analyses here.

P 4-58 - 4-60 Figure 4-13 - 4-15
Spiffy models and good graphics. How were the contour lines chosen? The progression between lines is not intuitive (i.e. They appear to increase by the powers of the square root of e) and interpolation of values between lines is difficult. I would like to see the models overlaid on a map showing positions of all sampling stations so I could see where the model was on solid ground and where it was guessing. I suspect the lines out of the harbor and deep within the harbor are guesses and maybe should be excised. Perhaps some of the lines should be a range of AETs. The lines do not seem to correspond to intertidal seep data from Yake and Norton.

P 4-66 4.2.1.2
"The June sampling focused on areas where benthic invertebrate AETs were exceeded, and concentrations of PAH were higher." This implies the differences between March and June were caused solely by sample location. There were other sources of differences. In the RI, figures on pp 6-90 and 6-91 show the differences occurred by site. Figure 2 and Table 2 attached compare TPAH in sediments from March and June samples at paired sites only. Obvious problems are apparent. The June samples reflect poorly the the March samples. Worse, for concentrations in the low AET range (near 20000) June samples are biased higher than March. These results further document problems with the HPLC data precision and probable accuracy. Please discuss these variations more thoroughly than in the one sentence quoted above.

P 4-79 Figure 4-28
Place all graphs on common scale. I would like a map showing only those sites where TPAH at depths were measured (including Hart-Crowser).

P 4-120 4.5.1
Second paragraph needs citation for assertion about the unique assemblage of chemicals found in comparison to other urban bays. My concern is that I don't know of other urban bays where NCAC were sought.

P 4-121 4.5.1 Use citations for assertions about the composition and effects of weathering of creosote.

P 4-123 4.5.1
"Their concentrations in sediments generally correlated with each other and with PAH levels" What does generally correlated mean? What were "r" values and "p" values? The PI could not correlate NCAC's with PAH.

P 4-124 4.5.3
"... had HPAH:LPAH ratios of 0.28 to 0.58, regardless of whether acenaphthalene, pyrene, benzo(g,h,i)perylene were included in the ratios." Clearer wording might be "depending on" instead of "regardless of".

P 4-127 4.5.3
Entire 2nd paragraph is speculation.

P 4-128 Figure 4-42
I am concerned that too much is being made of HPAH/LPAH ratios when they are based on imprecise data. Table 3 shows comparisons of these ratios for GCMS vs HPLC for the same samples. HPLC ratios are low compared to GCMS (SAS). Blind analytical replicates were not submitted to the laboratory (Blind replicates are recommended in Puget Sound Protocols; page 43 organics section) and thus analytical precision of HPLC analysis of field samples is unknown. Spike duplicates were analyzed and all 8 had high variation. With the problems of incomparable HPLC and GCMS analysis for the same samples and a lack of measurements of method precision, these ratios should be considered estimates with wide variance. This figure shows ratio steps between 1 and 2 and 3. I believe steps this small with data this uncertain merely portray noise.

P 4-134 Table 4-15
Why are ratios reported to 3 decimal places when they are based on estimates only? One significant figure please. Also, flag the data as the estimates they are.

Chapter 7 Conclusions

Page 7-4 7.2.1
"Deeper sediment from the "hot spot" had similarly large proportions of LPAH, which suggest that the source is (or was) creosote, not petroleum hydrocarbons."
Compared with Page 7-7 7.3.1
" ... but the dominance of HPAH in the samples suggest that PAH contribution from other petroleum products is probably small because the samples would have proportionally more LPAH..."

Please reconcile the disparity between the upper quote which suggests LPAH indicates creosote and the lower one which suggests other petroleum products are high in LPAH.

Table 1. Comparison of RI and PI sediment data for sites within 150 meters of each other. (ug/kg dry wt)

RI				PI	
SITE	March TPAH	June TPAH	Geo Mean TPAH	SITE	TPAH
633	3867	-	3867	2	2037
656	8774	10423	9563	4	76140
657	8923	24640	14828	3	2059
659	7467	24450	13512	5	25700
680	4748	74300	18782	7	38220
684	7104	6752	6926	6	9777
690	-	251250	251250	11	60620
693	7461	35313	16232	8	7E+06
703	-	34557	34557	12	18855
725	5241	42300	14889	13	9598
727	10431	17730	13599	14	22750
728	-	10608	10608	15	49870
751	8942	15660	11833	16	5772
752	8120	16400	11540	18	16990
762	6464	7880	7137	20	3645
765	7620	9846	8662	17	16578
775	5109	7650	6252	21	6868
788	65882	13022	29290	22	7720
812	5400	-	5400	23	2387
813	3396	-	3396	24	6614

All RI data are estimates.

Table 2. Comparisons of PAH concentrations between sites samples at different times. (ug/kg dry weight).

	Total PAH	
Site	March	June
617	3635	448
629	382	1345
656	8774	10423
657	8923	24640
658	6862	14070
659	7467	24450
680	4748	74300
682	24924	114290
684	7104	6752
688	87602	189373
689	194493	82123
693	7461	35312
701	307482	56137
713	8495	29189
714	10661	38966
715	12300	152869
716	29410	9577
717	4917	29085
725	5241	42300
727	10431	17730
750	6076	16580
751	8942	15660
752	8120	16400
762	6464	7880
765	7620	9846
775	5109	7650
787	7467	10346
788	65882	13022
789	12618	13950
801	4785	7567

All values are estimates only.

Table 3. Comparison of HPAH/LPAH ratios from two different methods on the same sample.

	HPLC			GCMS		
Site	LPAH	HPAH	H/LPAH	LPAH	HPAH	H/LPAH
629	263	1082	4.1	228	1337	5.9
656	163	10260	62.9	1407	5620	4.0
668	2400	13370	5.6	1742	8110	4.7
680	9030	65270	7.2	9590	137500	14.3
689	14223	82123	5.8	8880	73800	8.3
701	11813	54550	4.6	12920	97400	7.5
701B	58690	68100	1.2	8443	102900	12.2
701D	18569	42920	2.3	7997	108600	13.6
713	4768	24420	5.1	1907	15000	7.9
716	1917	7660	4.0	843	2870	3.4
725	6830	35470	5.2	562	2870	5.1
728	1078	9530	8.8	1039	10760	10.4
789	2400	11550	4.8	724	13650	18.9
PM-108	98	305	3.1	76	544	7.2
YH-120	0	0	0.0	45	232	5.2

All values are estimates.

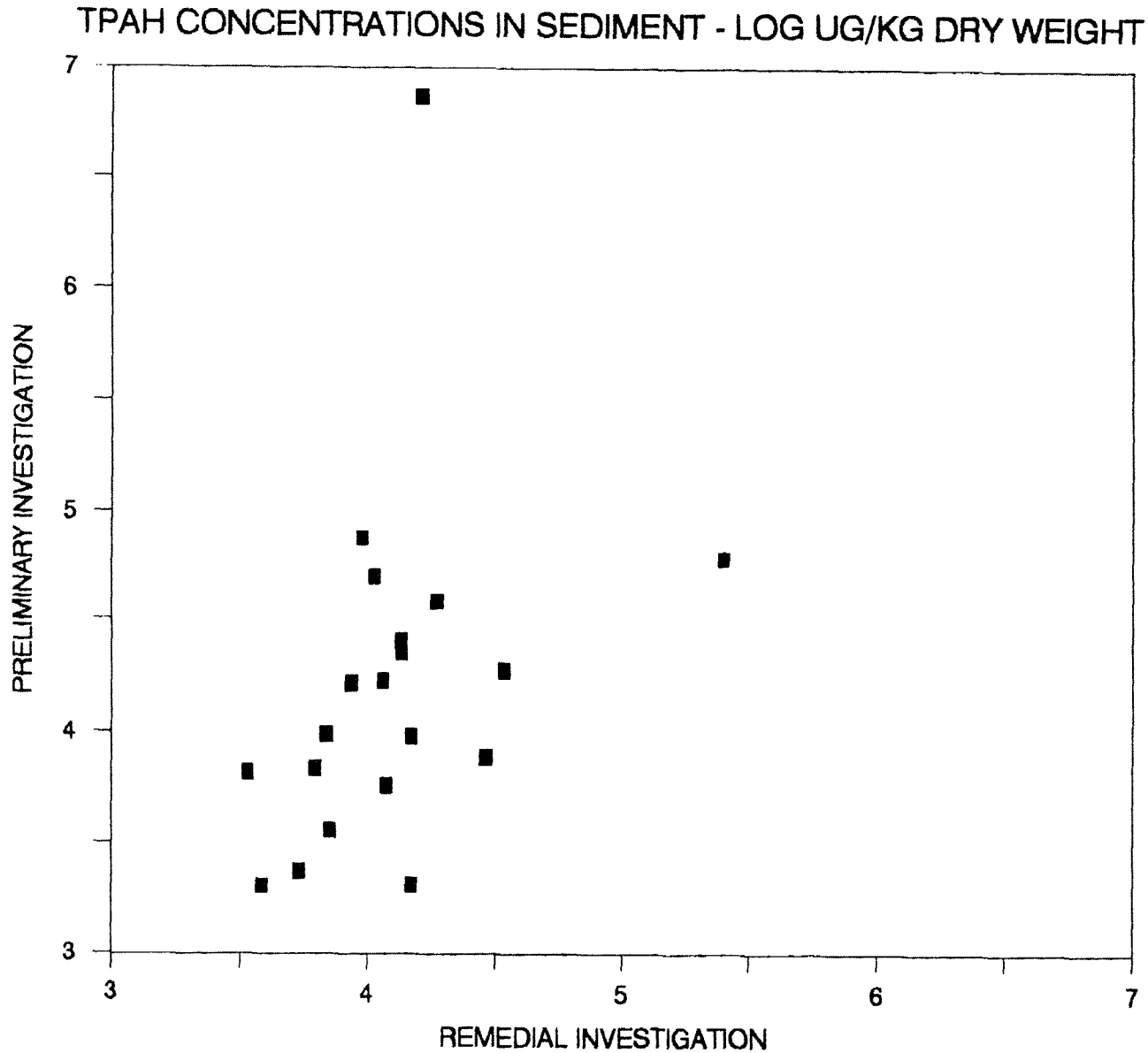

Figure 1. Correlation between samples taken at adjacent sites within 150 meters of each other in Eagle Harbor. ($r=.36$, $p>.05$ N.S.)

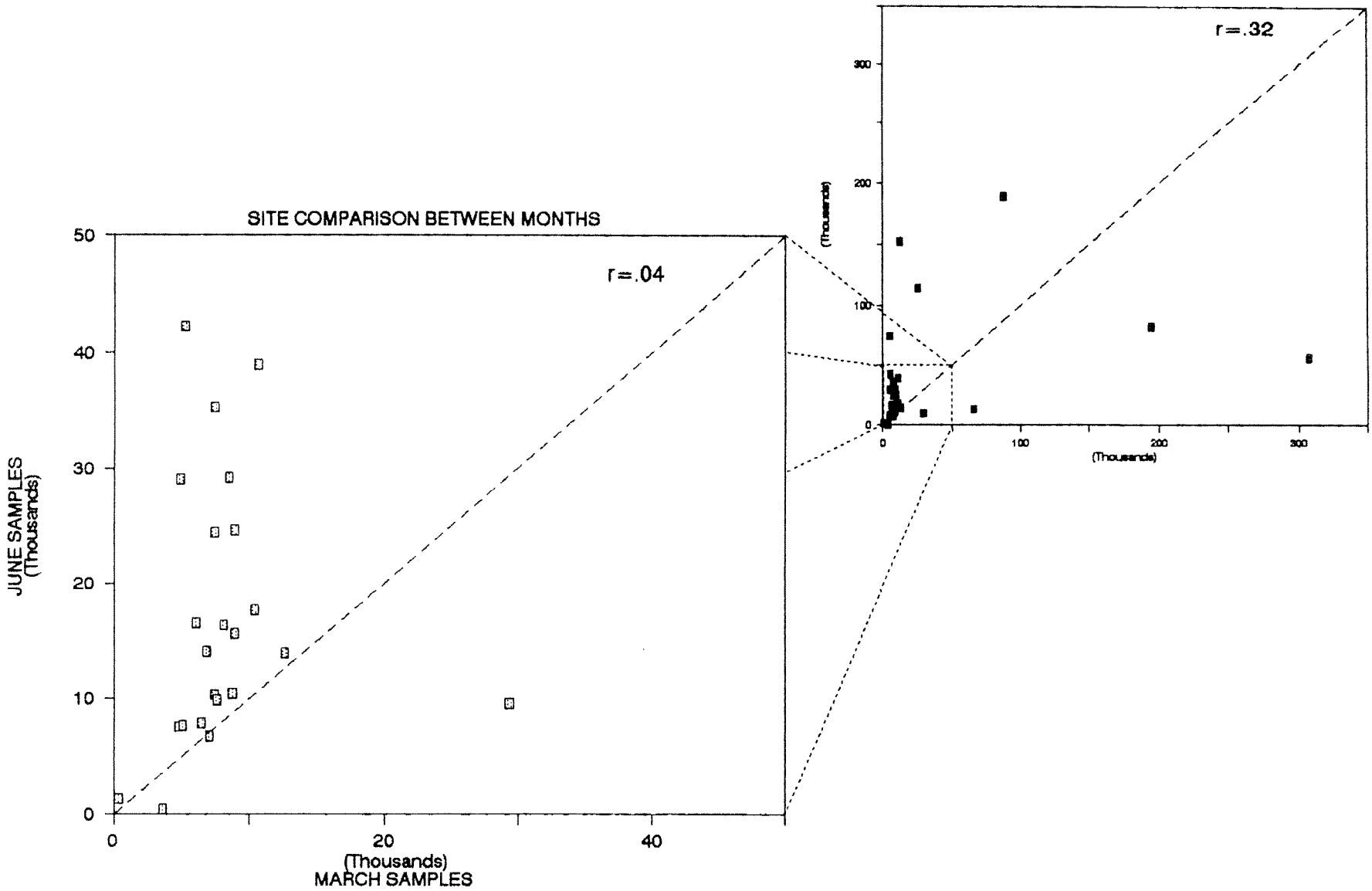

Figure 2 . Total PAH in sediments in Eagle Harbor analyzed by HPLC compared between two sampling months. All comparisons between the same sites. (ug/kg dry weight) (Source: 1989 RI from CH2MHill)