Potential Effects of Data Artifacts on Receptor Modeling Results

slides presented by R. Poirot, VT DEC at
EPA Workshop on Applying UNMIX and PMF to PM-2.5 Data, 2/14-16/99

Figure 1. Receptor Models identify sources of influence on the data. Data artifacts, including how labs report measurement uncertainty and MDL, and how data analysts use and interpret that information, can potentially represent a source of influence on receptor model results.

Figure 2. Ni & As are excellent "regional tracers" for rural VT sites, (for East Coast urban residual oil, and for Canadian smelter(s), respectively. But at Lye Brook IMPROVE site: Ni is < MDL 1/4 the time; As is < MDL half the time, and one or the other is < MDL 2/3 of the time.

Figure 3 (top). Censoring data (eliminating days when any input variable is below MDL) can create a small and biased subset of the original data.

Figure 3 (bottom). Ni and As are totally uncorrelated at Lye Brook (as we might expect since they come from different sources in opposite directions). However the reported measurement uncertainties for Ni and As exhibit a weak but significant positive correlation.

Figure 4 (top). Reported analytical uncertainties for Ni are typically less than 10% of reported concentration for concentrations above about 2 ng/m3, but increase substantially as concentrations approach the detection limits. Note the apparent "fingerprint" (ridge lines in this plot. What causes these?

Figure 4 (bottom). Illustration of use of random # generators to select concentrations between 0 and MDL for non-detects and within +- the reported uncertainty for above detects. Can't we improve on this? Pehaps by using inter-species, inter-site and time-lagged correlations, etc.

Figure 5. Although concentrations of Ni and As are uncorrelated, their MDLs are highly correlated, both as a function of (3) methods changes in different time periods, and also within each of 3 different reporting periods. Possibly due to common interferences or instrumental drift, but not due to changing ambient concentrations. Generally, in most long-term measurement programs both ambient concentrations and detection limits are likely to decrease over time, creating possibility of false positive correlations between source activity for some elements and lab activity for other elements.

Figure 6. Comparison of same-day, above-MDL As concentrations at Acadia and Mt Ranier IMPROVE sites exhibit no correlation (as we would expect given the continental distance between sites. However same-day As MDLs for these sites are correlated, generally due to "progress" (improving detection limits over time) in the 10+ year IMPROVE network.

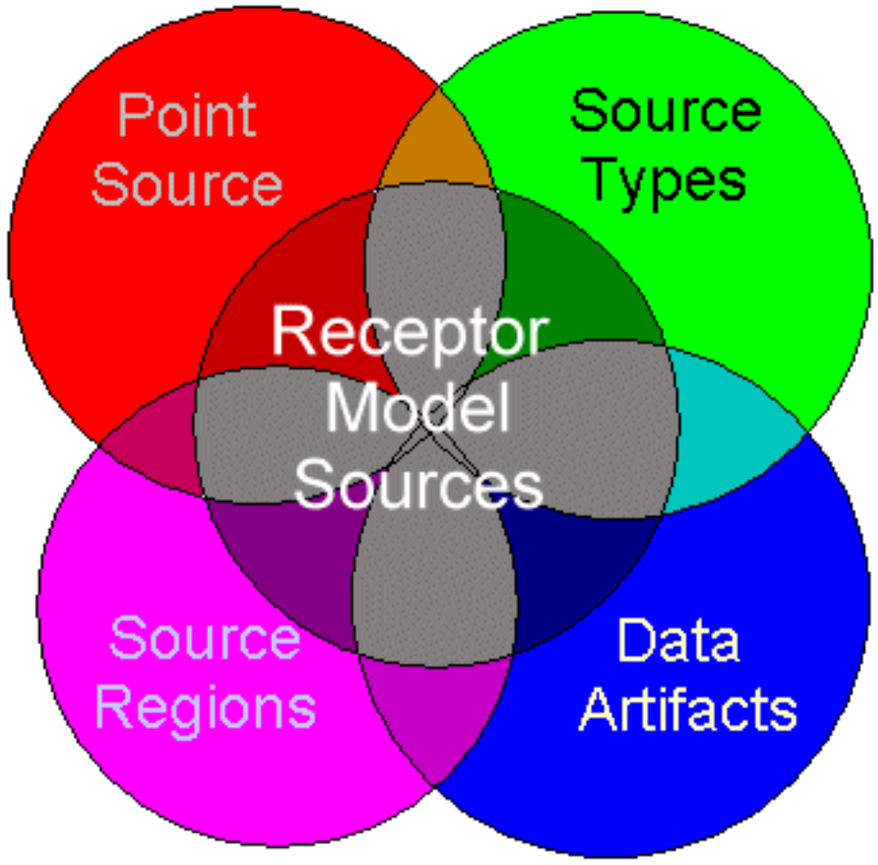
Figure 7. There appears to be a unique problem associated with the Al data at IMPROVE sites, which might be described as an "occasionally mis-quantified MDL for Al". For these samples it does not appear likely that actual Al concentrations were less than or equal to the reported MDL.

Figure 8. Comparative results from preliminary analysis of Underhill, VT IMPROVE-like data by UNMIX and PMF suggest that the 2 different models identified 3 common sources (despite wide differences in data preprocessing and model input). A 4th pair of sources (with high Se loading) showed differences and similarities. The split between the dissimilar and similar sources coincides with the addition of XRF analysis in 6/1/92, with lower MDL and uncertainties for Se and several other elements that the PIXE analysis employed prior to 6/1/92.

Figure 9. Same as above but displayed as time series with PMF contributions normalized to UNMIX units by regression. Data Artifacts, including MDLs and uncertainties as reported by labs and/or as processed by data analysts can & do influence receptor model results.

FIGURE 1

Receptor Models Identify Sources



What is a "Source"?

Data Artifacts are Sources of Influence on Results, and include:

- Measurement Errors, Uncertainties and Non-Detects, and
- How these Errors and MDLs are Reported by Labs and
- How these data features are Treated as Input to Model
- How these data features co-vary (or not) with other Sources

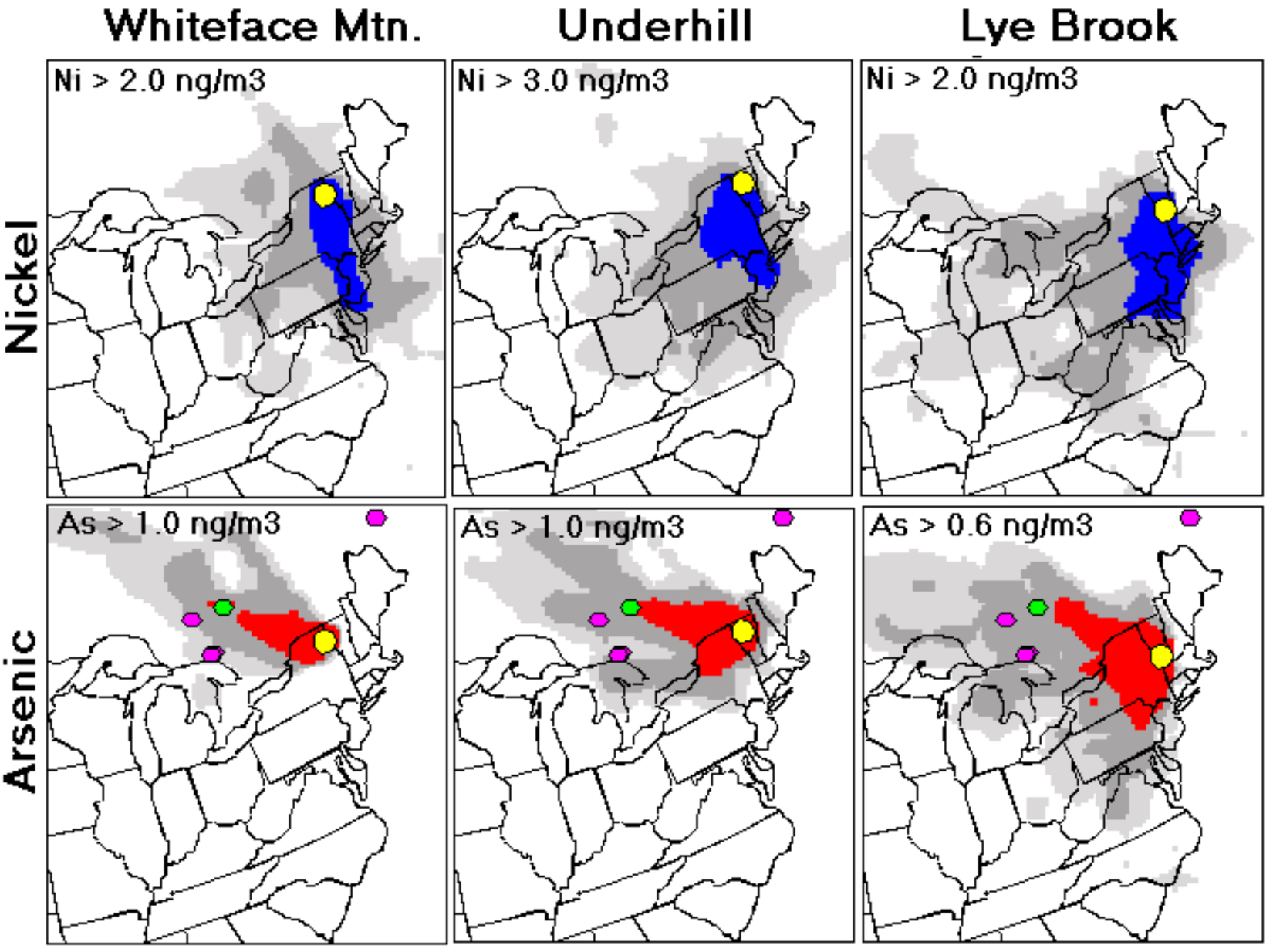
Approaches to The MDL Problem:

- Censor Input Data by screening out Samples or Variables
- Estimate Below Detects and Uncertainties from other Data

FIGURE 2

Nickel and Arsenic in VT are Excellent Tracers for East Coast Oil Combustion and Noranda Smelter.

Upwind Probability Fields for High Ni and As at VT Receptors

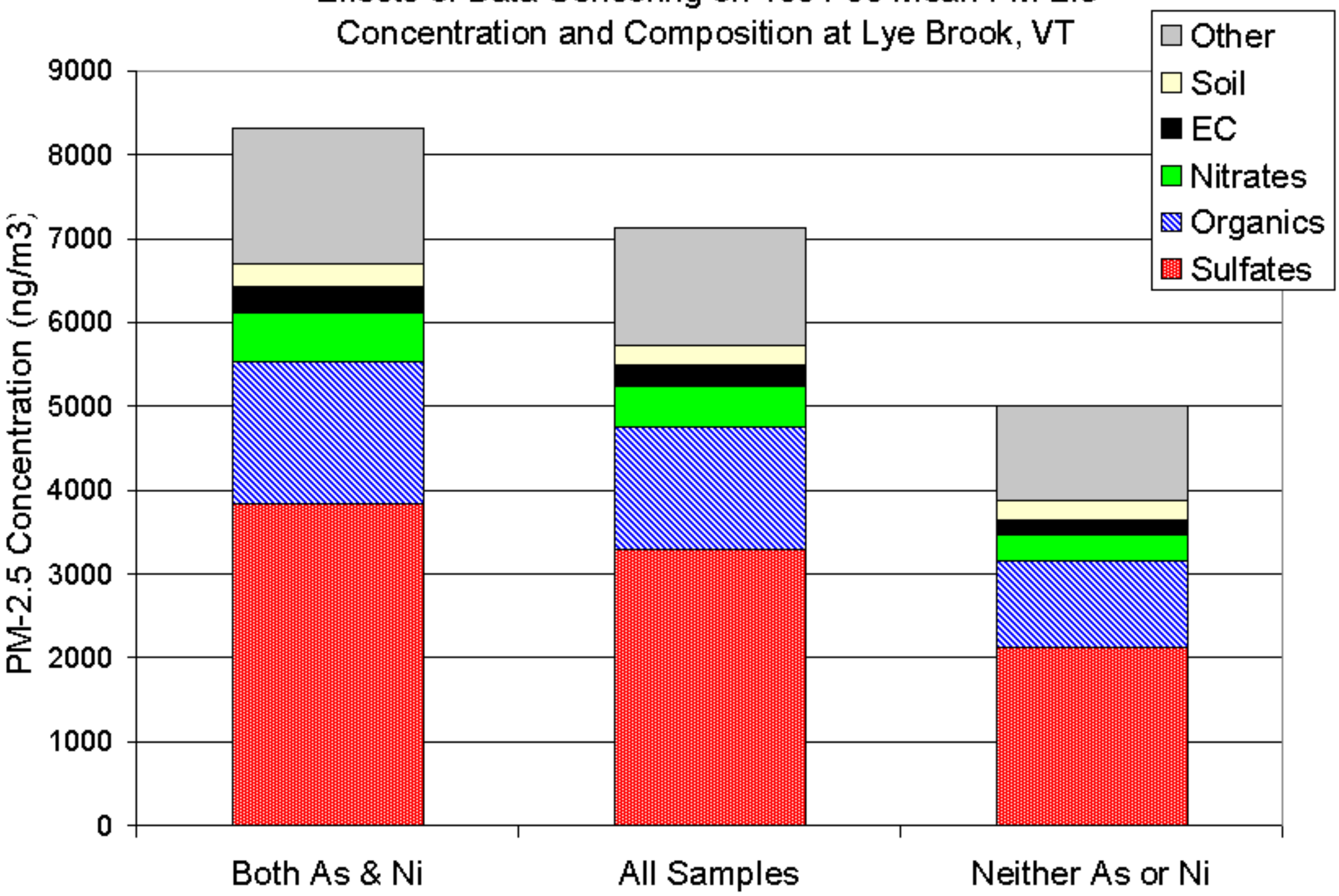


Ni and V are uncorrelated (R2 = 0.02) at Lye Brook, VT.

However:
Ni is Below MDL in 27% of Samples. and
As is Below MDL in 54% of Samples, and
As or Ni is Below MDL in 66% of Samples.

FIGURE 3

Effects of Data Censoring on 1991-99 Mean PM-2.5 Concentration and Composition at Lye Brook, VT



Reported Concentrations and Uncertainties for Ni and As at Lye Brook, VT: 1991-99

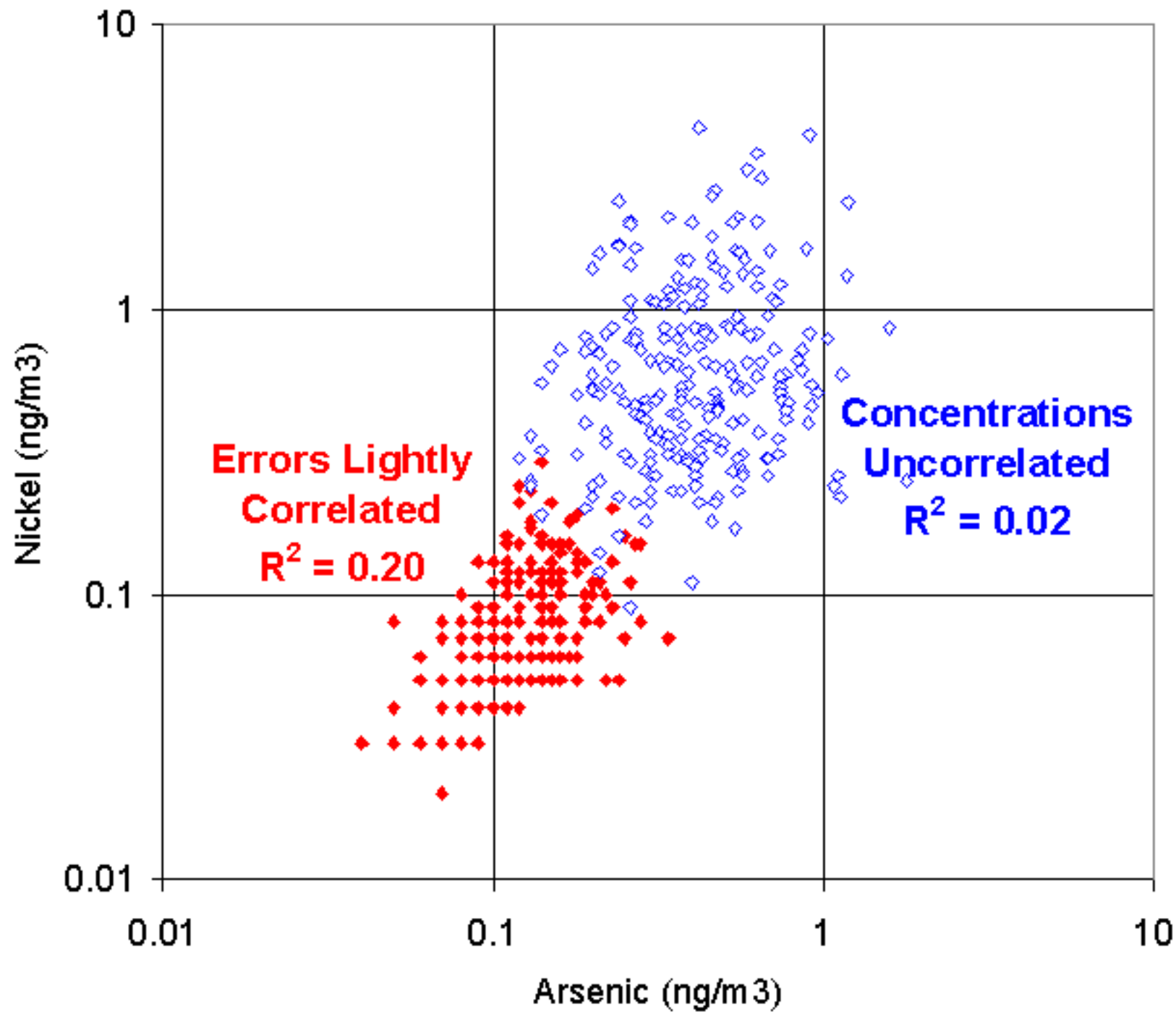
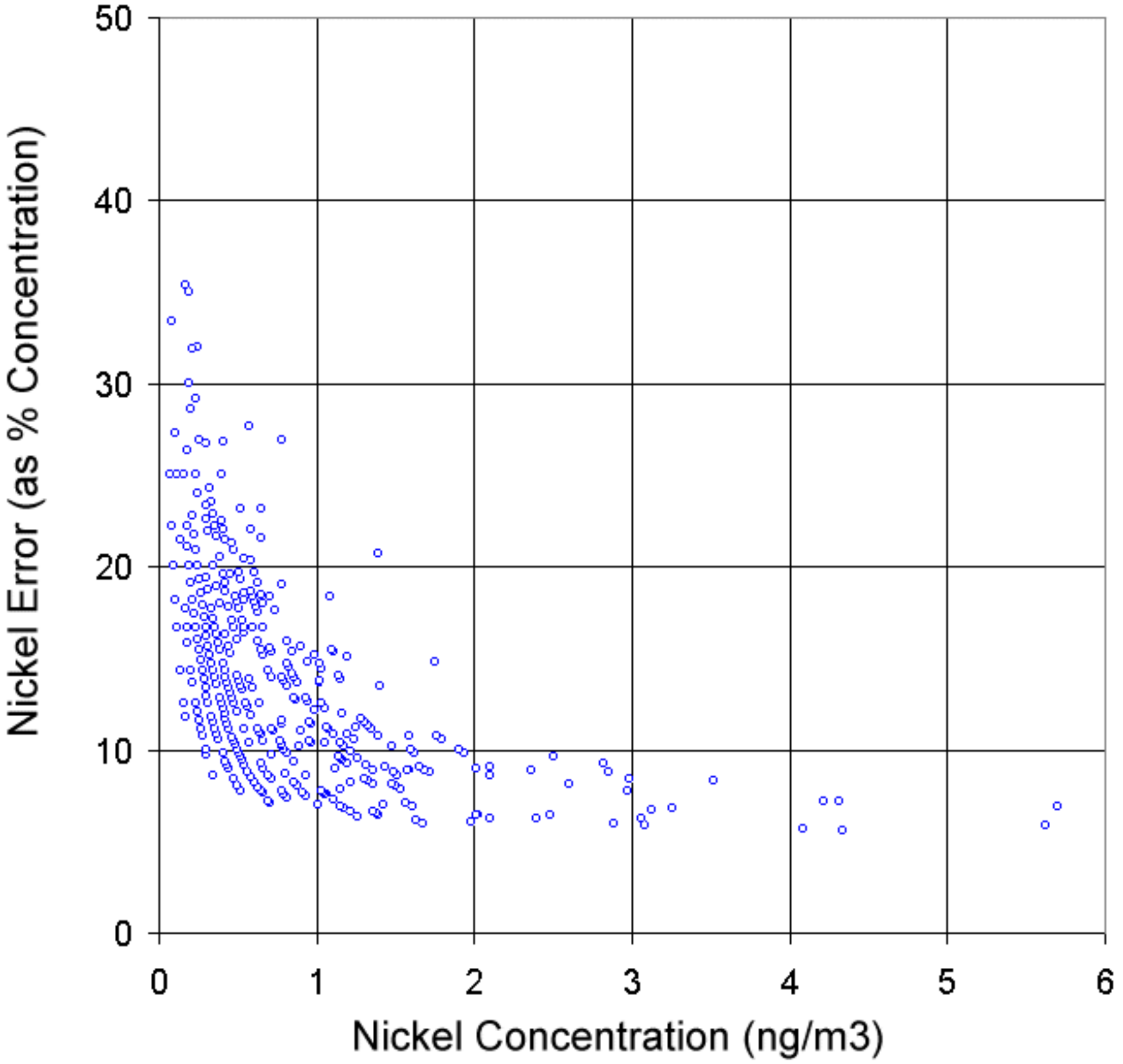


FIGURE 4

Relative Error Increases as Concentration Decreases



Random Extractions Incorporating MDL and Error

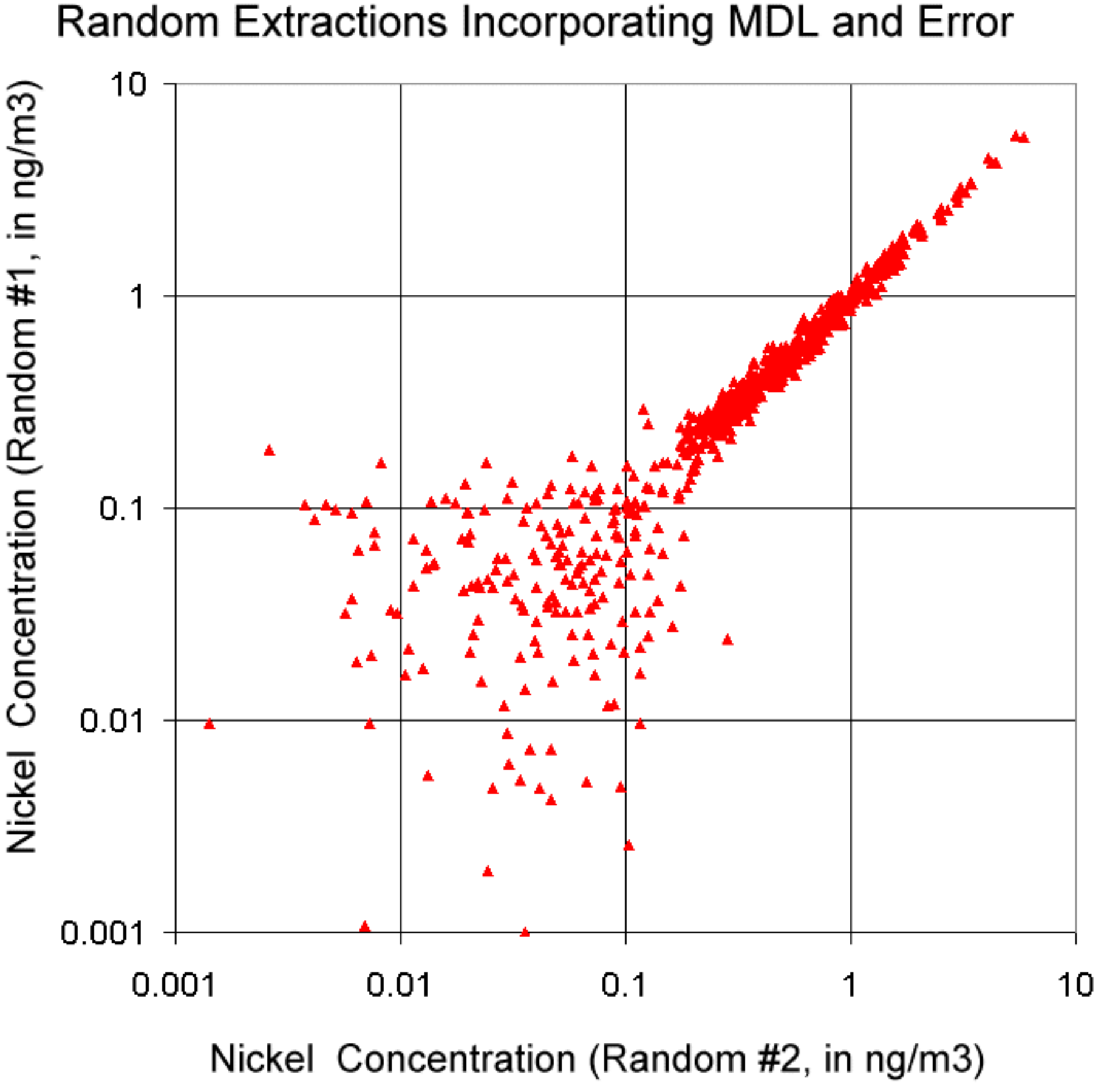
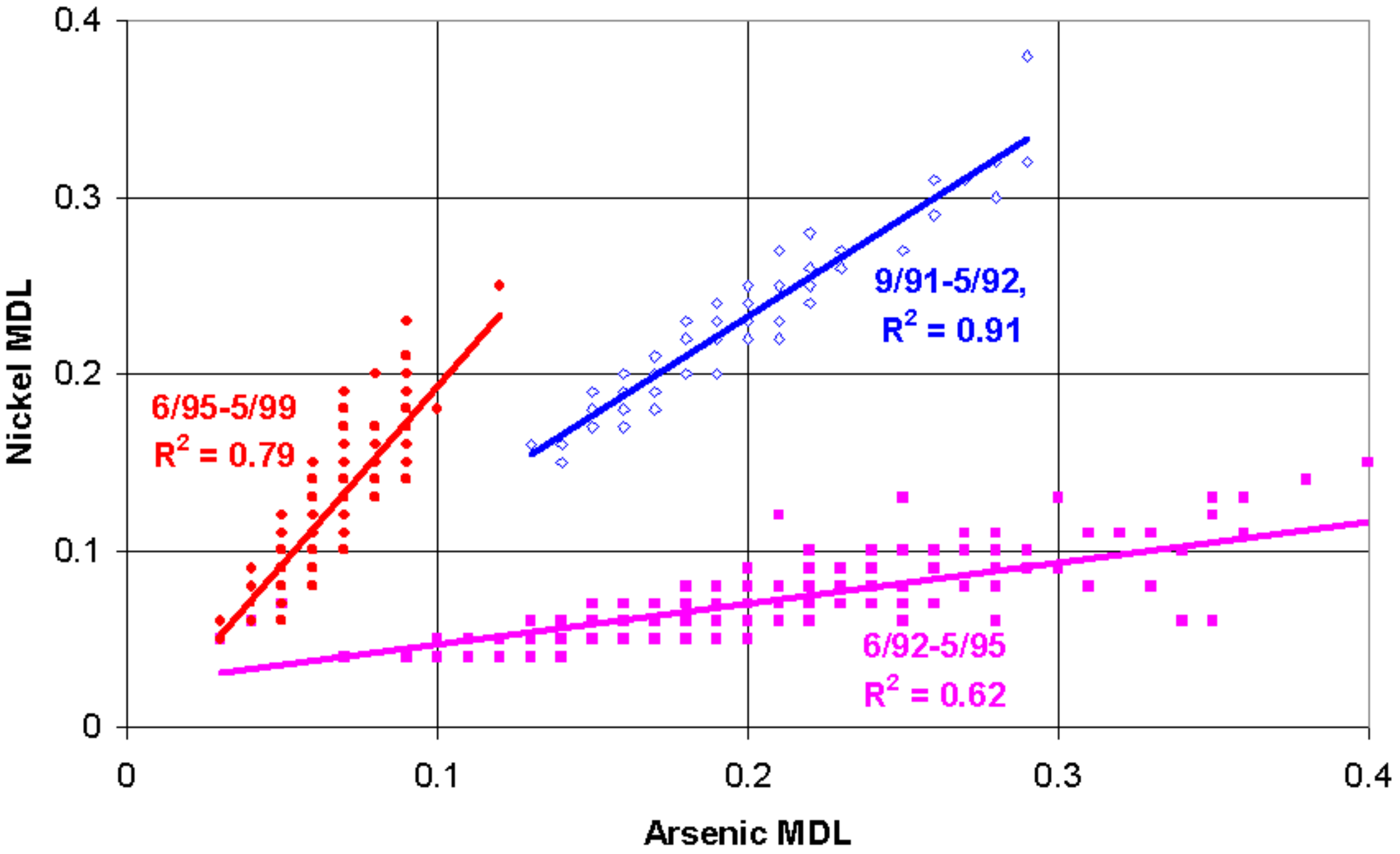


FIGURE 5

3 "Sources" of Arsenic and Nickel at Lye Brook, VT 9/91-5/99



Time Series of Arsenic & Nickel MDL at Lye Brook, VT: 9/91-5/99

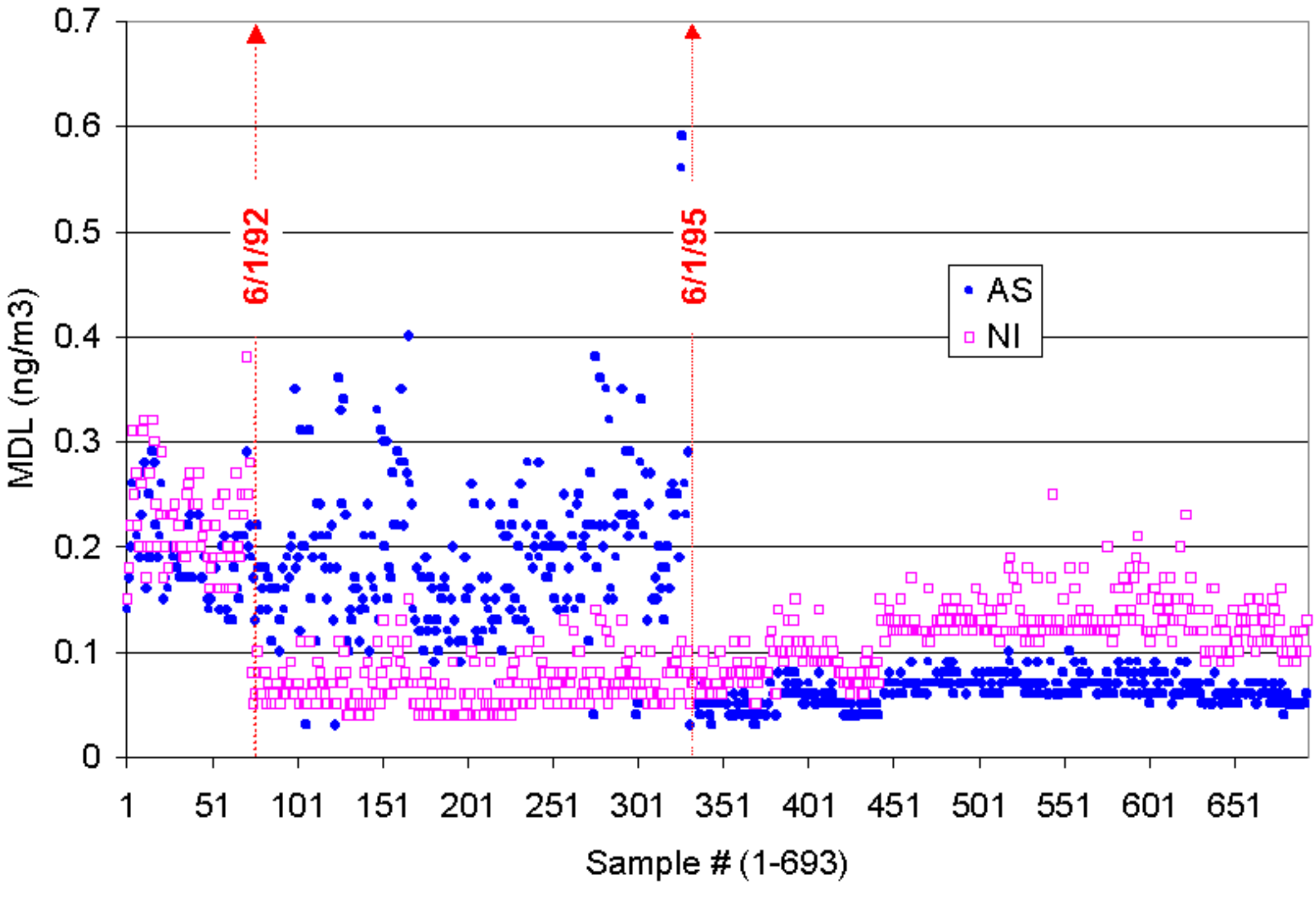
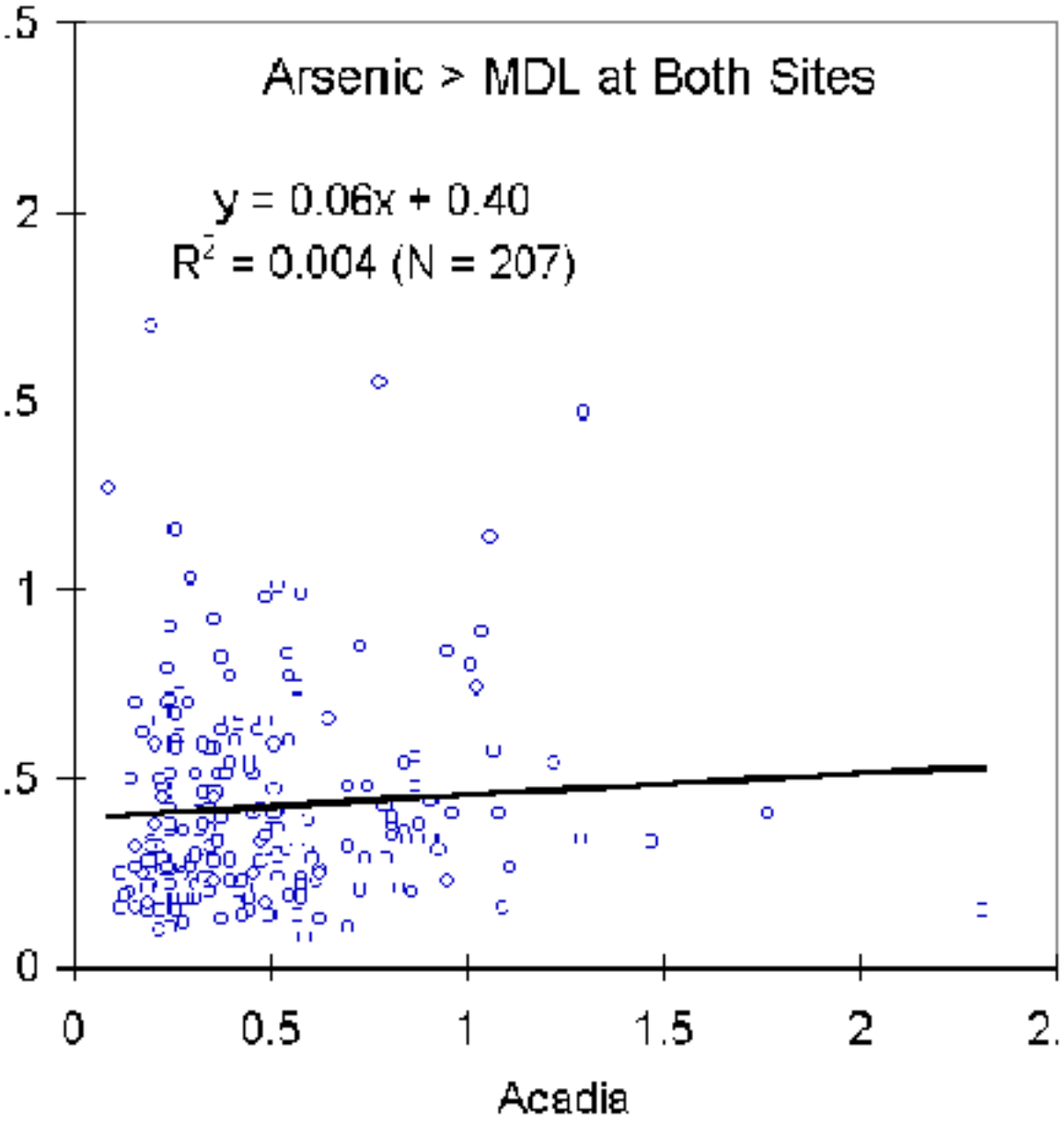
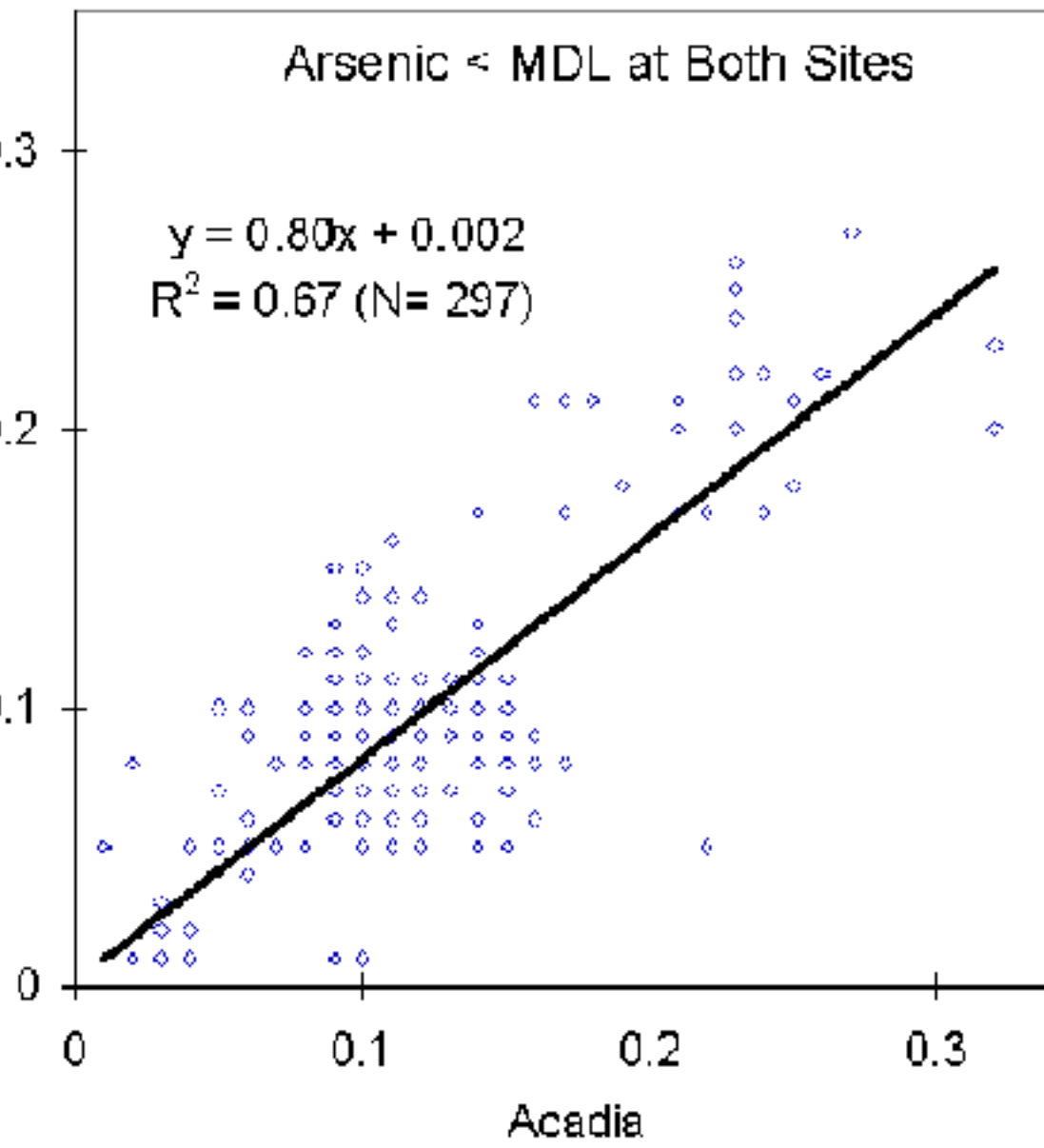


FIGURE 6


Arsenic > MDL at Both Sites
y = 0.06x + 0.40
R2 = 0.004 (N = 207)
Mt Ranier
Acadia
0
0.5
1
1.5
2
2.5


Arsenic < MDL at Both Sites
y = 0.80x + 0.002
R2 = 0.67 (N= 297)
Mt Ranier
Acadia
0
0.1
0.2
0.3

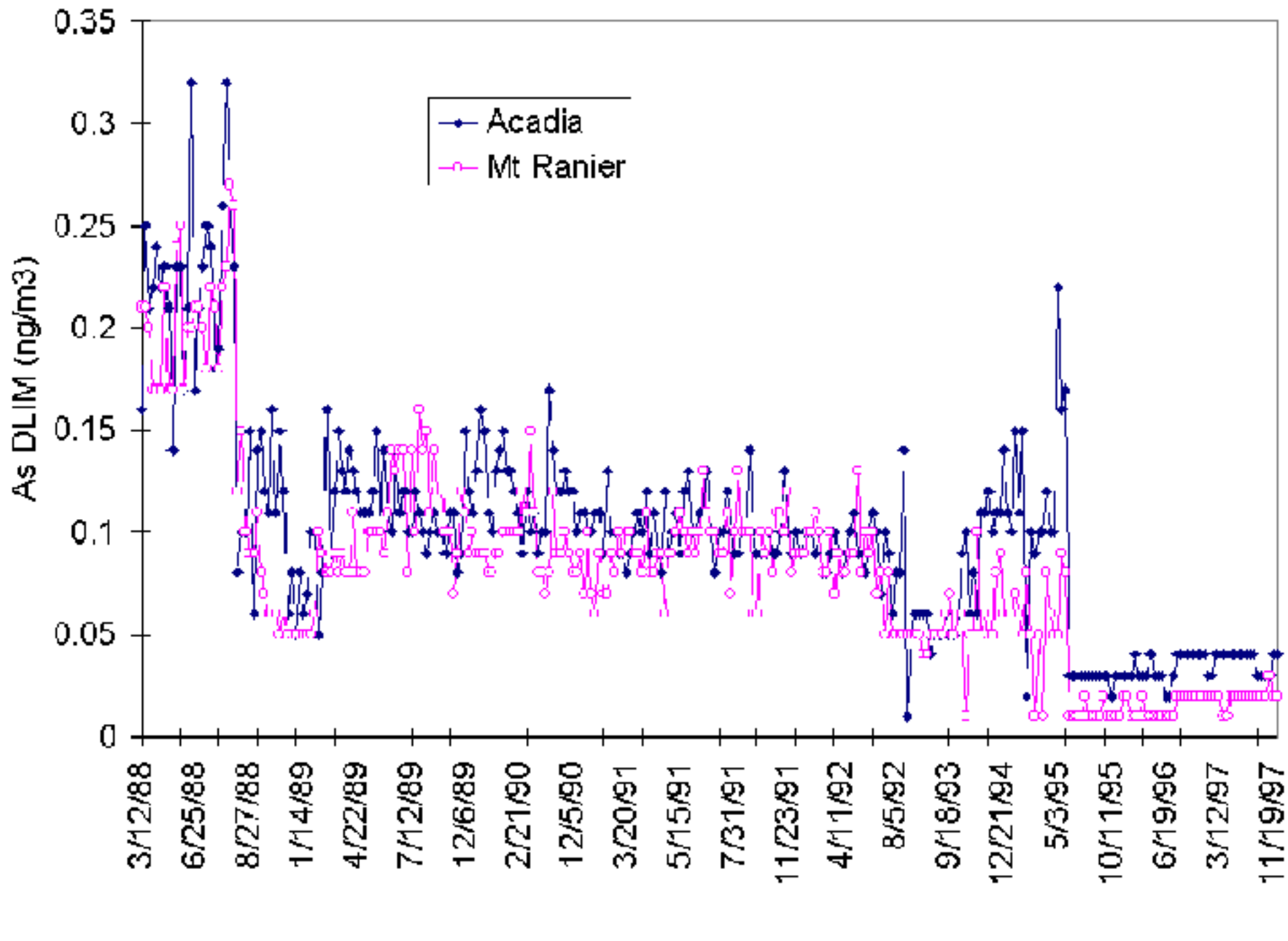
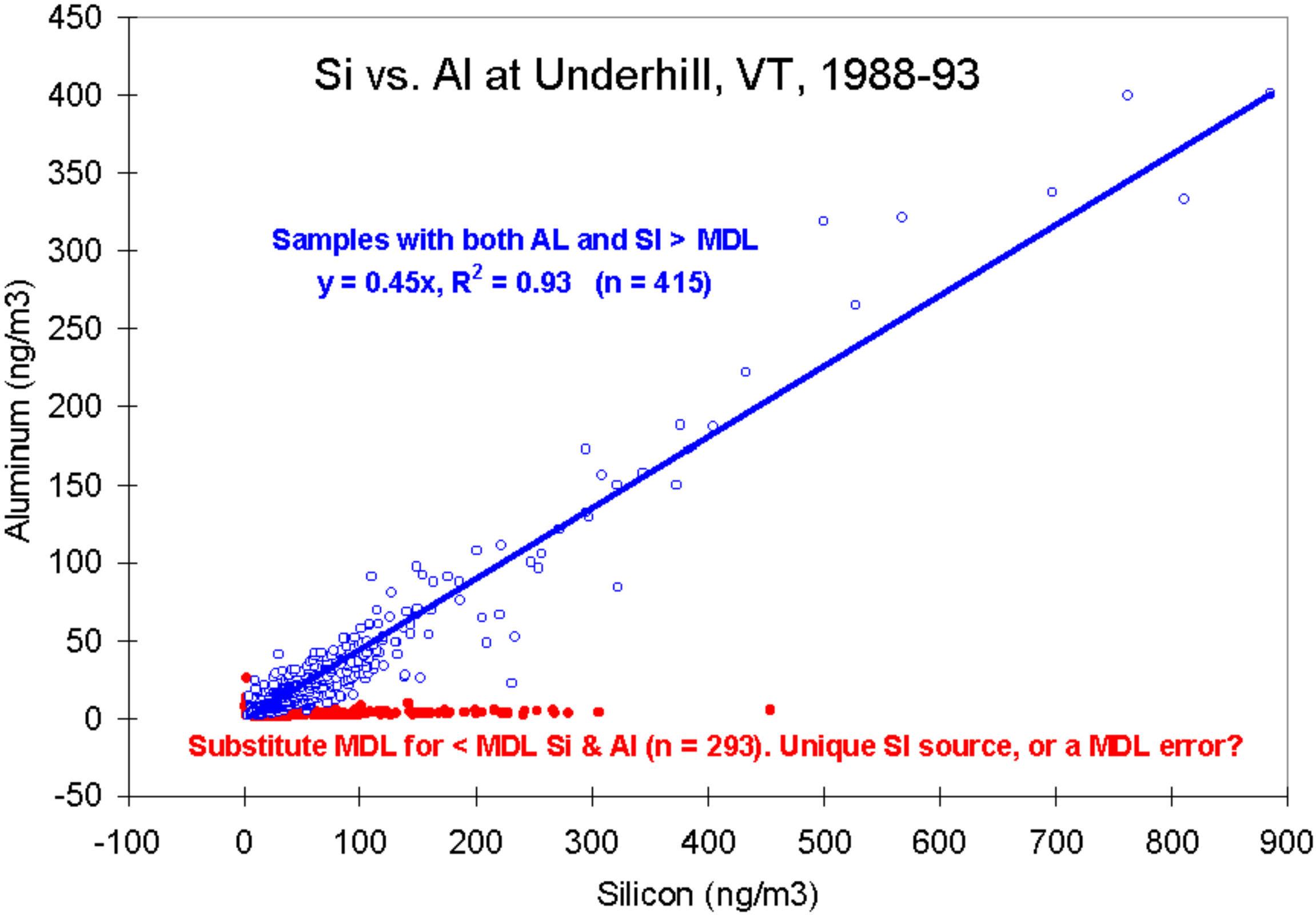

Acadia
Mt Ranier
As DLIM (ng/m3)
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
3/12/88
6/25/88
8/27/88
1/14/89
4/22/89
7/12/89
12/6/89
2/21/90
12/5/90
3/20/91
5/15/91
7/31/91
11/23/91
4/11/92
8/5/92
9/18/93
12/21/94
5/3/95
10/11/95
6/19/96
3/12/97
11/19/97

FIGURE 7



Aluminum and Silicon are well correlated when both are above MDL

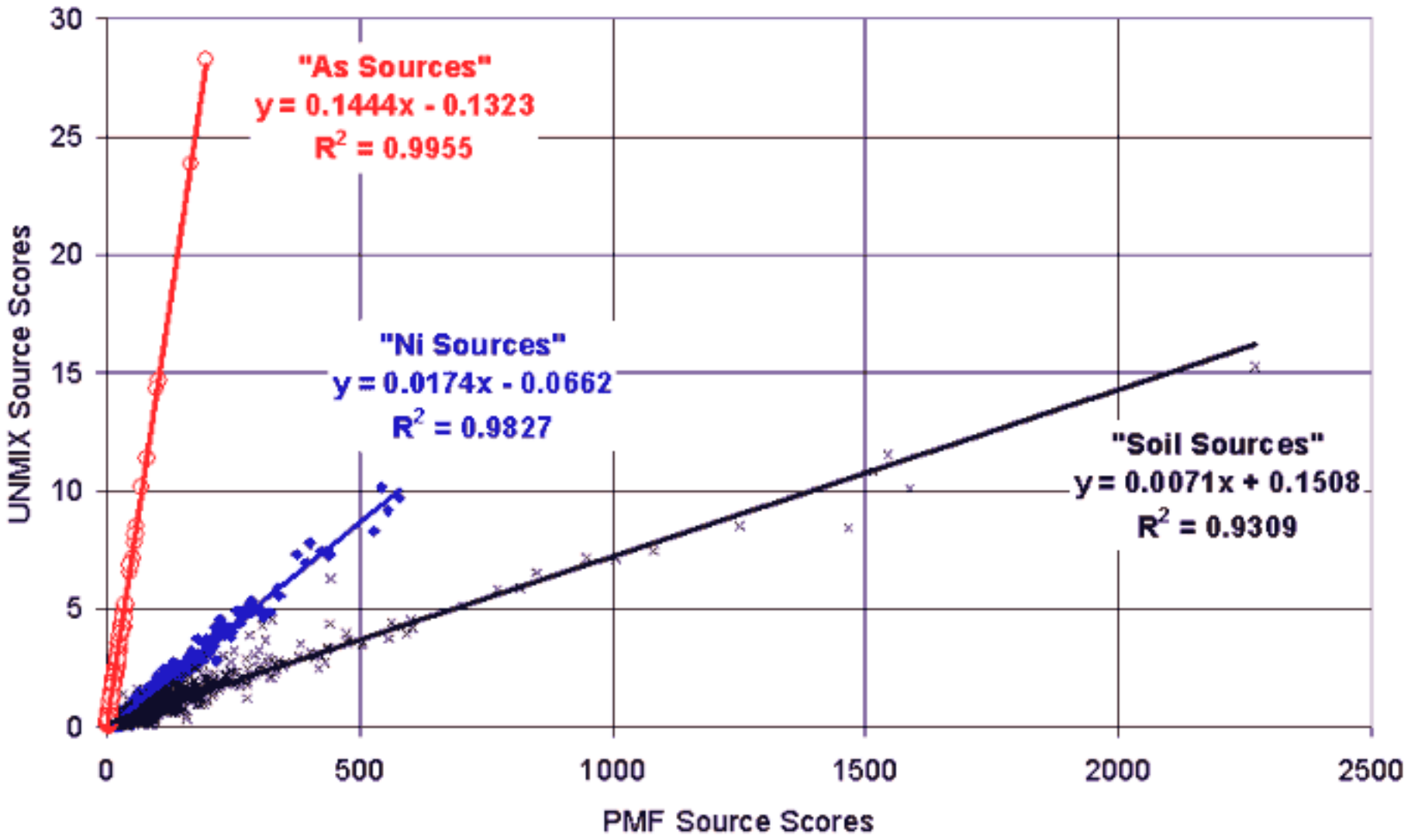
If the reported MDL for AL is used to estimate Al on below MDL days, the estimated values are clearly too low with respect to Si

So either there's a unique Si source with no Al (unlikely)
or the MDL for Al is mis-calculated (as a quantitative limit)

Although, it does appear to result in "wise" data censoring...?

FIGURE 8

Comparison of Daily UNMIX and PMF Source Scores at Underhill, VT (for "Sources" with Strong Loadings in As, Ni and Crustal elements)



Comparison of "Selenium Sources" at Underhill, VT by UNMIX and PMF

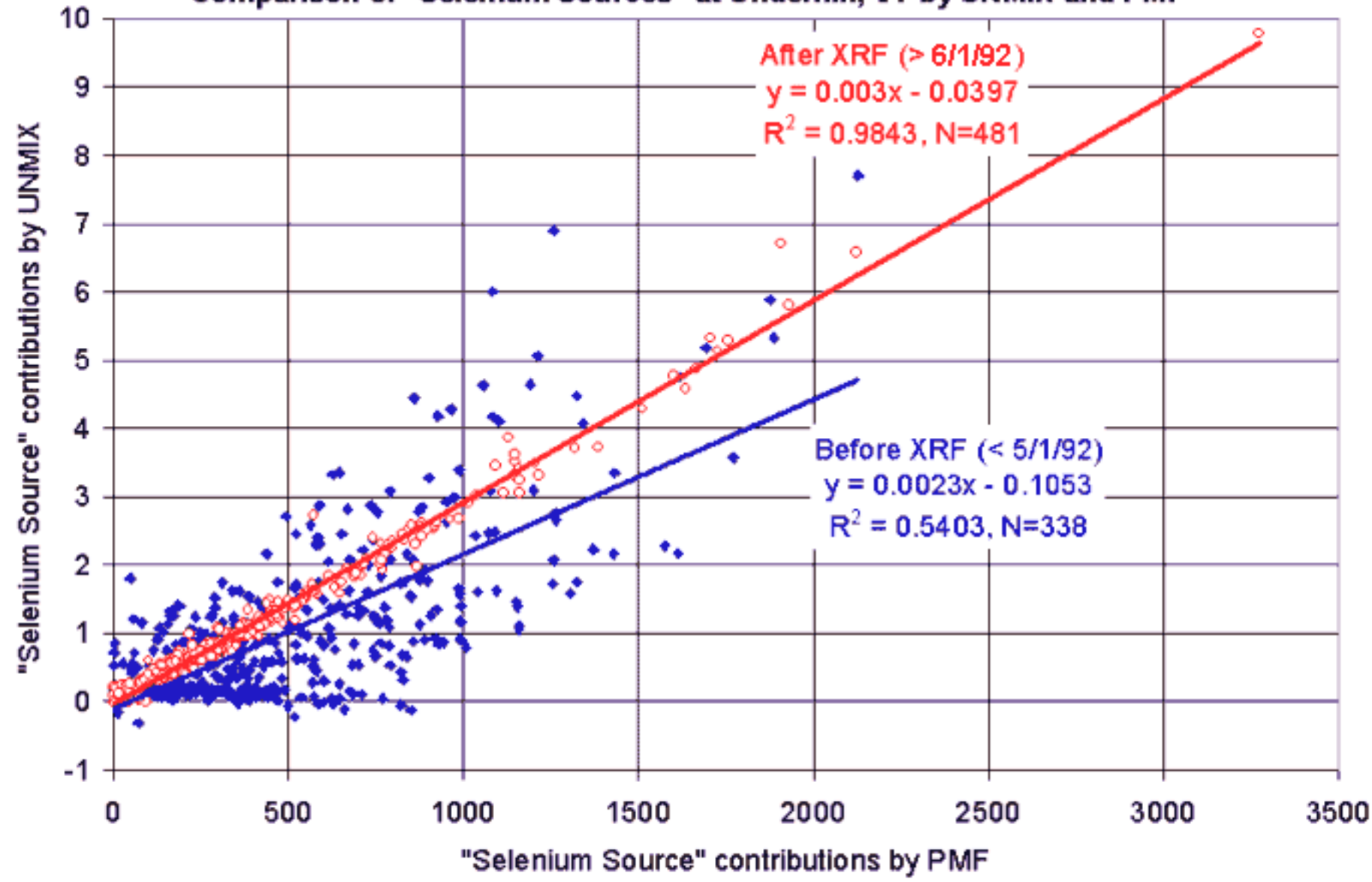
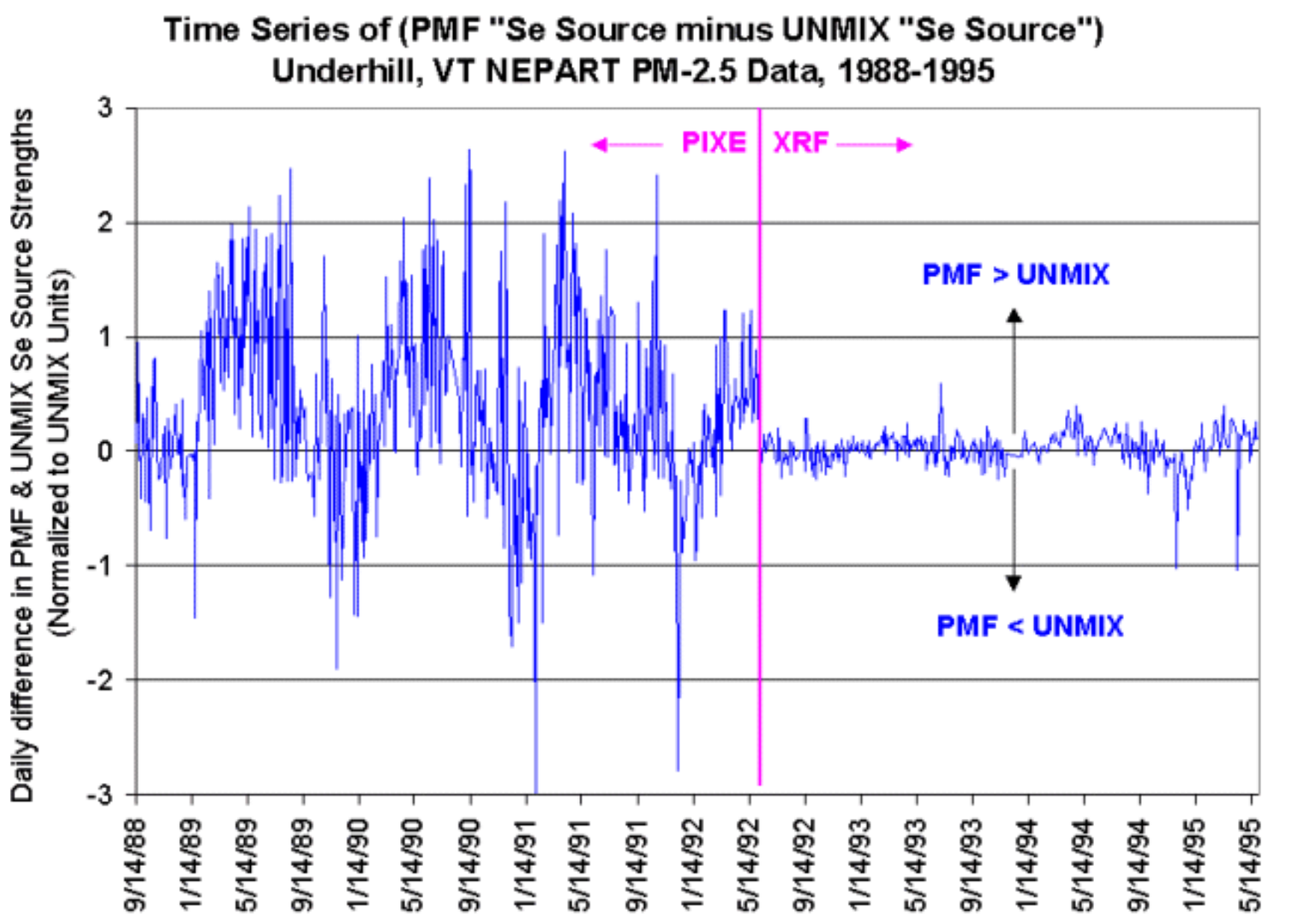
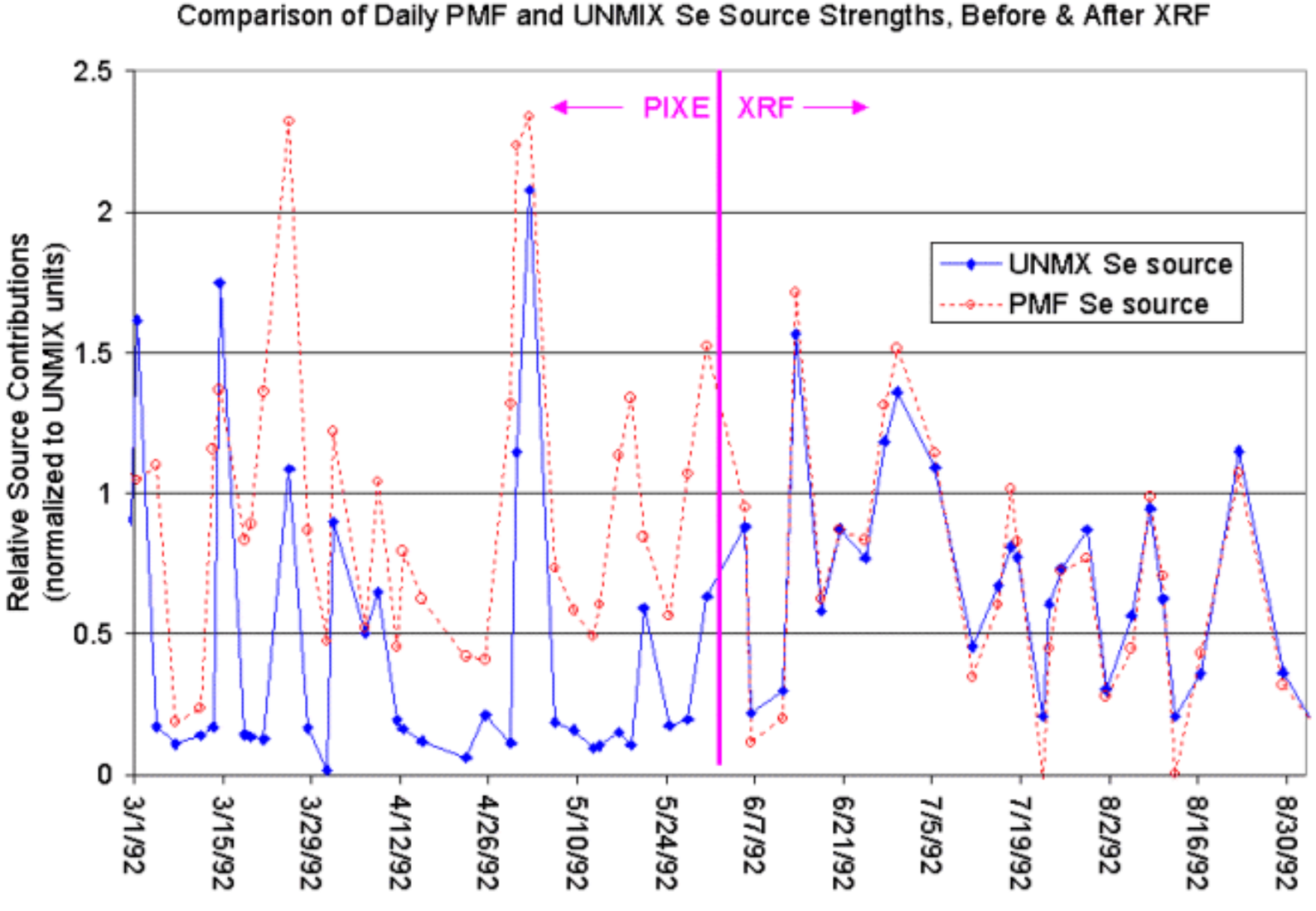


FIGURE 9


Time Series of (PMF "Se Source minus UNMIX "Se Source")
Underhill, VT NEPART PM-2.5 Data, 1988-1995
Daily difference in PMF & UNMIX Se Source Strengths
(Normalized to UNMIX Units)
PIXE
XRF
PMF > UNMIX
PMF < UNMIX
3
2
1
0
-1
-2
-3
9/14/88
1/14/89
5/14/89
9/14/89
1/14/90
5/14/90
9/14/90
1/14/91
5/14/91
9/14/91
1/14/92
5/14/92
9/14/92
1/14/93
5/14/93
9/14/93
1/14/94
5/14/94
9/14/94
1/14/95
5/14/95


Comparison of Daily PMF and UNMIX Se Source Strengths, Before & After XRF
Relative Source Contributions
(normalized to UNMIX units)
PIXE
XRF
UNMX Se source
PMF Se source
2.5
2
1.5
1
0.5
0
3/1/92
3/15/92
3/29/92
4/12/92
4/26/92
5/10/92
5/24/92
6/7/92
6/21/92
7/5/92
7/19/92
8/2/92
8/16/92
8/30/92