Subject: Evaluation of the Equation for Soil Composite   June 20, 2003
Internal Memo to IMPROVE Staff from Bob Eldred

Summary. The purpose of this report is to determine if we wish to revise the current equation for the soil composite. This report examines four aspects of the soil equation. The first is what to do when aluminum is missed because of x-ray interference. The second is how to treat soil potassium. The third is how to treat soil elements not included in the equation. The fourth is a suggested change in the equation for the uncertainty in the soil composite concentrations. My recommendation for the present is that for the equation, we only change the way we treat Al. This could be done either by CIRA or at UCD before the data are submitted.

Background. The soil equation was derived in the late 1980's based on data from the Western Fine Particle Network and from the soil composition table from F.J. Pettijohn, <u>Sedimentary Rocks</u>, (NY, Harper & Row, 1949, 1957, and 1975). This table has been included in the Handbook of Chemistry and Physics (see "earth, data" in the index). The current equation is based on five elements:

$$SOIL = 2.20*Al + 2.49*Si + 1.63*Ca + 2.42*Fe + 1.94*Ti$$

The assumptions that went into this equation are as follows.

1. Minerals present in soil. The Pettijohn table lists 15 minerals in sedimentary rocks. Since soil is formed from these rocks, we will assume that these are the primary constituents in soil. The minerals in decreasing importance in average sediment are: SiO_2, Al_2O_3, CaO, CO_2, Fe_2O_3, H_2O, K_2O, MgO, FeO, C, Na_2O, TiO_2, SO_2, P_2O_5, and Ba. Of the elements, we have selected five: Si, Al, Ca, Fe, K, and Ti. It should be noted that some of these elements can have a nonsoil source. For example, K can also come from smoke, Fe from industrial sources, Al from aluminum smelters (perhaps only at the two Columbia Gorge sites), and Ca from concrete dust.

2. Oxide Factors. The multiplicative factors to go from the element to the oxide can be calculated from the atomic weights. These are: $Al_2O_3 = 1.89$, $SiO_2 = 2.14$, $K_2O = 1.20$, $CaO = 1.40$, $TiO_2 = 1.67$, $Fe_2O_3 = 1.43$, and $FeO = 1.29$.

3. Iron. For most rock types, iron is approximately split equally between Fe_2O_3 and FeO. We chose to simply average the two oxide factors, giving Fe an oxide factor of 1.36. This is equivalent to all the iron in the form of Fe_4O_5. (For average sediment, the split is 60% and 40%. This gives a factor of 1.37, which is not significantly different.)

4. Potassium. We chose to correct for nonsoil sources for one of the elements, K. To correct for this, we used Fe as a surrogate. At the sites in the WFPN and in the original IMPROVE networks, there was little industrial Fe. Using ratio of the means for <u>coarse</u> particles (larger than 2.5 µm) measured in the NPS network by the stacked filter sampler between 1982 and 1986, we determined that soil K is 60% of Fe. The modified Fe factor is then $1.36+0.6*1.20 = 2.08$. K was then dropped from the equation.

5. Contributions of other elements. According to the Pettijohn table for average sediment, these five elements and their oxides account for only 86% of the mass. The above factors were then divided by 0.86.

6. If the element is not found in the x-ray spectrum, then one-half of the minimum detection limit (mdl) is substituted for the missing concentration. The assumption is that the element was missed because the concentration was below the mdl and not because of interference.

7. The uncertainty in the soil composite was derived using the quadratic sum of the constituent concentrations. This assumes that the uncertainties are all independent. If the element is not found, one-half of the mdl is used as the uncertainty. Denoting the uncertainty in Al by d(Al), the equation for the uncertainty in the soil composite is

$$[d(SOIL)]^2 = [2.20*d(Al)]^2 + [2.49*d(Si)]^2 + [1.63*d(Ca)]^2 + [2.42*d(Fe)]^2 + [1.94*d(Ti)]^2$$

Aluminum. Al is difficult to measure by x-ray analysis (PIXE or XRF), because of interference with the Si peak. Any degradation in the resolution of the detector can result in the spectral code not being able to identify Al. The historical detection fraction is shown in Figure 1. Note that the low values around 2001 were the result of running the PIXE and XRF samples for less than normal times in order to catch up in the analysis. For the entire time period, the found rate for Al has been 57%. This is considerably lower than for the other major soil elements (Si, Ca, K, and Fe) which ranged from 97% to 100%.
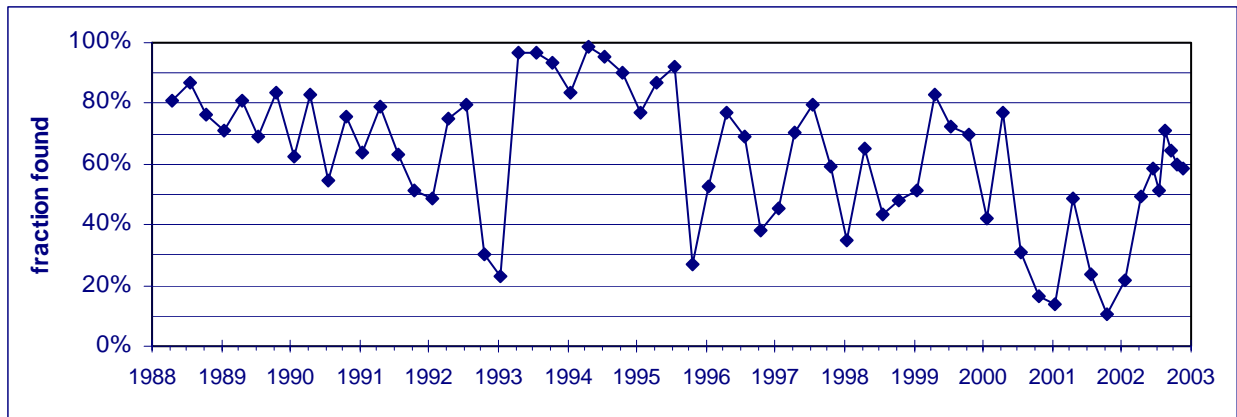

Figure 1. The fraction of cases where Al was detected for each analytical session (generally one quarter).

The impact of missing Al because of interference is significant. Without interference, we can assume that the true value is less than the mdl and one-half of the mdl can be substituted for the missing value. When an element is missed because of interference, the true value may be much larger than the mdl. It is not possible to calculate a more realistic minimum quantifiable limit, since unmeasured factors are involved. This would not be a major problem if Al were a minor element. Unfortunately, it is the second most common soil element, accounting for 22% of the

soil composite, based on the slope for samples when Al was found. Since the Al mdl is much less than the true value, when Al is missed because of interference, the soil composite is underestimated by about 22%. There is a much better substitution for the missing Al than the mdl. Figure 2 shows the excellent agreement between Al and Si for the most recent two years of data. The slope is 0.46 as is the ratio of means.
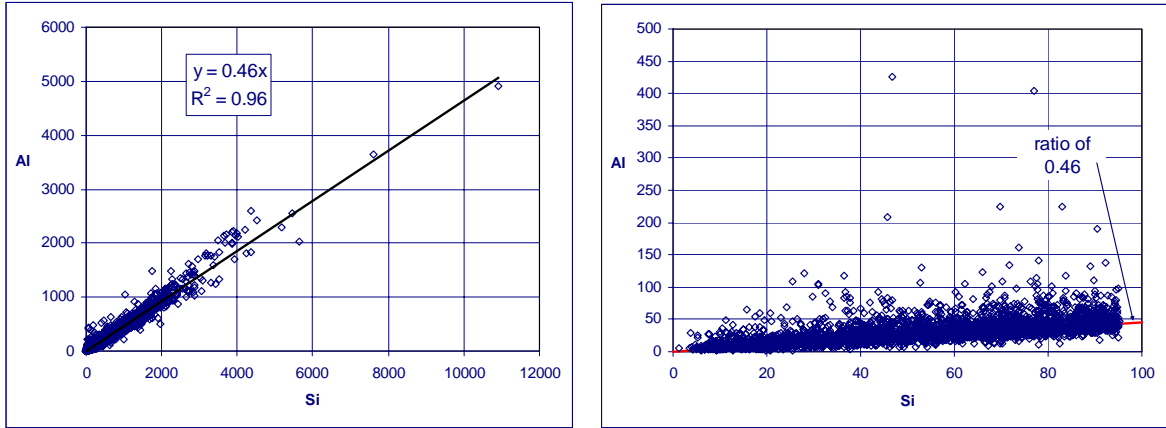

Figure 2. Comparison of Al and Si when both are found for all samples collected between December 1999 and November 2000. The right plot includes only the lower half of the data based on the soil composite. All units are ng/m^3.

For earlier years of the IMPROVE network the situation was complicated by the presence of Saharan dust. Throughout most of the 1990's, Saharan dust impacted Virgin Islands throughout the year and the continental U.S. for brief episodes during the summer. The comparisons are shown in Figure 4. The ratio for Saharan dust, found typically at the Virgin Islands site and occasionally in summer in the southeastern US, is approximately 0.60. The network showed only one probable impact in the continental US during the last four summers (1999 to 2002), at Big Bend on 8/4/99 .
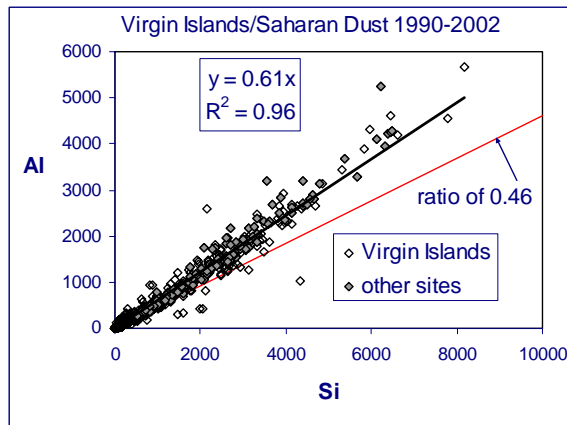

Figure 3. Comparison of Al and Si for samples at the Virgin Islands and at continental sites during selected summer episodes. All units are ng/m^3.

Figure 4 shows the relationship for 1988 to 1999 when these Saharan samples are removed. Also shown are samples collected at Sequoia in 1994 when the site was impacted by local construction.
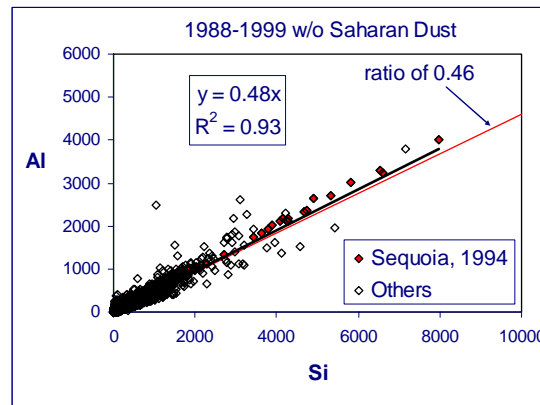

Figure 4. Comparison of Al and Si between 1988 and 1999 after removal of samples with probable Saharan dust impact. The shaded (red) diamonds were collected at Sequoia in 1994 when the site was impacted by construction. All units are ng/m^3.

The fact that Al and Si correlate so well allows us to use Si as a substitute for Al. It would be possible to eliminate Al completely and substitute 0.46 times Si for Al for every sample. This is certainly easiest and involves only changing the Al factor to zero and the Si factor from 2.49 to 2.49 + 0.46*2.20 = 3.50. However, this method would underestimate some high soil concentrations whenever the Al/Si ratio differs significantly from 0.46, such as at Virgin Islands and during Saharan episodes. A more conservative approach is to retain the measured Al when it was found and substitute 0.46 times Si only when it is missed. In general, Al is found for all samples with elevated soil concentrations. Thus it is unlikely that this method will produce an erroneous reconstructed extinction. In the rare case where Si is also missed, we again substitute one-half of the Al mdl. When both are missed, the chances are good that the true concentrations were below the mdl.

The current equation for the Al component is 2.20*MAX(AL, 0.5*AL_MDL). The revised equation could be written 2.20*MAX(IF(AL=0,0.46*SI, AL), 0.5*AL_MDL). The uncertainty calculation would use 0.46 times the Si uncertainty in this case.

Figure 5 compares the original and revised calculations of the soil composite. The two are equal when Al is found. The estimates differ when Al is missing. Since 0.46 times Si is generally much larger than 0.5 times the Al mdl, the revised calculation is considerable larger. Thus there are two lines differing by approximately 22%. The effect on reconstructed extinction will be less because soil plays a limited role in extinction. For all 30,300 samples in the data set, the arithmetic means differed by 8%. For the lowest 20% of the samples in terms of the original soil concentration, the means differed by 17%. This could make a significant difference in the reconstructed extinction.
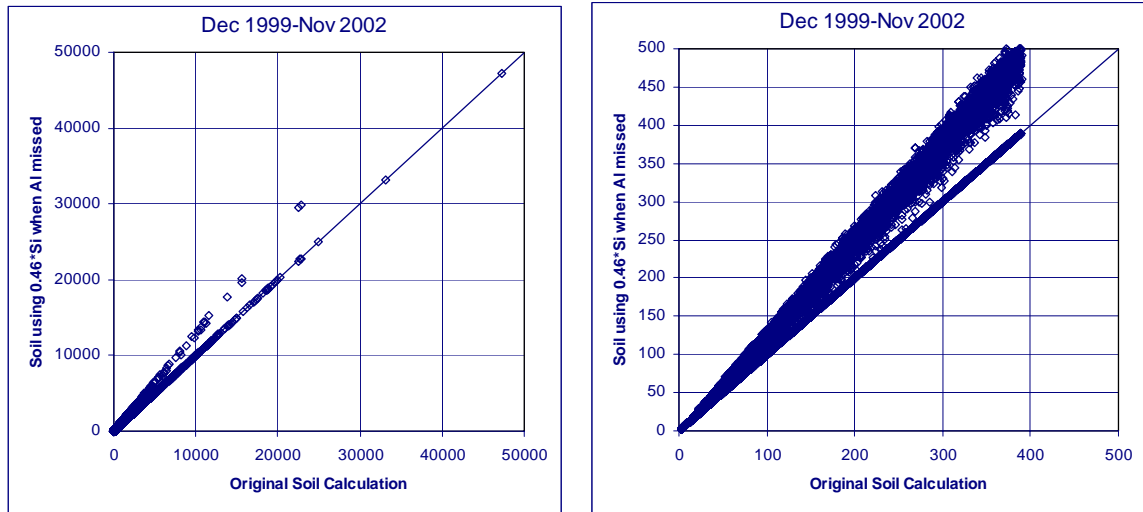
Figure 5. Comparison of original soil composite using mdl/2 when Al is missing with composite using 0.46 times Si when Al is missing. The right plot includes the lower half of the samples. All units are ng/m^3.

My conclusion is that it would be better to use 0.46 times Si instead of one-half the mdl when Al is missed. There are two ways to accomplish this. One is to revise the equations at CIRA used to calculate the soil parameter. The second is for UCD to make the substitution before providing the data. We would then need to either flag the data or provide a separate variable of measured Al. The advantage is that the current equation would remain the same.

Potassium. The existing equation uses only the soil component of K and estimates this using Fe. There are two issues here. First, should we use all of the measured K or continue to use only the soil fraction. We do not separate other elements such as Fe into soil and nonsoil components. The second issue is whether our original choice of Fe to estimate soil K was best. Because of the potential presence of nonsoil Fe, Si is a better surrogate for soil K. (See Figure 9 below.)

This gives us three choices:
    (1) Keep the existing equation.
    (2) Continue to use only the soil component of K but use Si instead of Fe. This changes the Si factor to 2.66 and the Fe factor to 1.58.
    (3) Include all of the K as K_2O. This changes the Fe factor to 1.58 and adds a K term at 1.40.

Both estimates in Figure 6 include only the soil component of K. The original equation uses 0.60*Fe to estimate soil K, while the revised equation uses 0.12*Si. There is very little difference in the soil composite. For all samples, the shift to Si will decrease the mean soil composite by 1%. For the lower 20% of samples based on the soil concentrations, decrease will be only 2%.
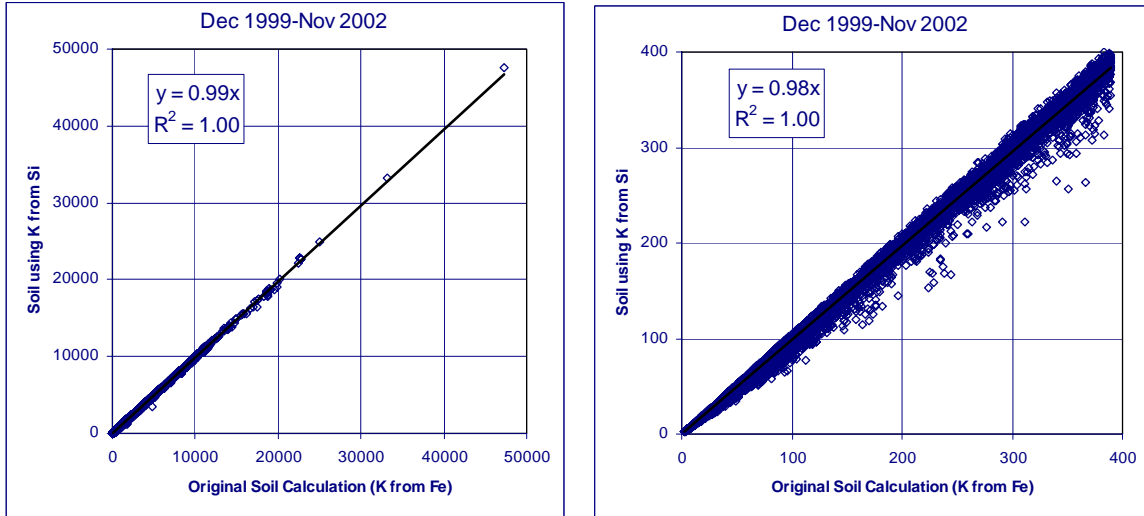
Figure 6. Comparison of the two soil composites using only a soil component of K. The x-variable uses Fe and the y-variable uses Si. The current composite is slightly larger because of the presence of nonsoil Fe. However, the slopes and ratio of means are close to unity. All units are ng/m^3.

Figure 7 shows the effect of using the entire measured K, including possible smoke components. For the total network, the shift to all K would increase the mean soil concentration by only 3%. However, the non-soil smoke component can make a difference for low soil concentrations. The revised soil composite increases by 20% or more for 15% of the samples and by 100% or more for 1% of the samples.
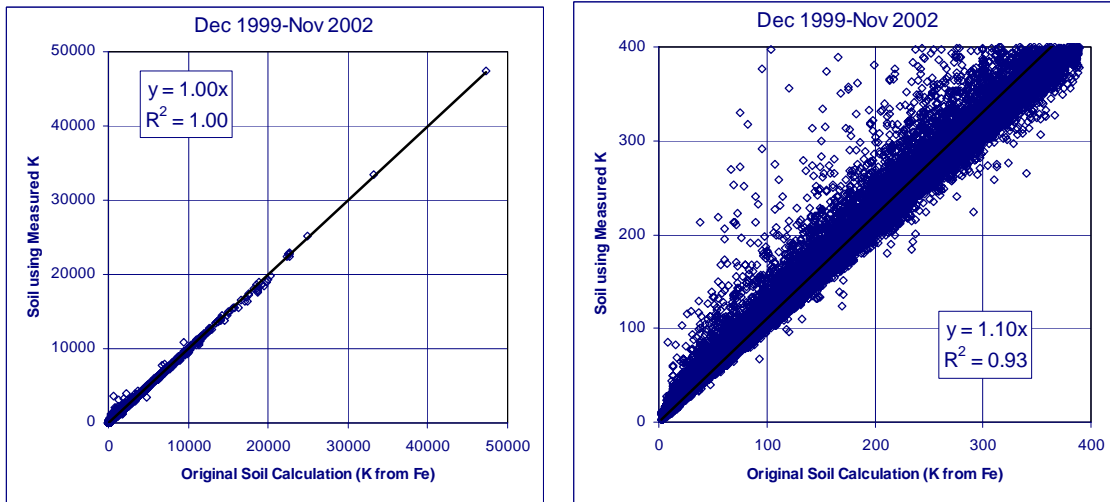

Figure 7. Effect of using measured K instead on only the non-soil fraction as estimated from Fe. The right plot includes the lower half of concentrations. All units are ng/m^3.

For the sake of continuity I recommend that we retain the current equation. Shifting the estimator from Fe to Si would be good, but not worth the change. I think our original choice to eliminate nonsoil K is still valid.

The 86% factor. Our choice of factor to account for unmeasured species is perhaps the most tenuous assumption in the equation. The 86% is based on average sediment, not on average airborne soil. It is obvious that there should be some factor, but is a factor of 86% reasonable?

One test would be to examine resuspended soil samples and compare the sum of the measured species to the gravimetric mass. We did this for soils in the San Joaquin valley, but unfortunately, the mass measurements were not sufficiently precise to provide any useful information. We could try to go back and re-examine other old data. DRI made resuspension measurements that they said required no extra factor.

One check from the IMPROVE measurements is to compare reconstructed and gravimetric mass for samples that are dominated by soil. The results for the most recent two years of data are shown in Figure 8. The reconstructed mass includes nitrate. Here, I selected only samples in which the soil composite was at least 50% of the total reconstructed mass. These samples were collected at 59 of the sites, so there is no indication of a bias toward specific sites. Of the 129 samples, 38% were from the desert southwest, 12% were from Virgin Islands, and the remaining 50% were scattered throughout the rest of the network. No correction was made for water on the particles at the time of gravimetric analysis. For all samples, the slope was 0.95. For the samples dominated by soil, the ratio was slightly smaller, 0.87. To get a slope of 0.95, the factor would be 0.76 rather than 0.86. That is, we would have to multiply the soil by a larger number. Using a factor of 1.0 (i.e. not correcting for minerals not measured) would move the slope even further from 0.95, dropping it to 0.79.
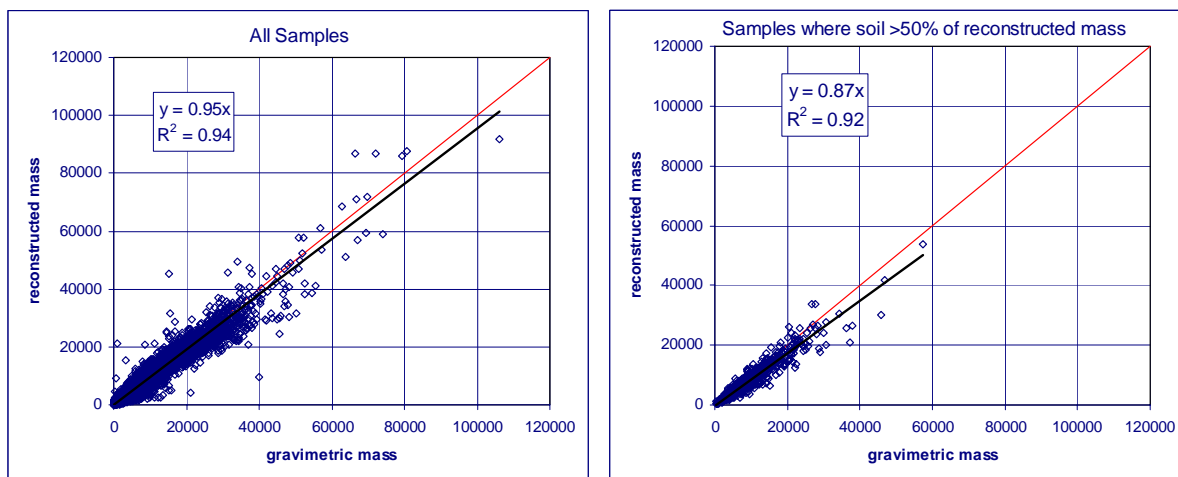

Figure 8. Comparison of gravimetric and reconstructed mass for samples collected between December 1999 and November 2002. The left plot is for all samples and the right plot for the small fraction of samples in which soil accounted for at least 50% of the reconstructed mass. All units are ng/m^3.

My recommendation is that we retain the 0.86 factor until we are able to determine a better value from resuspended soils.

The uncertainty in the soil composite. The current calculation of the uncertainty in the soil composite uses the quadratic sum of the uncertainties of the constituent concentrations.

$$[d(SOIL)]^2 = [2.20 * d(Al)]^2 + [2.49 * d(Si)]^2 + [1.63 * d(Ca)]^2 + [2.42 * d(Fe)]^2 + [1.94 * d(Ti)]^2$$

This assumes that the uncertainties are all independent. In reality, the uncertainties consist of an independent statistical component and shared fractional component of 5%, from flow rate and analytical calibration uncertainties. For very small concentrations, the independent statistical component dominates so the above equation is correct. For high concentrations, the estimated uncertainty is unreasonably low (about 3%). If we were to assume uncertainties were totally shared, the equation would be

$$d(SOIL) = 2.20*d(Al) + 2.49*d(Si) + 1.63*d(Ca) + 2.42*d(Fe) + 1.94*d(Ti)$$

This gives a reasonable uncertainty of 5% for high concentrations, but an unreasonably large estimate for concentrations. An alternative approach is separate the uncertainty for each element into independent and shared components. The independent components, calculated by removing the 5% term from each constituent uncertainty, are quadratically added. The 5% term is then quadratically added. The equation is

$$[d(SOIL)]^2 = 2.20^2 * \{[d(Al)]^2 - [0.05 * Al]^2\} + 2.49^2 * \{[d(Si)]^2 - [0.05 * Si]^2\} + ... + [0.05 * SOIL]^2$$

This equation gives the same result as the independent model at low concentrations, where we believe the independent model is correct. At large concentrations the uncertainties approach 5%, which is what we expect. I recommend we shift to this equation, even if it is more complicated.

Appendix. Relationship between the Soil Elements

Figure 9 shows the inter-relationship between Si and the other major soil elements. As discussed above, Al and Si correlate well. There is considerably more scatter with Ca and Fe. Some of the elevated Fe may be from nonsoil sources. The K plot clearly shows the nonsoil contributions. The points near the line at K/Si = 0.12 probably indicate that there is only a soil component. Points well above the line presumably have nonsoil components.

Because Si and Al correlate well and account for about 60% of the soil composite, there is a high correlation between Si and the soil composite. Figure 10 compares Si with the soil composite for samples where both Si and Al were found. It could be argued that Si alone is actually the best estimate for soil because some of the other elements have potential nonsoil sources.
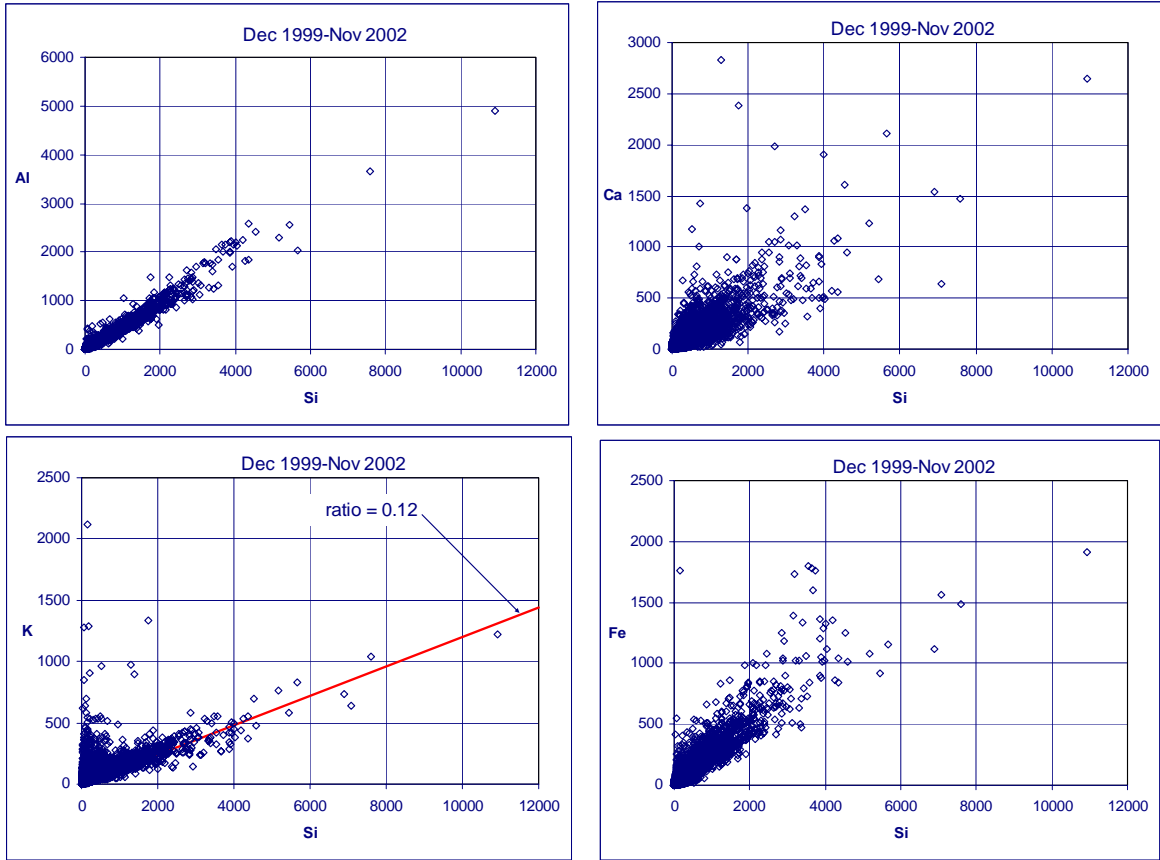
Figure 9. Comparison of the four major elements to Si. The line for K gives the presumed relationship for soil. The values above the line may be associated with nonsoil sources.
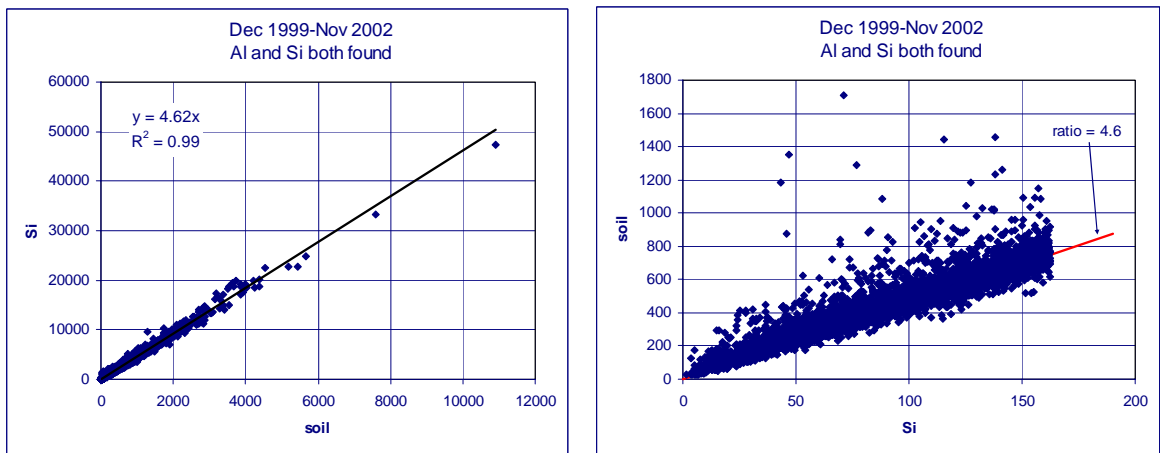

Figure 10. Comparison of Si and the soil composite for samples where both Si and Al were found. The right plot includes the lower half of concentrations. All units are ng/m³.