SRNL
SAVANNAH RIVER NATIONAL LABORATORY
TM

INTER-OFFICE MEMORANDUM

April 26, 2007

SRNL-SCS-2007-00011

To: B.T. Butcher, Jr., SRNL/Environmental Analysis & Performance
D. I. Kaplan, SRNL/Environmental Analysis & Performance
R. A. Hiergesell, SRNL/Environmental Analysis & Performance

From: E.P. Shine, SRNL/SCS

cc: SCS Files

4/25/2007
C.D. Harvel, Technical Reviewer

4/25/07
R.C.Tuckfield, Manager, Statistical Consulting Section

Preliminary Guidance for the Distribution of Cs, SR, and U Geochemical Input Terms to Stochastic Transport Models

Introduction

The performance assessment program will use a stochastic modeling approach to determine a range of risk posed by buried waste in E-area. This report documents lognormal fits of groundwater distributional parameters for cesium (Cs), strontium (Sr), and uranium (U). All statistical fitting was performed in JMP® Version 6.0.2 software from the SAS Institute, Inc.

The data listed in Appendix Table 1 consist of analyses of the distribution parameter Kd, as well as cation exchange capacity (CEC) and pH, taken at twenty depths along a single soil core that did not come from E-Area. At each of the depths a cross-section of the soil core was sampled, mixed, and analyzed. The data reflect the soil characteristics for this single available soil core, and the applicability to the soil beneath the E-area slit trenches is dependent on the similarity of the soil to this boring. Any relationships between pH or CEC and the Kd values were not studied in this investigation since their relationships, if any, are not exploited in the transport modeling.

The objective of this investigation is to determine whether a lognormal distribution appears to be a suitable choice for modeling the available borehole data. If such is the case, then it is consistent with the assumption that Kd values can be modeled as lognormal distributed parameters in the assessment of uncertainty for the transport model.

Examination of the Distributions of the Kd Parameters

A set of scatterplots for the natural logarithm of Cs, Sr, and U are presented in Appendix Figure 1. Sample 31A appears to be an outlier. It has a relatively high Kd value for Cs and a relatively low Kd value for U. From Appendix Table 1, the cation exchange capacity (CEC) of 10.98 mequ/100g is high relative to the other samples in this data set. So Sample 31A is a statistical outlier reflecting a different measurement profile than the remainder of the data from the sampled borehole. The inclusion or exclusion of this datum relies on a scientific assessment (outside the scope of this note) of the representativeness of the sampled borehole dataset or the subset for the zone beneath the E-Area slit trench. If Sample 31A is in a layer of soil from the sampled borehole with properties different from the soil beneath the trench, then it should be excluded from all analyses. Both the results with and without Sample 31A are given in the Appendix.

Exhibit 1A summarizes the fitting of a normal distribution to the Kd data for Cs. The mean of KD for Cs is 2,055 mL/g and is near the median of 2,034. The Kd distribution for Cs appears to have one or more outliers on the high side. The Wilk-Shapiro goodness of fit indicates that the normal distribution is a very poor fit for the distribution of Kd values (P-value = 0.0008). Removing the outlier from Sample 31A improves the fit (P-value = 0.3447), but examination of the normal quantile plot still indicates a high side outlier remains. The distribution of the natural logarithm of the Kd values for Cs is presented in Exhibit 1B. The goodness of fit test indicates that the fit is barely adequate (P-value = 0.0762) at $\alpha = 0.05$. Removing the largest Kd value from Sample 31A improves the fit considerably (P-value = 0.4347). a previous study, Kaplan and others, has suggested that the distribution of Kd values greater than 1,000 appears to be reasonably in accord with a lognormal distribution. The examination of an actual distribution of Kd parameters for Cs does not counter that insight. The existence of Kd outliers suggests patches or layers of soil that have different properties than most of the soil from the region of the sampled borehole. Since the outlier examined was high, the local effect would be a retardation of groundwater flow. The inclusion or exclusion of the Sample 31A outlier for Cs must be made based on a scientific assessment of its representativeness for the E-Area assessment and its impact on the transport model.

Exhibit1B presents a similar fit of a two-parameter lognormal distribution to the Kd data for Sr. The mean of the Kd values for Sr is 7.596 mL/g. The Wilk-Shapiro goodness-of-fit test is marginally adequate (P-value = 0.0839) for $\alpha = 0.05$ and improves slightly if Sample 31A is removed. However, the Kd value for Sample 31A is only one of the four largest Sr Kd values, none of which appears to be an outlier on the normal quantile plot. Consequently, its removal has no practical impact on the distribution of Kd values for Sr. Similar results are seen in Exhibit 2B for the natural logarithm of the SR Kd values, with the logarithm producing only a small improvement in the goodness of fit. The results indicate that either the normal or lognormal distribution would be an adequate fit for Kd values for Sr.

Exhibit 3A presents the statistics for the Kd values for U from the sampled borehole. The mean is 0.615 mL/g. The normal quantile plot shows one Kd value somewhat smaller and another Kd value larger than the bulk of measurements. The smallest Kd value is associated with Sample 31A. The goodness-of-fit to the normal distribution with all the data is marginal (P-value = 0.0668) and worsens when Sample 31A is removed (P-value = 0.0150). Application of the natural logarithm to the Kd values for U has mixed effects. The fit to the entire dataset is worse for the lognormal distribution than the normal distribution because the low outlier for Sample 31A is more pronounced after taking the log. With Sample 31A removed, the goodness-of–fit to the natural logarithm of Kd values appears to fit reasonably well (P-value = 0.1713).

The fits to the available data provide the lead to the following conclusions.

- The Kd data for Cs, Sr, and U cover three different ranges: over 1,000 mL/g, a relatively compact region from approximately 2.5 to 2.8 mL/g, and a region between 0.3 and 1 mL/g.

- The goodness-of-fit of either the normal or the lognormal distribution is chiefly influenced by the homogeneity of the sampled borehole. The distribution parameters appear very homogeneous for Sr and, consequently, either the normal or lognormal distributions provide an adequate fit.

- Sample 31A with relatively large CEC has a pronounced effect of the Kd fits for Cs and U. A normal distribution fit to all the Cs distribution parameters is very poor; removing Sample 31A makes the normal distribution an acceptable fit. The lognormal distribution improves the fit with and without Sample 31A. Sample 31A is the lowest Kd value for U.

Appendix

Table 1. Soil Parameters obtained from the core

Sample ID	Top Depth (m)	Bottom Depth (m)	Avg Depth	pH	CEC meq/ 100g	Cs (mL/g)	Cs +/-	Sr (mL/g)	Sr +/-	U (mL/g)	U +/-	Ln Cs	Ln Sr	Ln U
07A	14	14.6	14.3	8.54	5.07	2044.8	275.1	14.09	0.83	0.94	0.12	7.62306	2.64547	-0.06188
10A	17.6	18.2	17.9	8.80	4.73	2090.0	305.2	14.25	0.49	0.67	0.08	7.64492	2.65676	-0.40048
12A	21.2	21.6	21.4	8.77	4.60	2023.8	305.2	13.81	0.78	0.64	0.10	7.61273	2.62539	-0.44629
14A	24.5	25.2	24.85	8.73	4.62	1969.8	39.5	13.94	0.50	0.67	0.05	7.58569	2.63476	-0.40048
15A	27.6	28.3	27.95	8.75	4.11	1502.1	80.1	13.64	0.85	0.74	0.05	7.31462	2.61301	-0.30111
16A	30.6	31.4	31	8.77	2.32	1535.9	269.8	12.81	0.33	0.57	0.04	7.33687	2.55023	-0.56212
17A	33.5	34.2	33.85	8.52	4.98	2267.2	388.1	15.46	0.33	0.68	0.15	7.72630	2.73826	-0.38566
19A	36.9	37.6	37.25	8.50	4.72	2861.7	395.7	14.25	0.45	0.51	0.08	7.95917	2.65676	-0.67334
20A	39.5	40.3	39.9	8.52	4.67	2251.8	291.8	15.32	0.37	0.65	0.08	7.71949	2.72916	-0.43078
21A	43.1	43.9	43.5	8.56	4.56	2072.5	408.9	14.91	0.63	0.57	0.09	7.63651	2.70203	-0.56212
22A	46.3	47.1	46.7	8.94	7.33	1373.9	331.6	12.18	0.29	0.59	0.08	7.22541	2.49980	-0.52763
23A	48.9	49.7	49.3	8.82	8.41	2295.2	279.9	13.04	0.95	0.58	0.05	7.73858	2.56802	-0.54473
24A	55.1	55.7	55.4	8.81	9.03	2213.3	454.7	16.92	0.92	0.55	0.05	7.70224	2.82850	-0.59784
25A	57.8	58.5	58.15	8.89	6.63	1716.9	484.4	12.35	0.85	0.56	0.07	7.44828	2.51366	-0.57982
27A	59.2	61.4	60.3	8.88	8.36	1563.0	259.0	16.17	1.42	0.59	0.08	7.35436	2.78316	-0.52763
29A	64.3	64.4	64.35	8.84	7.77	2128.7	423.1	14.22	0.34	0.50	0.13	7.66327	2.65465	-0.69315
31A	65.4	67.8	66.6	8.56	10.98	4068.2	258.7	16.90	0.11	0.30	0.12	8.31096	2.82731	-1.20397
32A	69.8	70.5	70.15	8.93	8.39	1792.0	315.2	16.89	1.72	0.70	0.11	7.49109	2.82672	-0.35667
34A	72	72.6	72.3	8.92	6.21	1897.0	178.3	16.92	1.53	0.61	0.16	7.54803	2.82850	-0.49430
35A	73	73.6	73.3	8.89	6.65	1435.0	133.3	16.79	2.68	0.68	0.13	7.26892	2.82078	-0.38566

Exhibit 1. Scatterplots of Loge Cs, loge Sr, and loge U.

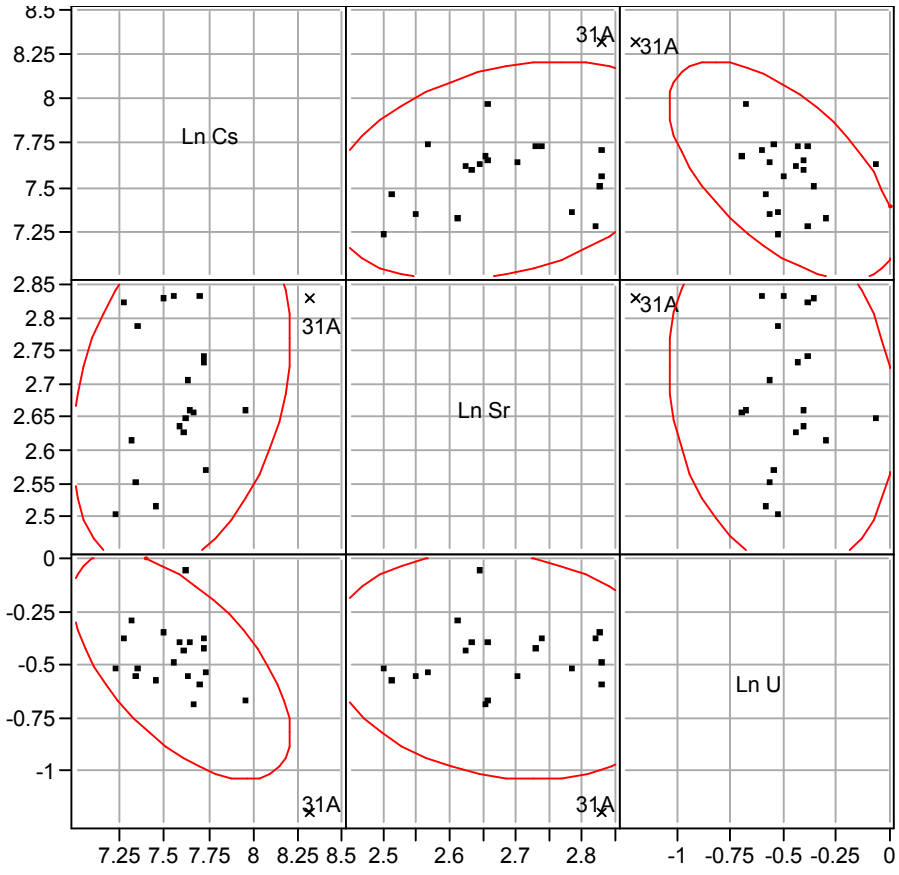
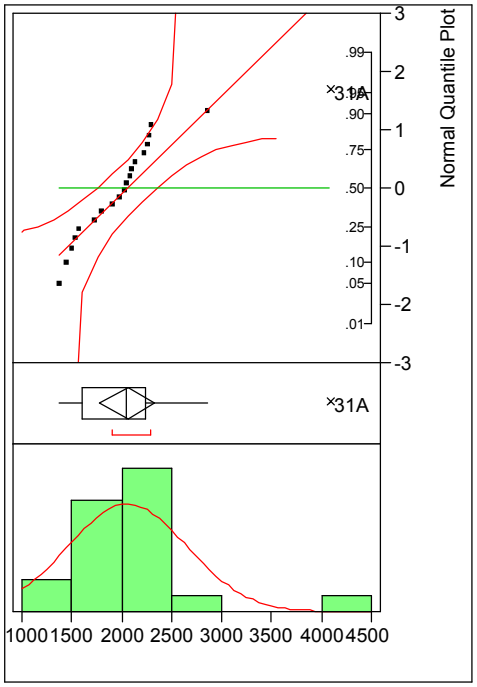


Exhibit 1A. Distribution of Kd for Cs

Cs (all data)



Quantiles

100.0%	maximum	4068.2
99.5%		4068.2
97.5%		4068.2
90.0%		2805.1
75.0%	quartile	2242.2
50.0%	median	2034.3
25.0%	quartile	1601.5
10.0%		1441.7
2.5%		1373.9
0.5%		1373.9
0.0%	minimum	1373.9

Moments

Mean	2055.14
Std Dev	596.45979
Std Err Mean	133.37246
upper 95% Mean	2334.2918
lower 95% Mean	1775.9882
Number of Samples	20
Variance	355764.28
Skewness	2.1146736
Kurtosis	6.425857
Coefficient of Variation	29.02283
N samples Excluded	0

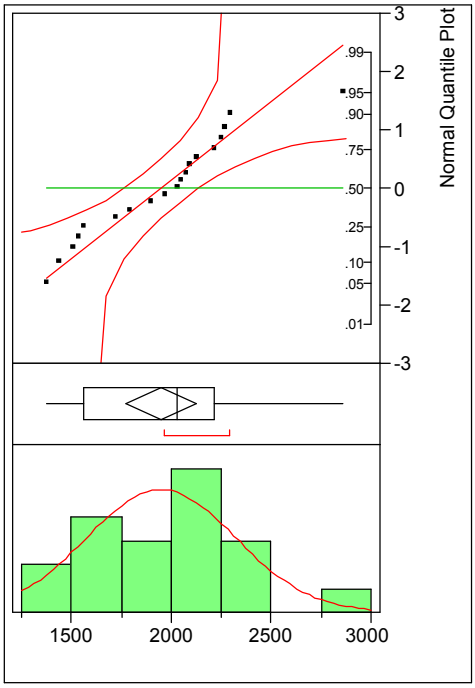
Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	2055.14	1775.9882	2334.2918
Dispersion	596.45979	453.60194	871.17225

Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.799404	0.0008

Cs (Excluding Sample 19A)



Quantiles

100.0%	maximum	2861.7
99.5%		2861.7
97.5%		2861.7
90.0%		2295.2
75.0%	quartile	2213.3
50.0%	median	2023.8
25.0%	quartile	1563.0
10.0%		1435.0
2.5%		1373.9
0.5%		1373.9
0.0%	minimum	1373.9

Moments

Mean	1949.1895
Std Dev	372.21752
Std Err Mean	85.392556
upper 95% Mean	2128.5926
lower 95% Mean	1769.7864
Number of Samples	19
Variance	138545.88
Skewness	0.3983362
Kurtosis	0.4993751
Coefficient of Variation	19.096015
N Samples Excluded	1

Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	1949.1895	1769.7864	2128.5926
Dispersion	372.21752	281.25226	550.44441

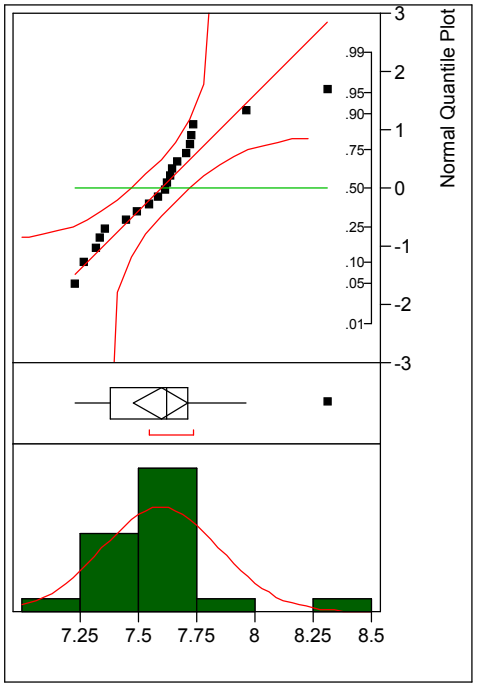
Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.946553	0.3447

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Exhibit 1B. Distribution of the Natural Logarithm of Kd for Cs

Ln Cs (all data)



Quantiles

100.0%	maximum	8.3110
99.5%		8.3110
97.5%		8.3110
90.0%		7.9371
75.0%	quartile	7.7152
50.0%	median	7.6179
25.0%	quartile	7.3778
10.0%		7.2735
2.5%		7.2254
0.5%		7.2254
0.0%	minimum	7.2254

Moments

Mean	7.5955237
Std Dev	0.2513888
Std Err Mean	0.0562122
upper 95% Mean	7.7131773
lower 95% Mean	7.4778701
Number of Samples	20
Variance	0.0631963
Skewness	1.0292603
Kurtosis	2.3466269
N samples Excluded	0

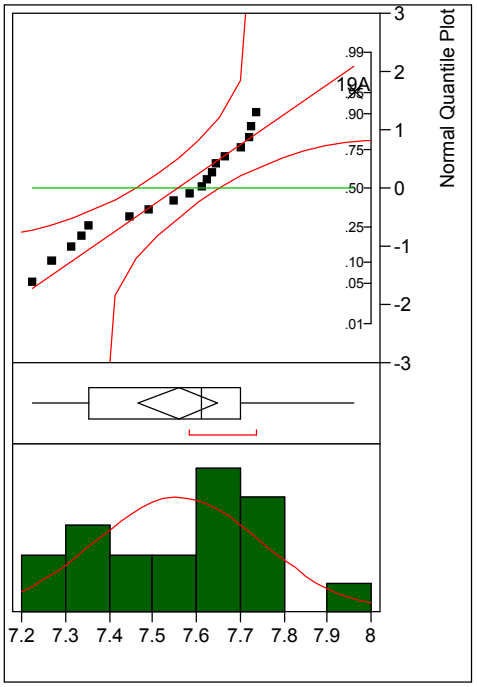
Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	7.5955237	7.4778701	7.7131773
Dispersion	0.2513888	0.1911788	0.3671714

Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.914051	0.0762

Ln Cs (Excluding Sample 19A)



Quantiles

100.0%	maximum	7.9592
99.5%		7.9592
97.5%		7.9592
90.0%		7.7386
75.0%	quartile	7.7022
50.0%	median	7.6127
25.0%	quartile	7.3544
10.0%		7.2689
2.5%		7.2254
0.5%		7.2254
0.0%	minimum	7.2254

Moments

Mean	7.5578694
Std Dev	0.1917678
Std Err Mean	0.0439945
upper 95% Mean	7.6502985
lower 95% Mean	7.4654403
Number of Samples	19
Variance	0.0367749
Skewness	-0.094957
Kurtosis	-0.352601
N Samples Excluded	1

Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	7.5578694	7.4654403	7.6502985
Dispersion	0.1917678	0.1449022	0.2835909

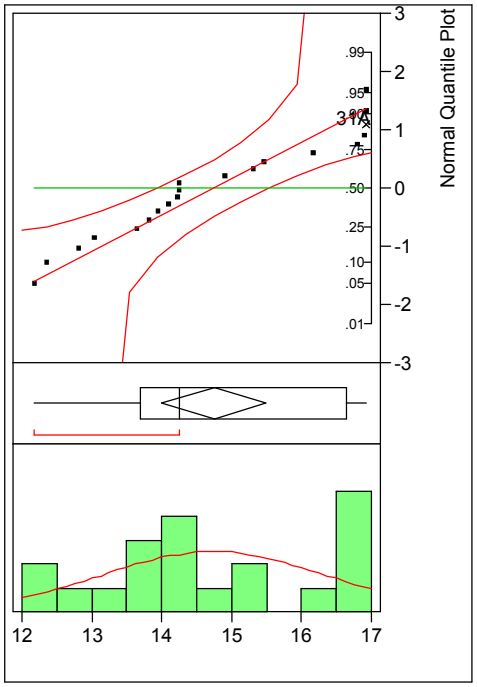
Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.952463	0.4347

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Exhibit 2A. Distribution of Kd for Sr

Sr (all data)



Quantiles

100.0%	maximum	16.920
99.5%		16.920
97.5%		16.920
90.0%		16.918
75.0%	quartile	16.635
50.0%	median	14.250
25.0%	quartile	13.683
10.0%		12.396
2.5%		12.180
0.5%		12.180
0.0%	minimum	12.180

Moments

Mean	14.743
Std Dev	1.6008619
Std Err Mean	0.3579636
upper 95% Mean	15.492226
lower 95% Mean	13.993774
Number of Samples	20
Variance	2.5627589
Skewness	0.0938186
Kurtosis	-1.215723
Coefficient of Variation	10.858454
N Samples Excluded	0

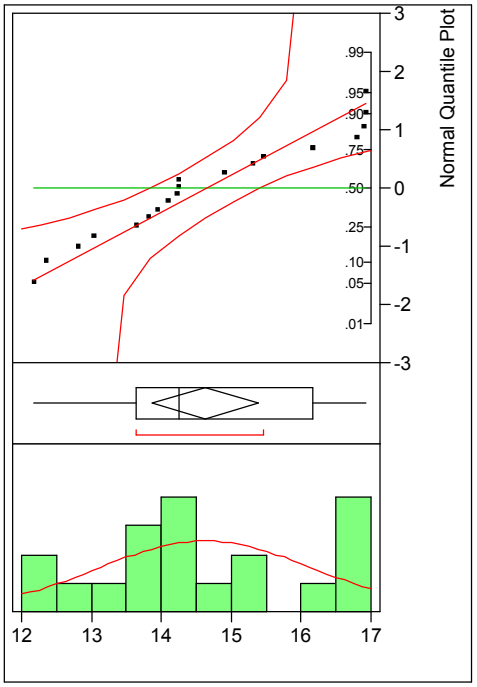
Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	14.743	13.993774	15.492226
Dispersion	1.6008619	1.2174401	2.3381735

Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.916233	0.0839

Sr (Excluding Sample 19A)



Quantiles

100.0%	maximum	16.920
99.5%		16.920
97.5%		16.920
90.0%		16.920
75.0%	quartile	16.170
50.0%	median	14.250
25.0%	quartile	13.640
10.0%		12.350
2.5%		12.180
0.5%		12.180
0.0%	minimum	12.180

Moments

Mean	14.629474
Std Dev	1.5598236
Std Err Mean	0.3578481
upper 95% Mean	15.381285
lower 95% Mean	13.877663
Number of Samples	19
Variance	2.4330497
Skewness	0.1879776
Kurtosis	-1.063749
Coefficient of Variation	10.662199
N Samples Excluded	1

Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	14.629474	13.877663	15.381285
Dispersion	1.5598236	1.1786224	2.3067054

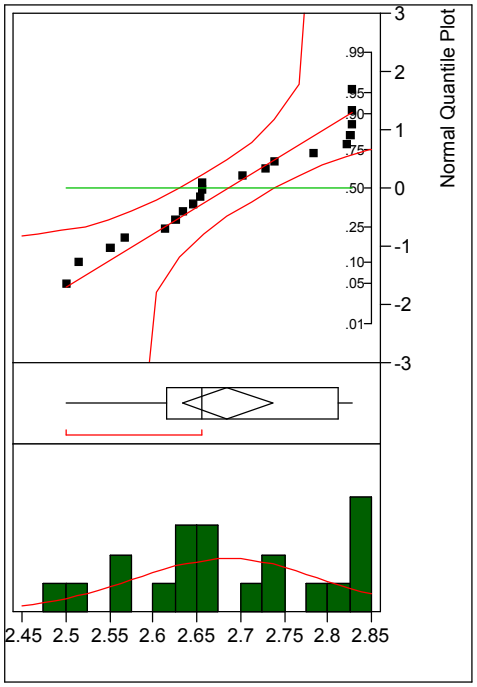
Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.928529	0.1627

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Exhibit 2B. Distribution of the Natural Logarithm of Kd for Sr

Ln Sr (all data)



Quantiles

100.0%	maximum	2.8285
99.5%		2.8285
97.5%		2.8285
90.0%		2.8284
75.0%	quartile	2.8114
50.0%	median	2.6568
25.0%	quartile	2.6161
10.0%		2.5173
2.5%		2.4998
0.5%		2.4998
0.0%	minimum	2.4998

Moments

Mean	2.6851453
Std Dev	0.108946
Std Err Mean	0.0243611
upper 95% Mean	2.7361336
lower 95% Mean	2.634157
Number of Samples	20
Variance	0.0118692
Skewness	-0.047354
Kurtosis	-1.122522
N Samples Excluded	0

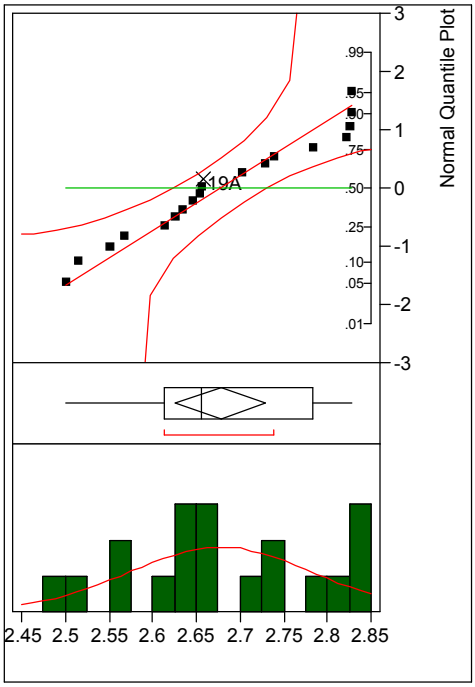
Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	2.6851453	2.634157	2.7361336
Dispersion	0.108946	0.0828524	0.1591235

Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.922898	0.1127

Ln Sr (Excluding Sample 19A)



Quantiles

100.0%	maximum	2.8285
99.5%		2.8285
97.5%		2.8285
90.0%		2.8285
75.0%	quartile	2.7832
50.0%	median	2.6568
25.0%	quartile	2.6130
10.0%		2.5137
2.5%		2.4998
0.5%		2.4998
0.0%	minimum	2.4998

Moments

Mean	2.6776628
Std Dev	0.1065207
Std Err Mean	0.0244375
upper 95% Mean	2.7290041
lower 95% Mean	2.6263214
Number of Samples	19
Variance	0.0113467
Skewness	0.0345268
Kurtosis	-1.007891
N Samples Excluded	1

Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	2.6776628	2.6263214	2.7290041
Dispersion	0.1065207	0.0804884	0.1575254

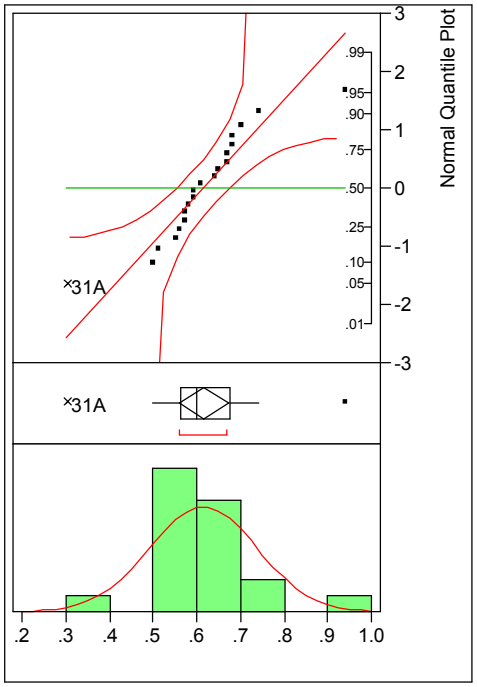
Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.935854	0.2218

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Exhibit 3A. Distribution of Kd for U

U (all data)



Quantiles

100.0%	maximum	0.94000
99.5%		0.94000
97.5%		0.94000
90.0%		0.73600
75.0%	quartile	0.67750
50.0%	median	0.60000
25.0%	quartile	0.56250
10.0%		0.50100
2.5%		0.30000
0.5%		0.30000
0.0%	minimum	0.30000

Moments

Mean	0.615
Std Dev	0.1218498
Std Err Mean	0.0272464
upper 95% Mean	0.6720275
lower 95% Mean	0.5579725
Number of Samples	20
Variance	0.0148474
Skewness	0.1035672
Kurtosis	3.5680583
Coefficient of Variation	19.812972
N Samples Excluded	0

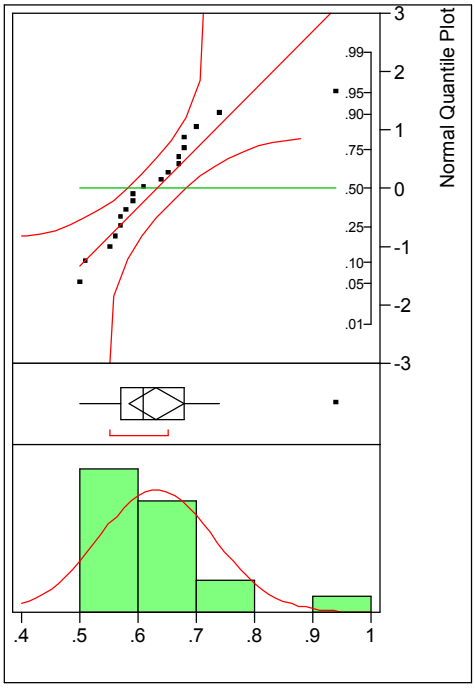
Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	0.615	0.5579725	0.6720275
Dispersion	0.1218498	0.0926656	0.1779703

Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.911053	0.0668

U (Exclude Sample 19A)



Quantiles

100.0%	maximum	0.94000
99.5%		0.94000
97.5%		0.94000
90.0%		0.74000
75.0%	quartile	0.68000
50.0%	median	0.61000
25.0%	quartile	0.57000
10.0%		0.51000
2.5%		0.50000
0.5%		0.50000
0.0%	minimum	0.50000

Moments

Mean	0.6315789
Std Dev	0.0993458
Std Err Mean	0.0227915
upper 95% Mean	0.6794621
lower 95% Mean	0.5836958
Number of Samples	19
Variance	0.0098696
Skewness	1.6171874
Kurtosis	4.2787279
Coefficient of Variation	15.729754
N Samples Excluded	1

Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	0.6315789	0.5836958	0.6794621
Dispersion	0.0993458	0.075067	0.146915

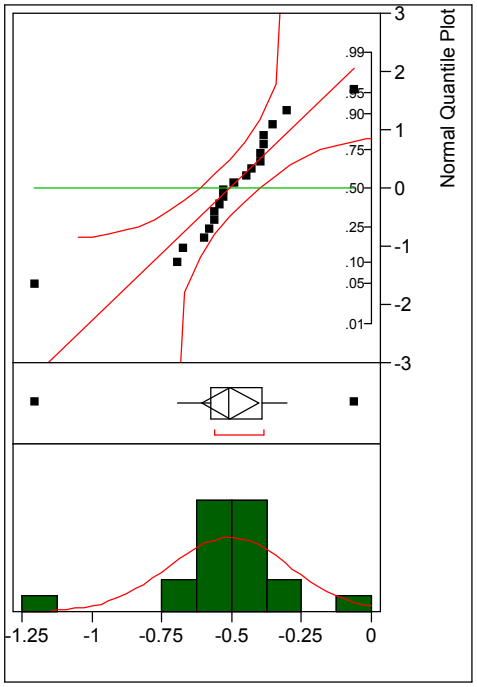
Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.871007	0.0150

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Exhibit 3B. Distribution of the Natural Logarithm of Kd for U

Ln U (all data)



Quantiles

100.0%	maximum	-0.062
99.5%		-0.062
97.5%		-0.062
90.0%		-0.307
75.0%	quartile	-0.389
50.0%	median	-0.511
25.0%	quartile	-0.575
10.0%		-0.691
2.5%		-1.204
0.5%		-1.204
0.0%	minimum	-1.204

Moments

Mean	-0.506783
Std Dev	0.2170276
Std Err Mean	0.0485288
upper 95% Mean	-0.405211
lower 95% Mean	-0.608355
Number of Samples	20
Variance	0.047101
Skewness	-1.400277
Kurtosis	5.5968888
N Samples Excluded	0

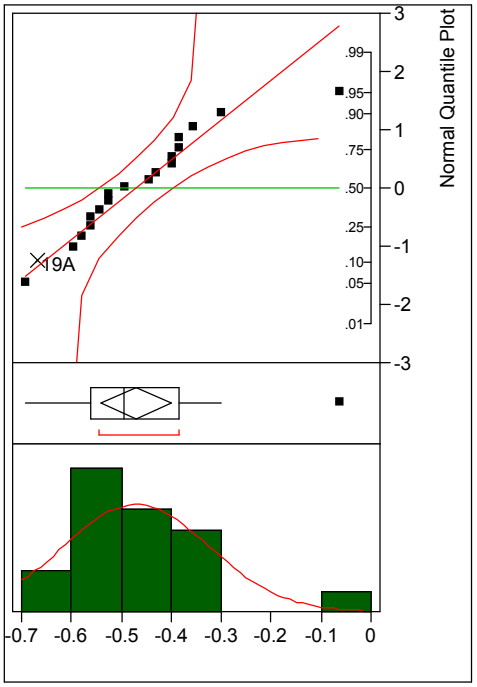
Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	-0.506783	-0.608355	-0.405211
Dispersion	0.2170276	0.1650474	0.3169844

Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.854936	0.0065

Ln U (Exclude Sample 19A)



Quantiles

100.0%	maximum	-0.0619
99.5%		-0.0619
97.5%		-0.0619
90.0%		-0.3011
75.0%	quartile	-0.3857
50.0%	median	-0.4943
25.0%	quartile	-0.5621
10.0%		-0.6733
2.5%		-0.6931
0.5%		-0.6931
0.0%	minimum	-0.6931

Moments

Mean	-0.470088
Std Dev	0.1459189
Std Err Mean	0.0334761
upper 95% Mean	-0.399758
lower 95% Mean	-0.540419
Number of samples	19
Variance	0.0212923
Skewness	1.0189951
Kurtosis	2.1942298
N Samples Excluded	1

Fitted Normal Parameter Estimates

Type	Estimate	Lower 95%	Upper 95%
Location	-0.470088	-0.540419	-0.399758
Dispersion	0.1459189	0.1102581	0.2157884

Shapiro-Wilk W Goodness-of-Fit Test

W	Prob<W
0.929737	0.1713

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.