

Supporting Information for

Analysis of Phosphorus Soil Sorption Data: Improved Results from Global Least-Squares Fitting by

Joel Tellinghuisen,*

Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235, United States

Email: joel.tellinghuisen@vanderbilt.edu. Phone: (615) 322-4873. Fax: (615) 343-1234

Paul Holford,

School of Science, Western Sydney University, Locked Bag 1797 Penrith 2751, Australia

Paul J Milham,

Hawkesbury Institute for the Environment, Western Sydney University, LB1797, Penrith 2751, Australia

ABSTRACT

Key soil properties are presented in Table S-1. The procedure for estimating variance functions for C and C_0 is described. The sorption data are presented in Table S-2, and the parameters from the alternative individual- Q analysis are in Table S-3. Note that the KA data for the highest P_f were excluded from the linear- Q global analysis summarized in Table 2 of the paper but were included in the individual- Q analysis. Results from several alternative linear- Q global fits for the KA data are given in Table S-4. The anomalies in the KA data are discussed, as is the unusual dependence of the MA global fit results on data selection for $P_f = 0$, shown in Figure S-2. Figure S-3 illustrates the exact agreement for analysis with Equations (3) and (4), taking S (or C_0) as dependent.

Soil Properties and Data Collection.

The soil properties are summarized in Table S-1. The soils are mainly drawn from the higher rainfall coastal zone of eastern Australia, where similar soils are common. All are acidic, support long-term pastures, and are low in available P. Approximately 30 kg subsamples of each soil were conditioned with differing amounts of $\text{Ca}(\text{H}_2\text{PO}_4)_2$ through multiple wetting/drying cycles before loading them into trays and conducting a surface runoff study under controlled conditions. The measured soil properties included organic carbon (OC); sand, silt, and clay size fractions; and total P (P_t). Samples were hand-crushed to pass a 2 mm sieve. The electrical conductivity (5 g: 25 mL water) of all samples was low (0.02–0.2 dS.m⁻¹).

Aliquots (5 g) of soil were suspended in 25 mL of 10 mM CaCl_2 , to which 50 μL of reagent grade chloroform was added (as an antimicrobial), and the suspensions conditioned by tumbling

end-over-end for 1 h at 14 revolutions min^{-1} . Up to 9 soil suspensions were spiked with varying amounts of PO_4 — as potassium dihydrogen phosphate dissolved in 10 mM CaCl_2 — and diluted to a final volume of 50 mL using 10 mM CaCl_2 . After spiking, suspensions were shaken as before for 17 h, after which aliquots of the supernatants were filtered (0.2 μm) and the concentrations of molybdate-reactive P in solution measured spectrophotometrically at 880 nm [29].

Table S-1. Initial soil characteristics.^a

Property	Camden	Flaxley	Glenmore	Moss Vale	Richmond
(units)	(KA)	(FL)	(MO)	(MA)	(D)
Site co-ordinates	-34.113181N, 150.701915E	-35.130802N, 138.815566E	-34.065972N, 150.595936E	-34.582482N, 150.534147E	-33.606433N, 150.734769E
Australian soil class	Brown Chromosol	Brown Chromosol	Red Chromosol	Brown Kurosol	Red Kandosol
US Taxonomy	Paleustalf	Haptustalf	Rhodustalf	Hapludult	Dystrudept
pH _{Ca}	5.1–5.4	5.3–5.5	4.8–5.0	4.3–4.5	5.0–5.3
Sand (g.kg^{-1})	520	520	230	270	850
Silt (g.kg^{-1})	170	190	340	300	50
Clay (g.kg^{-1})	310	290	430	430	100
Al _{ox} (mg.kg^{-1})	1300	2500	2900	4300	1100
Fe _{ox} (mg.kg^{-1})	3300	3800	5100	4400	970
P _{ox} (mg.kg^{-1})	230	330	480	160	135
OC (g.kg^{-1})	38	55	28	25	37
P _t (mg.kg^{-1})	510	690	1100	370	220

^aFrom Dougherty, *et al.* [28]. See also ref [25] in the paper.

Variance Function Estimation

The approach we have adopted is as follows: First we assume zero error in C and constant error in C_0 , and we analyze all the experimental data with the global model and collect the studentized residuals. We then repeat for zero error in C_0 and constant error in C . If either of these models should be correct, the residuals for the uncertain variable correctly treated as uncertain would support the constant variance model for that variable. Anticipating that this fortuitous situation does not occur, the next step is to run MC simulations on a representative data model, analyzed the same way — *i.e.*, in turn assuming constant uncertainty for C and C_0 — in a trial-and-error procedure to match the results for the actual data. For this purpose we devised two conventional NLS programs, one fitting C_0 as a function of C using the global version of Equation

(4), the other fitting C as a function of C_0 , using a numerical algorithm to solve iteratively (Newton's method) for each required value of C , again in the global model. In each case the weights were taken as 1.0, but the random error added to C and C_0 could be anything. For these simulations the data for the D soils (26 points, 3 P_{fs}) were taken as template.

Studentization of the residuals is accomplished using the "hat" matrix $\mathbf{H}$ to convert residual variance to approximate data variance through

$$\sigma_{\text{res},i}^2 = (1-H_{ii}) \sigma_i^2, \quad (\text{S-1})$$

where H_{ii} is the i th diagonal element of $\mathbf{H}$, given by

$$\mathbf{H} = \mathbf{XVX}^T\mathbf{W}. \quad (\text{S-2})$$

Here $\mathbf{V}$ is the covariance matrix and $\mathbf{W}$ is the diagonal weight matrix ($W_{ii} = w_i$). The elements of $\mathbf{X}$ are $X_{ij} = \partial y_i / \partial \beta_j$, where $y = C$ when C is fitted and C_0 when C_0 is fitted, and β_j is the j th adjustable parameter. ($\mathbf{X}$ is typically evaluated in solving the NLS equations.) The studentized residual is thus the observed residual δ_i divided by $(1-H_{ii})^{1/2}$.

The MC simulations used to estimate the data VFs typically comprised 1000 analyzed data sets. Random errors were assigned to C and C_0 in accord with the trial VFs, and the data were then fitted assuming either C or C_0 uncertain, with $w_i = 1$. Statistics were collected for the studentized residuals of representative C or C_0 values, and the SDs were compared with those obtained for the experimental data.

The simulations were conducted with assumed VFs an order of magnitude smaller than indicated by the SDs in Figure 5, to avoid problems that cause failure, like negative C values. To an adequate approximation, these results can be scaled into agreement with the observations. Figure S-1 illustrates the behavior for some simple assumed data VFs. In particular, note that even constant error in the "other" variable can manifest as apparent heteroscedasticity in the target variable. Indeed, most of the results in Figure S-1 are for assumed zero error in the target variable. In one exception to the use of constant weighting for the target variable, the results in (a-iv) were obtained using correct weighting, whereupon they confirm that the studentized residuals correctly reproduce the assumed VF.

The results in Figure S-1 show that the connection between the VFs and the residuals is not simple, as changes in the VF for one variable affect the residuals for both. After much trial-and-error effort to match the behavior of the experimental residuals in Figure 5, we settled on the VFs given in Equations (11) and included in the caption to Figure S-1.

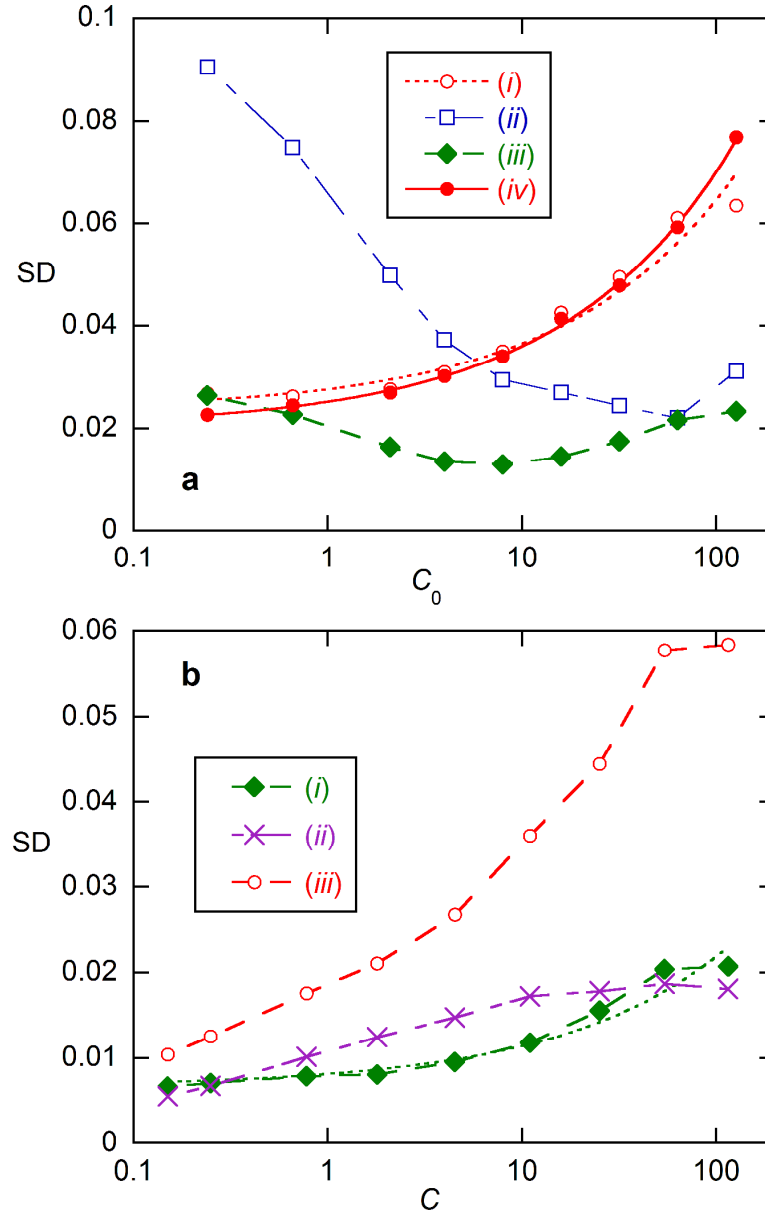


Figure S-1. Standard deviations from studentized residuals obtained from MC simulations fitting data to the global model for **(a)** assumed constant error in C_0 (except in *(iv)*) and zero error in C , and **(b)** constant error in C and zero error in C_0 . Each displayed point represents statistics for 1000 residuals, for which the RSD is $\sim(2000)^{-1/2}$. Random error was added to C and C_0 as follows: In **(a-i)**, **(a-iv)**, and **(b-iii)**: $\sigma_C = 0$, $\sigma_{C_0} = 0.02 + 0.005 C_0^{1/2}$; **(a-ii)**: $\sigma_C = 0.02$, $\sigma_{C_0} = 0$; **(a-iii)** and **(b-i)**: $\sigma_C = 0.005 + 0.002 C^{1/2}$, $\sigma_{C_0} = 0$; **(b-ii)** $\sigma_C = 0$, $\sigma_{C_0} = 0.02$. The curves in **(a-i)**, **(a-iv)**, and **(b-i)** (dotted) are results from NLS fits to the assumed VF; the agreement in **(a-iv)** was obtained by analyzing with correct weighting, confirming that the studentized residuals give the correct VF.

Table S-2. Sorption data^a. Omitted from global analysis of Table 2; omitted from alternative analyses of KA data (Table S-3).

Soil sample	C^b	X	C_0	P_f	Fe _{ox}	Al _{ox}	P _{ox}	pH
MA1	-0.01	4.9	0.48	0	4293	7460	157	4.25
	0.00	19.9	1.99					
	0.00	39.9	3.99					
	-0.01	40.1	4.00					
	-0.03	81.1	8.08					
	0.34	317.6	32.10					
	2.24	612.7	63.51					
	17.21	1083.5	125.56					
	91.50	1689.6	260.46					
	293.98	2240.3	518.01					
MA11	0.04	-0.4	0.00	604	4651	8110	941	4.54
	0.00	4.8	0.48					
	0.06	19.3	1.99					
	0.08	39.1	3.99					
	0.08	39.2	4.00					
	0.09	79.9	8.08					
	1.41	306.9	32.10					
	6.82	566.9	63.51					
	28.69	968.7	125.56					
MA15	0.60	-6.0	0.00	1322	4500	8223	1795	4.54
	0.55	-0.7	0.48					
	0.72	12.7	1.99					
	0.79	32.0	3.99					
	0.85	31.5	4.00					
	1.21	68.7	8.08					
	5.33	267.7	32.10					
	16.69	468.2	63.51					
	49.30	762.6	125.56					
D2	0.03	4.5	0.48	0	917	1767	135	4.96
	0.33	16.6	1.99					
	1.11	28.8	3.99					
	3.38	47.0	8.08					
	3.35	47.3	8.08					
	22.63	94.7	32.10					
	50.72	127.9	63.51					
	109.52	160.4	125.56					
D4	2.79	-27.9	0.00	1098	944	1924	665	5.30
	3.29	-22.8	1.01					
	4.00	-20.1	1.99					
	5.79	-18.0	3.99					
	8.66	-5.8	8.08					
	15.54	3.1	15.85					
	29.76	23.4	32.10					
	59.64	38.7	63.51					
	119.38	61.8	125.56					

D6	0.15	-1.5	0.00	81	805	1602	146	5.02
	0.25	2.3	0.48					
	0.78	12.1	1.99					
	1.80	21.9	3.99					
	4.53	35.5	8.08					
	10.98	48.7	15.85					
	25.05	70.5	32.10					
	54.19	93.2	63.51					
	115.19	103.7	125.56					
FL2	0.00	0.0	0.00	0	2999	3730	333	5.26
	0.00	4.8	0.48					
	0.10	18.9	1.99					
	0.36	36.3	3.99					
	0.33	36.7	4.00					
	4.05	118.0	15.85					
	11.97	201.3	32.10					
	31.24	322.7	63.51					
	84.70	408.6	125.56					
FL4	0.05	-0.5	0.00	96	3230	3805	369	5.40
	0.08	4.0	0.48					
	0.25	17.4	1.99					
	0.57	34.2	3.99					
	0.55	34.5	4.00					
	4.62	112.3	15.85					
	13.55	185.5	32.10					
	34.25	292.6	63.51					
	89.08	364.8	125.56					
FL9	0.18	-1.8	0.00	286	3248	3200	509	5.38
	0.20	2.8	0.48					
	0.41	15.8	1.99					
	0.73	32.6	3.99					
	0.80	32.0	4.00					
	4.80	110.5	15.85					
	13.05	190.5	32.10					
	34.03	294.8	63.51					
	86.38	391.8	125.56					
FL12	1.06	-10.6	0.00	658	3716	3757	777	5.51
	1.15	-6.7	0.48					
	1.61	3.8	1.99					
	2.27	17.2	3.99					
	3.44	46.4	8.08					
	4.02	40.6	8.08					
	18.86	132.4	32.10					
	39.89	236.2	63.51					
	96.98	285.8	125.56					

MO2	0.00	10.1	1.01	0	5951	6086	477	4.83
	0.02	19.7	1.99					
	0.06	39.3	3.99					
	0.07	39.3	4.00					
	0.27	78.1	8.08					
	4.89	272.1	32.10					
	18.91	446.0	63.51					
	54.35	712.1	125.56					
	168.53	919.3	260.46					
	399.71	1183.0	518.01					
MO6	0.05	9.6	1.01	71	5904	5711	539	4.87
	0.04	19.5	1.99					
	0.10	38.9	3.99					
	0.13	38.7	4.00					
	1.55	143.0	15.85					
	5.43	266.7	32.10					
	19.39	441.2	63.51					
	61.13	644.3	125.56					
MO9	0.13	-1.3	0.00	348	5832	5645	759	4.99
	0.20	8.1	1.01					
	0.25	17.4	1.99					
	0.46	35.3	3.99					
	0.45	35.5	4.00					
	0.85	72.3	8.08					
	7.97	241.3	32.10					
	24.94	385.7	63.51					
	68.38	571.8	125.56					
MO11	0.31	-3.1	0.00	602	5930	5579	882	4.99
	0.42	5.9	1.01					
	0.47	15.2	1.99					
	0.65	33.4	3.99					
	0.69	33.1	4.00					
	3.39	124.6	15.85					
	8.56	235.4	32.10					
	25.92	375.9	63.51					
	72.27	532.9	125.56					
MO5	1.33	-13.3	0.00	1248	6049	6179	1404	4.99
	1.53	-5.2	1.01					
	1.57	4.2	1.99					
	2.22	17.7	3.99					
	3.30	47.8	8.08					
	6.67	91.8	15.85					
	13.81	182.9	32.10					
	34.04	294.7	63.51					
	82.87	426.9	125.56					

KA1	0.00	4.8	0.48	0	3196	2244	227	5.13
	0.09	19.0	1.99					
	0.44	35.5	3.99					
	0.42	35.8	4.00					
	5.21	106.4	15.85					
	13.52	185.8	32.10					
	35.73	277.8	63.51					
	88.72	368.4	125.56					
KA5	0.07	−0.7	0.00	90	3111	2132	290	5.29
	0.11	3.7	0.48					
	0.30	16.9	1.99					
	0.57	34.2	3.99					
	0.64	33.6	4.00					
	5.54	103.1	15.85					
	14.08	180.2	32.10					
	35.95	275.6	63.51					
KA8	93.84	317.2	125.56	260	3256	2210	394	5.29
	0.35	−3.5	0.00					
	0.37	1.1	0.48					
	0.57	14.2	1.99					
	1.07	29.2	3.99					
	1.06	29.4	4.00					
	6.31	95.4	15.85					
	15.72	163.8	32.10					
KA11	38.28	252.3	63.51	599	3673	2358	701	5.39
	95.29	302.7	125.56					
	1.23	−12.3	0.00					
	1.20	−7.2	0.48					
	1.64	3.5	1.99					
	2.26	17.3	3.99					
	4.12	39.6	8.08					
	8.62	72.3	15.85					
KA15	18.63	134.7	32.10	1047	3814	2409	1052	5.48
	43.00	205.1	63.51					
	98.44	271.2	125.56					
	1.23	−12.3	0.00					
	1.27	−7.9	0.48					
	1.77	2.2	1.99					
	2.44	15.5	3.99					
	4.12	39.6	8.08					
	8.63	72.2	15.85					
	18.45	136.5	32.10					
	43.21	203.0	63.51					
	100.35	252.1	125.56					

^a Units: mg P/L — C , C_0 ; mg/kg X , P_f , Fe_{ox} , Al_{ox} , P_{ox} .

^b Values of $C \leq 0$ were converted to small positive values (0.0001) to avoid computational problems, including those associated with the use of finite-difference methods to calculate partial derivatives in evaluating weights.

Table S-3. Results from individual- Q global analyses of sorption data for five soils.^{a,b}

Parameter	Soil Type				
	MA	D	KA	FL	MO
RCS ^c	1.64	0.34	1.66	1.05	1.26
a_0	619(44)	46(5)	113(10)	86(9)	102(8)
a_1		1.38(24)	1.46(47)	-0.54(12)	-0.99(13)
a_2				3.6(1.5)	8.2(9) ^d
A	380	80	140	190	320
b_0	0.243(9)	0.297(20)	0.307(20)	0.325(12)	0.298(7)
b_1	0.308(20)	0.331(43)	0.424(114)		
q_0^e	137(45)	17(3)	52(7)	36(5)	77(7)
q_1^e	0.330(51)	0.073(7)	0.117(16)	0.094(13)	0.097(10)
RCS ^{c,e}	0.006	0.26	1.82	0.16	0.61

^a Quantities and units as in Table 2 in the paper. Q s were given there and are not repeated here.

^b 168 points fitted, excluding all green points in Table S-1, but including all KA15 data except next-to-last ($C = 43.21$).

^c Reduced chi-square: S_{TV} from Equation (5) divided by statistical degrees of freedom ν , = number of data points n (23-44) minus number of adjustable parameters p (7-10).

^d Contribution to a of form $a_2 \Delta[H^+]$, with $\Delta[H^+] = 10^{(6-pH)}$; pH ranges 4.83-4.99 (MO) and 5.26-5.51 (FL).

^e Results from weighted straight-line fits shown in Figure 6. SEs are prior.

Table S-4. Results from four linear- Q global analyses of sorption data for KA soil type.

Parameter	Table 2	Alt1 ^a	Alt2 ^b	Alt3 ^c
Reduced χ^2	1.61	1.07	1.91	7.95
a_0	107(10)	63(5)	65(7)	61(12)
a_1	0.97(26)		0.54(28)	-0.49(18)
A	140	140	140	140
b_0	0.312(21)	0.434(20)	0.438(27)	0.433(47)
b_1	0.286(86)	0.208(58)	0.32(10)	
q_0	40(9)	10(5)	15(6)	20(11)
q_1	0.160(12)	0.110(4)	0.072(7)	0.050(8)

^a Fitting KA data shown in Table S-1 for first 4 concentrations (P_f values) only, with inclusion of green points, exclusion of orange (31 points total).

^b Same data as for Alt1 for first 3 P_f , but replacement of 4th P_f data with 5th. All orange points excluded.

^c Data for all 5 P_f included, but all orange points excluded.

As has been noted in the paper, the KA data residuals exhibited outlier behavior that was more complex than expected for a single bad C_0 , as for example, can explain the FL data residuals. When the same data ($C_0 = 63.51$ mg/L) were excluded from the KA analysis, the residuals still showed systematic deviations. When alternatively the highest- C_0 data (“orange,” $C_0 = 125.56$ mg/L) were excluded, χ^2 for the linear- Q global fit (Alt1 in Table S-3) was actually smaller, but the fit parameters changed radically. For example the Q parameters became $q_0 = 10(5)$ and $q_1 = 0.110(4)$, as compared with the Table 2 values, $q_0 = 40(9)$ and $q_1 = 0.160(12)$. With exclusion of the orange points, all three alternative linear- Q fits gave $b_0 > 0.4$, the only values that high for any of the soils. In summary, the data for the 4th and 5th P_f values are statistically incompatible for the Freundlich model, as are the data for the highest two C_0 values for each P_f . As we note in the paper, such behavior could mean that the Freundlich model is just not a good fit for the KA data.

The surprisingly strong dependence of the MA global results on the data selection for $P_f = 0$ is illustrated in Figure S-2. There are no obvious outlier candidates among these data. However, we can note the following: (1) There are 5 measurements giving $C = 0$ within its uncertainty, meaning essentially all P in solution (C_0) was sorbed; only one of these is needed to hold the fit results at essentially the values for all points. (2) Deletion of all $C \approx 0$ points drops a_0 and the Q s by 2 σ or more. (3) Dropping the highest C , C_0 point gives a somewhat smaller reduction in a_0 and the Q s. (4) The effect on b is less significant, in the same direction for b_1 , opposite for b_0 .

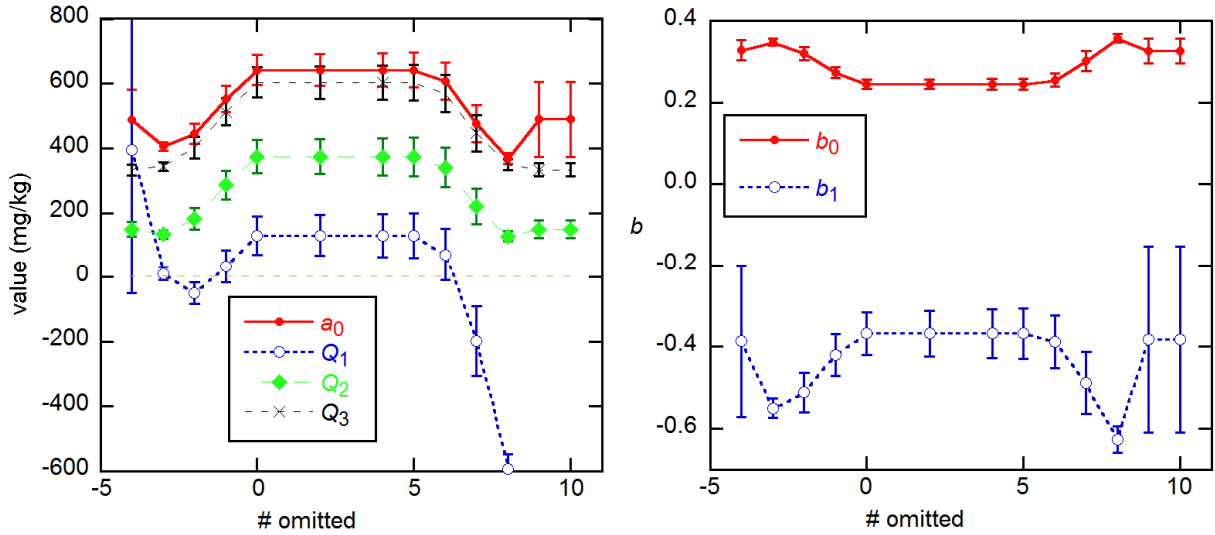


Figure S-2. Effect of deletion of $P_f = 0$ data for the MA soil type on least-squares results for the individual- Q global model. Positive numbers are for cumulative deletions from the lowest C_0 up, while negative are for highest C_0 down; e.g., -2 represents deletion of the points for $C_0 = 518$ and 260 mg/kg, while 5 denotes deletion of the first 5 values in Table S-1. (a_1 was statistically insignificant when all or most points were included so was set to zero for all of these results.)

Finally, we show in Figure S-3 results from fitting the KA data to the 3-parameter Freundlich model, for both Equations (3) and (4), showing the identical results obtained. Consistent with standard practice, all of these fits treat X , and equivalently C_0 as uncertain; they are inverse-variance weighted in accord with the derived VF for C_0 .

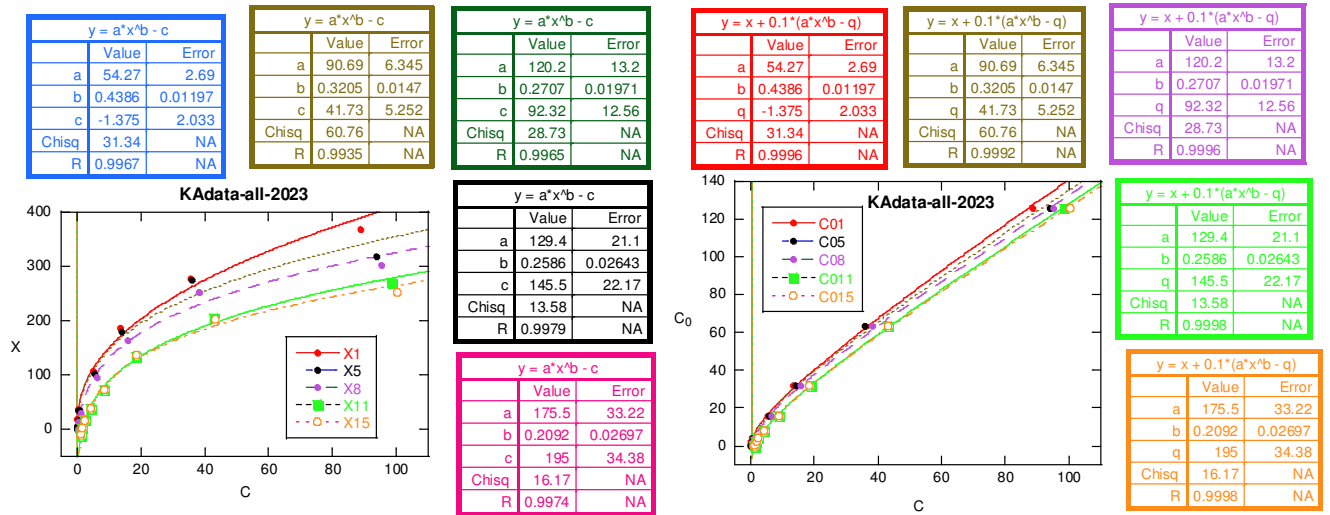


Figure S-3. Weighted fits of KA data to 3-parameter Freundlich model, as X vs C (left, Equation 3 in text) and C_0 vs C (Equation 4). Results are identical in the two fits in all cases.