Organic contaminant sorption parameters should only be compared across a consistent system of linear functions

Mark A. Chappell*, Jennifer M. Seiter, Haley M. West, Lesley F. Miller, Maria E. Negrete, Joshua J. LeMonte, Beth E. Porter, Cynthia L. Price, Matthew A. Middleton

U.S. Army Engineer Research & Development Center, 3909 Halls Ferry Rd, Vicksburg, MS, 39180, USA

ARTICLE INFO
Keywords:
Organic contaminants
Linear sorption modeling
Distribution coefficient
System of consistent linear equations
Homogeneous system of linear equations
Surface chemistry
Soil chemistry
Soil pollution
Contaminant transport
Environmental science

ABSTRACT
Modeling contaminant sorption data using a linear model is very common; however, the rationale for whether the y-intercept should be constrained or not remains a subject of debate. This article justifies constraining the y-intercept in the linear model to zero. By doing so, one imposes consistency on the system of linear equations, allowing for direct comparison of the sorption coefficients.

1. Introduction
Over the years, numerous reports published in the scientific literature have focused on various aspects of contaminant sorption modeling, from the development of new mathematical formulas to modifications of the more common linear and nonlinear sorption models (e.g., Freundlich, Langmuir). Sorption represents the simplest and most experimentally accessible of contaminant interactions with soils, providing a quantitative parameter to describe the transfer of a solute in bulk solution across the soil-water interface to be “solubilized” in the solid-phase. For this reason, sorption parameters have been expanded to incorporate kinetic information [1], describe partitioning across different soil chemical domains [2], and discriminate the particularities regarding the contaminant of interest.

With few exceptions [3], conventional contaminant sorption models are strictly empirical in nature given that the thermodynamic state of the soil surface currently remains impossible to define. The simplest of the models, linear partitioning, is:

\[ C_s = K_D C_w \]  

where, \( C_w \) = contaminant concentration in solution at equilibrium, \( C_s \) = sorbed concentration of contaminant on the soil surface, and \( K_D \) = the distribution coefficient of sorption referring to the affinity of the solid phase for the solute. Modeling sorption data with the linear sorption function may be conducted in two modes, either allowing the slope (representing the contaminant distribution coefficient or \( K_D \)) and y-intercept (which has unresolved physical and chemical meaning) to “float” independently (i.e., unconstrained) or, forcing the y-intercept through zero. Which of these approaches is preferred or entirely justifiable, to our knowledge, is not satisfactorily answered in the scientific literature. For example, abundant discussion exists among the sample of published studies surveyed for this paper [4, 5, 6, 7, 8, 9, 10, 11, 12, 13] regarding the appropriateness of the linear model and its mechanistic implications on contaminant sorption. However, most authors avoided any discussion regarding their treatment of the y-intercept although it is obvious from the sorption plots that this parameter was overwhelmingly forced through zero. Notable exceptions include Ruffino and Zanetti [14] and Donskova et al. [15], where the y-intercept was allowed to float during the linear modeling of the sorption isotherms. The author has largely preferred forcing the y-intercept through zero to give \( K_D \) estimates that are more readily comparable among different soils or
treatments [16, 17]. Overall, it may be concluded that setting the y-intercept equal to zero represents a theoretically reasonable assumption that the soil was not previously exposed to the contaminant. Thus, forcing the sorption model through zero strictly conforms to Eq. (1), where no y-intercept is depicted. However, this is far from satisfying response, especially when it is apparent from the sorption data that the plot unexplainably deviates substantially from zero sorption. Aside from its practical value, this short article explains the mathematical justification for controlling the y-intercept, and its necessity for making meaningful comparisons among contaminant K_D values.

In this paper, we propose that K_D values can only be compared among the different soils if the system of linear functions (by which K_D was extracted) describing sorption are consistent, which is defined as having at least one solution. For example, consider three linear functions with three unknowns [18]:

\[
\begin{align*}
x + 3y + 2z &= -3 \\
2x + 2y + 2z &= -2 \\
3x + 5y + 6z &= -5
\end{align*}
\]  

By simple elimination, a single solution exists for this system represented by a linear combination of three values (1,1,0). Graphically, this represents a single point in where the lines passing through three separate dimensions all intersect at (1,1,0). This is important as a consistent system of linear equations is considered independent, meaning that each function in the system is not proportional to any other of the functions, or in the same plane (such as for parallel functions). In practical terms, an independent equation provides unique information to the system that is not duplicated by any of the other equations. If a system of equations is consistent, then all equations in the system are independent. This means that all equations within the system are equally valuable for describing the system and directly comparable.

To our knowledge, a linear algebra-based justification for controlling the y-intercept has never been presented in linear sorption isotherm modeling.

2. Materials and methods

Here, this note draws on a recently published study [19] investigating the sorption of the insensitive munition compound, 2, 4-dinitroanisole (DNAN), on different taxonomic soil “types”. All sorption isotherms were modeled with the linear sorption model (Eq. 1) using the R programming language [20] via the RStudio interface [21]. Modeling this data without any constraints on the fitting generated a system of linear equations, in the form of \( y = Ax + C \), where \( A \) is the slope (representing the K_D) and \( C \) is the y-intercept or offset. System consistency was tested using using the ‘matlib’ package [22] for R.

3. Results and discussion

As evident in Figure 1A, the DNAN sorption isotherms appeared strongly linear among the different soils, exhibiting small deviations in the y-intercept away from zero. Both the slope and y-intercept parameters (Table 1) were statistically significant (\( p < 0.05 \)) for all soils. With the sorption curves parameterized, it was of interest to compare the K_D values in order to gain a relative sense of how the different soils ranked in terms of their preference for the contaminant.

Converting the functions into the standard form (\( Ax + By = C \)), the information was stored in a \( 2 \times 3 \) matrix and tested for consistency. To be considered as consistent, there must be a unique solution for the entire system. This is represented graphically by a single point where all of the lines intersect in two-dimensional space (given the two unknowns in the equations). Analysis [20] showed that the linear sorption isotherms intersected at multiple points (indicated by the circles in Figure 1B), suggesting that this system of equations was inconsistent. Given that each sorption curve exhibited its own unique slope and y-intercept, it is reasonable to assume that the sorption curves were consistent at least on a pairwise basis. However, there was no mathematical basis for comparing K_D values across the entire system without first imposing constraints on the fitted linear sorption model.

The problem of inconsistent linear functions can be circumvented by forcing the y-intercept to zero during the linear regression analysis. As a result, this gave a new \( 2 \times 3 \) matrix where the last column (C) was populated by zeroes, representing what is known as a homogeneous system of linear equations. Under this scenario, all of the curves intersected (Figure 1C) at a single point at (0,0), making the origin the unique solution for the entire system. Thus, statistically comparing the K_D values

![Figure 1](image-url)

**Figure 1.** DNAN sorption isotherms and tests for consistency based on constraints on the y-intercept (A) DNAN sorption data where modeling of the linear sorption parameters were un-restrained, and (B) corresponding consistency test results (C) Graphical analysis of the consistency of the linear functions after forcing the y-intercept through zero. In B and C, the linear sorption functions were extrapolated past the origin for comparing the curves graphically. Black circles represent all intersection points for the linear sorption functions.
across the entire system became mathematically viable with this step. As a consequence of constraining the y-intercept $\hat{a} = 0$, the $K_D$ values were on average 24% higher (representing an average 16% increase in the estimates' standard error), showing the “pull” of the y-intercept parameter on the overall linear regression (Table 2). However, all models remained highly significant, with the computed p-values for the F-statistic ranging from $10^{-9}$ – $10^{-17}$.

Not only does this simple approach allow for comparing $K_D$ values across the system discussed here, but also allows for expanding the size of the system to include legacy data or, conversely, new sorption information. On the odd chance that the system of linear functions are consistent, then forcing the y-intercept through zero would not be necessary. Alternatively, the y-intercept could be forced through a positive, non-zero value, but the value chosen for the offset would need to be determined as well as applicable across the entire system.

**Declarations**

**Author contribution statement**

Mark A Chappell: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Jennifer M Seiter, Haley M West, Lesley F Miller, Maria E Negrete, Beth E Porter, Matthew A. Middleton: Performed the experiments.

Joshua J LeMonte, Cynthia L Price: Analyzed and interpreted the data.

---

Table 1. Fit results and statistics for DNAN sorption modeling on different soil samples. Here, y-intercept was unconstrained in the linear regression modeling.

| Sample name | term  | estimate | SE   | t-stat | p0.05 | $R^2$ | F-stat | p0.05 |
|-------------|-------|----------|------|--------|-------|------|--------|-------|
| Dismal1     | y-int | 21.706   | 4.710| 4.608  | 4.902E-04 | 0.858| 78.509 | 7.183E-07 |
| Dismal1     | slope ($K_D$) | 6.116 | 0.690| 8.861  | 7.183E-07 |      |        |       |
| Dismal5     | y-int | 6.008    | 1.540| 3.900  | 1.825E-03 | 0.989| 1126.665 | 5.133E-14 |
| Dismal5     | slope ($K_D$) | 6.120 | 0.182| 33.566 | 5.133E-14 |      |        |       |
| FtPolk2     | y-int | 4.578    | 0.802| 5.710  | 7.173E-05 | 0.959| 303.132 | 2.168E-10 |
| FtPolk2     | slope ($K_D$) | 1.079 | 0.062| 17.411 | 2.168E-10 |      |        |       |
| Holmes3     | y-int | 3.343    | 0.789| 4.237  | 9.703E-04 | 0.860| 79.952 | 6.482E-07 |
| Holmes3     | slope ($K_D$) | 0.521 | 0.058| 8.942  | 6.482E-07 |      |        |       |
| Huntsville1 | y-int | 6.285    | 2.268| 2.771  | 1.588E-02 | 0.883| 97.791 | 2.048E-07 |
| Huntsville1 | slope ($K_D$) | 1.813 | 0.183| 9.589  | 2.048E-07 |      |        |       |
| Laurel3     | y-int | 14.391   | 1.582| 9.097  | 5.332E-07 | 0.994| 2147.739 | 8.007E-16 |
| Laurel3     | slope ($K_D$) | 15.256 | 0.329| 46.344 | 8.007E-16 |      |        |       |
| Laurel4     | y-int | 12.372   | 1.253| 9.878  | 2.075E-07 | 0.988| 1068.427 | 7.220E-14 |
| Laurel4     | slope ($K_D$) | 4.677 | 0.143| 32.687 | 7.220E-14 |      |        |       |
| Morrow1     | y-int | 14.373   | 1.908| 7.531  | 4.302E-06 | 0.981| 683.179 | 1.268E-12 |
| Morrow1     | slope ($K_D$) | 6.646 | 0.247| 26.138 | 1.268E-12 |      |        |       |
| Morrow3     | y-int | 9.367    | 1.070| 8.756  | 8.215E-07 | 0.984| 776.227 | 5.605E-13 |
| Morrow3     | slope ($K_D$) | 2.798 | 0.100| 27.861 | 5.605E-13 |      |        |       |
| Morrow5     | y-int | 4.814    | 0.919| 5.236  | 1.606E-04 | 0.983| 758.667 | 6.489E-13 |
| Morrow5     | slope ($K_D$) | 2.111 | 0.077| 27.544 | 6.489E-13 |      |        |       |
| Ohiopyle3   | y-int | 16.284   | 1.436| 11.342 | 4.098E-08 | 0.992| 1701.271 | 3.607E-15 |
| Ohiopyle3   | slope ($K_D$) | 8.348 | 0.202| 41.246 | 3.607E-15 |      |        |       |
| Stewart1    | y-int | 8.140    | 0.417| 19.507 | 5.202E-11 | 0.997| 4896.307 | 3.856E-18 |
| Stewart1    | slope ($K_D$) | 2.572 | 0.037| 69.974 | 3.856E-18 |      |        |       |
| Susquehanna1| y-int | 10.069   | 1.066| 9.443  | 3.485E-07 | 0.985| 836.766 | 3.465E-13 |
| Susquehanna1| slope ($K_D$) | 2.795 | 0.097| 28.927 | 3.465E-13 |      |        |       |
| Susquehanna3| y-int | 13.053   | 3.979| 3.280  | 5.969E-03 | 0.078| 1.095  | 3.144E-01 |
| Susquehanna3| slope ($K_D$) | 0.346 | 0.330| 1.046  | 3.144E-01 |      |        |       |
| Toledo1     | y-int | 7.802    | 2.027| 3.850  | 2.008E-03 | 0.945| 224.753 | 1.390E-09 |
| Toledo1     | slope ($K_D$) | 2.705 | 0.180| 14.992 | 1.390E-09 |      |        |       |
| Toledo2     | y-int | 7.445    | 2.152| 3.459  | 4.232E-03 | 0.881| 96.243 | 2.246E-07 |
| Toledo2     | slope ($K_D$) | 1.711 | 0.174| 9.810  | 2.246E-07 |      |        |       |
| Twin2       | y-int | 5.625    | 0.578| 9.725  | 2.484E-07 | 0.995| 2733.787 | 1.682E-16 |
| Twin2       | slope ($K_D$) | 2.663 | 0.051| 52.286 | 1.682E-16 |      |        |       |
| Woodlake1   | y-int | 10.144   | 2.748| 3.691  | 2.718E-03 | 0.871| 87.947  | 3.770E-07 |
| Woodlake1   | slope ($K_D$) | 2.360 | 0.252| 9.378  | 3.770E-07 |      |        |       |
| Woodlake2   | y-int | 4.744    | 1.855| 2.557  | 2.386E-02 | 0.922| 153.060 | 1.448E-08 |
| Woodlake2   | slope ($K_D$) | 1.986 | 0.161| 12.372 | 1.448E-08 |      |        |       |
| Woodlake6   | y-int | 2.328    | 1.041| 2.237  | 4.340E-02 | 0.917| 144.487 | 2.048E-08 |
| Woodlake6   | slope ($K_D$) | 0.939 | 0.078| 12.020 | 2.048E-08 |      |        |       |
Table 2. Fit statistics for the DNAN sorption modeling when the y-intercept when the y-intercept = 0.

| Individual parameters | Overall model |  
|-----------------------|---------------|
|                        | R²            | F-stat | p<0.05 |
| Dismal1 slope (Kₐ)    | 0.932466      | 193.302 | 1.38E-09 |
| Dismal5 slope (Kₐ)    | 0.993807      | 2271.945 | 6.79E-17 |
| FtPolk2 slope (Kₐ)    | 0.973869      | 521.7696 | 1.76E-12 |
| Holmes3 slope (Kₐ)    | 0.947681      | 253.5891 | 2.30E-10 |
| Huntville1 slope (Kₐ) | 0.960217      | 337.9125 | 3.36E-11 |
| Laurel3 slope (Kₐ)    | 0.988855      | 1241.592 | 4.51E-15 |
| Laurel4 slope (Kₐ)    | 0.978568      | 639.2423 | 4.39E-13 |
| Morrow1 slope (Kₐ)    | 0.978173      | 627.3933 | 4.99E-13 |
| Morrow3 slope (Kₐ)    | 0.977947      | 620.8228 | 5.37E-13 |
| Morrow5 slope (Kₐ)    | 0.988618      | 1215.995 | 5.21E-15 |
| Ohioype2 slope (Kₐ)   | 0.981586      | 746.3041 | 5.09E-14 |
| Stewart1 slope (Kₐ)   | 0.984241      | 874.3937 | 5.09E-14 |
| Susequehanna1 slope (Kₐ) | 0.976849 | 590.7205 | 7.54E-13 |
| Susequehanna3 slope (Kₐ) | 0.970258 | 27.7683 | 9.95E-05 |
| Toledo1 slope (Kₐ)    | 0.97507        | 547.5729 | 1.27E-12 |
| Toledo2 slope (Kₐ)    | 0.955188      | 298.4187 | 7.75E-11 |
| Twin2 slope (Kₐ)      | 0.991431      | 1616.42 | 7.24E-16 |
| Woodlake1 slope (Kₐ)  | 0.948601      | 258.3765 | 2.03E-10 |
| Woodlake2 slope (Kₐ)  | 0.974297      | 530.6801 | 1.57E-12 |
| Woodlake6 slope (Kₐ)  | 0.974457      | 534.0845 | 1.50E-12 |

Acknowledgements

The use of trade, product, or firm names in this report is for descriptive purposes only and does not imply endorsement by the U.S. Government. The tests described and the resulting data presented herein, unless otherwise noted, were obtained from research conducted under the Environmental Quality Technology Program of the US Army Corps of Engineers by the U.S. Army Engineer Research and Development Center (ERDC). Permission was granted by the Chief of Engineers to publish this information. The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents. The authors express gratitude to Dr. Elizabeth Ferguson, Technical Director of the U.S. Army ERDC Environmental Quality Technology Program for support of this research. The authors also express their gratitude to Mr. Matt Hathaway from BAE for assistance with acquiring the DNAN compound. We also express our appreciation to the following individuals that facilitated soil sampling at state parks, wildlife refuges, and military bases: Wayne Fariss (U.S. Army Fort Polk, LA); Ric LeGrange (South Toledo Bend State Park, Anacoco, LA); Greg Plump (Holmes County State Park, Durant, MS); Ray Black (Huntsville State Park, Huntsville, TX); Deloras Freeman (Great Dismal Swamp National Wildlife Refuge, Suffolk, VA); Michael Mumau (Laurel Hill State Park, Somerset, PA); Jeffery Davidson (Morrow Mountain State Park, Albemarle, NC); Ken Bisbee (Ohiopyle State Park, Ohiopyle, PA); Andrew Hangen (Susquehanna State Park, Havre De Grace, MD); Theresa Dufey (VA State Parks Dept. of Conservation & Recreation), and Phil Morgan (Twin Lakes State Park, Greenbay, VA).

References

[1] J. Skopp, Derivation of the Freundlich adsorption isotherm from kinetics, J. Chem. Educ. 86 (11) (2009) 1341.
[2] W. Huang, W.J. Weber, A distributed reactivity model for sorption by soils and sediments. 10. Relationships between desorption, hysteresis, and the chemical characteristics of organic domains, Environ. Sci. Technol. 31 (9) (1997) 2562–2569.
[3] M.A. Chappell, C.L. Price, B.E. Porter, B.A. Pettway, R.D. George, Differential kinetics and temperature dependence of abiotic and biotic processes controlling the environmental fate of TNT in simulated marine systems, Mar. Pollut. Bull. 62 (2011) 1736–1745.
[4] W. Huang, W.J. Weber Jr., A distributed reactivity model for sorption by soils and sediments. 10. Relationships between desorption, hysteresis, and the chemical characteristics of organic domains, Environ. Sci. Technol. 31 (1997) 2562–2569.
[5] A.L. Wood, D.C. Bouchard, M.L. Brusseau, P.S.C. Rao, Cosolvent effects on sorption and mobility of organic contaminants in soils, Chemosphere 21 (4) (1990) 575–587.
[6] C.T. Miller, W.J. Weber, Sorption of hydrophobic organic pollutants in saturated soil systems, J. Contam. Hydrod. 1 (1) (1986) 243–261.
[7] W.J. Weber, C.T. Miller, Modeling the sorption of hydrophobic contaminants by aquatic materials—I. Rates and equilibria, Water Res. 22 (4) (1988) 457–464.
[8] I.P. Breus, A.A. Mishchenko, Sorption of volatile organic contaminants by soils (a review), Eurasian Soil Sci. 39 (12) (2006) 1271–1283.
[9] R.A. Düring, S. Krahe, S. Gath, Sorption behavior of nonylphenol in terrestrial soils, Environ. Sci. Technol. 36 (19) (2002) 4052–4057.
[10] H. Li, G. Zheng, R.J. Teppen, C.T. Johnston, S.A. Boyd, Sorption and desorption of pesticides by clay minerals and humic acid-clay complexes, Soil Sci. Soc. Am. J. 67 (1) (2003) 122–131.
[11] C.T. Miller, W.J. Weber Jr., Modeling organic contaminant partitioning in groundwater systems, Groundwater 22 (5) (1984) 584–592.
[12] D.C. Bouchard, S.C. Mrazik, G.B. Smith, Benzene and naphthalene sorption on soil contaminated with high molecular weight residual hydrocarbons from unleaded gasoline, Chemosphere 21 (8) (1990) 975–989.
[13] W.J. Weber, P.M. McGinley, L.E. Katz, Sorption phenomena in subsurface systems: concepts, models and effects on contaminant fate and transport, Water Res. 25 (5) (1991) 499–528.
[14] B. Ruffino, M. Zanetti, Adsorption study of several hydrophobic organic contaminants on an aquatic material, Am. J. Environ. Sci. 5 (2009).
[15] K. Dontsova, S. Taylor, P. Perez-Rodriguez, M. Brusseau, J. Arthur, N. Mark, M.E. Walsh, J.H. Lever, J. Simunek, Dissolution of NTO, DNAN, and Insensitive Munition Formulations and Their Fates in Soils, US Army Engineer Research & Development Center, 2014.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.
[16] C.K. Katseanes, M.A. Chappell, B.G. Hopkins, B.S. Durham, C.L. Price, B.E. Porter, L.F. Miller, Multivariate functions for predicting the degradation kinetics of 2,4,6-trinitrotoluene (TNT) and 1,3,5-trinitro-1,3,5-tricyclohexane (RDX) among taxonomically distinct soils, J. Environ. Manag. 203 (2017) 383–390.

[17] C.K. Katseanes, M.A. Chappell, B.G. Hopkins, Multivariate functions for predicting the sorption of 2,4,6-trinitrotoluene (TNT) and 1,3,5-trinitro-1,3,5-tricyclohexane (RDX) among taxonomically distinct soils, J. Environ. Manag. 182 (2016) 101–110.

[18] G. Strang, Introduction to Linear Algebra, fourth ed., Wellesley - Cambridge Press, Wellesley, MA USA, 2009.

[19] M.A. Chappell, J.M. Seiter-Moser, H.M. West, B.D. Durham, B.E. Porter, C.L. Price, Predicting 2,4-dintroanisole (DANAN) sorption on soil using different compositional datasets, Geoderma 356 (2019) 111916.

[20] R. R Development Team, A Language and Environment for Statistical Computing, 2018. Vienna, Austria.

[21] RStudio Team, RStudio: Integrated Development for R, RStudio, Inc., Boston, MA, 2018.

[22] M. Friendly, J. Fox, P. Chalmers, G. Monette, G. Sanchez, Matlib: Matrix Functions for Teaching and Learning Linear Algebra and Multivariate Statistics, 9, 2019.