Langmuir
Article
pubs.acs.org/Langmuir

Number of Independent Parameters in the Potentiometric Titration of Humic Substances

Thomas Lenoir[†,‡] and Alain Manceau*[,†]

[†]*Mineralogy & Environments Group, LGCA, Université Joseph Fourier and CNRS, 38041 Grenoble Cedex 9, France, and* [‡]*Laboratoire Central des Ponts et Chaussées (LCPC), Route de Bouaye, BP 4129, 44341 Bouguenais Cedex, France*

Received September 10, 2009. Revised Manuscript Received November 16, 2009

With the advent of high-precision automatic titrators operating in pH stat mode, measuring the mass balance of protons in solid-solution mixtures against the pH of natural and synthetic polyelectrolytes is now routine. However, titration curves of complex molecules typically lack obvious inflection points, which complicates their analysis despite the high-precision measurements. The calculation of site densities and median proton affinity constants (pK) from such data can lead to considerable covariance between fit parameters. Knowing the number of independent parameters that can be freely varied during the least-squares minimization of a model fit to titration data is necessary to improve the model's applicability. This number was calculated for natural organic matter by applying principal component analysis (PCA) to a reference data set of 47 independent titration curves from fulvic and humic acids measured at $I = 0.1$ M. The complete data set was reconstructed statistically from pH 3.5 to 9.8 with only six parameters, compared to seven or eight generally adjusted with common semi-empirical speciation models for organic matter, and explains correlations that occur with the higher number of parameters. Existing proton-binding models are not necessarily overparametrized, but instead titration data lack the sensitivity needed to quantify the full set of binding properties of humic materials. Model-independent conditional pK values can be obtained directly from the derivative of titration data, and this approach is the most conservative. The apparent proton-binding constants of the 23 fulvic acids (FA) and 24 humic acids (HA) derived from a high-quality polynomial parametrization of the data set are $pK_{H,COOH}(FA) = 4.18 \pm 0.21$, $pK_{H,Ph-OH}(FA) = 9.29 \pm 0.33$, $pK_{H,COOH}(HA) = 4.49 \pm 0.18$, and $pK_{H,Ph-OH}(HA) = 9.29 \pm 0.38$. Their values at other ionic strengths are more reliably calculated with the empirical Davies equation than any existing model fit.

Introduction

Potentiometric titration is the method of choice for measuring equilibrium proton-binding constants in colloids and materials surface science, biology, and biochemistry. In environmental science, the titration method is commonly used to characterize the acid−base properties of natural organic matter (NOM). NOM is a complex, heterogeneous assemblage that includes many polyelectrolyte groups, such as carboxyl (COOH), phenolic (Ph−OH), amine, sulfhydryl, phosphate, and alcohol functional groups.[1] The first two types of ligands predominate and are the only ones considered in the modeling of titration data. Their amounts per unit mass of NOM, expressed as the total concentration of protons denoted by Q_{H1} and Q_{H2}, and their proton-binding constants K_{H1} and K_{H2} are the main intrinsic properties of a humic substance derived from a titration experiment. These two properties are also the most important because they determine the ion-binding capacity of NOM for hard and intermediate metals (e.g., Al(III), Fe(III), Cu(II), and Zn(II)).[2] For example, pK_H values are used to determine the extent of competition for binding sites between metal cations and protons at a given pH.

Proton-binding constants for fulvic (FA) and humic (HA) acids vary by at least one order of magnitude ($2.6 \le pK_{H1}(FA) \le 3.5$; $3.1 \le pK_{H1}(HA) \le 4.1$; $7.0 \le pK_{H2}(FA) \le 9.4$; and $7.7 \le pK_{H2}(HA) \le 8.9$) (Table 1). The large variability in pK_{Hi} obtained by titration either is real and results from the variability in composition and functionality of the two types of humic materials or is within the uncertainty and accuracy of the titration method, in which case model fits would not necessarily provide a realistic chemical description of the binding properties of humic substances. As an example, a variation of one pK unit on a data set causes an error in the acidity constant of carboxylic-type groups ($pK_{H1} \sim 3$) of $\sim^1/_3 = 33\%$ in log units. If we can determine the number of independent parameters that can be fit to the titration data, then we may be able to clarify the reason for the variability in pK_{Hi} values and advance the interpretation of acid−base potentiometric measurements of complex macromolecular NOM. We addressed this fundamental but still unresolved problem by using a principal component analysis (PCA) algorithm to calculate the number of uncorrelated abstract components that maximize the variance in a large data set of independent titration curves.[3,4] The number of statistically significant components obtained by this correlation-based analysis is equal to the number of independent parameters needed to reconstruct titration curves. To help follow the presentation and discussion of the PCA results, a brief technical description of the titration experiments and data analysis is first given below.

Derivation of Proton-Binding Parameters from Titration Data

Titration curves are usually acquired between pH 3 and 10 because of the difficulty in precisely measuring the variation in acidity at lower and higher pH and the risk of altering the original material at extreme pH.[5] Over this pH range, the distribution of

*Corresponding author. E-mail: alain.manceau@obs.ujf-grenoble.fr.

(1) Takács, M.; Alberts, J. J.; Egeberg, P. K. *Environ. Intern.* **1999**, *25*, 315–323.
(2) Smith, D. S.; Bella, R. A.; Kramerb, J. R. *Compar. Biochem. Phys.* **2002**, *C133*, 65–74.
(3) Malinowski, E. R. *Anal. Chem.* **1977**, *49*, 612–617.
(4) Malinowski, E. R. *Factor Analysis in Chemistry*, 2nd ed.; Wiley: New York, 1991.
(5) Santos, E. B. H.; Esteves, V. I.; Rodrigues, J. P. C.; Duarte, A. C. *Anal. Chim. Acta* **1999**, *392*, 333–341.

Table 1. Compilation of Intrinsic pK_{Hi} Values for Humic Materials as Determined by Potentiometric Titration

		pK_{H1} (COOH groups)		pK_{H2} (Ph−OH groups)	
electrostatic model	humics	mean	std dev	mean	std dev
NICA-Donnan	FA[a]	2.65	0.43	8.60	1.06
	HA[a]	3.09	0.51	7.98	0.96
	FA[b]	2.83	0.36	7.02	0.18
	HA[b]	3.74	1.09	7.68	0.60
	HA[c]	3.62	0.19	8.54	0.48
model VI[d]	FA	3.20	0.13	9.40	0.78
	HA	4.10	0.16	8.80	0.23
Stockholm humic model[e]	FA	3.50	0.40	8.75	0.30
	HA	4.10	0.20	8.95	0.15

[a] Milne et al.[21] [b] Plaza et al.[44] [c] Drosos et al.[45] [d] Tipping.[15] [e] Gustafsson.[18]

acid−base reactive sites in NOM is fundamentally bimodal, with the first maximum of the derivative obtained at pH 3−5 from the deprotonation of carboxylic-type groups and the second obtained at pH 8−10 from phenolic-type groups (Figure 1a).[6] Because protons tend to be electrostatically retained by COO^- and $Ph-O^-$ groups, more base has to be added to remove them from the negatively charged NOM than from a neutral surface. Consequently, inflection points on titration curves are apparent pK values (pK_{Hiapp}) that are shifted to higher pH relative to intrinsic binding constants (pK_{Hi}) by a value not directly accessible by experiment but instead estimated by fitting the data to an electrostatic model (Table 1, Figure 1b,c). In the standard Donnan-type equilibrium model originally developed by Marinsky and coworkers,[7,8] the ionic strength dependence of the surface charge is captured simply with a single electrostatic interaction parameter b.[9−12] Similarly, for two reasons the concentrations of carboxylic- (Q_{H1}) and phenolic-type (Q_{H2}) groups are not obtained directly from titration measurements but instead by fitting the data to an ion-binding model such as the NICA model.[6,13−15] First, the low-pH plateau of titration curves, which defines the initial charge of the reactant, is below pH 3, and the high-pH plateau, which defines the final protonation state ($Q_{H1} + Q_{H2}$) of NOM, is above pH 10. As Westall et al.[16] stated, the "spectral window" is wider than the "data window". Second, the final protonation state of carboxylic-type groups (Q_{H1}), which is also the initial protonation state of phenolic-type groups, is masked by minor circumneutral groups with pK_H values between 5 and 8, such as phosphate H_2PO_4, amine NH_2, thiol R−SH, and anilium $C_6H_5NH_3^+$.[1]

In total, a minimum of seven parameters are required to fit a single titration curve of NOM to the combined NICA-Donnan (N-D) model: two pK_{Hi} values, two affinity distribution parameters (m_{Hi}) accounting for the variability in binding strength (i.e., in equilibrium constants), two Q_{Hi} values, and b.[17] When the

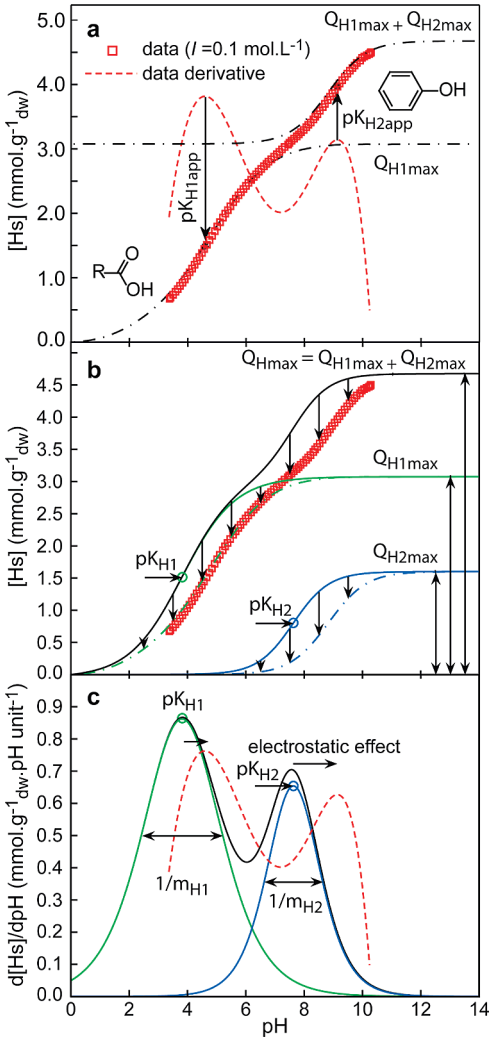

Figure 1. Main physicochemical parameters for natural organic matter derived from a titration curve obtained by measuring the number of protons released in solution [Hs] per gram of dry matter. (a) The red dotted line is the first derivative. Inflection points are the apparent (also called conditional) dissociation constants for carboxylic- and phenolic-type sites. The dashed lines in black are the optimal bimodal (R−COOH, Ph−OH) fit obtained with the N-D FIT code.[11] (b) Same simulation as previously but without an electrostatic effect. The new pK_H values are intrinsic values. The inclusion of an electrostatic correction does not change Q_{Himax}. Vertical arrows point to the direction in which the deprotonation curves shift because of electrostatic effects. (c) First derivatives of the previous deprotonation curves and modeling of the pseudo-normal affinity distribution with two Gaussians of $\sigma = 1/m_{Hi}$ width, as in the NICA model.[13,17]

(6) Dudal, Y.; Gérard, F. *Earth Sci. Rev.* **2004**, *66*, 199–216.
(7) Marinsky, J. A.; Ephraim, J. H. *Environ. Sci. Technol.* **1986**, *20*, 349–354.
(8) Marinsky, J. A.; Gupta, S.; Schindler, P. *J. Colloid Interface Sci.* **1982**, *89*, 412–426.
(9) Benedetti, M. F.; Van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1996**, *30*, 1805–1813.
(10) Kinniburgh, D. G.; Milne, C. J.; Benedetti, M. F.; Pinheiro, J. P.; Filius, J.; Koopal, L. K.; Van Riemsdijk, W. H. *Environ. Sci. Technol.* **1996**, *30*, 1687–1698.
(11) Kinniburgh, D. G. *FIT User Guide*; British Geological Survey: Keyworth, England, 1999.
(12) Avena, M. J.; Koopal, L. K.; Van Riemsdijk, W. H. *J. Colloid Interface Sci.* **1999**, *217*, 37–48.
(13) Koopal, L. K.; Van Riemsdijk, W. H.; De Wit, J. C. M.; Benedetti, M. F. *J. Colloid Interface Sci.* **1994**, *166*, 51–60.
(14) Benedetti, M. F.; Milne, C. J.; Kinniburgh, D. G.; Van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1995**, *29*, 446–457.
(15) Tipping, E. *Aquat. Geochem.* **1998**, *4*, 3–47.
(16) Westall, J. C.; Jones, J. D.; Turner, G. D.; Zachara, J. M. *Environ. Sci. Technol.* **1995**, *29*, 951–959.
(17) Koopal, L. K.; Saito, T.; Pinheiro, J. P.; Riemsdijk, W. H. V. *Colloids Surf., A* **2005**, *265*, 40–54.

initial charge is treated as an adjustable parameter,[10] the total number is eight. In model VI[15] and the Stockholm humic model,[18] Q_{H2} is deduced from Q_{H1}, thus reducing the number of degrees of freedom by one. Because titration curves for humic substances are rather featureless, increasing monotonically with pH, and the range of data is at best over seven pH units, the question arises as to the number of independent parameters that can be uniquely fit to the data.[17,19,20] As will be shown below by PCA, the limit is six but may be even lower if the data window is smaller.

PCA Theory

PCA is an eigenvector-based multivariate analysis tool that reduces a set of overdetermined vectors (titration curves) to a subset of linear independent orthogonal vectors, called principal or abstract components (PCs), needed to describe the complete set of data within the error. The conversion from titration curves to PCs can be expressed as the matrix equation $\mathbf{D} = \mathbf{E}\mathbf{V}^{1/2}\mathbf{W}^{\mathbf{T}}$

$$\underbrace{\begin{bmatrix} data_{11} & \cdots & data_{1n} \\ \vdots & \ddots & \vdots \\ data_{m1} & \cdots & data_{mn} \end{bmatrix}_{u*v}}_{data} = \underbrace{\begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mn} \end{bmatrix}}_{e_{i(1\le i\le n)}} \times \underbrace{\begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}^{1/2}}_{\lambda_i} \times \underbrace{\begin{bmatrix} \rho_{e_1,data_1} & \cdots & \rho_{e_1,data_n} \\ \vdots & \ddots & \vdots \\ \rho_{e_n,data_1} & \cdots & \rho_{e_n,data_n} \end{bmatrix}}_{correlations}$$

where $\mathbf{D}(m \times n)$ is the data matrix whose columns are titration curves and rows are data points (i.e., [Hs] values in Figure 1a), $\mathbf{E}(m \times n)$ is a column-orthogonal matrix (eigenvectors), $\mathbf{V}(n \times n)$ is a diagonal matrix (eigenvalues λ), and $\mathbf{W}^{\mathbf{T}}(n \times n)$ is the transpose of an $n \times n$ orthogonal matrix. The eigenvectors are the PCs, and the eigenvalues represent variance. They are used to rank PCs according to their importance in reproducing data. Thus, the result of applying PCA to titration data is a set of orthogonal PCs equal to the number of curves (n), and the analysis seeks the number of significant PCs ($j < n$) that is necessary and sufficient to reconstruct all data within error:

$$rec(data)_{j\mathrm{pc}} = \sum_{i=1}^{j} \lambda_i e_i \rho_{e_i,\,data}$$

Because j is the number of uncorrelated linear combinations of PCs that maximizes the variance in the data set and hence explains the variation among all observables, its value represents the number of independent parameters contained in a titration curve.

The fit quality of the reconstructions was evaluated with the normalized sum-square ($NSS_{i\mathrm{pc}} = \sum_i (y_{\mathrm{exp}} - y_{\mathrm{fit}})^2/y_{\mathrm{exp}}^2$) parameter (Figure S1 in Supporting Information). Because NSS decreases when the number of component fits (i) increases, a best-fit criterion had to be defined to estimate j. We considered that $i = j$ when $NSS_{i\mathrm{pc}} \approx NSS_{\mathrm{data}}$, that is, when the precision of the reconstruction was comparable to the dispersion of the data points. In fact, the reconstruction by PCA can be seen as a noise-filtering technique because the least important components ($i > j$) represent noise and other errors.

Analytical Treatment

Statistically, the precision of j increases with n and m, that is, with the total number of data points. To maximize n and in the meantime to be able to discuss the PCA results in the context of proton-binding constants from the literature, the extended data set analyzed by Milne et al.[21] was chosen for this study. For consistency, the original names of the titration data were preserved (i.e., HHx and FHx, where the first letter stands for humic (H) or fulvic (F) acid, the second letter stands for a proton, and x stands for the sample number). Because all curves were not measured over the same pH interval, there is a trade off between optimizing n and m. Enlarging the common pH interval by extrapolation of the data points is not an option because the statistical analysis would lose robustness and is likely to compromise the results.[4] The pH interval retained is [3.5−9.8] and comprises 23 FA and 24 HA curves all at $I = 0.1$ M (Figures S2 and S3 in Supporting Information). No data that satisfied the two conditions were deliberately omitted. Replicate measurements (denoted HH/FHx_y) were considered to be independent. The validity of this hypothesis was verified a posteriori. PCA was conducted on three data subsets (FA, HA, and FA + HA) over two pH intervals ([3.5−5.5] and [3.5−9.8]) for each group. The basis data were interpolated on two uniform grids starting at pH 3.5:

- ΔpH = 0.02 for 3.5 ≤ pH ≤ 5.5 (101 points)
- ΔpH = 0.1 for 3.5 ≤ pH ≤ 9.8 (62 points)

The primary aim of the parametrization is to obtain the best possible analytical expression for the titration data, this unified representation being necessary for statistical analysis. Although Nederlof et al.[22] recommended using a spline function to parametrize titration data on a common pH grid, a polynomial interpolation was preferred for two reasons. First, polynomials are more suitable for PCA because they can be expressed as a sum of orthogonal components. Second, $pK_{Hi\mathrm{app}}$ is calculated simply by derivation of the polynomial. Finally, all titration curves were translated to the same origin of [3.5, 0.0] before numerical analysis. Thus, the real number of independent parameters is $j + 1$. Interpolations were performed with MATLAB 7 and PCA with the homemade Labview software described in Manceau et al.[23]

The normalized sum-square difference between data and the polynomial fit (NSS_{inter}) became quite sensitive to experimental errors above the sixth order (Tables S1 and S2). Therefore, this order was retained for the interpolation (i.e., $NSS_{\mathrm{inter}} = NSS_{\mathrm{data}}$). Upon inspecting Tables S1 and S2 and Figures S2 and S3, the reader may easily convince himself that neither unwanted nor unacceptable errors were introduced by these interpolations, thus possible errors are negligible. The precision of [Hs] derived from NSS_{inter} varies from 10^{-7} to 10^{-5} (Table S2), and the uncertainty in $pK_{H1\mathrm{app}}$ deduced from these values is 0.057 ± 0.040 for FA and 0.026 ± 0.014 for HA. These calculated uncertainties are similar to experimental errors reported in the literature.[24,25] In addition, the $pK_{H1\mathrm{app}}$ values calculated for all FA and all HA followed a normal distribution, as shown numerically in Table S3 and graphically in Figure 2. This result provides a stringent test of the quality of our polynomial parametrization and also demonstrates

(18) Gustafsson, J. P. *J. Colloid Interface Sci.* **2001**, *244*, 102–112.
(19) Kinniburgh, D. G.; Van Riemsdijk, W. H.; Luuk, K.; Koopal, C.; Borkovec, M.; Benedetti, M. F.; Avena, M. J. *Colloids Surf., A* **1999**, *151*, 147–166.
(20) Christl, I.; Milne, C. J.; Kinniburgh, D. G.; Kretzschmar, R. *Environ. Sci. Technol.* **2001**, *35*, 2512–2517.
(21) Milne, C. J.; Kinniburgh, D. G.; Tipping, E. *Environ. Sci. Technol.* **2001**, *35*, 2049–2059.
(22) Nederlof, M. M.; Van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1994**, *28*, 1037–1047.
(23) Manceau, A.; Marcus, M. A.; Tamura, N. Quantitative Speciation of Heavy Metals in Soils and Sediments by Synchrotron X-ray Techniques. In *Applications of Synchrotron Radiation in Low-Temperature Geochemistry and Environmental Science*; Fenter, P. A., Rivers, M. L., Sturchio N. C., Sutton, S. R., Eds.; Reviews in Mineralogy and Geochemistry; Geochemical Society and Mineralogical Society of America: Washington, DC, 2002; Vol. *49*, pp 341−428.
(24) Cabaniss, S. E.; Mcvey, I. F. *Spectrochim. Acta, Part A* **1995**, *51*, 2385–2395.
(25) Koort, E.; Herodes, K.; Pihl, V.; Leito, I. *Anal. Bioanal. Chem.* **2004**, *379*, 720–729.

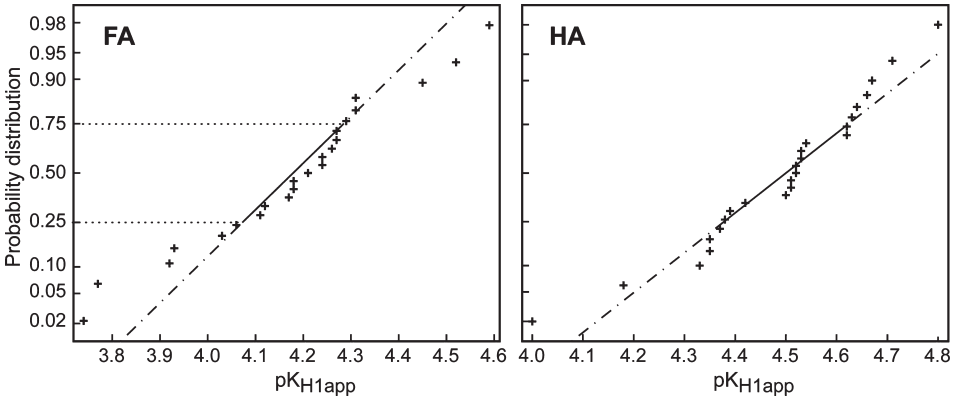

Figure 2. Probability distribution of pK_{H1app} values represented in the form of a cumulative distribution function. The lines are the cumulative functions for a random variable (i.e., normal distribution) with mean values of 4.18 and $\sigma = 0.21$ (FA = FH subdata set) and 4.49 and $\sigma = 0.18$ (HA = HH subdata set). The straight lines represent the middle half of the distribution or the interquartile range, defined as the spread in the distribution between the first and third quartiles (dashed-dotted lines) with respect to the median. For the most part, the pK_{H1app} values follow a normal distribution, meaning that all titration curves are independent data.

Table 2. PCA Results in the [3.5−5.5] pH Interval

FA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
min	2.64×10^{-7}	5.77×10^{-6}	7.75×10^{-8}	1.30×10^{-11}
max	8.30×10^{-5}	2.30×10^{-3}	1.76×10^{-4}	4.93×10^{-7}
$NSS_{inter} < NSS_{jpc}$		22	16	0
HA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
min	1.67×10^{-7}	2.08×10^{-5}	3.15×10^{-7}	1.54×10^{-9}
max	2.49×10^{-4}	9.95×10^{-3}	1.36×10^{-4}	1.99×10^{-7}
$NSS_{inter} < NSS_{jpc}$		23	12	0
FA + HA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}
min	1.67×10^{-7}	2.17×10^{-5}	9.34×10^{-8}	5.46×10^{-12}
max	2.49×10^{-4}	1.67×10^{-2}	1.99×10^{-4}	4.96×10^{-7}
$NSS_{inter} < NSS_{jpc}$		47	35	0

our initial assumption about the lack of correlation between multiple measurements of the same sample in the data set. The distribution probability of pK_{H2app} could not be calculated because 25 curves have their second inflection point at pH >9.8 (i.e., only 50% of phenolic-type sites had been deprotonated).

Apparent pK_{Hi} Values

The mean values of the apparent proton-binding constants are pK_{H1app}(FA) = 4.18 ± 0.21, pK_{H1app}(HA) = 4.49 ± 0.18, pK_{H2app}(FA) = 9.29 ± 0.33, and pK_{H2app}(HA) = 9.29 ± 0.38 (Table S3). These values coincide with those reported in the literature: $4.20 \le pK_{H1app} \le 4.28$ and $9.27 \le pK_{H2app} \le 9.80$ for NOM at $I = 0.1$ M;[1] pK_{H1app}(FA) = 3.80 ± 0.11, pK_{H1app}(HA) = 4.38 ± 0.13, pK_{H2app}(FA) = 9.78 ± 0.48, and pK_{H2app}(HA) = 9.72 ± 0.23 for NOM standards from the International Humic Substances Society (IHSS) at $I = 0.1$ M using the NICA model (called the modified Henderson−Hasselbalch in the source article);[26] pK_{H1app}(HA) = 4.60 ± 0.17 and pK_{H2app}(HA) = 8.87 ± 0.23 for vermicompost humic acids at $I = 0.1$ M;[27] and pK_{H1app} = 3.3/4.8 (two-component fit) and pK_{H2app} = 9.60 for dissolved organic matter at $I = 0.01$ M.[28] They are also consistent with the compilation of Perdue,[29] who showed that carboxylic- and phenolic-type groups have a Gaussian distribution of pK values centered at pK_{H1app} = 4.5 and pK_{H2app} = 10.

PCA Results over the [3.5−5.5] pH Interval

As mentioned previously, NSS_{ipc} decreases continuously when n increases. PCA results (Tables 2 and S4) show that the reconstructed signal falls progressively within the noise level for $2 < n < 3$ because $NSS_{3pc} < NSS_{inter} \approx NSS_{2pc} < NSS_{1pc}$. For $n = 2$, NSS_{ipc} and NSS_{inter} have the same scale class (hereafter noted $\langle NSS \rangle$) when all data are taken together. However, an examination of individual reconstructions (Table S4) shows that 16 out of the 23 FA (70% of data), 12 out of the 24 HA (50%), and 35 out of 47 FA + HA (75%) do not satisfy the $NSS_{npc} \approx NSS_{inter}$ criterion. Thus, the minimum number of PCs required to reproduce the data set within error in the [3.5−5.5] pH interval is three.

With two PCs, fewer FAs (30%) are reproduced than HAs (50%). This result is consistent with the difference in variability of the pK_{H1app} values for the two humic materials noted previously (σ(FA) = 0.21 and σ(HA) = 0.18). It agrees also with the generic values of the site distribution parameter m_{Hi} in the N-D model because $m_{H1(FA)} = 0.38$ and $m_{H1(HA)} = 0.50$[21] with m_{Hi} being inversely proportional to heterogeneity (Figure 1c).

PCA Results over the [3.5−9.8] pH Interval

In the [3.5−9.8] pH interval, every j increment decreases $\langle NSS_{npc} \rangle$ by about one order of magnitude up to $j = 6$ (Tables 3 and S5). For $j = 6$, $\langle NSS_{npc} \rangle \approx 0$ because a sixth-degree polynomial is used to interpolate titration curves. A sixth-order

(26) Ritchie, J. D.; Perdue, E. M. *Geochim. Cosmochim. Acta* **2003**, *67*, 85–96.
(27) Masini, J. C.; Abate, G.; Lima, E. C.; Hahn, L. C.; Nakamura, M. S.; Lichtig, J.; Nagatomy, H. R. *Anal. Chim. Acta* **1998**, *364*, 223–233.
(28) Lu, Y.; Allen, H. E. *Water Res.* **2002**, *36*, 5083–5101.
(29) Perdue, E. In *Humic Substances in Soil, Sediment, and Water: Geochemistry, Isolation, and Characterization*; Aiken, G. R., McKnight, D. M., Wershaw, R. L., McCarthy, P., Eds.; John Wiley and Sons: New York, 1985; Vol. *1*, pp 493−526.

Table 3. PCA Results in the [3.5−9.8] pH Interval

FA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
min	2.89×10^{-7}	3.90×10^{-5}	3.14×10^{-6}	2.64×10^{-7}	2.74×10^{-8}	5.35×10^{-9}	2.18×10^{-17}
max	3.55×10^{-5}	1.40×10^{-2}	3.27×10^{-4}	1.40×10^{-4}	1.61×10^{-5}	5.84×10^{-6}	1.31×10^{-16}
$NSS_{inter} < NSS_{jpc}$		23	23	21	10	3	0
HA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
min	4.09×10^{-7}	8.70×10^{-5}	2.17×10^{-6}	8.78×10^{-7}	4.86×10^{-7}	$4.98 \times 10^{-1}0$	2.82×10^{-17}
max	7.83×10^{-5}	3.50×10^{-5}	7.95×10^{-4}	2.68×10^{-4}	5.84×10^{-5}	3.67×10^{-6}	2.54×10^{-16}
$NSS_{inter} < NSS_{jpc}$		24	24	19	9	1	0
FA + HA	NSS_{inter}	NSS_{1pc}	NSS_{2pc}	NSS_{3pc}	NSS_{4pc}	NSS_{5pc}	NSS_{6pc}
min	2.89×10^{-7}	1.23×10^{-4}	1.98×10^{-5}	9.96×10^{-7}	3.56×10^{-8}	$4.95 \times 10^{-1}0$	3.98×10^{-17}
max	7.83×10^{-5}	5.29×10^{-2}	5.19×10^{-3}	5.77×10^{-4}	6.12×10^{-5}	5.73×10^{-6}	2.61×10^{-16}
$NSS_{inter} < NSS_{jpc}$		47	47	45	27	4	0

polynomial has seven degrees of freedom, but here this number is reduced to six because all curves were normalized to the same origin. For the three data sets, $\langle NSS_{inter} \rangle \approx \langle NSS_{4pc} \rangle$. With four PCs, 10 FA (44%), 9 HA (38%), and 27 FA + HA (57%) fail to satisfy the reconstruction criterion. With five PCs, all curves meet the criterion except for three FAs (13%), one HA (4%), and four FA + HAs (9%). Upon closer examination, in fact it appears that the three FAs (FH19, 20, and 21) have an NSS_{5pc}/NSS_{inter} ratio ≤3.14, which is small on the logarithmic scale (Table S5). Therefore, we conclude that j(FA) = 5. The HA that is not reproduced correctly is HH12_2. Its NSS_{5pc}/NSS_{inter} ratio is $3.12 \times 10^{-6}/4.09 \times 10^{-7} = 7.6$, a relatively high value. However, this curve has the lowest NSS_{inter} value of the 24 curves in the HA data set, meaning that the quality of this fit is probably too high and thus the error in this data is underestimated. We conclude that this data is extreme in terms of the experimental error, thus j(HA) = 5. Lastly, $1.1 < NSS_{5pc}/NSS_{inter} < 3.8$ for three out of the four FA + HAs (FH211, FH22, and HH8), which is small, and the fourth, which has the highest ratio (6.9), is again HH12_2. Thus, j(FA+HA) = 5.

In summary, the number of significant independent PCs accounting for the variance in the titration data is five from pH 3.5 to 9.8 and three from pH 3.5 to 5.5. The number of three at low pH means that the reactivity of carboxylic-type groups can be described (i) with a continuous probability distribution of binding constants, such as a normal distribution with a maximum density at the mean (pK_{Hi}), (ii) a pK_{Hi} ($1/m_{Hi}$) distribution equal to σ at the 68% confidence level, and (iii) a total concentration of sites (Q_{Hi}) equal to the integrated area (Figure 1c). The lower sensitivity of titration data to phenolic-type groups (two degrees of freedom instead of three) is explained by the fact that the experimental pK_{H2app} values (9.29 ± 0.33 for FA and 9.29 ± 0.38 for HA, Table S3) are close to the upper limit of the titration curves (pH_{max} 9.8). Clearly, the titration data lack sensitivity to pK_{H2}. Thus, the higher variability of phenolic constants reported in the literature may not always be real, and pK_{H2} values are likely underestimated for $pH_{max} < 10$.

Concluding Remarks

Because of the bimodal character of the proton affinity for NOM, two Gaussian functions are typically needed to describe the full acid−base properties of the reactant. By adding electrostatic effects and the initial charge offset,[30] seven to eight parameters are optimized to simulate a single data set. The number of unknown parameters exceeds by at least one the number of independent parameters (six) that can be freely varied during data fitting and up to three if the pH span is insufficient. Thus, the flexibility of proton-binding models is in general too high to provide robust chemical solutions because a single data set can be fit with an infinite number of numerical combinations. Obviously, some constraints have to be enforced to decrease correlations between parameters, as in model VI[15] and the Stockholm humic model;[18] otherwise, the results obtained represent essentially subjective choices. This does not mean that the existing models are overparametrized but that the titration data lack sensitivity to quantify the full binding properties of humic substances and that independent scientific and statistical knowledge has to be applied to constrain parameters in order to obtain meaningful values from measurements.

The considerable covariance between fit parameters has often been mentioned in the literature (e.g., refs 16 and 19), but the maximum number of parameters that can be refined was always unknown. In fact, modelers are cautioned against the existence of correlations in the NICA-Donnan User Guide:[11] "High correlations indicate that a minimum is being sought in a deep flat-bottomed valley: there is no problem finding a low point but you may have to wander down the valley a long way to find the exact lowest point." The PCA results show that the task at hand is statistically impossible unless one decreases the number of adjustable parameters in the least-squares fit of the experimental data. Sometimes the electrostatic b parameter of the N-D model is determined independently by measuring data at multiple ionic strengths and calculating a master curve.[31,32] Still, "in practice, it is not easy to derive a unique master curve," states Kinniburgh et al.[19] This difficulty can be explained by the linear relationship between the logarithms of the Donnan volume (V_D) and the electrolyte concentration in the Donnan electrostatic model as $\log V_D = b(1 - \log I) - 1$. Therefore, varying I does not solve the problem, and if it did, the fit model to the master curve still would be underconstrained by one degree in most cases.

The only way to improve tangibly the applicability of potentiometric titration models for NOM is to gather independent estimates of some proton-binding parameters by other means in order to reduce further the number of parameters adjusted by regression analysis. In a study of the binding of Cu(II) and Pb(II) to FA and HA, the number of phenolic-type sites was determined by ^{13}C NMR and the Donnan volume was estimated from peak elution times measured by size exclusion chromatography.[20,33]

(30) Milne, C. J.; Kinniburgh, D. G.; De Wit, J. C. M.; Van Riemsdijk, W. H.; Koopal, L. K. *Geochim. Cosmochim. Acta* **1995**, *59*, 1101–1112.

(31) De Wit, J. C. M.; Van Riemsdijk, W. H.; Nederlof, M. M.; Kinniburgh, D. G.; Koopal, L. K. *Anal. Chim. Acta* **1990**, *232*, 189–207.

(32) De Wit, J. C. M.; Van Riemsdijk, W. H.; Koopal, L. K. *Environ. Sci. Technol.* **1993**, *27*, 2005–2014.

(33) Christl, I.; Kretzschmar, R. *Environ. Sci. Technol.* **2001**, *35*, 2505–2511.

Avena et al.[12] investigated the possibility of using hydrodynamic radii measured by viscometry as the Donnan radius and came to the conclusion that the thus-obtained Donnan volume was too small to account for ionic strength effects on the surface charge (or the gel volumes used in the N-D model were unrealistically high, which is the same). More recently, semi-empirical pK_{Hi} values were estimated from structural models using the linear free-energy relationships (LFER) of Hammett and were used to reduce the number of fitting parameters.[34] The predicted pK_{Hi} values obtained by this approach are 3.73 ± 0.13 and 9.83 ± 0.23 for HA and 3.80 ± 0.20 and 9.87 ± 0.31 for FA. The predicted values for carboxylic-type groups compare well with results from potentiometric titrations analyzed using model VI and the Stockholm humic model (Table 1) but are about one log unit higher than the generic N-D values.[21] In contrast, phenolic-type sites have titration constants that are all lower than the predicted values, in part because it is complicated to measure the proton balance at high pH, as discussed previously. Given the difficulty in safely deducing intrinsic pK values from titration data only, publishing apparent (i.e., conditional) values obtained directly from the first derivative[35,36] of the titration data seems to be the most conservative approach. Then their dependence on ionic strength can be estimated empirically from the Davies equation[37,38] with reasonably good precision because log K_H typically varies by only 0.3 to 0.4 unit in the $0 \leq I \leq 2.0$ interval.[34]

The accuracy of acidic−basic parameters also depends on the accuracy of the titration measurement. Existing models suppose that protonation equilibrium is reached between subsequent aliquot additions of acid and base, which may not be the case when the initially rapid proton uptake is followed by a slower diffusion process. Suggestive evidence for kinetic effects are provided by titration hysteresis.[39−43] Lastly, systematically publishing the data from titration measurements in electronic annexes also would be helpful in the communal effort to improve the applicability of the existing electrostatic models and in turn the reliability of published intrinsic constants. Because the intrinsic constants are a property of the humic substance alone, they should not vary with solution composition nor the unpredictability of a minimization code.[19]

Acknowledgment. We thank Dr. M. F. Benedetti for providing the titration database analyzed in this work and Drs. A. Matynia and B. Causse for fruitful scientific discussions.

Supporting Information Available: Method used to calculate normalized sum-square values (*NSS*), plots of all titration curves with best-fit sixth-order polynomials, numerical results of all polynomial fits and coefficients of the sixth-order polynomial fits, $pK_{Hi\mathrm{app}}$ values for all titration curves, and numerical PCA results. This material is available free of charge via the Internet at http://pubs.acs.org.

(34) Matynia, A.; Lenoir, T.; Causse, B.; Spadini, L.; Jacquet, T.; Manceau, A. *Geochim. Cosmochim. Acta*, 2010, doi: 10.1016/j.gca.2009.12.022.
(35) Prélot, B.; Charmas, R.; Zarzycki, P.; Thomas, F.; Villiéras, F.; Piasecki, W.; Rudzinski, W. *J. Phys. Chem.* **2002**, *106*, 13280–13286.
(36) Charmas, R.; Zarzycki, P.; Villiéras, F.; Thomas, F.; Prélot, B.; Piasecki, W. *Colloids Surf., A* **2004**, *244*, 9–17.
(37) Davies, C. W. *Ion Association*; Butterworths: Washington, DC, 1962.
(38) Stumm, W.; Morgan, J. J. *Aquatic Chemistry*, 3rd ed. Wiley: New York, 1996.
(39) Marshall, S. J.; Young, S. D.; Gregson, K. *Eur. J. Soil Sci.* **1995**, *46*, 471–480.
(40) Duc, M.; Adekola, F.; Lefevre, G.; Fedoroff, M. *J. Colloid Interface Sci.* **2006**, *303*, 49–55.
(41) Lefevre, G.; Duc, M.; Fedoroff, M. Accuracy in the Determination of Acid-Base Properties of Metal Oxides Surfaces. In *Surface Complexation Modeling*; Lützenkirchen, J., Ed.; Elsevier: Amsterdam, 2006; pp 35−66.
(42) Cooke, J. D.; Hamilton-Taylor, J.; Tipping, E. *Environ. Sci. Technol.* **2007**, *41*, 465–470.
(43) Zarzycki, P.; Rosso, K. M. *Langmuir* **2009**, *25*, 6841–6848.
(44) Plaza, C.; Senesi, N.; Polo, A.; Brunetti, G. *Environ. Sci. Technol.* **2005**, *39*, 7141–7146.
(45) Drosos, M.; Jerykiewicz, M.; Deligiannakis, Y. *J. Colloid Interface Sci.* **2009**, *332*, 78–84.