Supporting Information for:

Determination of Aluminum and Iron Complexation by Water-Soluble Organic Matter Using Landscape Fluorescence Spectroscopy and Multi-way Analysis

Tsutomu Ohno,* Aria Amirbahman, and Rasmus Bro

Site Description, Sampling, and Extraction.   Bear Brook Watershed in Maine (BBWM) is a site of a long-term, paired-watershed experimental acidification study that includes both deciduous and coniferous stands (1,2). O-horizons of deciduous and coniferous stands were collected from the reference watershed at BBWM. Soils supporting forests at the BBWM are freely drained Spodosols [FAO classification Podzols], typical of northern New England, and there is no history of physical disturbance or cultivation at this site. Soils are in the Turnbridge and Rawsonville soil series (loamy, mixed, frigid Typic Haplorthods) with well-developed spodic horizons. Organic horizons were excavated using a 15 x 15 cm frame and the material was collected to the surface of the underlying mineral horizon. All samples were air-dried. Two types of vegetation predominate in the watersheds. The upper elevations have coniferous stands, composed primarily of red spruce (*Picea rubens* Sarg.) and balsam fir (*Abies balsamea* L. Mill.), while the lower elevations have deciduous stands, comprised primarily of American beech (*Fagus grandifolia* Ehrh.), yellow birch (*Betula alleghaniensis* Britt.), red maple (*Acer rubrum* L.) and sugar maple (*Acer saccharum* Marsh.)

Soil pH was measured using distilled water at a 1 g:10 mL ratio. The acid functional group content of each material was determined by potentiometric titration in a glass reaction beaker maintained at 25.0 ± 0.1 °C.   Titration solutions were adjusted to 20 mmol L^{-1} C_{TS} and 20 mmol L^{-1} ionic strength using 1M KCl solution. The solution was adjusted to pH 3 using 0.10 mol L^{-1} HCl , N_2 was bubbled through the solutions for

15 min prior to titration to minimize CO_2 contamination, and titrated with standardized 0.05 mol L^{-1} NaOH to pH 10. The moles of titrant consumed between the operational beginning and end points of pH 3 and pH 8 was taken to be equal to the carboxyl-group content. Phenolic-group content was taken to be equal to twice the titrant consumed between pH 8 and 10. Blank corrections were made by subtracting the quantity of base consumed in titrating 0.02 M KCl solutions in from pH 3 to 10.

WSOM Properties. The coniferous O-horizon had over twice the WSOM content as the deciduous O-horizon (Table S1). Ultra-violet molar absorptivity at 280 nm and E2/E3 ratio has been used to probe the aromaticity and humification status of WSOM. Both UV spectrometric indices indicate that the deciduous WSOM was about 20% more humified than the coniferous WSOM (Table S1). Carboxyl- and phenolic-acidic functional group contents measured by potentiometric titration showed that the deciduous WSOM contained about 20% more acidic functional groups than the coniferous WSOM. These differences between the two WSOM extracts may affect their ability to bind metals since acidic functional group content and aromaticity are important chemical factors controlling metal complexation by organic matter ligands.

Table 1. Selected chemical properties of the water-soluble organic matter from the reference deciduous and coniferous stands at Bear Brook Watershed in Maine study site.

Parameter	Deciduous		Coniferous
Water-Soluble Organic C, g kg^{-1} soil	2.09		5.48
280 nm Absorptivity, L mol^{-1} cm^{-1}	104		83
E2/E3 Absorbance Ratio	5.3		7.1
Carboxyl-group Content, mmol g^{-1} C	8.0		6.1
Phenolic-group Content, mmol g^{-1} C	2.7		1.9
Total Acidity Content, mmol g^{-1} C	10.7		8.0

PARAFAC Modeling.   Essentially a PARAFAC model of a set of EEM landscapes provides an estimate of the number of fluorophores as well as the excitation and emission spectrum of these. It also provides the relative concentration of each fluorophore in each sample. Each fluorescence landscape measurement provides an EEM and when several of these are combined, they can be held in a so-called three-way array of size $I \times J \times K$, where I is the number of samples, J the number of emission wavelengths, and K the number of excitation wavelengths. Such a three-way array can not directly be modeled by standard multivariate analysis tools because these only work on two-way matrices. The PARAFAC model is specifically made to deal with such three-way data and can be viewed as an extension of principal component analysis (3). Unlike principal component analysis, the PARAFAC model is uniquely identified without additional orthogonality constraints (4). This means that if the underlying structure of the three-way data coincides with the PARAFAC model, then the parameters of the PARAFAC model will reflect the true underlying parameters. For fluorescence data, this is ideally the case. Each fluorophore will give rise to one PARAFAC component and each such component consists of an estimated emission spectrum, and estimated excitation spectrum and a score vector where each element is the relative concentration of the fluorophore.

The PARAFAC modelling was conducted with MATLAB Release 14 (Mathworks, Natick, MA) using PLS_Toolbox version 4.0 (Eigenvector Research, Manson, WA). A non-negativity constraint was applied to the parameters to allow only chemically relevant results. The PARAFAC models with two to six components were computed for the pooled (Fe and Al titration EEM and replicates) deciduous and coniferous O-horizon data sets separately. The determination of the correct number of components in the data set was assessed by the core consistency diagnostic score which should be close to 100 % for appropriate models. The core consistency provides an estimate of how well the model captures trilinear information and if the consistency turns low, i.e. towards zero, it is a strong indication that the model is invalid (5). The number of components was

further validated by visual inspection of the estimated parameters and additional model diagnostics.

Several pre-processing steps were used to minimize the influence of scatter lines and other attributes of the EEM landscape that are due to the background solution matrix prior to PARAFAC modeling. Subtraction of a control DI-H_2O EEM from sample EEM was used to remove the lower intensity Raman scatter lines. The higher intensity Rayleigh scatter lines were removed by replacing the fluorescence intensity values with missing values in the region immediately adjacent to where emission wavelength was equal to one and two times the excitation wavelength. In addition, the EEM spectra had a triangular shaped region where the emission wavelength was less than that of the excitation wavelength. Such a characteristic is a physical impossibility and thus these data pairs were set to zero (6).

References:

1) Ohno, T.; Fernandez, I.J.; Hiradate, S.; Sherman, J.F. Effects of soil acidification and forest type on water soluble soil organic matter properties. G*eoderma* **2007**, 140, 176-187.

2) Norton, S.A.; Fernandez, I.J.; Kahl, J.S.; Reinhardt, R.L. Acidification trends and the evolution of neutralization mechanisms through time at the Bear Brook Watershed in Maine (BBWM), U.S.A. *Water Air Soil Pollut. Focus* **2004,** 4, 289-310.

3) Harshman, R.A. Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis. UCLA Working Papers in Phonetics 1970,16, 1-84.

4) Sidiropoulos, N.D.; Bro, R. On the uniqueness of multilinear decomposition of n-way arrays. J.Chemometrics **2000**, 14, 229-239.

5) Bro, R.; Kiers, H.A.L. A new efficient method for determining the number of components in PARAFAC models. *J. Chemometrics* **2003**, 17, 274-286.

6)   Thygesen, L.G.; Rinnan, A.; Barsberg, S.; Møller, J.K.S. Stabilizing the PARAFAC

        decomposition of flourescence spectra by insertion of zeros outside the data

        area. *Chemometrics Intelligent Lab. Syst.* **2006**, 81, 107-108.