Fluorescence-based proxies for lignin in freshwater dissolved organic matter

Peter J. Hernes,[1] Brian A. Bergamaschi,[2] Robert S. Eckard,[1] and Robert G. M. Spencer[1,3]

Received 16 January 2009; revised 8 July 2009; accepted 28 July 2009; published 4 November 2009.

[1] Lignin phenols have proven to be powerful biomarkers in environmental studies; however, the complexity of lignin analysis limits the number of samples and thus spatial and temporal resolution in any given study. In contrast, spectrophotometric characterization of dissolved organic matter (DOM) is rapid, noninvasive, relatively inexpensive, requires small sample volumes, and can even be measured in situ to capture fine-scale temporal and spatial detail of DOM cycling. Here we present a series of cross-validated Partial Least Squares models that use fluorescence properties of DOM to explain up to 91% of lignin compositional and concentration variability in samples collected seasonally over 2 years in the Sacramento River/San Joaquin River Delta in California, United States. These models were subsequently used to predict lignin composition and concentration from fluorescence measurements collected during a diurnal study in the San Joaquin River. While modeled lignin composition remained largely unchanged over the diurnal cycle, changes in modeled lignin concentrations were much greater than expected and indicate that the sensitivity of fluorescence-based proxies for lignin may prove invaluable as a tool for selecting the most informative samples for detailed lignin characterization. With adequate calibration, similar models could be used to significantly expand our ability to study sources and processing of DOM in complex surface water systems.

Citation: Hernes, P. J., B. A. Bergamaschi, R. S. Eckard, and R. G. M. Spencer (2009), Fluorescence-based proxies for lignin in freshwater dissolved organic matter, *J. Geophys. Res.*, *114*, G00F03, doi:10.1029/2009JG000938.

1. Introduction

[2] The biogeochemistry of dissolved organic matter (DOM) has emerged as a critical component of the global carbon cycle, capable of transferring significant amounts of reduced carbon between global reservoirs while fueling food webs in aquatic environments. Processes that affect DOM cycling are intimately linked to DOM composition and the reactivity of individual molecular structures within the DOM pool. Biomarker analytical techniques such as CuO oxidation for lignin [*Hedges and Ertel*, 1982] are important tools in DOM research because they provide insight into the molecular world of DOM, and thus are crucial toward understanding DOM reactivity.

[3] Lignin provides important source information about the vascular plant or terrigenous component of DOM, but also has the capability for capturing diagenetic history. Measurements of dissolved lignin have been utilized to show that terrigenous OM is just a few percent of the oceanic DOM pool despite the fact that riverine DOM fluxes to the ocean are greater than average turnover of the oceanic DOM pool [*Hernes and Benner*, 2006; *Opsahl and Benner*, 1997]. These findings indicate rapid loss rates of terrigenous DOM in marine environments. One likely sink, photochemical oxidation, imparts a chemical signature to the lignin component of DOM, namely elevated acid:aldehyde ratios along with decreasing ratios of syringyl:vanillyl phenols in the high molecular weight (~1 to 100 nm) fraction [*Hernes and Benner*, 2003; *Opsahl and Benner*, 1998]. In freshwater systems, concentrations and compositions of dissolved lignin demonstrate the strong influence of local landscape-scale features on DOM [*Eckard et al.*, 2007] as well as the hydrology of the system [*Dalzell et al.*, 2007; *Hernes et al.*, 2008]. Furthermore, carbon-normalized lignin yields have been used to show that a significant fraction of riverine DOM may not be vascular-plant derived as commonly thought [*Hernes et al.*, 2007; *Spencer et al.*, 2008]. Together, these studies point toward vascular plant-derived DOM as a dynamic pool that undergoes large changes both spatially and temporally in aquatic environments, and demonstrate that lignin is a powerful tracer of significant environmental processes. However, for many studies that would benefit from the use of lignin analyses, the analytical difficulty and expense of CuO oxidation significantly impedes our abilities to achieve sufficient spatial and temporal resolution to capture the dynamics of natural systems.

[4] In stark contrast to time-consuming and complex molecular biomarker measurements, spectrophotometric

[1]Department of Land, Air, and Water Resources, University of California, Davis, California, USA.
[2]U.S. Geological Survey, Sacramento, California, USA.
[3]Department of Plant Sciences, University of California, Davis, California, USA.

Copyright 2009 by the American Geophysical Union.
0148-0227/09/2009JG000938$09.00

techniques such as DOM fluorescence can be rapidly measured on benchtop machines or collected nearly continuously by in situ instruments. Although the information obtained through fluorescence measurements is not as definitive as molecular biomarker analyses, fluorescence data are still directly tied to the chemical composition of dissolved constituents, and as such can be used to determine DOM source and processing history. A number of studies have related fluorescence properties of natural waters to the concentration, source and composition of DOM [e.g., *Bergamaschi et al.*, 2005; *McKnight et al.*, 2001; *Stedmon et al.*, 2003]. For example, certain features in fluorescence excitation emission matrices (EEMs) such as hypsochromic shifts in peaks attributed to humic and fulvic-like material have been attributed to a breakdown in aromaticity [*Blough and Del Vecchio*, 2002; *Coble*, 1996] and thus have been used to distinguish between allochthonous and autochthonous-derived DOM [*Spencer et al.*, 2007a]. The ability to use spectrophotometric measurements to distinguish between DOM of different sources and observe DOM processing coupled with the emerging ability to conduct these measurements in situ leads to the potential for high-frequency, real-time monitoring of aquatic systems.

[5] *Hedges et al.* [1997] hypothesized that detailed surveys based on spectrophotometric measurements might allow DOM dynamics (e.g., photochemical breakdown) to be observed in the coastal ocean. Recent studies have indeed shown the potential of high-frequency spectrophotometric measurements to discriminate DOM of different sources and to monitor changes in DOM composition over short timescales in both marine [*Coble*, 2007] and riverine systems [*Spencer et al.*, 2007b]. The focus of this study was to construct fluorescence-based models for lignin concentrations and compositions measured in a diverse estuarine environment as a potential tool for capturing the dynamics of lignin cycling with greater sensitivity and temporal/spatial coverage afforded by fluorescence measurements.

2. Methods

2.1. Sampling and Analyses

[6] Fifty-eight water samples (0.2 μm filtered) were collected from 11 stations within the Sacramento River/San Joaquin River Delta (hereafter referred to as the Delta) over the course of two annual hydrological cycles, December 1999 to May 2001 (Figure 1 [*Stepanauskas et al.*, 2005]). The Delta is the tidal freshwater portion of the San Francisco Bay Estuary, located in central California, United States, and consists of many contrasting habitats and land uses. The 11 stations sampled included wetlands, rivers, channelized waterways, and open water sites. All Delta samples were analyzed within 24 h for fluorescence properties and DOM concentrated by XAD8 and XAD4 according to the methods of *Aiken et al.* [1992], then freeze-dried [*Kraus et al.*, 2008]. As these measurements were part of a much larger study, XAD was used in order to isolate enough material for several different analyses, as opposed to C18 which is used more routinely for just lignin measurements. However, XAD recovers lignin in freshwater with ~90% efficiency with minimal compositional fractionation compared to rotary evaporated water or C18 (unpublished data), thus the lignin measurements on XAD are essentially whole-water measurements. Lignin analyses on XAD extracts were done by alkaline CuO oxidation at 155°C in Monel reaction vessels, extracted by ethyl acetate, and analyzed by GC-MS using a five-point calibration scheme with a precision of ~10% for individual lignin phenols [*Eckard et al.*, 2007; *Hernes and Benner*, 2002]. These lignin measurements together with the fluorescence spectra served as the basis for constructing our models. In addition to the Delta samples, 25 water samples (0.2 μm filtered) were collected from the San Joaquin River at Crow's Landing (~20 miles upstream of the Delta) at 2-h time intervals starting from noon, 28 July 2005 through noon, 30 July 2005 as described by *Spencer et al.* [2007b]. The San Joaquin River drains the southern portion of the Central Valley in California, is disconnected from its source during summer months due to water diversions, and drains predominantly row crop agriculture, orchard and wetlands during this period [*Kratzer et al.*, 2004]. Samples at Crow's Landing were analyzed for fluorescence properties within 24 h of collection, which were then used to model lignin compositions and concentrations.

[7] Fluorescence was measured on filtered water from both study sites using SPEX model Fluoromax 2 (Delta samples) and 3 (Crow's Landing samples) spectrophotometers (Horiba Jobin Yvon, Ltd., Japan) at room temperature in a 10 mm quartz cuvette. Excitation occurred over the wavelengths 250–440 nm at 10 nm intervals; for each excitation wavelength, emissions were detected within the 300–600 nm range at 1 nm intervals, but the lowest wavelength utilized in this study was always 50 nm higher than excitation, which avoids Rayleigh scattering features. All resulting excitation-emission matrices (EEMs) were water blank-corrected, Raman-normalized, and corrected for instrument biases using correction factors supplied by the manufacturer. Thus, fluorescence spectra from the Fluoromax 2 and Fluoromax 3 are directly comparable. Fluorescence spectra were also corrected for inner filter effects (IFEs) [*Spencer et al.*, 2007a], Lignin phenols measured included three vanillyl (V) phenols (vanillin, acetovanillone, vanillic acid), three syringyl (S) phenols (syringaldehyde, acetosyringone, syringic acid), and two cinnamyl (C) phenols (*p*-coumaric acid, ferulic acid) (Figure 2).

2.2. Modeling and Application

[8] Partial Least Squares (PLS) models were constructed using The Unscrambler software (Camo) to predict lignin concentrations and compositions in the Delta from fluorescence EEMs (models generated with 10 nm emission steps performed identically to 1 nm steps, we chose the former to decrease run times, thus using ~350 emissions per EEM). All measured excitation/emission pairs were used in the model with no prior parameterization in order to take full advantage of all information in the EEM. PLS is similar to principle component analysis (PCA) in that a large number of variables are reduced to orthogonal principle components (PCs). However, whereas PCA is used to find overall structure between all variables in a single matrix, PLS iteratively computes PCs for both input (fluorescence) and output (lignin) variable matrices while utilizing linear regression to maximize the covariance between the two matrices. Thus, PLS is a more appropriate choice when building predictive models. Further information about PLS modeling can be found in the tutorial by *Geladi and Kowalski* [1986]. All

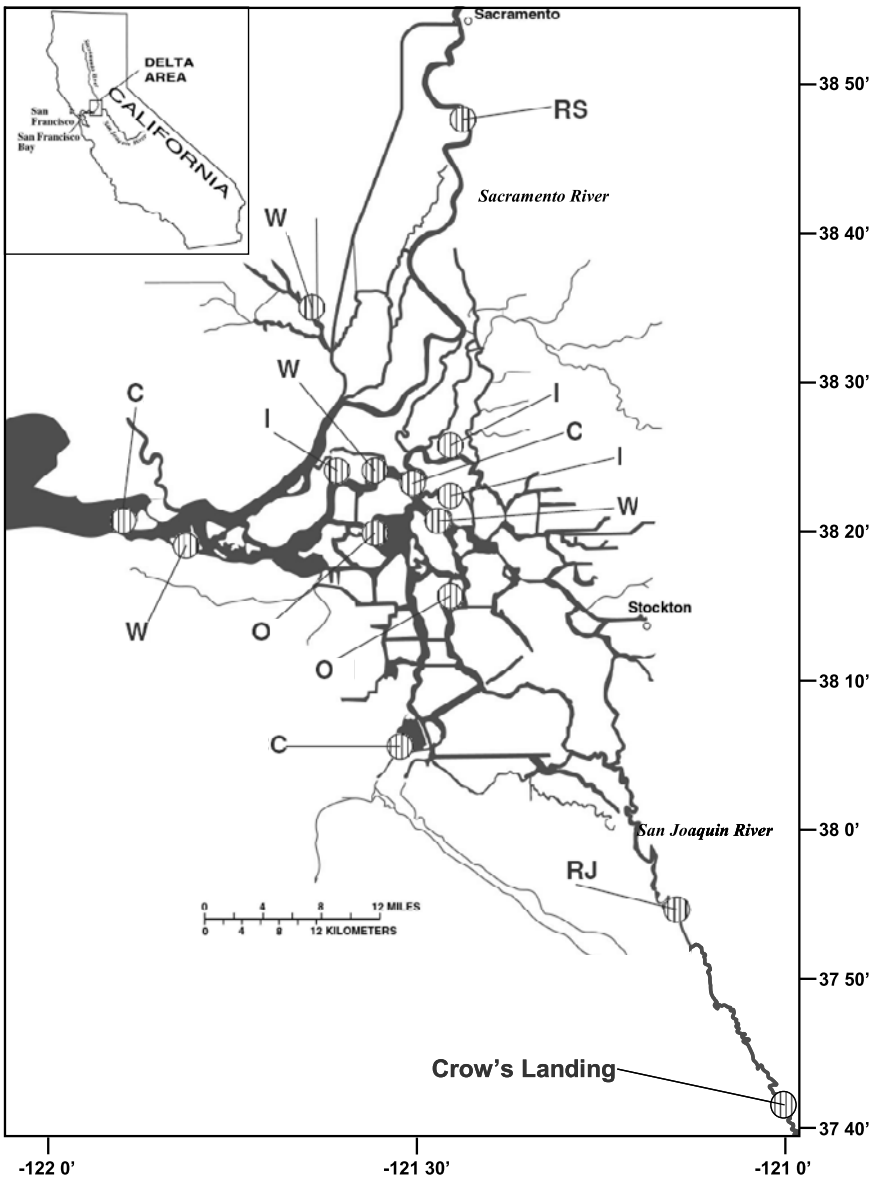

Figure 1. Map of Sacramento–San Joaquin Delta indicating sampling station, identified according to habitat or land use type: C, channel; W, wetland; RS, Sacramento River; RJ, San Joaquin River; O, open water; I, island drain. Our model application site, Crow's Landing, is identified by full name. Island drain measurements were not included in the model due to a different lignin source (peat soils) along with concentrations 5–10X greater than the median of the other samples.

fluorescence and lignin data were normalized by one standard deviation for consistency; however, we observed that normalization of the lignin data had no effect on the resulting model. All variables were mean centered for the actual models, although this data pretreatment was eliminated for ease of interpretation when investigating the regions of the EEMs with the most predictive power. Mean centering was observed to improve model fits by ~5%. All models underwent "leave-one-out" cross validation, in which each sample is sequentially left out and then predicted from a model calibrated with the remaining samples. A generated model consists of regression coefficients in a matrix identical to the EEMs along with a baseline value for a specific lignin parameter, the latter which is similar to an intercept in linear regression. When mean centering is applied, this baseline value is approximately equal to the average of all the sample lignin values. Predictions are made by multiplying the regression coefficients pairwise with any identically sized EEM, summing all the products, then combining that sum with the baseline value. Residuals between predicted and observed are used to evaluate model performance. Fifteen PCs were calculated for each model; however, only the first 3–7 were significant. We generated model relationships for lignin concentrations (μg L^{-1}), carbon-normalized lignin yields (mg 100 mg OC^{-1}), ratios of cinnamyl to vanillyl phenols, C:V, ratios of syringyl to vanillyl phenols, S:V, and ratios of vanillic acid to vanillin, (Ad:Al)v. Subsequently, these model relationships were used to predict lignin concentrations and compositions from a diurnal study in the San Joaquin River at Crow's Landing [*Spencer et al.*, 2007b].

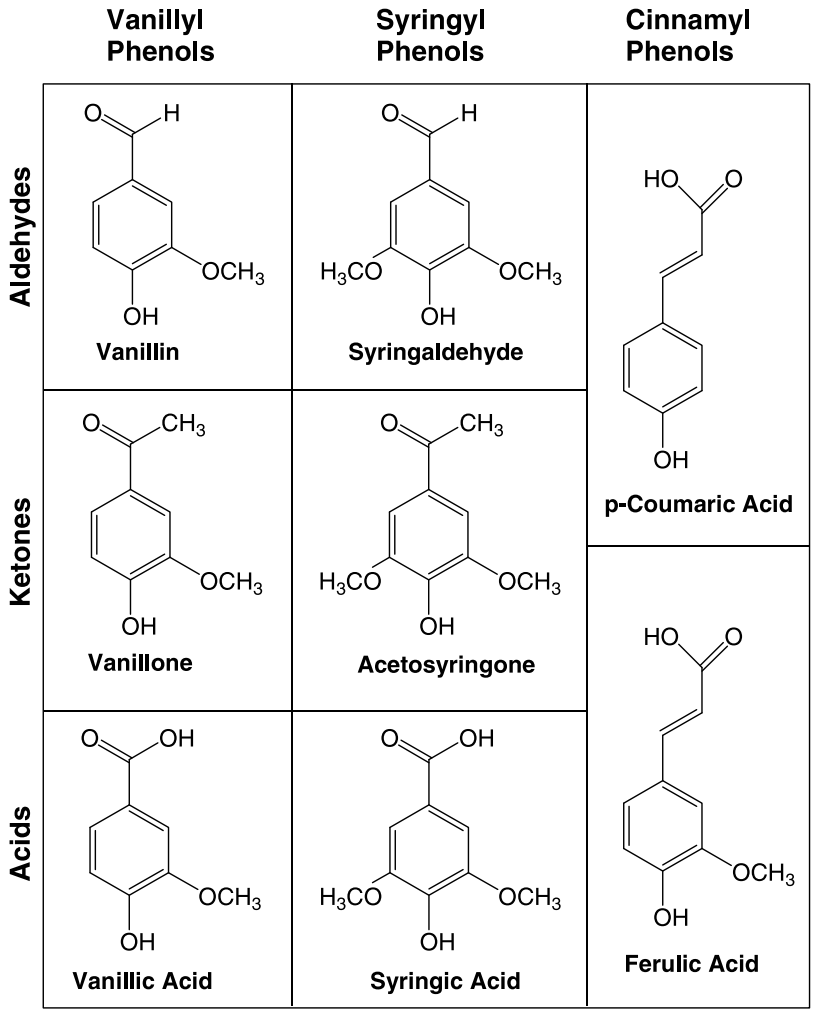

Figure 2. Structures of lignin monomers released by alkaline CuO oxidation.

[9] Given the growing interest in parallel factor analysis (PARAFAC) as a data reduction technique for fluorescence EEMs, we also generated 3-, 5-, and 7-component PARAFAC models for Delta fluorescence samples to test whether any of the components would correlate with our lignin parameters. PARAFAC was conducted with MATLAB using the N way toolbox version 3.10 [*Andersson and Bro*, 2000] on 58 samples. All models were verified by comparing the measured and modeled EEMs for all samples and observing the intensity of the residual EEMs. Split half analysis was not chosen for these models because of the limited number of samples.

3. Results

[10] All lignin data for this modeling exercise has been previously published in the work of *Eckard et al.* [2007]. To briefly summarize (Table 1), lignin concentrations in the modeled samples varied from 3.0 to 36.8 μg L^{-1}, with the highest average concentrations (18.5 μg L^{-1}) in the wetlands and the lowest average concentrations (7.7 μg L^{-1}) in the rivers. Carbon-normalized yields, Λ_8, ranged from 0.13 mg 100 mg OC^{-1} to 0.846 mg 100 mg OC^{-1}, with the highest average values (0.48) again in the wetlands while the lowest average values (0.29) were measured in the open water sites. Ratios of syringyl to vanilly phenols, S:V, were highest on average (1.28) in the open water sites and lowest (0.94) in the river samples while ranging from 0.59 to 1.53. Ratios of cinnamyl to vanillyl phenols, C:V, varied from 0.13 to 0.85 with similar distributions to S:V–highest average values (0.79) in the open water sites and lowest average values (0.29) in the rivers. Finally, the vanillic acid:vanillin ratios, (Ad:Al)v, modeled from this data set varied nearly fourfold from 0.83 to 3.21, with highest average values in the open water (1.44) and river (1.42) sites and lowest average values in the wetland (1.11) and Delta channel sites (1.12).

[11] Fluorescence EEMs for the four sample types were quite similar (see Figure 3 for an example of each). Primary distinguishing features include a hypsochromic shift in the humic-like peak A (see *Stedmon et al.* [2003] for a summary of EEM peaks) in the river sample (centered at 250/450 in the river versus 250/470 in the remaining samples), and a generally overall higher fluorescence intensity in the wetland sample (max intensity of ~1 Raman Unit, R.U., versus 0.8 for the open water and 0.6 for the river and channel).

[12] Our modeling of the five principal lignin parameters within the Delta data set exhibited varying degrees of success. Predicted lignin concentrations for the leave-one-out validation samples were well correlated ($p < 0.0001$,

Table 1. Summary of Sacramento River/San Joaquin River Delta Lignin Phenol Data From *Eckard et al.* [2007] Used to Build Partial Least Squares Models and Ancillary Optical Data

	DOC (mg L^{-1})	Λ_8 (mg 100 mg OC^{-1})	Σ_8 (μg L^{-1})	C:V	S:V	[Ad:Al]v	[Ad:Al]s	a_{254}[a] (m^{-1})
Maximum	6.0	0.85	36.8	1.17	1.54	3.21	3.02	50.7
Minimum	1.3	0.13	3.0	0.29	0.59	0.81	0.66	8.3
Average	3.0	0.37	12.0	0.57	1.06	1.18	1.06	23.8

[a]Absorption coefficient at 254 nm is included as an indicator of the relative color of the water samples used in this study. Wetland samples dominated the high a_{254} values.

student *t* test) with observed measurements (r^2 = 0.91, Figure 4a), as were carbon-normalized yields (r^2 = 0.79; Figure 4b), S:V (r^2 = 0.74, Figure 4c), and C:V (r^2 = 0.50, Figure 4d). However, the model was not successful at predicting (Ad:Al)v (r^2 = 0.09, plot not shown). In addition to correlation coefficients, the number of significant PCs gives an indication of the overall robustness of individual models, in general, fewer PCs indicate more robust models. The number of significant PCs for the four successful models ranged from three (C:V) to seven (Λ_8) (Figure 4). There is currently no universally accepted approach for estimating prediction error with PLS modeling [*Nadler and Coifman*, 2005; *Zhang and Garcia-Munoz*, 2009], but the most intuitive relate the residuals (the difference between predicted and observed values) to either the total number of samples (standard error of prediction, SEP) or the average of the observed values (average relative error, ARE). SEP (square root of the sum of squares of the residuals divided by total number of samples) provides reasonable estimates of the absolute error near the average of the range of values, but overestimates the absolute error in the lower range and underestimates the upper range. ARE (average residual expressed as a percent of average observed value) likely underestimates the true error, but expressed as a percentage is a more consistent representation of prediction error throughout the entire range of values, thus we report AREs for this study, which ranged from 9% for the S:V model to 21% for

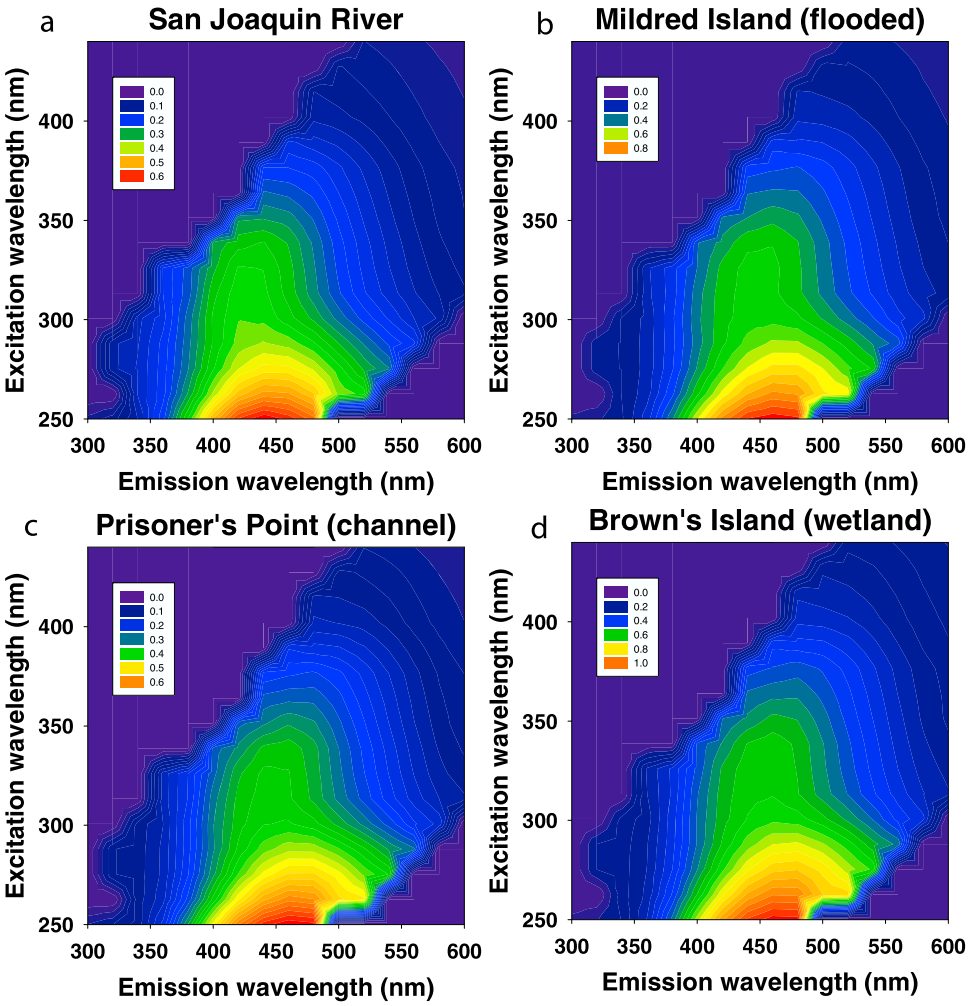

Figure 3. Representative fluorescence EEMs of the four sample types included in this study: (a) San Joaquin River, (b) Mildred Island (flooded lake-like), (c) Prisoner's Point (Delta channels), and (d) Brown's Island (wetland).

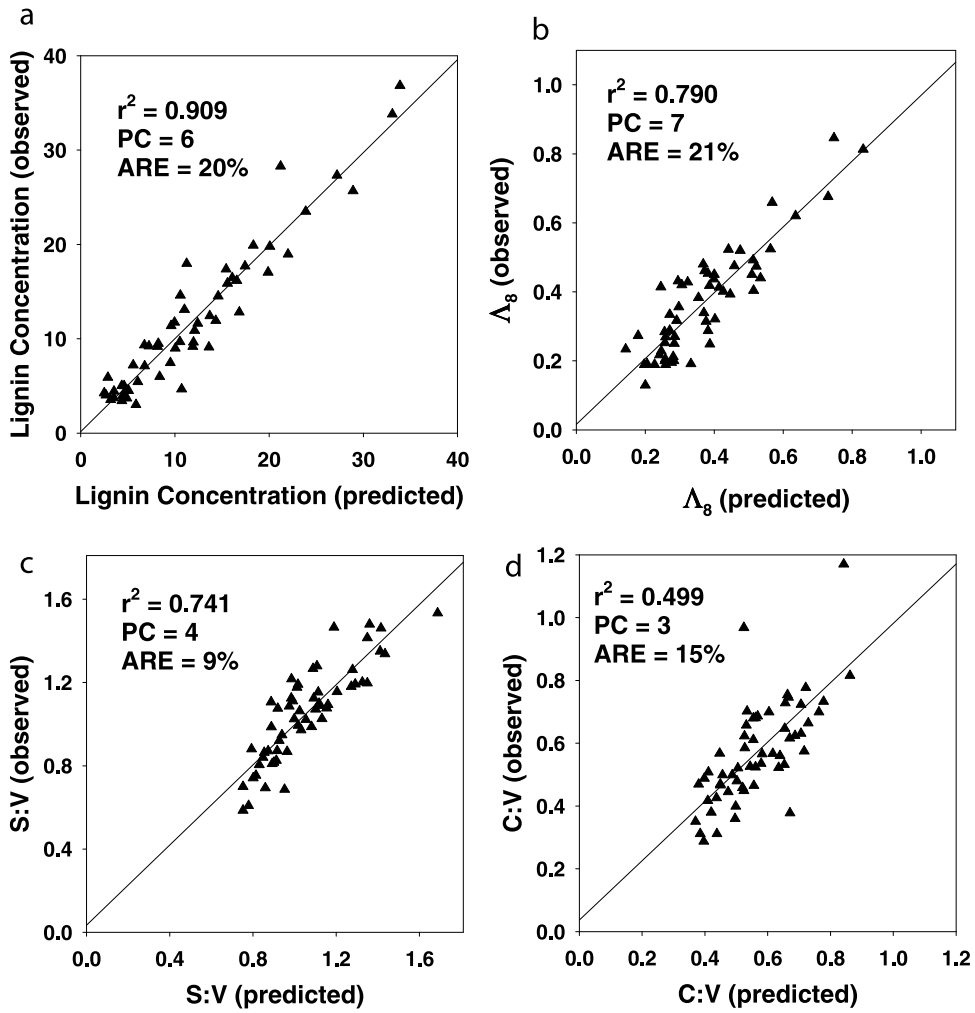

Figure 4. Observed measurements versus model-predicted values in the Delta for (a) lignin concentrations (μg L^{-1}), (b) carbon-normalized lignin yields (mg 100 mg OC^{-1}), (c) syringyl to vanillyl ratios, S:V, and (d) cinnamyl to vanillyl ratios, C:V.

Λ_8 (Figure 4). Expressing SEP as a percentage of the average gives values ~2% higher than ARE.

[13] PARAFAC analysis generally produced varied fluorescence components dominated by the humic Peak A or Peak C (Table 2). There was no significant correlation between any of the components and lignin concentrations or compositions, with r^2 ranging from 0 to 0.14.

[14] Predicted lignin concentrations at Crow's Landing over the 48 h sampling period ranged from 5.9 to 10.7 μg L^{-1} and exhibited a diurnal pattern with maximum concentrations midmorning and minimum concentrations midafternoon (Figure 5a). Predicted carbon-normalized lignin yields followed an identical pattern and ranged from 0.04 to 0.21 mg 100 mg OC^{-1} (Figure 5b). In contrast, predicted S:V and C:V ratios demonstrated little variability over the course of the 48 h, with S:V averaging 1.25 and C:V averaging 0.81 (Figures 5c and 5d).

4. Discussion

[15] The efficacy of lignin as a tracer for DOM cycling hinges upon the diverse and specific information that can be

Table 2. Results From PARAFAC Modeling of Delta Fluorescence

	Component						
	1	2	3	4	5	6	7
Maximum peak intensity	320/420	260/425	260/495				
Peak[a]	C	A	A				
Maximum peak intensity	270/440	330/425	260/490	290/355	260/520		
Peak[a]	A	C	A	C	unknown		
Maximum peak intensity	260/445	330/430	260/520	280/500	260/490	290/390	290/410
Peak[a]	A	C	unknown	unknown	A	M	M

[a]Peak designations are according to classifications summarized by *Stedmon et al.* [2003].

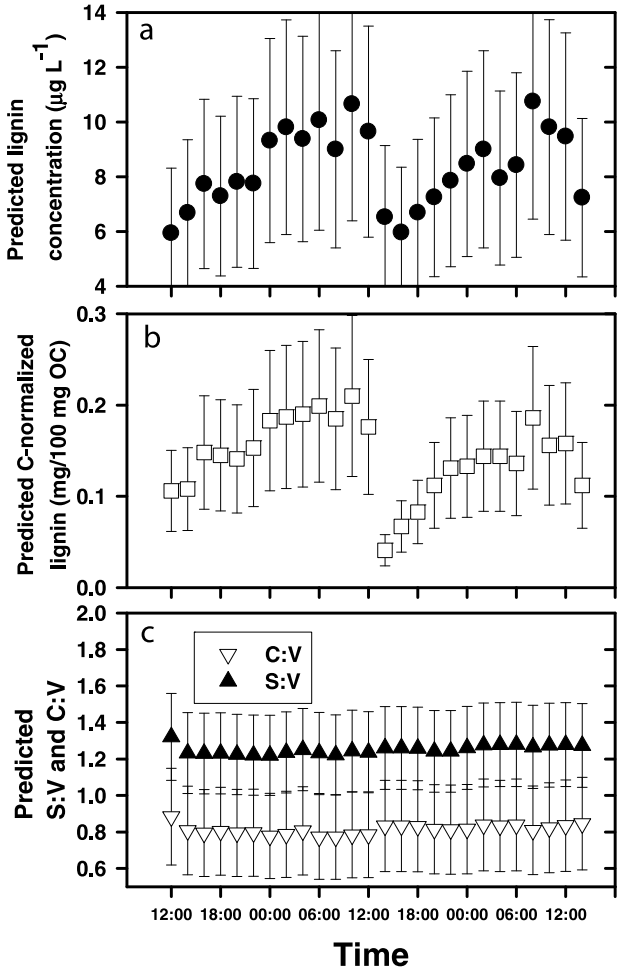

Figure 5. Model-predicted values at Crow's Landing, California, on the San Joaquin River for (a) lignin concentrations (μg L^{-1}), (b) carbon-normalized lignin yields (mg 100 mg OC^{-1}), and (c) syringyl to vanillyl ratios, S:V, and cinnamyl to vanillyl ratios, C:V. Samples were collected every 2 h from noon on 28 July 2005 until noon on 30 July 2005. Error bars represent two times the average relative error (ARE) as reported in Figure 4 applied to each sample value, which approximates a 95% confidence interval.

obtained from a single analysis. The proportion of DOM that is vascular-plant derived or terrigenous in origin can be quantified using lignin concentrations and carbon-normalized yields [*Hernes and Benner*, 2002; *Hernes et al.*, 2007; *Opsahl and Benner*, 1997]. Lignin ratios, S:V and C:V, can be used to distinguish between woody and nonwoody angiosperm and gymnosperm tissues [*Hedges and Mann*, 1979]. At the landscape scale, S:V and C:V signatures have been used to quantify landscape-scale sources of DOM in the Delta [*Eckard et al.*, 2007]. In addition, S:V and (Ad:Al)v in DOM are both sensitive to photooxidation and thus can capture diagenetic history [*Hernes and Benner*, 2003; *Opsahl and Benner*, 1998; *Spencer et al.*, 2009b]. Our results suggest great potential for using fluorescence EEMs as a proxy for lignin concentration, carbon-normalized yields, S:V ratios, and to a lesser extent C:V ratios in systems for which fluorescence EEMs have been well calibrated with lignin measurements. Although our generated models resulted in highly significant correlations between predicted and observed values for these four parameters, the diversity of samples and sources used in this modeling exercise (i.e., wetland dissolved lignin from monocots is different from riverine lignin derived from forests and agricultural lands) likely contributed to a lower overall model fit to the data. In other words, fluorescence proxies for lignin are likely to be even more successful when the overall source of lignin to a set of samples is more homogeneous, as has been demonstrated for modeling lignin concentrations with absorbance data in the Yukon River system [*Spencer et al.*, 2009a].

[16] Our modeling results are consistent with theoretical relationships between fluorescence and chemical functionality. While fluorescence intensity at any given wavelength pairing for excitation-emission depends on the concentration of fluorescing functional groups, the pattern observed in EEMs relates to functional group diversity. Thus, concentrations of fluorescing compounds like lignin should be strongly related to fluorescence intensity, while modeling success for compositional ratios will be dependent on relative differences in chemical structure between the components of the ratios. In the case of carbon-normalized yields, the difference in functionality related to fluorescence between polyphenolic lignin and bulk DOM is significant, hence a strong predictive capability. In contrast, (Ad:Al)v is derived from similar propylphenol structures within the polyphenolic structure that likely differ by only a single oxygen on the propyl side chain (e.g., Figure 2). Syringyl phenols differ from vanillyl phenols by a single methoxy group on the aromatic ring, while differences between cinnamyl and vanillyl phenols occur on both the aromatic ring (number of methoxy groups) and the propyl side chain (presence of carbon-carbon double bonds in the cinnamyl phenols) (Figure 2). The strong predictive capability for C:V and S:V in comparison to (Ad:Al)v would suggest that methoxyl substitution patterns on the aromatic ring and double bonding in the propyl side chain lead to more unique fluorescence patterns than oxygenation of the side chain. One caveat to this interpretation is that lignin is not solely responsible for all fluorescence properties. However, it seems apparent from the strong modeled relationships for both concentration and composition that the lignin contribution to the overall fluorescence signature is unique, significant, and quantifiable.

[17] In order to explore excitation-emission (Ex/Em) pairs with the greatest predictive capacity, we generated 3D contour plots in the same 3D space generally used for plotting EEMs by multiplying model regression coefficient matrices by a sample EEM – in essence performing the first step in using a PLS model for generating a prediction. For this exercise, we chose the sample that demonstrated the tightest overall fits to all four successfully modeled lignin parameters, a San Joaquin River sample (Station RJ in Figure 1, sample EEM presented in Figure 3a) collected in mid July. We also reran the models without centering the data which simplifies data interpretation. Regression models generated from centered data include both positive and negative coefficients since the starting value for predicting a lignin parameter is approximately the average value for the whole sample set, then regression coefficients have the effect of adjusting that value up or down. Regression models generated from uncentered data primarily "buildup" from zero.

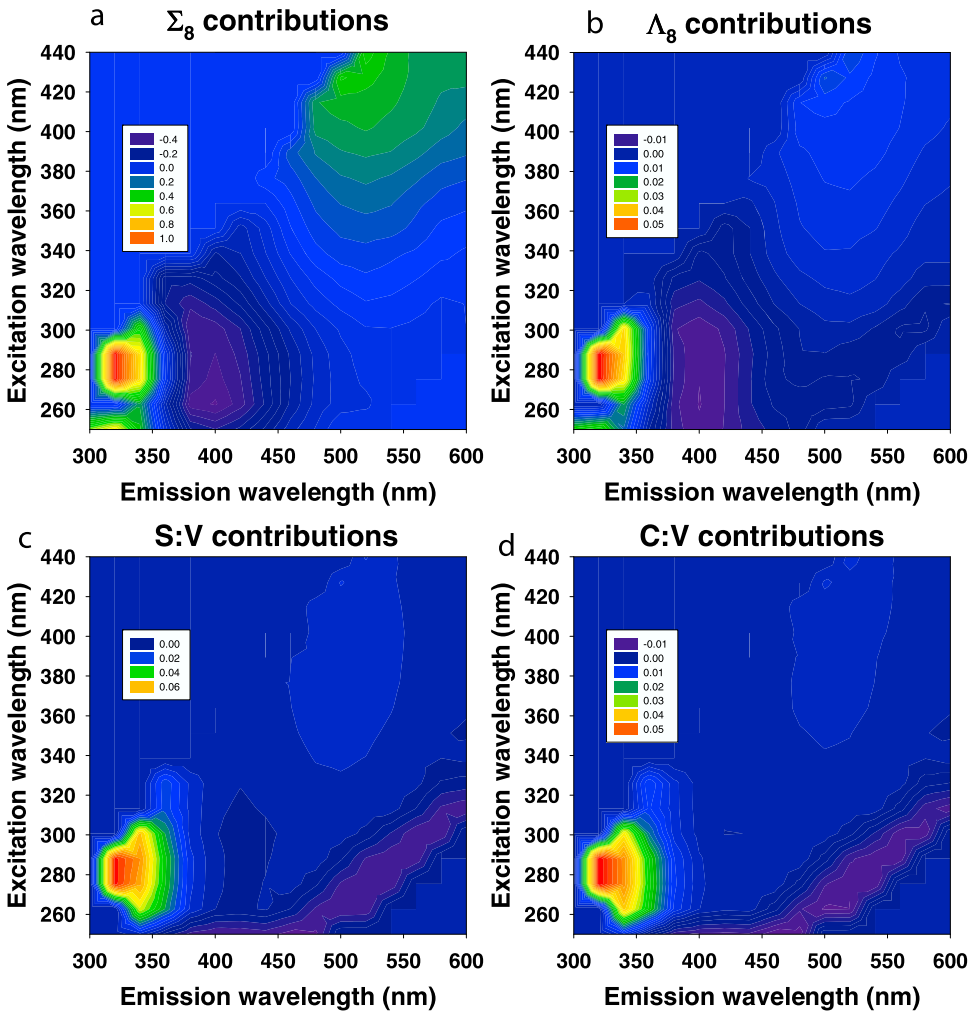

Figure 6. An example of EEM regions that contained the most significant predictive capability for lignin parameters, generated by multiplying model regression coefficient matrices times a fluorescence EEM for the best modeled sample in the Delta data set, a San Joaquin River sample collected in July 2000. Important regions of this EEM in terms of predictive capability for lignin parameters are highlighted for (a) total lignin concentration for the eight phenols, Σ_8 in μg L^{-1}, (b) carbon-normalized yields for the eight phenols, Λ_8 in mg 100 mg OC^{-1}, (c) ratios of syringyl to vanillyl phenols, S:V, and (d) ratios of cinnamyl to vanillyl phenols, C:V. The scale of each plot is dependent on the scale of the modeled lignin parameter and not directly comparable to a fluorescence EEM.

[18] Lignin is often considered a component of humic substances, yet the region of the EEMs with greatest predictive capability did not fall within the traditionally defined humic-like regions as summarized by *Stedmon et al.* [2003] (Ex = 260 nm/Em = 400–460 nm, and Ex = 320–360 nm/Em = 420–460 nm), but rather within the region generally attributed to aromatic amino acids like tryptophan or tyrosine (Figure 6). Structurally, the propylphenol monomers that make up lignin are similar to tyrosine and tryptophan, and another class of polyphenols, tannin, have been shown to fluoresce in the same region of the EEM [*Maie et al.*, 2007]. It is noteworthy that neither the protein-like regions that contribute the most toward predicting lignin parameters, nor the high Ex/Em region that contributes toward lignin concentration (Figure 6a) contain prominent features in the overall EEM (Figure 3). Thus, one important conclusion from this exercise is that the entire EEM contains valuable information and not just the prominent features that are typically identified and interpreted. This is reinforced by the PARAFAC results, which isolated the prominent A and C peaks, neither of which have significant predictive power for lignin in this system.

[19] The ability to predict lignin concentrations and compositions using EEMs has the potential to greatly extend research capabilities in terrestrial freshwater systems as well as coastal zones with large riverine input. In the Delta study, for instance, lignin compositions were used to demonstrate significant changes in DOM sources over the course of one hydrologic season as well as the strong influence of localized landscape features on DOM composition [*Eckard et al.*, 2007]. However, the limitations of sampling and analytical throughput permitted only six snapshots. With the possibility of field autosamplers and benchtop analyses of fluorescence or even in situ monitoring of fluorescence, future molecular-level studies in the Delta could be greatly expanded with proper molecular-level calibration beyond six snapshots to

Table 3. Comparison of Predicted Lignin Parameters on the San Joaquin River at Crow's Landing to Measured Values at Vernalis[a]

Station	Date	Σ_8 (μg L^{-1})	Λ_8 (mg 100 mg OC^{-1})	S:V	C:V
Vernalis	9 Dec 1999	6.0	0.25	1.18	0.75
Vernalis	19 Mar 2000	11.9	0.39	0.87	0.31
Vernalis	16 Jul 2000	7.1	0.27	0.92	0.53
Vernalis	15 Oct 2000	9.6	0.32	1.11	0.97
Vernalis	4 Feb 2001	7.5	0.20	1.22	0.53
Vernalis	20 May 2001	5.4	0.25	0.81	0.40
Crow's Landing	28 Jul 2005 to 30 Jul 2005	5.9–10.7	0.04–0.21	1.22–1.32.	0.77–0.88

[a]Vernalis data from *Eckard et al.* [2007]. Abbreviations: Σ_8, sum of concentrations of eight lignin phenols; Λ_8, carbon-normalized yield of eight lignin phenols; S:V, ratio of syringyl phenols to vanillyl phenols; C:V, ratio of cinnamyl phenols to vanillyl phenols.

provide a much more detailed picture of vascular plant-derived DOM cycling.

[20] In addition to much greater sampling throughput, spectrophotometric properties of DOM can typically be measured with much greater sensitivity and reproducibility than discrete molecular measurements. Thus, well-calibrated fluorescence models for lignin or other biomarkers have the potential to reveal fine-scale changes that would be challenging from discrete biomarker measurements alone. For example, we applied our Delta-derived models to a 48-h diurnal study that was conducted during the summer on the San Joaquin River (Crow's Landing) approximately 20 miles upriver from the San Joaquin River site (Vernalis) included in the modeled Delta data set (Figure 1). Predicted lignin parameters at Crow's Landing compared favorably to those measured at Vernalis, with the Crow's Landing lignin concentrations and C:V falling within the range of the six measurements taken at Vernalis throughout the seasons (Table 3), while carbon-normalized yields were slightly lower and S:V slightly higher. Given the significant influence that local landscape features can exert on lignin concentrations and compositions [*Eckard et al.*, 2007], these predicted values are entirely reasonable. However, the accuracy of the predictions in this example is not as important as the relative trends that are revealed in Figure 5 since it is the trends that are indicative of short-term processing and not the absolute values of the lignin parameters.

[21] Apparent diurnal trends in lignin concentration and carbon-normalized yields represent an intriguing finding as to the nature of DOM cycling in riverine systems. Unchanging S:V and C:V values indicate a similar source for dissolved lignin throughout the 48 h sampling period. Total flow at Crow's Landing over the course of the 48 h changed by less than 1%, precluding any hydrologic pulses of lignin-rich waters. Other factors that could come into play include diurnal variability in amounts of lignin released from the riparian zone, photobleaching of lignin, or diurnal cycling of other compounds that fluoresce in the predictive regions highlighted in Figure 6. Clearly, the diurnal trends predicted for lignin at Crow's Landing must be reproduced with properly calibrated models or discrete lignin measurements before they can be accepted as real. Ultimately, however, the power of this example is that even without prior calibration of a system with actual lignin data, we can use our generated models as a screening tool in any system that has similar EEMs data to point us toward processes that merit further research with discrete biomarker analyses: in essence, a fluorescence-guidance system for our molecular toolkit.

[22] Continued progress in DOM cycling research will require both the specificity of biomarker measurements along with the high spatial and temporal resolution capability that spectrophotometric measurements offer. The vast richness of information contained at the molecular level in DOM remains a virtually untapped resource, largely because of the difficulty and expense of making enough measurements to adequately address issues of temporal and spatial scaling. Coupling fluorescence and biomarker analytical approaches with PLS or similar models provides a powerful new tool that significantly expands the scope of what is possible for future studies. Emerging in situ capabilities for fluorescence points toward real time monitoring of biomarkers such as lignin in the near future, while the more general development of biomarker proxies using optical measurements ultimately could lead to remote sensing capabilities: precisely the kind of significant advances needed to move the field of DOM research forward.

[23] **Acknowledgments.** We would like to gratefully acknowledge the California Bay Delta Authority Ecosystem Program and Drinking Water Program for their support (grant B-17). We thank Bryan Downing for assistance with EEM processing and PARAFAC analyses. We would also like to thank the Aqueous Organic Geochemistry group at UC Davis, the R. Benner research group at University of South Carolina, the AE, and an anonymous reviewer for helpful comments on this manuscript.

References

Aiken, G. R., et al. (1992), Isolation of hydrophilic organic acids from water using nonionic macroporous resins, *Org. Geochem.*, *18*(4), 567–573, doi:10.1016/0146-6380(92)90119-I.

Andersson, C. A., and R. Bro (2000), The N-way toolbox for MATLAB, *Chemom. Intell. Lab. Syst.*, *52*, 1–4, doi:10.1016/S0169-7439(00)00071-X. (Toolbox available at http://www.models.life.ku.dk/source/nwaytoolbox)

Bergamaschi, B. A., et al. (2005), An assessment of optical properties of dissolved organic material as quantitative source indicators in the Santa Ana River Basin, southern California, *U.S. Geol. Surv. Sci. Invest. Rep., 5152*.

Blough, N. V., and R. Del Vecchio (2002), Chromophoric DOM in the coastal environment, in *Biogeochemistry of Marine Dissolved Organic Matter*, edited by D. A. Hansell and C. A. Carlson, pp. 509–546, Academic, San Diego, Calif.

Coble, P. G. (1996), Characterization of marine and terrestrial DOM in seawater using excitation emission matrix spectroscopy, *Mar. Chem.*, *51*(4), 325–346, doi:10.1016/0304-4203(95)00062-3.

Coble, P. G. (2007), Marine optical biogeochemistry: The chemistry of ocean color, *Chem. Rev.*, *107*(2), 402–418, doi:10.1021/cr050350+.

Dalzell, B. J., et al. (2007), The role of hydrology in annual organic carbon loads and terrestrial organic matter export from a midwestern agricultural watershed, *Geochim. Cosmochim. Acta*, *71*(6), 1448–1462, doi:10.1016/j.gca.2006.12.009.

Eckard, R. S., et al. (2007), Landscape scale controls on the vascular plant component of dissolved organic carbon across a freshwater delta, *Geochim. Cosmochim. Acta*, *71*(24), 5968–5984, doi:10.1016/j.gca.2007.09.027.

Geladi, P., and B. R. Kowalski (1986), Partial least squares regression: A tutorial, *Anal. Chim. Acta*, *185*, 1–17, doi:10.1016/0003-2670(86)80028-9.

Hedges, J. I., and J. R. Ertel (1982), Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products, *Anal. Chem.*, *54*(2), 174–178, doi:10.1021/ac00239a007.

Hedges, J. I., and D. C. Mann (1979), Characterization of plant tissues by their lignin oxidation products, *Geochim. Cosmochim. Acta*, *43*(11), 1803–1807, doi:10.1016/0016-7037(79)90028-0.

Hedges, J. I., et al. (1997), What happens to terrestrial organic matter in the ocean?, *Org. Geochem.*, *27*, 5–6.

Hernes, P. J., and R. Benner (2002), Transport and diagenesis of dissolved and particulate terrigenous organic matter in the North Pacific Ocean, *Deep Sea Res., Part I*, *49*(12), 2119–2132, doi:10.1016/S0967-0637(02)00128-0.

Hernes, P. J., and R. Benner (2003), Photochemical and microbial degradation of dissolved lignin phenols: Implications for the fate of terrigenous dissolved organic matter in marine environments, *J. Geophys. Res.*, *108*(C9), 3291, doi:10.1029/2002JC001421.

Hernes, P. J., and R. Benner (2006), Terrigenous organic matter sources and reactivity in the North Atlantic Ocean and a comparison to the Arctic and Pacific oceans, *Mar. Chem.*, *100*(1–2), 66–79, doi:10.1016/j.marchem.2005.11.003.

Hernes, P. J., et al. (2007), Fractionation of lignin during leaching and sorption and implications for organic matter "freshness," *Geophys. Res. Lett.*, *34*, L17401, doi:10.1029/2007GL031017.

Hernes, P. J., et al. (2008), The role of hydrologic regimes on dissolved organic carbon composition in an agricultural watershed, *Geochim. Cosmochim. Acta*, *72*(21), 5266–5277, doi:10.1016/j.gca.2008.07.031.

Kratzer, C. R., et al. (2004), Sources and transport of nutrients, organic carbon, and chlorophyll-a in the San Joaquin River upstream of Vernalis, California, during summer and fall, 2000 and 2001, *U.S. Geol. Surv. Water Resour. Invest. Rep., 03-4127*, 124 pp.

Kraus, T. E. C., et al. (2008), Assessing the contribution of wetlands and subsided islands to dissolved organic matter and disinfection byproduct precursors in the Sacramento-San Joaquin River Delta: A geochemical approach, *Org. Geochem.*, *39*(9), 1302–1318, doi:10.1016/j.orggeochem.2008.05.012.

Maie, N., et al. (2007), Composition of a protein-like fluorophore of dissolved organic matter in coastal wetland and estuarine ecosystems, *Water Res.*, *41*(3), 563–570, doi:10.1016/j.watres.2006.11.006.

McKnight, D. M., et al. (2001), Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity, *Limnol. Oceanogr.*, *46*(1), 38–48.

Nadler, B., and R. R. Coifman (2005), The prediction error in CLS and PLS: The importance of feature selection prior to multivariate calibration, *J. Chemometr.*, *19*(2), 107–118, doi:10.1002/cem.915.

Opsahl, S., and R. Benner (1997), Distribution and cycling of terrigenous dissolved organic matter in the ocean, *Nature*, *386*(6624), 480–482, doi:10.1038/386480a0.

Opsahl, S., and R. Benner (1998), Photochemical reactivity of dissolved lignin in river and ocean waters, *Limnol. Oceanogr.*, *43*(6), 1297–1304.

Spencer, R. G. M., et al. (2007a), Discriminatory classification of natural and anthropogenic waters in two UK estuaries, *Sci. Total Environ.*, *373*(1), 305–323, doi:10.1016/j.scitotenv.2006.10.052.

Spencer, R. G. M., et al. (2007b), Diurnal variability in riverine dissolved organic matter composition determined by in situ optical measurement in the San Joaquin River (California, USA), *Hydrol. Process.*, *21*(23), 3181–3189, doi:10.1002/hyp.6887.

Spencer, R. G. M., et al. (2008), Seasonal and spatial variability in dissolved organic matter quantity and composition from the Yukon River basin, Alaska, *Global Biogeochem. Cycles*, *22*, GB4002, doi:10.1029/2008GB003231.

Spencer, R. G. M., et al. (2009a), Utilizing chromophoric dissolved organic matter measurements to derive export and reactivity of dissolved organic carbon exported to the Arctic Ocean: A case study of the Yukon River, Alaska, *Geophys. Res. Lett.*, *36*, L06401, doi:10.1029/2008GL036831.

Spencer, R. G. M., et al. (2009b), Photochemical degradation of dissolved organic matter and dissolved lignin phenols from the Congo River, *J. Geophys. Res.*, *114*, G03010, doi:10.1029/2009JG000968.

Stedmon, C. A., et al. (2003), Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy, *Mar. Chem.*, *82*(3–4), 239–254, doi:10.1016/S0304-4203(03)00072-0.

Stepanauskas, R., et al. (2005), Sources, bioavailability, and photoreactivity of dissolved organic carbon in the Sacramento-San Joaquin River Delta, *Biogeochemistry*, *74*(2), 131–149, doi:10.1007/s10533-004-3361-2.

Zhang, L., and S. Garcia-Munoz (2009), A comparison of different methods to estimate prediction uncertainty using Partial Least Squares (PLS): A practitioner's perspective, *Chemom. Intell. Lab. Syst.*, *97*, 152–158, doi:10.1016/j.chemolab.2009.03.007.

B. A. Bergamaschi, U.S. Geological Survey, Placer Hall, M/S 6129, 6000 J Street, Sacramento, CA 95819, USA.

R. S. Eckard, P. J. Hernes, and R. G. M. Spencer, Department of Land, Air, and Water Resources, University of California, Davis, CA 95616, USA. (pjhernes@ucdavis.edu)