

# Controls of dissolved organic matter quality: evidence from a large-scale boreal lake survey

DOLLY N. KOTHAWALA<sup>1</sup>, COLIN A. STEDMON<sup>2</sup>, ROGER A. MÜLLER<sup>1</sup>,  
GESA A. WEYHENMEYER<sup>1</sup>, STEPHAN J. KÖHLER<sup>3</sup> and LARS J. TRANVIK<sup>1</sup>

<sup>1</sup>Evolutionary Biology Centre, Department of Ecology and Genetics/Limnology, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden, <sup>2</sup>National Institute of Aquatic Resources, Technical University of Denmark, Kavalergården 6, 2920 Charlottenlund, Denmark, <sup>3</sup>Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Box 7050 Lennart Hjelmns väg 9, 750 07 Uppsala, Sweden

## Abstract

Inland waters transport large amounts of dissolved organic matter (DOM) from terrestrial environments to the oceans, but DOM also reacts en route, with substantial water column losses by mineralization and sedimentation. For DOM transformations along the aquatic continuum, lakes play an important role as they retain waters in the landscape allowing for more time to alter DOM. We know DOM losses are significant at the global scale, yet little is known about how the reactivity of DOM varies across landscapes and climates. DOM reactivity is inherently linked to its chemical composition. We used fluorescence spectroscopy to explore DOM quality from 560 lakes distributed across Sweden and encompassed a wide climatic gradient typical of the boreal ecozone. Six fluorescence components were identified using parallel factor analysis (PARAFAC). The intensity and relative abundance of these components were analyzed in relation to lake chemistry, catchment, and climate characteristics. Land cover, particularly the percentage of water in the catchment, was a primary factor explaining variability in PARAFAC components. Likewise, lake water retention time influenced DOM quality. These results suggest that processes occurring in upstream water bodies, in addition to the lake itself, have a dominant influence on DOM quality. PARAFAC components with longer emission wavelengths, or red-shifted components, were most reactive. In contrast, protein-like components were most persistent within lakes. Generalized characteristics of PARAFAC components based on emission wavelength could ease future interpretation of fluorescence spectra. An important secondary influence on DOM quality was mean annual temperature, which ranged between  $-6.2$  and  $+7.5$  °C. These results suggest that DOM reactivity depends more heavily on the duration of time taken to pass through the landscape, rather than temperature. Projected increases in runoff in the boreal region may force lake DOM toward a higher overall amount and proportion of humic-like substances.

**Keywords:** boreal lakes, climate, dissolved organic matter, fluorescence spectroscopy, PARAFAC components, precipitation, water retention times

Received 16 May 2013; revised version received 14 November 2013 and accepted 19 November 2013

## Introduction

Dissolved organic matter (DOM) has a central role in biogeochemical processes within lakes and has an active role in the global carbon cycle (Cole *et al.*, 2007; Battin *et al.*, 2009; Tranvik *et al.*, 2009). Lakes receive and process DOM generated from catchment areas substantially greater than the area of the lake itself (Prairie, 2008). The mineralization of organic matter stored within soils is inhibited by numerous stabilization processes (Kögel-Knabner *et al.*, 2008), however as organic matter is mobilized from catchment soils to the aquatic environment, it becomes far more reactive (Kalbitz *et al.*, 2005) with the potential to become an

important source of CO<sub>2</sub> (Prairie *et al.*, 2002; Sobek *et al.*, 2005). The molecular composition of lake DOM is highly variable and dynamic, reflecting the source, amount of time processed within lakes, and biological reactivity (Berggren *et al.*, 2007; Cory *et al.*, 2007; Jaffe *et al.*, 2008). While lake DOM inputs are largely of terrestrial origin, inputs of DOM from algae and macrophytes can also be highly relevant energy sources for the microbial loop (Yamashita & Tanoue, 2003; Lapierre & Frenette, 2009; Guillemette & Del Giorgio, 2011; Karlsson *et al.*, 2012; Catalán *et al.*, 2013). Thus, DOM can be utilized and transformed by numerous in-lake processes including microbial respiration, photolytic degradation, and flocculation (Bertilsson & Tranvik, 2000; Cory *et al.*, 2007; Von Wachenfeldt & Tranvik, 2008; Koehler *et al.*, 2012). Variability in the hydrological connectivity of different water sources may also have a regulating

Correspondence: Dolly N. Kothawala, tel. +46 18 471 2701, fax. +46 18 53 11 34, e-mail: dolly.kothawala@ebc.uu.se

influence on the composition of DOM (Stedmon & Markager, 2003; Massicotte & Frenette, 2011).

Currently, there is widespread awareness that concentrations of DOM exported to aquatic ecosystems have increased in parts of Northern Europe and North America (De Wit *et al.*, 2007; Monteith *et al.*, 2007; Couture *et al.*, 2012). Associated changes to the molecular composition of DOM with increased concentrations remain unclear as monitoring programs have traditionally ignored or included limited measures of DOM quality. Consequently, the task of identifying mechanisms for increased DOM fluxes has been difficult (Tranvik & Jansson, 2002; Roulet & Moore, 2006). Previous links to declining sulfate deposition have relied on correlative analysis (Monteith *et al.*, 2007), yet, recent experimental work has provided support that increasing dissolved organic carbon (DOC) export from soils is associated with decreasing soil acidity (Evans *et al.*, 2012). It is expected that catchment soils will continue to recover from acidification in the upcoming decade; however, regions, such as the boreal ecozone anticipating increasing precipitation levels, may also observe continued inputs of additional DOM fluxes due to greater runoff (Hongve *et al.*, 2004; Eimers *et al.*, 2008; Löfgren & Zetterberg, 2011).

Increasing DOM fluxes are particularly problematic for water utilities because there is a higher cost associated with removing DOM for drinking water purposes. Untreated DOM can contribute to poor taste and odor problems, increase the potential to form harmful disinfection by-products and contribute to biological regrowth within the distribution system (Matilainen *et al.*, 2011). Thus, there is a growing demand to identify key DOM components found in surface waters that are most difficult to treat, as well as those that are easily removed during water treatment (Bierzoza *et al.*, 2009; Murphy *et al.*, 2011).

The concentration of lake water DOM is influenced by a multitude of drivers varying in scale from regional climate influences, to catchment-specific variables, and lake water chemistry. Lake DOM concentration has been linked to mean annual temperature (Weyhenmeyer & Karlsson, 2009; Laudon *et al.*, 2012), water discharge (Eimers *et al.*, 2008), and hydrological connectivity between key landscape features such as wetlands and streams (Schiff *et al.*, 1997; Köhler *et al.*, 2008; Laudon *et al.*, 2011). Increases in precipitation and discharge are anticipated in northern boreal regions (Kundzewicz *et al.*, 2007), which would lead to shorter lake water retention times. For this reason, it is particularly relevant to understand how altered lake water retention times could influence the quality of DOM. At the catchment scale, DOM concentration has been strongly linked to land cover, particularly wetland area

and the percentage of other water bodies in the catchment (Kortelainen, 1993; Dillon & Molot, 1997), water retention times (Hanson *et al.*, 2011), differences in topographical gradients (Creed *et al.*, 2003), catchment to lake area, landscape positioning, slope, and catchment soil characteristics (Trumbmore *et al.*, 1992; Kalbitz *et al.*, 2000). Accordingly, a multitude of factors may influence the overall concentration of DOM spatially and temporally.

Studies dedicated to examining how landscape and climate controls the quality of DOM are also emerging. Land use can influence DOM quality. For example, more bioavailable DOM has been found in streams draining agricultural (Wilson & Xenopoulos, 2009; Williams *et al.*, 2010) and urbanized areas (Baker & Inverarity, 2004). DOM draining forested landscapes has been identified as more bioavailable than DOM draining wetland-dominated catchments (Berggren *et al.*, 2007). A recent detailed hydrological study of the St. Lawrence river identified the spatial connectivity of mixing water bodies as an important regulator of DOM composition (Massicotte & Frenette, 2011). In addition, studies on arctic rivers have found that hydrological transport, as well as photochemical (Cory *et al.*, 2007), and biological processes (Mann *et al.*, 2012), control the chemical composition of DOM. Changes to the composition of DOM can have societally relevant consequences. For instance, shifts in DOM from terrestrial to algal-derived composition within a drinking water reservoir has been linked to disinfection by-product formation potential (Kraus *et al.*, 2011). A detailed understanding of how different DOM pools react over time can provide unique insight into how the biogeochemical function of DOM is also dynamic (Miller *et al.*, 2009). During transport from terrestrial sources to the sea, DOM has been described as undergoing a shift in composition analogous to weathering (Smith & Benner, 2005) and typically becomes less colored during passage through the landscape (Weyhenmeyer *et al.*, 2012).

Fluorescence spectroscopy is a relatively easy and cost-efficient method to characterize DOM in natural waters, whereby the fluorescence properties reflect differences in the composition of organic matter (Hudson *et al.*, 2007). Fluorescence spectroscopy generates valuable insight into DOM, the source and level of processing and provides great promise for use in routine monitoring programs as well as for drinking water purposes, due to its selectivity and sensitivity (Baker & Inverarity, 2004; Jaffe *et al.*, 2008; Henderson *et al.*, 2009; Fellman *et al.*, 2010). An especially useful advance has been to decompose the fluorescence spectra into underlying fluorescent components with multivariate data analysis, such as parallel factor analysis, (PARAFAC) (Bro, 1997; Stedmon & Bro, 2008) and

subsequently identify patterns with these underlying fluorescence components (Stedmon & Markager, 2005; Walker *et al.*, 2009).

The aim of this study was to link the fluorescence characteristics of lake water DOM to lake chemistry, catchment, and climate characteristics across a wide range of land-use and climate gradients spanning the full geographical area of Sweden. By using multivariate statistical approaches, we aim to decipher the relative importance of climate, catchment, and lake water chemistry in driving DOM quality. Our final goal was also to identify unique PARAFAC components that could be useful proxies for DOM reactivity, by recognizing patterns between components most readily lost within lakes and those found to be most persistent within lakes. These findings are useful for predicting the effects of changing climate patterns and land-use on lake DOM quality and to assess its treatability for drinking water purposes.

## Materials and methods

### Study sites

Lakes were sampled within the annual Swedish lake water-monitoring program conducted by the Swedish University of Agricultural Sciences (SLU). Of the approximately 100 000 lakes in Sweden larger than 1 hectare, more than 1000 lakes are sampled annually. In 2009 and 2010, every third lake visited by the national monitoring program was analyzed for DOM fluorescence and absorbance characteristics. A total of 560 samples were measured, spanning across the 13° latitudinal gradient of Sweden (Fig. 1). All sampling was performed by helicopter, whereby lake water was collected at a depth of ca. 1 m near the center of the lake during the autumn when lakes are well-mixed, starting from northern to southern Sweden ([webstar.vatten.slu.se/db.html](http://webstar.vatten.slu.se/db.html)). We did not monitor fluxes of water or DOM quality from sources of lake inputs (e.g., streams, groundwater, and precipitation), and we assumed water and DOM within lakes were well mixed. Each lake was sampled once, and lakes were not necessarily hydrologically connected to each other. Sampling occurred between September 6th and November 11, 2009 and between September 26th and November 25, 2010. Lake water samples were transported to the laboratory within 2 days of collection, where they were stored at 4 °C in the dark. All analyses including water chemistry and spectroscopic analyses were conducted within days, and no longer than 2 weeks from collection.

### Climate and geographical variables

Latitude is negatively correlated with the wide range of mean annual temperatures (MAT) included in this study, spanning from -6.1 in the north, to 6.5 in the south. MAT was adjusted for altitude with a -0.6 °C correction per 100 m (Livingstone *et al.*, 1999). In addition to MAT, we used mean growing

degree days (GDD) corresponding to the number of days in the year above 0 °C. GDD ranged from 100 days in the north, to 210 days in the south. Mean annual precipitation (MAP) and GDD were derived from the Swedish Meteorological and Hydrological Institute (SMHI) ([www.SMHI.se](http://www.SMHI.se)) as a mean between the years 1961 and 1990. Estimates of the mean long-term lake discharge (Q) [ $\text{m}^3 \text{s}^{-1}$ ] from the reference period January 1, 1961 to December 31, 1990, and lake volumes (V) were available from SMHI. Where lake volumes were missing (73% of the lakes), we estimated lake volumes from lake area and maximum slope within a 50 m buffer (Sobek *et al.*, 2011). Lake water retention times (WRT) were calculated as  $V/Q$ . The altitude of lakes was provided by Lantmäteriet ([www.lantmateriet.se](http://www.lantmateriet.se)). Land cover data for lake catchments was generated from the CORINE database (Hagner *et al.*, 2005) developed by SLU, and expressed as percentage forest (conifer, deciduous, and mixed), wetland (open wetland, conifer/deciduous, and mixed forest on wetland), agriculture (landscaped green space, cultivated areas, pasture, and arable land), other (urban, poor vegetation, exploited land, and logging), and water (the surface area of the catchment covered by open water, not including the lake itself). Percentage agricultural coverage was converted to a ranked variable, because >50% of sites had <5% agriculture.

### Lake water chemistry

Routine chemical analysis was conducted by the Department of Aquatic Sciences and Assessment (Swedish Agricultural University, SLU) a nationally accredited laboratory using standard protocols for the analyses of lake water (SLU, <http://>

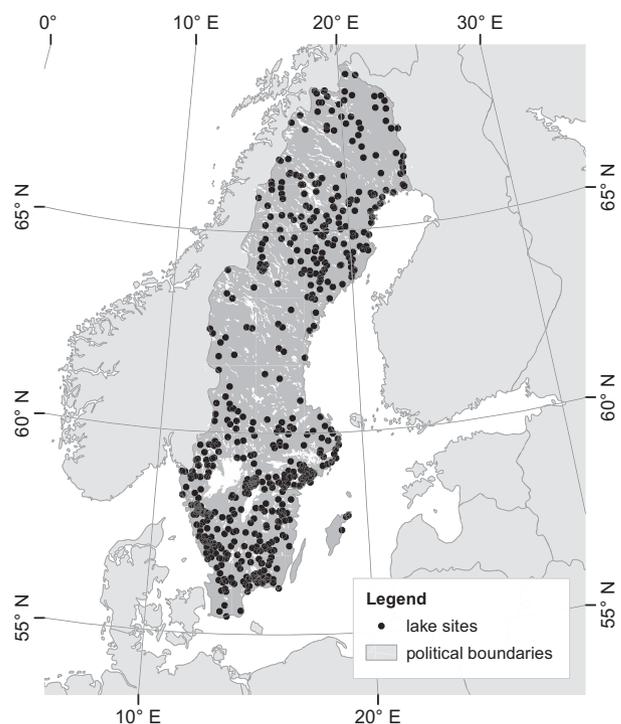


Fig. 1 Spatial distribution of 560 study lakes across Sweden.

www.slu.se/vatten-miljo). Chemical variables included total organic carbon (TOC), pH, conductivity, alkalinity, the sum of base cations (BC;  $\text{Ca}^{+2}$ ,  $\text{Mg}^{+2}$ ,  $\text{Na}^{+1}$ , and  $\text{K}^{+1}$ ), anions ( $\text{SO}_4^{-2}$ ,  $\text{F}^-$ , and  $\text{Cl}^-$ ), total nitrogen (TN), total inorganic nitrogen (TIN =  $\text{NO}_3^{-}\text{-N}$  +  $\text{NH}_4^{+}\text{-N}$ ), total organic nitrogen (TON=TIN-TIN), total phosphorus (TP), total heavy metals ( $\text{Cu}^{+2}$ ,  $\text{Zn}^{+2}$ ,  $\text{Cd}^{+2}$ ,  $\text{Pb}^{+2}$ ,  $\text{Ni}^{+2}$ , and  $\text{Co}^{+2}$ ), aluminum (Al), and iron (Fe). Lake water was not filtered before the analysis of TOC, TN or TIN; however, the vast majority of organic matter in boreal lakes is in the dissolved state (>90%) (Köhler *et al.*, 2002; Kortelainen *et al.*, 2006; Eimers *et al.*, 2008; Von Wachenfeldt & Tranvik, 2008). Thus, TOC and TON are used here as measures of dissolved organic carbon (DOC) and dissolved organic nitrogen (DON). All samples were filtered prior to spectral analysis using a 0.45- $\mu\text{m}$  filter.

#### UV-visible absorbance

The absorbance spectra of filtered DOM were measured across a wavelength range from 200 to 600 nm, at 1 nm intervals at a scan speed of 240 nm  $\text{min}^{-1}$  and slit width of 2 nm using a Lambda 40 UV-visible spectrophotometer (Perkin Elmer). Samples were measured in a 1 cm quartz cuvette, and Milli-q water was used as the blank. The shape of the absorption spectrum between 250 and 600 nm was characterized with an exponential model ( $a_\lambda = a_{\lambda_0} e^{S(\lambda_0 - \lambda)}$ );  $a_\lambda$  ( $\text{m}^{-1}$ ) is the absorbance coefficient at the wavelength  $\lambda$  (nm), and  $a_{\lambda_0}$  is the absorbance at a reference wavelength,  $\lambda_0$  (250 nm). We used the measured decadal absorbance at 254 (A254) and 420 nm (A420). The slope ( $S_{250-600}$ ) is a measure of how steeply the absorbance decreases with increasing wavelength. SUVA ( $\text{m}^2 \text{g}^{-1} \text{C}$ ) was calculated as the DOC normalized absorbance at 254 nm.

#### Fluorescence measurements

Excitation-emission matrices (EEM) from lake water samples were analyzed in sample mode (S) using a fluorescence spectrophotometer (SPEX FluoroMax-2, Horiba Jobin Yvon) and a 1 cm quartz cuvette. The sample signal was then corrected for the reference (R) lamp signal, to get S/R. The excitation wavelengths ( $\lambda_{\text{ex}}$ ) spanned from 250 to 445 nm, at increments of 5 nm, while emission wavelengths ( $\lambda_{\text{em}}$ ) ranged from 300 to 600 nm, at increments of 4 nm. Excitation and emission slit widths were set to 5 nm, and the integration time was 0.1 s. All EEMs were blank-subtracted using the EEM of Milli-q pure water run on the same day. Manufacturer supplied instrument correction factors were used to correct for instrument-specific biases (using  $S_c/R_c$ ). Spectra were corrected for inner filter effects using the absorbance-based approach (Lakowicz, 2006; Kothawala *et al.*, 2012), which was tested to be effective within the absorbance range used in this study (Kothawala *et al.*, 2013). Fluorescence intensity was calibrated to Raman units by dividing the intensity by the Raman area of pure water integrated at an excitation of 350 nm, and over an emission range of 380 to 420 nm. Any residual traces of a Rayleigh peak were removed before PARAFAC analysis by transforming negative values to zero.

#### PARAFAC analysis

Parallel factor analysis was applied to characterize the DOM fluorescence signal using MATLAB software (MATLAB® 7.7.0, The MathWorks, Natick, USA, 2008) and the DOMFluor toolbox (Stedmon & Bro, 2008). PARAFAC decomposes the EEMS dataset into a series of tri-linear components, which explains the variability in fluorescence characteristics of the dataset. Prior to PARAFAC modeling, data in the region of second-order scatter was deleted (emission wavelength >500 nm). First-order scattering from water was removed by inserting a band of zeros where excitation was greater than emission, starting at 298 nm. Steps to perform PARAFAC analysis included the identification of outliers, minimization of sum of squares residuals, identification of the appropriate number of PARAFAC components, random initialization with 10 iterations, and model validation using split-half analysis (Stedmon & Bro, 2008). Accordingly, six PARAFAC components were found to provide a robust description of DOM fluorescence within the dataset. Four of the six components had multiple excitation loadings, and all emission loadings had smooth uni-modal peaks, as expected for fluorophores (Andersen & Bro, 2003). In addition to assessing the intensity of each PARAFAC component ( $C_i$ ), the relative fluorescence intensity of each component was expressed as a percentage (%  $C_i$ ) of the sum of all six component intensities ( $\sum C_i$ ) using, %  $C_i = C_i / \sum C_i \times 100\%$ . To evaluate the robustness of the six-component PARAFAC model in describing fluorescence characteristics of lake DOM across the Swedish landscape, we calculated the accumulated sum of squares error across excitation and emission wavelengths for three distinct subregions of the country and compared the results. The three subregions were as follows: Northern Alpine, Northern Lowlands, and Southern Lowlands.

#### Testing the robustness of PARAFAC Components

A total of six fluorescing PARAFAC components were identified (Fig. S1; Tables S1 and S2). Visual inspection of residual EEM spectra ensured that unaccounted fluorescence was near the baseline for each sample, with residual fluorescence <10% of measured fluorescence intensity. In addition, we did not find any systematic pattern of residual fluorescence by visually examining the average root mean squared error summed across excitation ( $n = 40$ ) and emission ( $n = 76$ ), even when the dataset was divided into three distinct subregions of Sweden (Fig. S2). The analysis of residuals along with split-half validation (Stedmon & Bro, 2008) ensured the six-component PARAFAC model was a robust description of lake DOM fluorescence.

#### Principle component analysis (PCA) and partial least squares (PLS) analyses

Principle component analysis (PCA) was used initially as a diagnostic tool to examine key relationships and to identify strongly correlated variables, excluding optical measurements. Partial least squares (PLS) was further applied to predict PARAFAC components (Y response variables) from lake

water chemistry, catchment characteristics, and climate variables (X predictor variables listed in Table 1). The PLS approach was chosen due to its robustness to collinearity of multiple variables by identifying an optimal set of orthogonal principle components able to explain the maximum variability between X and Y variables. Samples with a disproportionately strong leverage on the overall model were identified as outliers (>95% confidence limit) using the Hotelling's  $T^2$  analysis. To ensure assumptions of the model (PCA and PLS) were met, data were normalized using log-transformations where necessary, with skewness targeted to a min/max <0.1 (Eriksson *et al.*, 2001). PLS models were constructed using all six PARAFAC components as the Y variables, expressed either as fluorescence intensities ( $C_i$ ), or as percentages (% $C_i$ ). Climate, catchment, and lake water chemistry were X variables.

To interpret loadings plots of PLS analysis, the distance and positioning of Y variables (PARAFAC components) relative to X variables was revealing of how well, or poorly, they related to each other. The greater the distance a variable (X or Y) was from the origin, the greater its overall influence. Variables situated close together on the PLS plot were positively correlated, while variables opposite to each other were negatively correlated. We performed internal cross-validation, to test the repeatability of the analysis, by removing a random subset of data (1/7th of the samples) to be used as the response dataset, while parallel models were run on the reduced calibration dataset. A comparison of predicted values from the calibration and response datasets allowed computation of the predictive residual sum of squares, expressed as a  $Q^2Y$ . Overall, PLS model performance was based on the cumulative goodness of fit ( $R^2Y$ , explained variation), and the cumulative goodness of prediction (or  $Q^2Y$ , predicted variation) for each model. The variable influence on projections (VIP scores) provided a means of interpreting the importance of X variables on the overall model. X-predictor variables with a VIP score  $\geq 1$  were considered highly influential, between 0.8 and 1.0 moderately influential, and <0.8 less influential predictors.

A permutation test was performed to estimate how much of the explained variability could be attributed to random chance. After creating a PLS model for each Y variable separately, the Y variable was randomly permuted, and re-run. The difference between y-intercepts of  $R^2$  and  $Q^2$  regressions passing through the unpermuted dataset and 100 permuted datasets provided an estimate of the background correlation due to chance for each Y variable. Small model background correlations indicated robust models (Eriksson *et al.*, 2001). All PCA and PLS analyses were carried out in SIMCA version 13.0.2 (Umetrics AB, Umeå, Sweden). Once key relationships between Y and X variables were identified, simple linear and exponential correlation analyses were performed independently of PLS analysis using JMP version 10.0 (SAS Institute, Cary, NC).

## Results

### Prediction of PARAFAC components

The relative positioning of Y and X variables differed substantially when Y variables were expressed as

PARAFAC component intensities (C1 to C6; referred to as the C-intensity model) (Fig. 2a), or as a percentage (%C1 to %C6; referred to as the C-percentage model) (Fig. 2b). Insight from both PLS models provided a more robust interpretation of potential factors driving DOM quality.

The C-intensity model resulted in four principle components, explaining a high amount of the overall variation ( $R^2Y = 0.61$ ), with good predictive ability ( $Q^2Y = 0.61$ ) (Table S3) and low-background correlations (Table S4). C3, C4, and C1 clustered near C2 situated near highly influential X predictor variables, including DOC concentration and covarying variables (A420, SUVA, Fe, Al, metals, HIX) (Fig. 2a). C6 was situated furthest away from highly influential X predictor variables and C5 between C6 and C2 (Fig. 2a). The strength of explained variation ( $R^2Y$ ) and the goodness of prediction ( $Q^2Y$ ) for C1 to C6 in the C-intensity model (Table S4) were consistent with their relationships to DOC concentration (Table S5), being strongest for C3 ( $R^2 = 0.81$ ,  $P < 0.0001$ ), and weakest for C6 ( $R^2 = 0.05$ ,  $P < 0.0001$ ) (Table S5). Consequently, the C-intensity model resulted in poor separation between Y response variables, with little explanation by X predictor variables, due to the strong influence of DOC concentration on PARAFAC component intensities.

For the C-percentage model, two broad-scale patterns were evident. Firstly, the relative positioning of both X and Y variables situated along the first-principle component (horizontal axis, 26%) was a function of DOC concentration (Fig. 2b) and correlated variables (Fe, Al, pH, SUVA). In addition, the %Water in the catchment, and DOC / DON were highly influential X variables (VIP > 1.0). Secondly, separation along the second-principle component (vertical axis, 10%) corresponded to MAT and correlated variables (e.g., GDD, TN and TP) (Fig. 2b). In addition, PARAFAC components with longer  $\lambda_{em}$  (red-shifted) were positioned to the right on the first-principle component and hence were most abundant in lakes with higher DOC concentration, while shorter  $\lambda_{em}$  components were associated with lower DOC concentration. Overall, the C-percentage model resulted in four principle components, with an explained variance of  $R^2Y = 0.48$  and predictive ability of  $Q^2Y = 0.45$  (Table S3), with low-background correlation (0.08 to 0.11) (Table S4), and the first- and second-principle components axes could be related to DOC concentration and MAT, respectively (Fig. S3).

In contrast to C6 in the C-intensity model, %C6 was the Y response variable with greatest amount of overall explained variance and predictability ( $R^2Y = 0.75$ ,  $Q^2Y = 0.73$ ), followed by %C3 ( $R^2Y = 0.64$ ,  $Q^2Y = 0.62$ ) (Table S4). Accordingly, %C6 and %C3 were situated at opposing ends of the loadings plot (x axis) (Fig. 2b),

**Table 1** Distribution of all variables included in partial least squares analysis including in-lake chemistry, catchment, climate and optical characteristics for boreal lakes included in the study ( $n = 560$ )

Variables	Abbreviation	Mean $\pm$ Std Dev.	Min	Max
<b>In-Lake Chemistry</b>				
DOC (mg l <sup>-1</sup> )	DOC	13.2 $\pm$ 6.08	2.4	32.4
pH	pH	6.5 $\pm$ 0.7	4.52	8.01
Total nitrogen (mg l <sup>-1</sup> )	TN	0.46 $\pm$ 0.20	0.12	1.20
Total phosphorous ( $\mu$ g l <sup>-1</sup> )	TP	13.1 $\pm$ 8.8	2.0	48.0
DOC : DON	DOC : DON	34 $\pm$ 9	15	60
Sum of Ca <sup>+2</sup> ,Mg <sup>+1</sup> ,K <sup>+1</sup> ,Na <sup>+1</sup> (meq l <sup>-1</sup> )	BC	0.47 $\pm$ 0.31	0.04	1.88
Dissolved inorganic nitrogen	DIN	0.07 $\pm$ 0.07	0.00	0.30
Dissolved organic nitrogen	DON	0.39 $\pm$ 0.16	0.09	1.00
Total Fe (mg l <sup>-1</sup> )	Fe	0.66 $\pm$ 0.60	0.01	3.30
Total Al (mg l <sup>-1</sup> )	Al	0.14 $\pm$ 0.12	0.00	0.75
Total Mn (mg l <sup>-1</sup> )	Mn	0.05 $\pm$ 0.05	0.00	0.31
Sum of SO <sub>4</sub> <sup>-2</sup> , Cl <sup>-1</sup> (meq l <sup>-1</sup> )	Anions	0.12 $\pm$ 0.15	0.01	0.68
Sum of Cu, Zn, Cd, Pb,Cr, Ni, Co, As, V ( $\mu$ g l <sup>-1</sup> )	Metals	5.33 $\pm$ 2.91	0.94	14.16
<b>Catchment Characteristics</b>				
Water retention time (years)	WRT	0.86 $\pm$ 0.89	0.00	3.92
Altitude (m)	Alt.	215 $\pm$ 161	7	848
Slope (°)	Slope	8 $\pm$ 4	1	29
Lake Area : Catchment Area	Lk : Catch	0.15 $\pm$ 0.07	0.06	0.52
Runoff (m <sup>3</sup> sec <sup>-1</sup> )	Runoff	1.195	0	55.546
Water (% of Catchment)	% Water	9.4 $\pm$ 7.1	0	48
Forest (% of Catchment)	% Forest	63 $\pm$ 15	0	98
Wetland (% of Catchment)	% Wetland	12 $\pm$ 12	0	87
Agriculture (% of Catchment)*	% Agr	5 $\pm$ 11	0	97
Other (% of Catchment)	% Other	12 $\pm$ 13	0	91
<b>Climate Variables</b>				
Mean annual temperature (°C)†	MAT	2.5 $\pm$ 3.5	-6.2	7.5
Mean annual precipitation (mm)	MAP	792 $\pm$ 150	450	1250
Mean growing degree days	GDD	175 $\pm$ 28	110	210
<b>Absorbance Indexes</b>				
Specific absorbance at 254 nm (L mg <sup>-1</sup> C m <sup>-1</sup> )	SUVA	3.64 $\pm$ 0.77	1.42	5.52
Absorbance at 420 nm (m <sup>-1</sup> )	A420	0.05 $\pm$ 0.03	0.00	0.14
Spectral slope (nm <sup>-1</sup> )	S250–600	0.013 $\pm$ 0.001	0.011	0.017

\*% Agr converted to a ranked variable for PLS analysis.

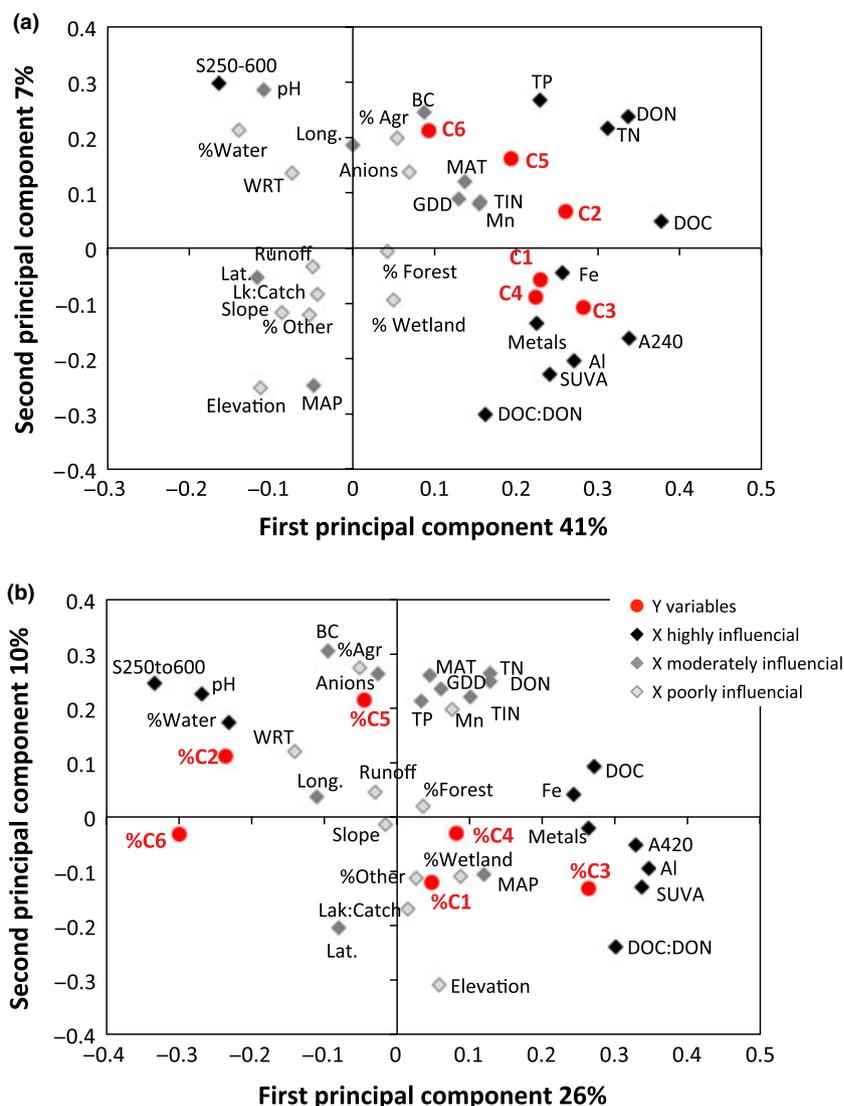
†MAT corrected for altitude.

and their relationships with X predictor variables were likewise opposite to each other. C3 is a humic-like component and C6 is a protein-like component. The explained variation and predictability of %C2 was also moderate ( $R^2Y = 0.52$ ,  $Q^2Y = 0.49$ ) and positively correlated with %C6 (Fig. 2b).

There was a group of highly influential X predictor variables clustered near %C3, with strong positive correlations including DOC concentration (Fig. 3e), and covariates (Fe, Al, metals, and SUVA) (Fig. 2b). DOC concentration and lake water pH were negatively correlated based on simple linear regressions ( $R^2 = 0.16$ ,  $P < 0.0001$ ), as evident from the PCA loadings plot, where they appear opposite to each other (Fig. S4). Accordingly, these two X variables are situated at opposing ends of the loadings plot with %C6 situated

near lake water pH. DOC / DON was strongly influential in predicting %C6 and %C3 (Fig. 2b), as confirmed in a scatter plot (Fig. 3g and h). Simple absorbance metrics, SUVA and  $S_{250-600}$ , were also highly influential X predictors. SUVA had a strong positive relationships with %C3, while  $S_{250-600}$  was negatively related to %C6 (Fig. 2b). The removal of absorbance metrics from PLS analysis reduced the  $R^2Y$  on the first-principle component axis by 0.01, yet we include them to simply show how %C3 and %C6 are paired with these commonly used metrics.

%Water was identified as a highly influential X variable in the C-percentage model. Lakes in this study with a high % Water in the surrounding catchment also tended to have a high WRT (WRT =  $27.9 \times$  %Water + 40.4,  $R^2 = 0.36$ ,  $P < 0.0001$ ), despite being



**Fig. 2** Partial least squares loadings plots predicting the variability of six PARAFAC components (Y variables) expressed as (a) fluorescence intensities, and (b) as a percentage, with several climate, geographical, in-lake chemistry and spectral characteristics as predictors (X variables).

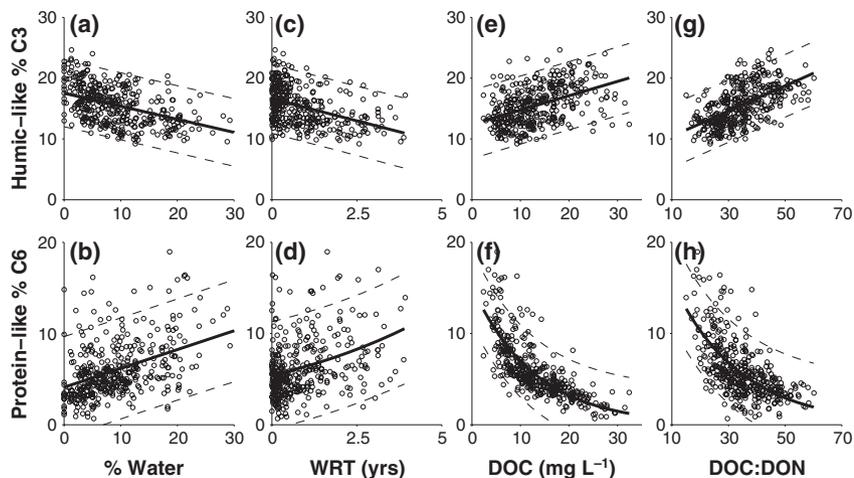
measures that are independent of each other, and thus, while weaker, there were significant relationships between WRT and both %C3 (negative) and %C6 (positive) (Fig. 3c, d). When we examined how the intensity of C3 and C6 varied directly with %Water, WRT, DOC, and DOC/DON (Fig. 4), we found a highly significant exponential decrease in humic-like C3 with greater %Water, and likewise with longer WRT (Fig. 4a and c). In contrast, with increasing %Water and longer WRT, the amount of C6 fluorescence remained stable (Fig. 4b and d). Regression statistics for relationships presented in Figs 3 and 4 are provided in Table 2.

Along the second-principle component axis, %C5 was found to cluster with several variables, particularly MAT, which is strongly correlated to GDD, ions (anions

and base cations), nutrients (TON, TIN, TN, TP), and % agriculture (Fig. 2b). While relationships derived along the second-principle component were not as strong as the first, there was a significant relationship between %C5 and MAT (%C5 = 0.32 MAT + 0.00,  $R^2 = 0.10$ ,  $P > 0.0001$ ).

## Discussion

Using a multivariate approach comparing 560 lakes spanning a geographical area of approximately 450 000 km<sup>2</sup> across Sweden, we were able to discriminate between the relative influence of lake water chemistry, catchment characteristics, and climatic factors, on the quality of lake DOM across a broad geographical scale.



**Fig. 3** Relationships between the relative abundance of two key PARAFAC components %C3 and %C6, and the percentage of water in the surrounding catchment (%Water) (a, b), lake water retention time (WRT) (c, d), dissolved organic carbon (DOC) (e, f), and DOC to dissolved organic nitrogen (DOC:DON) (g, h) (Regression statistics in Table 2).

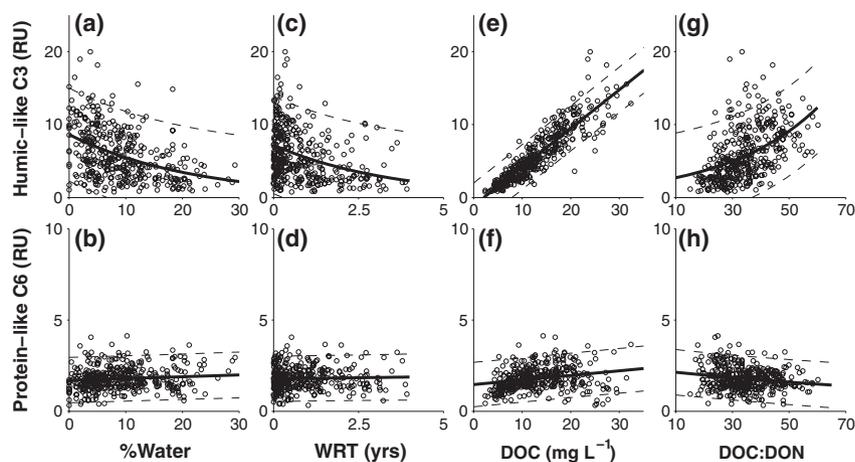
The most highly influential land cover variable predicting the composition of DOM was the percentage of water in the surrounding catchment (%Water), as revealed from both PLS analysis and direct regression analysis (Fig. 2b, Fig. 3a, b). %Water was defined as the relative areal coverage of open water in the catchment not including the lake itself. So, for instance, aside from small ponds in the catchment, headwater lakes generally had a negligible %Water ( $\approx 0\%$ ). The %Water has been previously identified as the main feature explaining variation in DOC concentration across several Finnish lake survey studies (Kortelainen & Mannio, 1988; Rantakari *et al.*, 2004; Mattsson *et al.*, 2005). As with the Finnish study, this study also finds lower DOC concentrations in lakes with a higher %Water, likely due to the greater possibility for DOM to be lost, produced, and transformed during in-lake processing in upstream water bodies (Kortelainen, 1993). Lakes with a higher %Water will have a lower areal transport of DOM derived from terrestrial sources. With increasing %Water and longer WRT, we found the quality of DOM shifted, with greater proportions of a protein-like component (%C6), and lower proportions of the terrestrially derived humic-like component (%C3) (Fig. 3). Based on component intensities, C3 was lost, but the amount of C6 remained stable regardless of %Water and WRT (Fig. 4). The hydrological retention time of lake water has been identified as an important factor influencing the loss of terrestrial DOM (Hanson *et al.*, 2011). The net loss of DOC and preferential loss of colored DOM (based on SUVA, Fig. 2b) in lakes with longer WRT has been documented previously and attributed to losses from microbial and photodegradation, as well as losses to flocculation, and new algal inputs from primary produc-

**Table 2** Regression statistics for relationships between the intensity (C3 and C6) and relative intensity of PARAFAC components (%C3 and %C6) with select variables (Figs 3 and 4).

	%Water	WRT	DOC	DOC : DON
%C3	$R^2 = 0.23$ (-0.2, 17.5) $P < 0.0001$	$R^2 = 0.17$ (-1.5, 16.6) $P < 0.0001$	$R^2 = 0.18$ (0.2, 12.6) $P < 0.0001$	$R^2 = 0.32$ (0.2, 8.4) $P < 0.0001$
%C6	$R^2 = 0.20^*$ (0.2, 4.2) $P < 0.0001$	$R^2 = 0.12$ (1.2, 5.1) $P < 0.0001$	$R^2 = 0.54$ (-0.38, 11.1) $P < 0.0001$	$R^2 = 0.35$ (-0.2, 13.5) $P < 0.0001$
C3	$R^2 = 0.19$ (-0.04, 8.6) $P = 0.01$	$R^2 = 0.12$ (-0.3, 7.2) $ns$	$R^2 = 0.81^*$ (0.5, -1.1) $P < 0.0001$	$R^2 = 0.24$ (0.03, 2.0) $P < 0.0001$
C6	$R^2 = 0.01$ (0.01, 1.7) $P = 0.01$	$R^2 = 0.00$ (0.01, 1.8) $ns$	$R^2 = 0.05$ (0.03, 1.46) $P < 0.0001$	$R^2 = 0.03$ (-0.01, 2.3) $P < 0.0001$

The adjusted coefficient of variation ( $R^2$ ), slope, and intercept ( $k$ ,  $b$ ) provided in brackets, and significance level is indicated with  $p$ -value;  $ns$  indicate nonsignificant relationships. All relationships are exponential ( $Y = b \cdot e^{k \cdot X}$ ) with the exception of linear relationships ( $Y = K \cdot X + b$ ) marked with\*.

tivity (Curtis & Schindler, 1997; Kraus *et al.*, 2011; Weyhenmeyer *et al.*, 2012; Köhler *et al.*, 2013). However, traditional calculations of lake WRT neglect to consider the retention time of water in upstream water bodies. In fact, a recent study revealed that the exponential relationship used to describe the loss of colored DOM with longer WRT using conventional measures almost doubled ( $R^2$  of 0.11 to 0.21) when the accumulated WRT of upstream inputs were included (Müller *et al.*, 2013). Hence, while traditional WRT estimates only consider water retention within the specific lake of interest,



**Fig. 4** Relationships between the intensity of PARAFAC components C3 and C6 in Raman units (RU), with the percentage of water in the surrounding catchment (%Water) (a, b), lake water retention time (WRT) (c, d), dissolved organic carbon (DOC) (e, f), and DOC to dissolved organic nitrogen (DOC : DON) (g, h) (Regression statistics in Table 2).

%Water is a better proxy for the total time DOM is exposed to in-lake processing within the catchment. This study confirms that WRT influences DOM quality, but further emphasizes the importance of considering the catchment as a whole when considering potential drivers of DOM quality.

The inverse patterns of %C3 and %C6 across gradients of %Water, WRT, and DOC was also reflected in the DOC/DON. This may be a result of N-rich compounds comprising the protein-like component C6 being selectively retained. DON can be retained efficiently in lakes with increasing WRT (Windolf *et al.*, 1996). Likewise, during passage through the landscape DON may have been retained and even recycled, despite continued net utilization of DOC. An arctic stream study also noticed that more processed DOM contained more DON, while terrestrial sources of DOM were generally poorer in DON (Cory *et al.*, 2007). The wide range of DOC/DON observed in this particular dataset (Table 1) is attributed to the inclusion of hyper-oligotrophic lakes common to northern Sweden, along with southern eutrophic lakes where historical N deposition was greater (Bergström *et al.*, 2005).

Secondary influence on DOM quality was the MAT and a group of associated geographical variables (Fig. 2b). As MAT lies along the secondary principle components axis, results from this study suggest that %Water and WRT are more influential to DOM quality than regional-scale MAT. Warmer temperatures, longer growing degree days, and a greater percentage of agriculture increased the relative abundance of C5. The positioning of C5 on the EEM is in the region of Peak A (Coble, 1996), a very common and highly prevalent region associated with humic-like fluorescence from terrestrial sources. Regional-scale associations between

MAT and DOM concentration were expected, because optimal conditions for DOM production in soils and export to inland waters are between 0 and +3 °C (Laudon *et al.*, 2012). MAT may have an indirect influence on DOM quality due to a wide range of potential factors including shifts in vegetation cover and litter quality, the balance between primary productivity and decomposition rates, length of the ice-free season, and even the length of exposure to solar radiation, only to mention a few.

The actual and relative intensity of two PARAFAC components, C3 (%C3) and C6 (%C6), provided the most insight into DOM quality. The positioning of C6 resembles that of the amino acid tryptophan, which indicates autochthonous algal and microbially derived proteins (Yamashita & Tanoue, 2003); however, C6 may also include some other polyphenolic compounds including tannins and lignins (Maie *et al.*, 2007; Hernes *et al.*, 2009). Controlled bottle incubation studies have found fluorescence in the protein-like region to be biologically reactive, both produced and degraded over time (Lønborg *et al.*, 2010; Guillemette & Del Giorgio, 2011). The stable levels of C6 (and increasing %C6) in lakes with greater %Water and longer WRT suggests that it may be more persistent component, constantly being renewed despite the loss of other components. The negative association of %C6 with A420 and SUVA, and positive association with  $S_{250-600}$  observed here, provides further evidence of being weakly colored by-products of microbial degradation, photo-bleaching (Galgani *et al.*, 2011), and production from algae.

The intensity and percentage of C3 was greatest in lakes with lower %Water and shorter WRT. This suggests that C3 was most susceptible to in-lake loss processes and may have been lost rapidly from the water

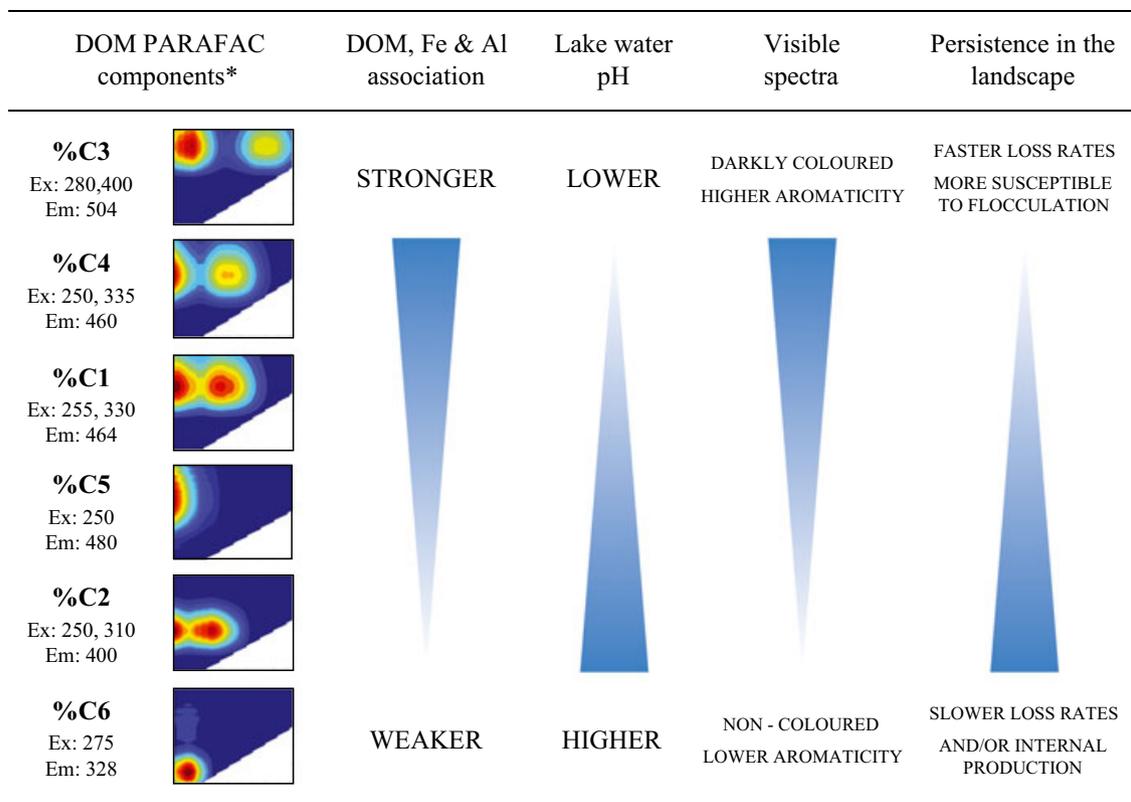
column by mineralization to CO<sub>2</sub>, flocculation, or transformation into other DOM by-products. A potentially important loss process for C3 may have been flocculation. A substantial proportion of DOM (8 to 22%), generally of terrestrial origin, can be lost from the water column solely from flocculation to lake sediments (Von Wachenfeldt *et al.*, 2008). The abundance of free cations (iron, aluminum, hydrogen, and metals) strongly associated with C3 increases the opportunity for ionic associations such as ligand exchange or cation bridging, reducing the solubility of DOM comprising C3, and improving its ability to naturally flocculate out of the water column (Tipping, 1981; Lofts & Tipping, 1998; Luider *et al.*, 2003). C3 fluorescence was situated at the longest emission wavelength of the humic-like components indicating molecules with higher molecular weight, containing more conjugated structures and being more hydrophobic (Wu *et al.*, 2003; Hunt & Ohno, 2007). C3 also had the strongest positive relationship with SUVA, an indicator of darker DOM with aromatic structures (Weishaar *et al.*, 2003) (Fig. 3). Other studies have likewise found that the most aromatic, highly colored fractions with large molecular weight DOM have shorter half-lives in aquatic systems (Weyhenmeyer *et al.*, 2012; Köhler *et al.*, 2013).

The potential susceptibility of C3 to flocculate out of lake water is of direct relevance to drinking water treatment, particularly where iron or aluminum is used as a flocculating agent. An extensive study examining changes during water treatment found that two protein-like PARAFAC components increased relative to humic-like components (Beggs & Summers, 2011). While the two humic-like components ( $\lambda_{\text{ex}} : \lambda_{\text{em}}$  of 240, 340:464 and 240, 300:420) identified by Beggs & Summers (2011) are not directly comparable with C3 in this study ( $\lambda_{\text{ex}} : \lambda_{\text{em}}$  of 280,400 : 504), the observation that protein-like components (such as C6) were more resistant to flocculation is consistent with predictions arising from our nationwide lake survey study. From an ecological and biological reactivity perspective, lakes with higher concentrations of DOC and higher %C6 may be important sources of labile and biologically reactive DOM. The overall structural characteristics of C6, being less aromatic than other components, suggest that it would be most resistant to flocculation in the natural environment, relative to other PARAFAC components. In a recent review, three reoccurring PARAFAC components were identified (Ishii & Boyer, 2012). Two of which were identified in this study, corresponding to C5 and C3. The review study also reports C3 to be of terrestrial origin, with strong sorption characteristics to gibbsite, goethite, and sediments (Ishii & Boyer, 2012). This supports our

findings of C3 being strongly associated with aluminum and iron, which likely included gibbsite and goethite, resulting in sorption and flocculation to sediments. Stedmon *et al.* (2005) also report a component similar to C3 to be most abundant in terrestrial systems, less abundant in estuaries and missing entirely in the open ocean, which also suggest preferential loss of C3 with progressive internal processing.

The relative intensity of PARAFAC components (%C's) was particularly useful to assess DOM quality independently of DOC concentration. The separation of %C's along the first-principle component axis (Fig. 2b) revealed that emission wavelength positioning could be a useful general indicator of physico-chemical properties (Fig. 5). Components at longer emission wavelengths, or red-shifted components, were more conjugated (higher SUVA) (Weishaar *et al.*, 2003) and more colored (higher A420). Red-shifted components were also most readily lost from lake water and thus were most reactive in the environment (Fig. 5). Consequently, emission wavelength positioning may be a simple way to characterize the complexity and reactivity of PARAFAC components (Fig. 5), particularly to distinguish between components that would otherwise be vaguely classified as humic-like (e.g., %C1 to %C5).

DOC concentration was a primary predictor of DOM quality based on the relative abundance of PARAFAC components. Numerous studies have found relationships between DOC concentration and DOM quality measures using fluorescence and absorbance metrics (Yamashita *et al.*, 2010; Erlandsson *et al.*, 2012; Inamdar *et al.*, 2012). However, these findings are not universal. For example, a study examining fluorescence characteristics across multiple ecosystems and across freshwater and estuarine environments did not find a relationship between DOM quality and concentration (Jaffe *et al.*, 2008). Several lake water chemistry variables including pH, iron, aluminum, and metals were strongly correlated with lake water DOC concentration (Fig. S3), appearing to have relationships with both %C3 and %C6 (Fig. 2b); however, their influence may be complex. Water chemistry may have influenced DOM fluorescence intensity due to static quenching by iron (Fe<sup>+3</sup>), metals, and H<sup>+</sup> (Miller, 1981; Lakowicz, 2006; Kelton *et al.*, 2007). Upon investigating the linearity of relationships between fluorescence and absorbance in an earlier study, it was clear that the protein-like region of EEMs (near C6) was particularly attenuated for acidic lakes (4.6 to 6.0) (Kothawala *et al.*, 2013). Fe<sup>+3</sup> is a particularly effective quencher at low pH (Cabaniss, 1992) but is only abundant at pH below 4.5 (Neubauer *et al.*, 2013), which contributed to the poor observed relationship for C6 (Kothawala *et al.*, 2013). However, it



**Fig. 5** Conceptual diagram of key physicochemical properties associated with the relative abundance of PARAFAC components contributing to dissolved organic matter (DOM) quality across boreal lakes, including their level of reactivity within the landscape. \*PARAFAC components are sorted from longer to shorter emission wavelengths (on vertical axis), detailed spectra are presented in Fig. S1.

remains unclear which regions of the EEM were affected most severely by  $\text{Fe}^{+3}$  quenching. As this study focused on broad-scale controls on PARAFAC components, only the most dominant overlying relationships were revealed here. Sample manipulation to adjust pH or convert  $\text{Fe}^{+3}$  to  $\text{Fe}^{+2}$  (Mcknight *et al.*, 2001; Doane & Horwath, 2010) was not possible prior to analysis in this case. Thereby, results are representative of whole DOM under ambient conditions.

Some fraction of the unexplained variation in our study may be attributed to hydrological mixing and dilution of PARAFAC components originating from different sources. For instance, some components may have been transported in part from topsoils (e.g., C3), or groundwater inputs (e.g., C6). The hydrological connectivity of mixing water bodies can play an important role in regulating DOM quality (Massicotte & Frenette, 2011). In this nationwide study, we are unable to monitor the relative quality of DOM from streams and groundwaters. Yet, our results show that annual lake runoff was one of the poorest influencing variables in this dataset (Fig. 2a and b). Studies including a network of hydrologically connected lakes over a smaller spatial scale may be necessary to fully evaluate the relative influence of catchment and climate characteristics along

with hydrological mixing and dilution on lake DOM quality.

This is the first comprehensive study of PARAFAC components representing boreal lakes across a wide geographical area. While previous studies have examined relationships between DOM quality and drivers unique to a particular site, this study identified the relative importance of lake chemistry, catchment, and climate drivers over an extremely large spatial scale. Consequently, these results should be interpreted bearing in mind that many variables are spatially correlated within this dataset (Fig. S3). For example, southern Sweden is warmer, with greater agricultural activity and more nutrients, while more mountainous lakes are located in the north. Thus, the relative influence of nutrients and altitude are inherently linked to the overlying influence of temperature in this dataset. In addition, most lakes included in the survey are relatively pristine oligotrophic lakes. Thus, there is some uncertainty as to how universal these results are to the entire boreal landscape. As this study is correlative, we are unable to specifically distinguish between the relative importance of in-lakes processes (e.g., primary production, flocculation, decomposition, photodegradation) and hydrological dilution, which may have collectively

contributed to observed shifts in DOM quality. Alternative drivers are almost certainly responsible for DOM quality at more localized scales; however, this study included a vast number of overlapping gradients, chemically, geographically and climatically, and we aimed to identify only the most dominant controls.

This study highlights the influence of land cover, particularly the percentage of water in the catchment and lake water retention time in regulating DOM quality with the mean annual temperature being a secondary influence. These results are particularly relevant in light of projected increases in precipitation and runoff for boreal ecosystems (Kundzewicz *et al.*, 2007). Warmer and wetter conditions may lead to increased DOM inputs from soils and peatlands (Neff & Hooper, 2002; Pastor *et al.*, 2003; Davidson & Janssens, 2006). Shifts in the timing of peak spring melt and frequency of rain on snow events could alter the source and quality of terrestrial DOM inputs (Sebestyen *et al.*, 2008; Casson *et al.*, 2012). Greater annual runoff could result in shorter lake water retention times, decreasing the opportunity for in-lake processing and primary production, shifting the quality of DOM to contain a greater proportion of humic-like C3, relative to protein-like C6. This shift in DOM composition is likely to increase the prevalence of humic and aromatic DOM structures with darker color that we reason here to be easier to treat with flocculation for drinking water purposes (Weishaar *et al.*, 2003). We expect the most aromatic and colored compounds (e.g., C3) to be most effective at absorbing solar radiation (Moran & Zepp, 1997; Vähatalo *et al.*, 2003) and may potentially be a primary source of CO<sub>2</sub> due to photomineralization. From an ecological perspective, a higher prevalence of humic lakes in the boreal ecozone could stimulate greater food web subsidies from energy mobilized by aquatic bacteria (Jansson *et al.*, 2007).

### Acknowledgements

We thank Jan Johansson for the analysis of fluorescence and absorbance scans, Christian Demandt and analytical chemists at the Swedish Agricultural University (SLU) for all data on lake water chemistry, Jens Fölster (SLU) for the organization of samples for spectral analysis. We thank the Swedish Meteorological and Hydrological Institute (SMHI) for hydrological and climate data, Jakob Nisell (SLU) for providing GIS, land cover, and CORINE database support, and Lantmäteriet for maps and geographical data. Valuable analytical support and statistical advice were provided from Sebastian Sobek, Anne Kellerman, Hannes Peter, Silke Langenheder, Jürg Brendan Logue, Jason Xing Ji and Cristian Gudasz. This study was funded by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning, FORMAS, and is part of the project 'The color of water – interplay with climate, and effects on drinking water supply.' Funding for G.A.W. was provided

by the Swedish Research Council (VR), Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) and the Nordic Centre of Excellence "CRAICC" supported by NordForsk. We thank four reviewers for their valuable contributions, which helped improve the manuscript.

### References

- Andersen CM, Bro R (2003) Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics*, **17**, 200–215.
- Baker A, Inverarity R (2004) Protein-like fluorescence intensity as a possible tool for determining river water quality. *Hydrological Processes*, **18**, 2927–2945.
- Battin TJ, Luysaert S, Kaplan LA, Aufdenkampe AK, Richter A, Tranvik LJ (2009) The boundless carbon cycle. *Nature Geoscience*, **2**, 598–600.
- Beggs KMH, Summers RS (2011) Character and chlorine reactivity of dissolved organic matter from a mountain pine beetle impacted watershed. *Environmental Science & Technology*, **45**, 5717–5724.
- Berggren M, Laudon H, Jansson M (2007) Landscape regulation of bacterial growth efficiency in boreal freshwaters. *Global Biogeochemical Cycles*, **21**, doi 10.1029/2006gb002844.
- Bergström AK, Blomqvist P, Jansson M (2005) Effects of atmospheric nitrogen deposition on nutrient limitation and phytoplankton biomass in unproductive Swedish lakes. *Limnology and Oceanography*, **50**, 987–994.
- Bertilsson S, Tranvik LJ (2000) Photochemical transformation of dissolved organic matter in lakes. *Limnology and Oceanography*, **45**, 753–762.
- Bierozza M, Baker A, Bridgeman J (2009) Relating freshwater organic matter fluorescence to organic carbon removal efficiency in drinking water treatment. *Science of the Total Environment*, **407**, 1765–1774.
- Bro R (1997) PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, **38**, 149–171.
- Cabaniss SE (1992) Synchronous fluorescence spectra of metal-fulvic acid complexes. *Environmental Science & Technology*, **26**, 1133–1139.
- Casson NJ, Eimers MC, Watmough SA (2012) Impact of winter warming on the timing of nutrient export from forested catchments. *Hydrological Processes*, **26**, 2546–2554.
- Catalán N, Obrador B, Felip M, Pretus JL (2013) Higher reactivity of allochthonous vs. autochthonous DOC sources in a shallow lake. *Aquatic Sciences*, **75**, 581–593.
- Coble PG (1996) Characterization of marine and terrestrial DOM in seawater using excitation emission matrix spectroscopy. *Marine Chemistry*, **51**, 325–346.
- Cole JJ, Prairie YT, Caraco NF *et al.* (2007) Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems*, **10**, 171–184.
- Cory RM, Mcknight DM, Chin YP, Miller P, Jaros CL (2007) Chemical characteristics of fulvic acids from Arctic surface waters: microbial contributions and photochemical transformations. *Journal of Geophysical Research-Biogeosciences*, **112**, G04s51.
- Couture S, Houle D, Gagnon C (2012) Increases of dissolved organic carbon in temperate and boreal lakes in Quebec, Canada. *Environmental Science and Pollution Research*, **19**, 361–371.
- Creed IF, Sanford SE, Beall FD, Molot LA, Dillon PJ (2003) Cryptic wetlands: integrating hidden wetlands in regression models of the export of dissolved organic carbon from forested landscapes. *Hydrological Processes*, **17**, 3629–3648.
- Curtis PJ, Schindler DW (1997) Hydrologic control of dissolved organic matter in low-order Precambrian Shield Lakes. *Biogeochemistry*, **36**, 125–138.
- Davidson E, Janssens I (2006) Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, **440**, 165–173.
- De Wit HA, Mulder J, Hindar A, Hole L (2007) Long-term increase in dissolved organic carbon in streamwaters in Norway is response to reduced acid deposition. *Environmental Science & Technology*, **41**, 7706–7713.
- Dillon PJ, Molot LA (1997) Effect of landscape form on export of dissolved organic carbon, iron, and phosphorus from forested stream catchments. *Water Resources Research*, **33**, 2591–2600.
- Doane TA, Horwath WR (2010) Eliminating interference from iron (III) for ultraviolet absorbance measurements of dissolved organic matter. *Chemosphere*, **78**, 1409–1415.
- Eimers MC, Buttle J, Watmough SA (2008) Influence of seasonal changes in runoff and extreme events on dissolved organic carbon trends in wetland- and upland-draining streams. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 796–808.

- Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (2001) *Multi- and Megavariate Data Analysis*. Principles and Applications, Umeå, Sweden, Umetrics Academy.
- Erlandsson M, Futter MN, Kothawala DN, Köhler SJ (2012) Variability in spectral absorbance metrics across boreal lake waters. *Journal of Environmental Monitoring*, **14**, 2643–2652.
- Evans CD, Jones TG, Burden A *et al.* (2012) Acidity controls on dissolved organic carbon mobility in organic soils. *Global Change Biology*, **18**, 3317–3331.
- Fellman JB, Hood E, Spencer RGM (2010) Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: a review. *Limnology and Oceanography*, **55**, 2452–2462.
- Galgani L, Tognazzi A, Rossi C *et al.* (2011) Assessing the optical changes in dissolved organic matter in humic lakes by spectral slope distributions. *Journal of Photochemistry and Photobiology B-Biology*, **102**, 132–139.
- Guillemette F, Del Giorgio PA (2011) Reconstructing the various facets of dissolved organic carbon bioavailability in freshwater ecosystems. *Limnology and Oceanography*, **56**, 734–748.
- Hagner O, Nilsson M, Reese H, Egberth M, Olsson H (eds.) (2005) *Procedure for classification of forests for CORINE land cover in Sweden: New strategies for European Remote Sensing*. Millpress, Rotterdam.
- Hanson PC, Hamilton DP, Stanley EH, Preston N, Langman OC, Kara EL (2011) Fate of allochthonous dissolved organic carbon in lakes: a quantitative approach. *PLoS One*, **6**, e21884.
- Henderson RK, Baker A, Murphy KR, Hambly A, Stuetz RM, Khan SJ (2009) Fluorescence as a potential monitoring tool for recycled water systems: a review. *Water Research*, **43**, 863–881.
- Hernes PJ, Bergamaschi BA, Eckard RS, Spencer RGM (2009) Fluorescence-based proxies for lignin in freshwater dissolved organic matter. *Journal of Geophysical Research-Biogeosciences*, **114**, doi: 10.1029/2009jg000938.
- Hongve D, Riise G, Kristiansen JF (2004) Increased colour and organic acid concentrations in Norwegian forest lakes and drinking water - a result of increased precipitation? *Aquatic Sciences*, **66**, 231–238.
- Hudson N, Baker A, Reynolds D (2007) Fluorescence analysis of dissolved organic matter in natural, waste and polluted waters - A review. *River Research and Applications*, **23**, 631–649.
- Hunt JF, Ohno T (2007) Characterization of fresh and decomposed dissolved organic matter using excitation-emission matrix fluorescence spectroscopy and multiway analysis. *Journal of Agricultural and Food Chemistry*, **55**, 2121–2128.
- Inamdar S, Finger N, Singh S *et al.* (2012) Dissolved organic matter (DOM) concentration and quality in a forested mid-Atlantic watershed, USA. *Biogeochemistry*, **108**, 55–76.
- Ishii SKL, Boyer TH (2012) Behavior of reoccurring PARAFAC components in fluorescent dissolved organic matter in natural and engineered Systems: a critical review. *Environmental Science & Technology*, **46**, 2006–2017.
- Jaffe R, Mcknight D, Maie N, Cory R, McDowell WH, Campbell JL (2008) Spatial and temporal variations in DOM composition in ecosystems: the importance of long-term monitoring of optical properties. *Journal of Geophysical Research-Biogeosciences*, **113**, doi: 10.1029/2008jg000683.
- Jansson M, Persson L, De Roos AM, Jones RI, Tranvik LJ (2007) Terrestrial carbon and intraspecific size-variation shape lake ecosystems. *Trends in Ecology & Evolution*, **22**, 316–322.
- Kalbitz K, Solinger S, Park JH, Michalzik B, Matzner E (2000) Controls on the dynamics of dissolved organic matter in soils: a review. *Soil Science*, **165**, 277–304.
- Kalbitz K, Schwesig D, Rethemeyer J, Matzner E (2005) Stabilization of dissolved organic matter by sorption to the mineral soil. *Soil Biology & Biochemistry*, **37**, 1319–1331.
- Karlsson J, Berggren M, Ask J, Bystrom P, Jonsson A, Laudon H, Jansson M (2012) Terrestrial organic matter support of lake food webs: evidence from lake metabolism and stable hydrogen isotopes of consumers. *Limnology and Oceanography*, **57**, 1042–1048.
- Kelton N, Molot LA, Dillon PJ (2007) Spectrofluorometric properties of dissolved organic matter from Central and Southern Ontario streams and the influence of iron and irradiation. *Water Research*, **41**, 638–646.
- Koehler B, Von Wachenfeldt E, Kothawala D, Tranvik LJ (2012) Reactivity continuum of dissolved organic carbon decomposition in lake water. *Journal of Geophysical Research-Biogeosciences*, **117**, G01024.
- Kögel-Knabner I, Ekschmitt K, Flessa H, Guggenberger G, Matzner E, Marschner B, Von Lützw M (2008) An integrative approach of organic matter stabilization in temperate soils: linking chemistry, physics, and biology. *Journal of Plant Nutrition and Soil Science*, **171**, 5–13.
- Köhler S, Hruska J, Jonsson J, Lovgren L, Lofts S (2002) Evaluation of different approaches to quantify strong organic acidity and acid-base buffering of organic-rich surface waters in Sweden. *Water Research*, **36**, 4487–4496.
- Köhler SJ, Buffam I, Laudon H, Bishop KH (2008) Climate's control of intra-annual and interannual variability of total organic carbon concentration and flux in two contrasting boreal landscape elements. *Journal of Geophysical Research-Biogeosciences*, **113**, G03012.
- Köhler SJ, Kothawala DN, Futter MN, Tranvik LJ (2013) In-lake processes offset increased terrestrial inputs of dissolved organic carbon and color to lakes. *PLoS One*, **8**, e70598.
- Kortelainen P (1993) Content of total organic carbon in Finnish lakes and its relationship to catchment characteristics. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 1477–1483.
- Kortelainen P, Mannio J (1988) Natural and anthropogenic acidity sources for Finnish lakes. *Water Air and Soil Pollution*, **42**, 341–352.
- Kortelainen P, Mattsson T, Finer L, Ahtiainen M, Saukkonen S, Sallantausta T (2006) Controls on the export of C, N, P and Fe from undisturbed boreal catchments, Finland. *Aquatic Sciences*, **68**, 453–468.
- Kothawala D, Von Wachenfeldt E, Koehler B, Tranvik L (2012) Selective loss and preservation of lake water dissolved organic matter fluorescence during long-term dark incubations. *The Science of the Total Environment*, **433**, 238–246.
- Kothawala D, Murphy KR, Stedmon C, Weyhenmeyer GA, Tranvik LJ (2013) Inner filter correction of dissolved organic matter fluorescence. *Limnology and Oceanography - Methods*, **11**, 616–630.
- Kraus TEC, Bergamaschi BA, Hernes PJ, Doctor D, Kendall C, Downing BD, Losee RF (2011) How reservoirs alter drinking water quality: organic matter sources, sinks, and transformations. *Lake and Reservoir Management*, **27**, 205–219.
- Kundzewicz ZW, Mata LJ, Arnell NW *et al.* (2007) Freshwater resources and their management. In: *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. (ed Parry ML, Canziani OF, Palutikof JP, Van Der Linden PJ, Hanson CE) pp 173–210. Cambridge University Press, Cambridge, UK.
- Lakowicz JR (2006) *Principles of Fluorescence Spectroscopy*, Springer-Verlag, New York, NY.
- Lapierre J-F, Frenette JJ (2009) Effects of macrophytes and terrestrial inputs on fluorescent organic matter in a large river system. *Aquatic Sciences*, **71**, 15–24.
- Laudon H, Berggren M, Agren A *et al.* (2011) Patterns and dynamics of dissolved organic carbon (DOC) in boreal streams: the role of processes, connectivity, and scaling. *Ecosystems*, **14**, 880–893.
- Laudon H, Buttle J, Carey SK *et al.* (2012) Cross-regional prediction of long-term trajectory of stream water DOC response to climate change. *Geophysical Research Letters*, **39**, L18404.
- Livingstone DM, Lotter AF, Walker IR (1999) The decrease in summer surf are water temperature with altitude in Swiss Alpine lakes: a comparison with air temperature lapse rates. *Arctic Antarctic and Alpine Research*, **31**, 341–352.
- Löfgren S, Zetterberg T (2011) Decreased DOC concentrations in soil water in forested areas in southern Sweden during 1987–2008. *Science of the Total Environment*, **409**, 1916–1926.
- Lofts S, Tipping E (1998) An assemblage model for cation binding by natural particulate matter. *Geochimica Et Cosmochimica Acta*, **62**, 2609–2625.
- Lønborg C, Álvarez-Salgado XA, Davidson K, Martínez-García S, Teira E (2010) Assessing the microbial bioavailability and degradation rate constants of dissolved organic matter by fluorescence spectroscopy in the coastal upwelling system of the Ria de Vigo. *Marine Chemistry*, **119**, 121–129.
- Luider C, Petticrew E, Curtis PJ (2003) Scavenging of dissolved organic matter (DOM) by amorphous iron hydroxide particles Fe(OH)(3(s)). *Hydrobiologia*, **494**, 37–41.
- Maie N, Scully NM, Pisani O, Jaffe R (2007) Composition of a protein-like fluorophore of dissolved organic matter in coastal wetland and estuarine ecosystems. *Water Research*, **41**, 563–570.
- Mann PJ, Davydova A, Zimov N *et al.* (2012) Controls on the composition and lability of dissolved organic matter in Siberia's Kolyma River basin. *Journal of Geophysical Research-Biogeosciences*, **117**, G01028.
- Massicotte P, Frenette JJ (2011) Spatial connectivity in a large river system: resolving the sources and fate of dissolved organic matter. *Ecological Applications*, **21**, 2600–2617.
- Matilainen A, Gjessing ET, Lahtinen T, Hed L, Bhatnagar A, Sillanpää M (2011) An overview of the methods used in the characterisation of natural organic matter (NOM) in relation to drinking water treatment. *Chemosphere*, **83**, 1431–1442.
- Mattsson T, Kortelainen P, Raika A (2005) Export of DOM from boreal catchments: impacts of land use cover and climate. *Biogeochemistry*, **76**, 373–394.

- Mcknight DM, Boyer EW, Westerhoff PK, Doran PT, Kulbe T, Andersen DT (2001) Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. *Limnology and Oceanography*, **46**, 38–48.
- Miller JN (1981) General Considerations on Fluorescence Spectrometry. In: *Standards in Fluorescence Spectrometry, Ultraviolet Spectrometry Group*. (ed Miller JN). New York, NY, Chapman and Hall/Methuen, Inc.
- Miller MP, Mcknight DM, Chapra SC, Williams MW (2009) A model of degradation and production of three pools of dissolved organic matter in an alpine lake. *Limnology and Oceanography*, **54**, 2213–2227.
- Monteith DT, Stoddard JL, Evans CD *et al.* (2007) Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry. *Nature*, **450**, 537–540.
- Moran MA, Zepp RG (1997) Role of photoreactions in the formation of biologically labile compounds from dissolved organic matter. *Limnology and Oceanography*, **42**, 1307–1316.
- Müller RA, Futter MN, Sobek S, Nisell J, Bishop K, Weyhenmeyer GA (2013) Water renewal along the aquatic continuum offsets cumulative retention by lakes: implications for the character of organic carbon in boreal lakes. *Aquatic Sciences*, **75**, 535–545.
- Murphy KR, Hambly A, Singh S, Henderson RK, Baker A, Stuetz R, Khan SJ (2011) Organic Matter Fluorescence in Municipal Water Recycling Schemes: toward a Unified PARAFAC Model. *Environmental Science & Technology*, **45**, 2909–2916.
- Neff JC, Hooper DU (2002) Vegetation and climate controls on potential CO<sub>2</sub>, DOC and DON production in northern latitude soils. *Global Change Biology*, **8**, 872–884.
- Neubauer E, Köhler SJ, Von Der Kammer F, Laudon H, Hofmann T (2013) Effect of pH and Stream Order on Iron and Arsenic Speciation in Boreal Catchments. *Environmental Science & Technology*, **47**, 7120–7128.
- Pastor J, Solin J, Bridgman SD, Updegraff K, Harth C, Weishampel P, Dewey B (2003) Global warming and the export of dissolved organic carbon from boreal peatlands. *Oikos*, **100**, 380–386.
- Prairie YT (2008) Carbo-centric limnology: looking back, looking forward. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 543–548.
- Prairie YT, Bird DF, Cole JJ (2002) The summer metabolic balance in the epilimnion of southeastern Quebec lakes. *Limnology and Oceanography*, **47**, 316–321.
- Rantakari M, Kortelainen P, Vuorenmaa J, Mannio J, Forsius M (2004) Finnish lake survey: the role of catchment attributes in determining nitrogen, phosphorous, and organic carbon concentrations. *Water Air and Soil Pollution*, **4**, 683–699.
- Roulet N, Moore TR (2006) Browning the waters. *Nature*, **444**, 283–284.
- Schiff SL, Aravena R, Trumbore SE, Hinton MJ, Elgood R, Dillon PJ (1997) Export of DOC from forested catchments on the Precambrian Shield of Central Ontario: clues from <sup>13</sup>C and <sup>14</sup>C. *Biogeochemistry*, **36**, 43–65.
- Sebestyen SD, Boyer EW, Shanley JB, Kendall C, Doctor DH, Aiken GR, Ohte N (2008) Sources, transformations, and hydrological processes that control stream nitrate and dissolved organic matter concentrations during snowmelt in an upland forest. *Water Resources Research*, **44**, doi: 10.1029/2008wr006983.
- Smith EM, Benner R (2005) Photochemical transformations of riverine dissolved organic matter: effects on estuarine bacterial metabolism and nutrient demand. *Aquatic Microbial Ecology*, **40**, 37–50.
- Sobek S, Tranvik LJ, Cole JJ (2005) Temperature independence of carbon dioxide supersaturation in global lakes. *Global Biogeochemical Cycles*, **19**, Gb2003.
- Sobek S, Nisell J, Folster J (2011) Predicting the volume and depth of lakes from map-derived parameters. *Inland Waters*, **1**, 177–184.
- Stedmon CA, Bro R (2008) Characterizing dissolved organic matter fluorescence with parallel factor analysis: a tutorial. *Limnology and Oceanography-Methods*, **6**, 572–579.
- Stedmon CA, Markager S (2003) Behaviour of the optical properties of coloured dissolved organic matter under conservative mixing. *Estuarine Coastal and Shelf Science*, **57**, 973–979.
- Stedmon CA, Markager S (2005) Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. *Limnology and Oceanography*, **50**, 686–697.
- Tipping E (1981) The adsorption of aquatic humic substances by iron oxides. *Geochimica Cosmochimica Acta*, **45**, 191–199.
- Tranvik LJ, Jansson M (2002) Climate change - Terrestrial export of organic carbon. *Nature*, **415**, 861–862.
- Tranvik LJ, Downing JA, Cotner JB *et al.* (2009) Lakes and reservoirs as regulators of carbon cycling and climate. *Limnology and Oceanography*, **54**, 2298–2314.
- Trumbore SE, Schiff SL, Aravena R, Elgood R (1992) Sources and transformations of dissolved organic carbon in the Harp Lake forested catchment: the role of soils. *Radiocarbon*, **34**, 626–635.
- Vähatalo AV, Salonen K, Munster U, Jarvinen M, Wetzel RG (2003) Photochemical transformation of allochthonous organic matter provides bioavailable nutrients in a humic lake. *Archiv Fur Hydrobiologie*, **156**, 287–314.
- Von Wachenfeldt E, Tranvik LJ (2008) Sedimentation in boreal lakes - The role of flocculation of allochthonous dissolved organic matter in the water column. *Ecosystems*, **11**, 803–814.
- Von Wachenfeldt E, Sobek S, Bastviken D, Tranvik LJ (2008) Linking allochthonous dissolved organic matter and boreal lake sediment carbon sequestration: the role of light-mediated flocculation. *Limnology and Oceanography*, **53**, 2416–2426.
- Walker SA, Amon RMW, Stedmon C, Duan SW, Louchouart P (2009) The use of PARAFAC modeling to trace terrestrial dissolved organic matter and fingerprint water masses in coastal Canadian Arctic surface waters. *Journal of Geophysical Research-Biogeosciences*, **114**, G00f06.
- Weishaar JL, Aiken GR, Bergamaschi BA, Fram MS, Fujii R, Mopper K (2003) Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. *Environmental Science & Technology*, **37**, 4702–4708.
- Weyhenmeyer GA, Karlsson J (2009) Nonlinear response of dissolved organic carbon concentrations in boreal lakes to increasing temperatures. *Limnology and Oceanography*, **54**, 2513–2519.
- Weyhenmeyer GA, Froberg M, Karlton E, Khalili M, Kothawala D, Temnerud J, Tranvik LJ (2012) Selective decay of terrestrial organic carbon during transport from land to sea. *Global Change Biology*, **18**, 349–355.
- Williams CJ, Yamashita Y, Wilson HF, Jaffe R, Xenopoulos MA (2010) Unraveling the role of land use and microbial activity in shaping dissolved organic matter characteristics in stream ecosystems. *Limnology and Oceanography*, **55**, 1159–1171.
- Wilson HF, Xenopoulos MA (2009) Effects of agricultural land use on the composition of fluvial dissolved organic matter. *Nature Geoscience*, **2**, 37–41.
- Windolf J, Jeppesen E, Jensen JP, Kristensen P (1996) Modelling of seasonal variation in nitrogen retention and in-lake concentration: a four-year mass balance study in 16 shallow Danish lakes. *Biogeochemistry*, **33**, 25–44.
- Wu FC, Evans RD, Dillon PJ (2003) Separation and characterization of NOM by high-performance liquid chromatography and on-line three-dimensional excitation emission matrix fluorescence detection. *Environmental Science & Technology*, **37**, 3687–3693.
- Yamashita Y, Tanoue E (2003) Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids. *Marine Chemistry*, **82**, 255–271.
- Yamashita Y, Maie N, Briceno H, Jaffe R (2010) Optical characterization of dissolved organic matter in tropical rivers of the Guayana Shield, Venezuela. *Journal of Geophysical Research-Biogeosciences*, **115**, G00f10.

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Excitation loading values for PARAFAC components.

**Table S2.** Emission loading values for PARAFAC components.

**Table S3.** Summary of PLS model performance.

**Table S4.** Results of PLS permutations test.

**Table S5.** Linear regression relationships between DOC and PARAFAC components.

**Figure S1.** Loadings plots of PARAFAC components.

**Figure S2.** Residuals of PARAFAC analysis from subregions of dataset.

**Figure S3.** Relationship between scores of the first- and second-principle components axes (PC1 and PC2) from Fig. 2b with DOC and MAT, respectively.

**Figure S4.** PCA plot of all variables including PARAFAC components.

# **Supplementary Information**

## **Controls of dissolved organic matter quality: Evidence from a large-scale boreal lake survey**

**Dolly N. Kothawala, Colin Stedmon, Roger A. Müller, Gesa A. Weyhenmeyer, Stephan Köhler, and Lars J. Tranvik**

### **Table of Contents:**

S. Table 1: Excitation loading values for PARAFAC components

S. Table 2: Emission loading values for PARAFAC components

S. Table 3: Summary of PLS model performance

S. Table 4: Results of PLS permutations test

S. Table 5: Linear regression relationships between DOC and PARAFAC components

S. Fig 1: Loadings plots of PARAFAC components

S. Fig 2: Residuals of PARAFAC analysis from sub-regions of dataset

S. Fig 3: Relationship between scores of the first and second principle components axes (PC1 and PC2) from Fig 2b with DOC and MAT, respectively

S. Fig 4: PCA plot of all variables including PARAFAC components

**S. Table 1:** Excitation loadings for six PARAFAC components

Wavelength	C1	C2	C3	C4	C5	C6
250	0.246962	0.291595	0.19378	0.325018	0.523545	0.223084
255	0.249937	0.272274	0.206104	0.280731	0.463169	0.287063
260	0.244009	0.245061	0.231245	0.244679	0.40051	0.310308
265	0.227484	0.211314	0.230935	0.213318	0.340908	0.345004
270	0.211474	0.197637	0.253604	0.184857	0.29238	0.383278
275	0.192314	0.191815	0.254696	0.158029	0.231646	0.384502
280	0.182179	0.207701	0.266848	0.137617	0.19408	0.364943
285	0.174971	0.229545	0.259675	0.126958	0.155778	0.310267
290	0.155474	0.229446	0.221903	0.107725	0.10988	0.234053
295	0.158494	0.24928	0.201922	0.10871	0.090932	0.192422
300	0.155437	0.247294	0.164974	0.110059	0.068158	0.141553
305	0.177747	0.2703	0.146972	0.133872	0.060862	0.109962
310	0.194979	0.270783	0.121296	0.154769	0.052736	0.06413
315	0.206963	0.254287	0.096479	0.173841	0.036227	0.016383
320	0.217207	0.230366	0.075625	0.194278	0.022675	0
325	0.219667	0.200773	0.06003	0.208106	0.015088	0
330	0.222249	0.173665	0.051826	0.215005	0.004773	0
335	0.215257	0.144842	0.048017	0.217823	0.001783	0
340	0.204109	0.117595	0.049649	0.214495	0	0
345	0.20129	0.097623	0.056627	0.216908	0	0
350	0.189213	0.073868	0.063558	0.212996	0	0
355	0.178843	0.05293	0.074395	0.20878	0	0
360	0.159212	0.032505	0.084591	0.19196	0.002992	0
365	0.145941	0.016208	0.098826	0.181811	0.005599	0
370	0.125804	0.002471	0.110257	0.161556	0.007872	0
375	0.108569	0	0.133019	0.141137	0.003627	0
380	0.086887	0	0.14822	0.115615	0	0
385	0.070829	0	0.160719	0.095894	0	0
390	0.053757	0	0.165464	0.075901	0	0
395	0.039784	0	0.167076	0.059017	0	0
400	0.029028	0	0.170426	0.045684	0	0
405	0.020242	0	0.170135	0.033225	0	0
410	0.011837	0	0.167283	0.022174	0	0
415	0.004996	0	0.164865	0.01307	0	0
420	0	0	0.157492	0.004941	0	0
425	0	0	0.142098	0	0	0
430	0	0	0.128774	0	0	0
435	0	0	0.109894	0	0	0
440	0	0	0.095511	0	0	0
445	0	0	0.084458	0	0	0

**S. Table 2:** Emission loadings for six PARAFAC components

Wavelength	C1	C2	C3	C4	C5	C6
300	0	0	0	0.000115	0	0.171886
304	0	0	0	0	0	0.206877
308	0	0	0	0	0	0.231076
312	0	0	0	0	0	0.248632
316	0	0	0	0	0.00149	0.257416
320	0	0	0	0	0.005102	0.260012
324	0	0	0	0	0.006768	0.268458
328	5.79E-05	0	1.68E-05	0	0.012153	0.277496
332	0	0.005217	0.000724	0	0.013254	0.273729
336	0	0.015415	0.000651	0	0.01127	0.262168
340	0	0.025365	0.000776	0	0.013417	0.250729
344	0	0.036384	0.001271	0	0.014472	0.242533
348	0	0.050162	0.001685	0	0.016557	0.228062
352	0	0.066288	0.001816	0	0.018803	0.210977
356	0	0.084256	0.001811	0	0.021622	0.192643
360	0	0.103398	0.000999	0.003027	0.027608	0.169891
364	0	0.123948	0	0.006781	0.031668	0.149439
368	0	0.145626	0	0.012933	0.03948	0.125003
372	0	0.169298	0	0.019846	0.046991	0.103271
376	0.004155	0.187395	0	0.02698	0.054196	0.08688
380	0.010812	0.202538	0	0.035174	0.063793	0.071235
384	0.019563	0.218875	0	0.044253	0.07211	0.06451
388	0.028577	0.233076	0	0.054138	0.082308	0.050395
392	0.038586	0.243964	0	0.064836	0.091419	0.041288
396	0.049571	0.251694	0	0.076138	0.100564	0.033176
400	0.061184	0.256432	0	0.087793	0.109276	0.026708
404	0.073824	0.255549	0	0.100532	0.117896	0.023088
408	0.087099	0.253262	0	0.113236	0.125913	0.02435
412	0.100908	0.247409	0	0.126974	0.135044	0.027619
416	0.115113	0.23816	0.001029	0.140898	0.144539	0.031349
420	0.129047	0.226881	0.008195	0.153268	0.151988	0.032156
424	0.141687	0.213244	0.01817	0.164831	0.159306	0.031273
428	0.154134	0.198667	0.029779	0.175434	0.16566	0.031059
432	0.16578	0.182781	0.042297	0.18473	0.171156	0.031764
436	0.175915	0.166891	0.055859	0.192638	0.175675	0.031333
440	0.185075	0.150897	0.070436	0.199289	0.180289	0.031093
444	0.191997	0.134978	0.085602	0.203116	0.181488	0.031943
448	0.198558	0.120086	0.100474	0.20753	0.184844	0.032126
452	0.203777	0.105812	0.114527	0.210161	0.18618	0.033865
456	0.207124	0.092149	0.128351	0.211228	0.18669	0.034839
460	0.209716	0.079392	0.142295	0.211469	0.186577	0.038644
464	0.210788	0.067252	0.156798	0.209091	0.185158	0.041274
468	0.210773	0.056511	0.169366	0.207127	0.184483	0.042933
472	0.209023	0.04723	0.181846	0.202992	0.183188	0.040369
476	0.206133	0.03885	0.19305	0.197824	0.180702	0.037596
480	0.202136	0.032661	0.202228	0.194804	0.18832	0.023527
484	0.198085	0.026293	0.211346	0.18653	0.176564	0.036351
488	0.193081	0.020884	0.218037	0.179719	0.173091	0.035326
492	0.186825	0.016837	0.222482	0.171999	0.169584	0.034478
496	0.18019	0.013312	0.22548	0.164062	0.164621	0.030608
500	0.173595	0.009114	0.225898	0.158077	0.174979	0.019344
504	0.166945	0.008047	0.227283	0.148901	0.157566	0.025421
508	0.160279	0.006158	0.226952	0.141275	0.15217	0.02416
512	0.153141	0.004836	0.224926	0.133789	0.147821	0.020724
516	0.14581	0.003764	0.221441	0.125959	0.142615	0.018022
520	0.139063	0	0.216392	0.120013	0.160102	0.015229
524	0.131406	0.002377	0.212049	0.111545	0.133253	0.014761
528	0.124742	0.002323	0.206712	0.104851	0.126945	0.012719
532	0.118058	0.002667	0.201047	0.09883	0.121521	0.008174
536	0.111684	0.002841	0.194388	0.092571	0.114279	0.008576
540	0.105666	0	0.187419	0.087372	0.134367	0.022338

544	0.099358	0.003291	0.180643	0.081583	0.104919	0.003958
548	0.093997	0.003562	0.17382	0.076482	0.099992	0.003757

**S. Table 3.** Summary of PLS model performance as a function of model complexity (number of PLS components), with the cumulative explained variation ( $R^2Y$ ), the cumulative goodness of prediction ( $Q^2Y$ ) and explained variance of predicted variables ( $R^2X$ ) for each PLS component derived for models predicting the variability among six PARAFAC components (Y-variables), as an intensity (C1 to C6) in the C-intensity model, or as a percentage (% C1 to % C6) in the C-percentage model. Both models used in-lake chemistry, geographical, climate and absorbance indexes listed in Table 1 as X variables.

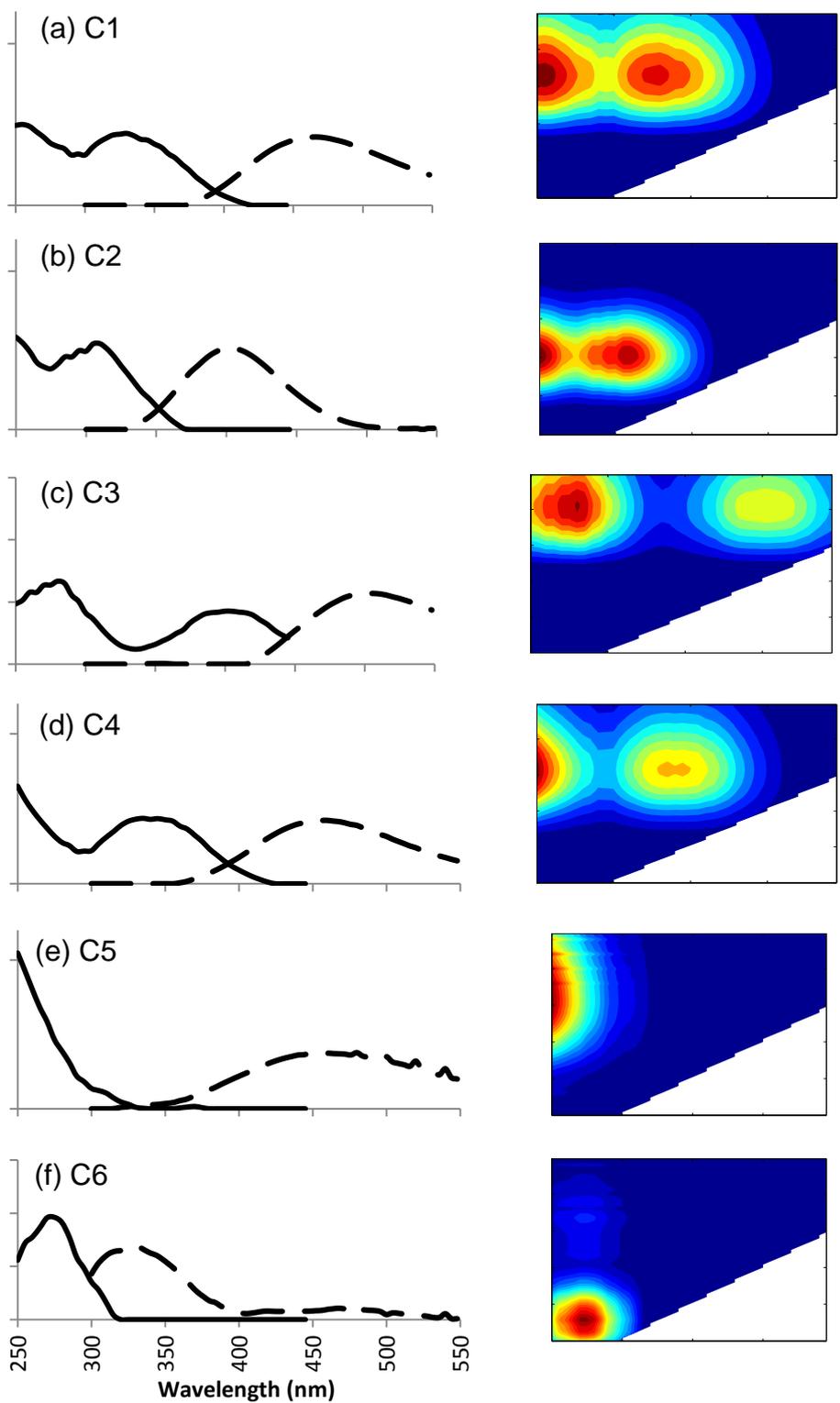
Model (Y Variables)	PLS component	$R^2Y$	$Q^2Y$	$R^2X$
C-intensities (C1, C2, C3, C4, C5 and C6)	1	0.41	0.40	0.29
	2	0.48	0.48	0.44
	3	0.55	0.55	0.58
	4	0.61	0.61	0.62
C-percentages (%C1, %C2, %C3, %C4, %C5 and %C6)	1	0.26	0.26	0.25
	2	0.36	0.35	0.47
	3	0.42	0.40	0.58
	4	0.48	0.45	0.62

**S. Table 4.** Permutation test results including the cumulative amount of explained variation ( $R^2Y$ ) and goodness of prediction ( $Q^2Y$ ) for each PARAFAC component used as Y response variable, either as an intensity (C1 to C6) in the C intensity model or as a percentage (%C1 to %C6) in the C percentage model. Lake water chemistry, catchment and climate were X variables (listed in Table 1). The difference between y-intercepts of  $R^2$  and  $Q^2$  regressions passing through an un-permuted dataset and 100 permuted datasets was used to derive the background correlation.

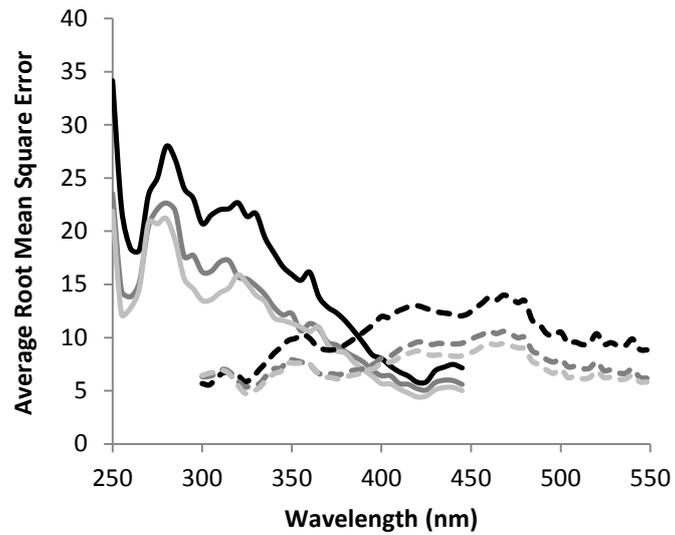
	C intensity model		C percentage model		
	$R^2Y$ , $Q^2Y$	Background	$R^2Y$ , $Q^2Y$	Background	
C1	0.78, 0.76	0.14	% C1	0.29, 0.23	0.08
C2	0.76, 0.74	0.16	% C2	0.52, 0.49	0.09
C3	0.81, 0.84	0.15	% C3	0.64, 0.62	0.11
C4	0.45, 0.42	0.12	% C4	0.27, 0.21	0.11
C5	0.57, 0.54	0.12	% C5	0.42, 0.40	0.08
C6	0.35, 0.33	0.09	% C6	0.75, 0.73	0.08

**S. Table 5.** Linear relationships between DOC and the fluorescence intensity of six PARAFAC components. The significance of the intercept differing from zero is indicated with \*, \*\*, or \*\*\*, representing  $p < 0.05$ , 0.01 and 0.001, ns represent a non-significant difference from zero. Significance that the slope differed from zero was  $< 0.0001$  for all relationships, and residuals were normally distributed.

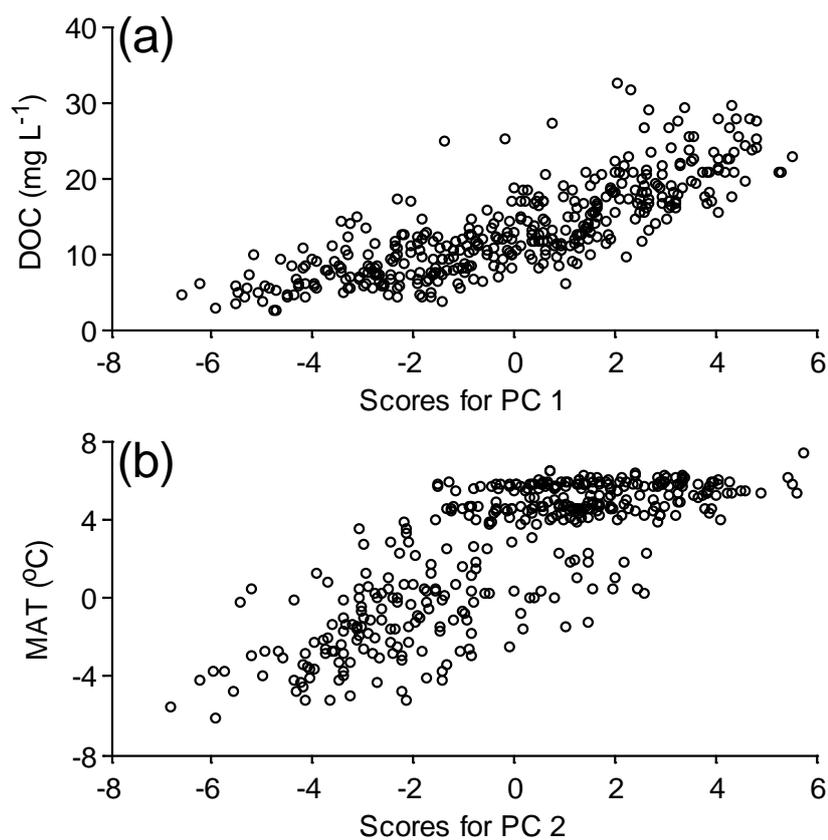
	slope	intercept	R <sup>2</sup>
C1	0.82	-0.73, ns	0.63
C2	0.54	0.71, ns	0.71
C3	0.53	-1.10, ***	0.81
C4	0.45	-0.01, ns	0.39
C5	0.46	-1.15**	0.43
C6	0.03	1.46***	0.05



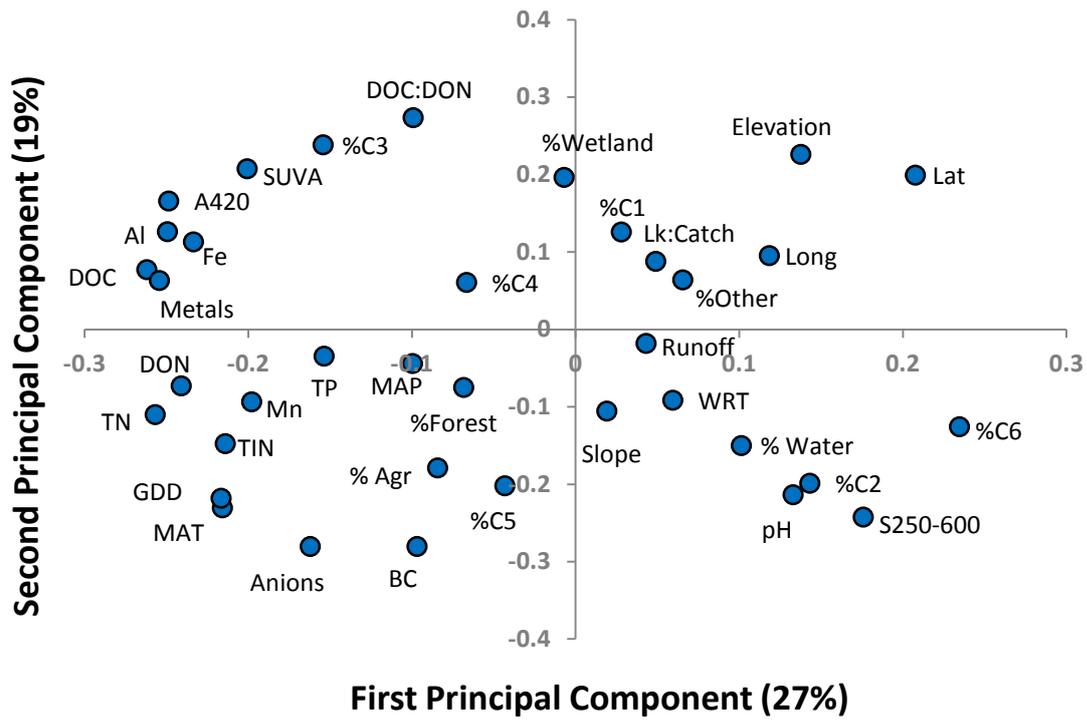
**S. Fig 1.** Positioning of six PARAFAC components generated from 560 boreal Swedish lakes. Excitation (solid lines) and emission (dashed lines) loadings plots shown at left, with contour plots of full excitation (x-axis)-emission (y-axis) matrices to right.



**S.Fig. 2.** Comparison of residuals from PARAFAC analysis across excitation-emission spectra for the sub-set of lakes from Northern Alpine (dark lines,  $n = 128$ ), Northern Lowlands (grey lines,  $n = 80$ ), and Southern Lowlands (light grey lines,  $n = 324$ ) based on the average root square error summed across excitation (solid lines) and emission (dashed lines) spectra.



**S. Fig. 3.** Relationships between scores (from percentage PLS model, Fig 2b) for lake samples along the first and second principal component axes (PC 1 and PC 2), with dissolved organic carbon (DOC) and mean annual temperature (MAT), respectively.



**S. Fig. 4.** Loadings plot of principle component analysis (PCA) for all variables included in Table 1.