Article (refereed) - postprint

Carter, Heather; Tipping, Edward; Koprivnjak, Jean-Francois; Miller, Matthew P.; Cookson, Brenda; Hamilton-Taylor, John. 2012 Freshwater DOM quantity and quality from a two-component model of UV absorbance. *Water Research*, 46 (14). 4532-4542. 10.1016/j.watres.2012.05.021

.

Second revision, submitted to Water Research, May 2012

Freshwater DOM Quantity and Quality from a Two-Component Model of UV Absorbance

Heather T. Carter[A,D], Edward Tipping[A,E], Jean-Francois Koprivnjak[B], Matthew P. Miller[C], Brenda Cookson[D] and John Hamilton-Taylor[D].

[A] Centre for Ecology and Hydrology, Lancaster Environment Centre, Bailrigg, Lancaster, LA1 4AP, United Kingdom

[B] Environmental and Resource Studies Program, Water Quality Centre, Trent University, K9J 7B8, Canada

[C] U.S. Geological Survey Water Science Center, 121 West 200 South Moab, UT 84532, USA

[D] Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, United Kingdom

[E] Corresponding author Email: et@ceh.ac.uk

Highlights

- UV optical absorbance by DOM can be explained by two end-member components (A & B).
- We analysed a data set of *c.* 1700 samples to derive the spectrum of each component.
- Fractions of A and B can be obtained from optical absorbance at two wavelengths.
- The results permit DOC concentration to be estimated accurately and without bias.
- The fractional contributions provide information on DOM quality.

Keywords

Dissolved organic carbon, dissolved organic matter, model, UV absorbance

ABSTRACT

We present a model that considers UV-absorbing dissolved organic matter (DOM) to consist of two components (A and B), each with a distinct and constant spectrum. Component A absorbs UV light strongly, and is therefore presumed to possess aromatic chromophores and hydrophobic character, whereas B absorbs weakly and can be assumed hydrophilic. We parameterised the model with dissolved organic carbon concentrations, [DOC], and corresponding UV spectra for c. 1700 filtered surface water samples from North America and the United Kingdom, by optimising extinction coefficients for A and B, together with a small constant concentration of non-absorbing DOM (0.80 mg DOC L^{-1}). Good unbiased predictions of [DOC] from absorbance data at 270 and 350 nm were obtained (r^2=0.98), the sum of squared residuals in [DOC] being reduced by 66% compared to a regression model fitted to absorbance at 270 nm alone. The parameterised model can use measured optical absorbance values at any pair of suitable wavelengths to calculate both [DOC] and the relative amounts of A and B in a water sample, i.e. measures of quantity and quality. Blind prediction of [DOC] was satisfactory for 9 of 11 independent data sets (181 of 213 individual samples).

1. Introduction

Dissolved organic matter (DOM), comprising the partial decomposition products of plant and other biological materials, is ubiquitous in surface, soil and ground waters (Perdue and Gjessing, 1990; Kullberg et al., 1993; Hessen and Tranvik, 1998). It has numerous ecological and geochemical functions, including light absorption, pH buffering, interactions with metals and organic contaminants, adsorption to surfaces and photochemical activity. It plays a role in the terrestrial and aquatic carbon cycles, so that monitoring its concentrations and fluxes can aid in understanding the effects of land use change, acidification reversal and climatic warming (Pastor et al., 2003; Worrall et al., 2004; Monteith et al., 2007;). Because differences in source material, rate and extent of decomposition, and fractionation processes, are likely to generate substantial variability in DOM concentrations and properties, monitoring studies often require measurements not only of concentration, but also one or more indicators of quality. Currently this is most often performed by measuring the concentration of dissolved organic carbon - [DOC] - together with spectroscopic properties.

The main techniques for measuring [DOC] are persulphate oxidation or high temperature combustion methods (Menzel and Vaccaro, 1964; Chen et al., 2002), including either the removal or measurement of inorganic carbon. Both methods are fairly time-consuming and costly, and require the transfer of samples to a laboratory. Qualitative information about DOM is commonly sought from UV-visible absorption and fluorescence spectroscopy. For example specific UV absorbance (SUVA), measured at 254 or 280 nm, is a measure of aromaticity (Chin et al., 1994; Weishaar et al., 2003), absorbance slopes and slope ratios provide information about DOM sources and properties (Helms et al., 2008; Loiselle et al., 2009). Fluorescence spectra are widely and increasingly used to compare DOM in time and space (Cabaniss and Shuman, 1987; Coble et al., 1990; Chen et al., 2003; Cory and McKnight, 2005).

Since DOM varies in spectroscopic properties, it follows that, in general, measurement of absorbance at a single wavelength cannot give accurate estimates of [DOC] (Tipping et al., 2009). However within an individual environmental system (stream, lakewater) it is often the case that temporal variability is sufficiently small for absorbance to provide a useful estimate of DOC concentration (Mattson et al., 1974; Lewis and Canfield, 1977; Grieve, 1984; Moore, 1987) Furthermore, data for more than one wavelength have been utilised in several studies to improve conversions from absorbance to concentration (Downing et al.,

2009; Fichot and Benner, 2011). A variant on this approach calibrates a multi-wavelength detector against a DOM standard formulated in the laboratory (Sandford et al., 2010).

In the multiple wavelength methods (Mattson et al., 1974; Downing et al., 2009), empirical approaches were taken to extract [DOC] from the spectra. However, Tipping et al. (2009) showed in a preliminary study that a formal two-component (A and B) model could account for absorbance data at two wavelengths (254 and 340 nm), and thereby provide accurate estimates of [DOC]. Using a modest dataset of 48 samples from unpolluted surface- and ground-waters, extinction coefficients for the two components were optimised and an excellent model fit was obtained, with r^2=0.997 and root-mean-squared deviation (*rmsd*) of 0.7 mg L^{-1} in [DOC]. The much greater absorption of UV light by component A indicates greater aromaticity and hydrophobicity (Thacker et al., 2005, 2008), so that knowing its fractional amount provides a simple measure of DOM quality.

However, deviations from model predictions were evident in samples from waters draining urban and industrial areas, and it was speculated that these arose from the presence of anthropogenic non-UV-absorbing DOM.

The aim of the present study was to test and parameterise the two-component approach more thoroughly, by making new parallel measurements of [DOC] and optical absorbance on UK surface waters, and combining the results with data already obtained for sites in Canada (Koprivnjak et al., 2010) and the USA (Miller and McKnight, 2010). The UK sampling was designed to cover a wide range of water types, in an attempt to maximise variability in DOM. We also tested the parameterised model on independent data sets.

2. Methods

In all, 427 samples were collected from UK sites, mainly from the North of England, but also from East Anglia and Scotland. Most samples were from streams or rivers, but one sampling site was a eutrophic lake. About one-third were repeat samples from 6 sites, and the rest were single samples from different waters. The land-use types in the drainage areas included upland and lowland peats, arable farmland, intensive pasture, urban and industrial areas, coniferous and broadleaved forest. Samples for analyses of [DOC], absorbance, conductivity and iron were collected in acid-washed polyethylene sampling vessels of 500 cm^3 capacity. A separate sample was collected in a 100 cm^3 Pyrex bottle, fully filled, for pH determination. The reported results refer to measurements made within 5 days (usually two days) of collection. To test for stability, in some cases a second 500 cm^3 sample was taken and stored in a refrigerator (5^{o}C) before analysis.

The UK samples were passed through GF/F glass-fibre filters with a nominal size cut-off of 0.7 μm. Filtered samples were analysed for conductivity (Jenway 4510 instrument) and total iron using ferrozine (Lofts et al., 2008), and for [DOC] by combustion to CO_2 using a Shimadzu TOC-VCPH instrument, after acidification and purging with nitrogen to remove dissolved inorganic carbon (DIC). The instrument was calibrated with phthalic acid solutions (0 - 40 mg C L^{-1}). Suwannee River Fulvic Acid (SRFA) and sodium bicarbonate solutions were used for quality control, the latter to check that the acidification-purge procedure completely eliminated DIC. Measurements were made in triplicate, with random ordering within each batch.

Absorption spectra in the UV-visible wavelength range (200-900 nm) were recorded first for unfiltered surface water samples (without settling of particulates), and then for the same samples after filtration. Measurements were also made on blanks and Suwannee River Fulvic Acid as a quality control. An Agilent 8453 diode array instrument was used, with 10 mm quartz cuvettes, which were washed with 18.2 mΩ water and then rinsed with sample before each measurement. All samples were allowed to reach room temperature before recording spectra, to avoid condensation on the cuvette. Spectra were recorded in triplicate in random order with quality standards at regular intervals. Sample pH was measured with a glass electrode using a Radiometer instrument. The pH electrodes and the conductivity meter were calibrated at the start of each set of samples, the spectrophotometer calibration was checked monthly. To obtain absorbance values in the UV range for modelling, we routinely subtracted the value of A_{700} from the measured values.

In Colorado, 116 samples were collected from May-September 2006 from oligotrophic alpine and subalpine stream and lake sites in the Green Lakes Valley and adjacent Como Creek watershed. The Green Lakes Valley is part of the Niwot Ridge Long-term Ecological Research (NWTLTER) site and is not influenced by direct human impacts. The system is characterized by a pulse of DOC during snowmelt in late May or early June, followed by a gradual return to lower concentrations during baseflow (Miller & McKnight, 2010). Samples were filtered with GF/F glassfibre filters of 0.7-µm nominal pore size (Whatman, c/o GE Healthcare Bio-Sciences Corporation, Piscataway, NJ, USA), DOC was measured by high temperature catalytic oxidation with a Shimadzu 5050A TOC Analyzer (Columbia, MD, USA), and absorption spectra were measured on an Agilent 8453 UV-visible spectroscopy system (Santa Clara, CA, USA).

Trent University collected stream and lake samples from a forested region of the Precambrian Shield in Ontario from November 2007 to October 2008. The lakes are oligotrophic to mesotrophic and are all headwater lakes, with one exception. Wetlands, primarily Sphagnum-conifer swamps or beaver ponds, comprise from 0 to 25% of the catchment areas. All streams were sampled near the mouth and all lakes at the outflow. Samples were filtered with Millipore 0.45-µm membrane filters, and analysed for DOC (Shimadzu TOC-VPH, Columbia, MD, USA) and optical absorbance (Cary 59 UV/Vis spectrophotometer, Varian, Palo Alto, CA, USA). We assume that the different filters used at Trent University and in the other two laboratories produce negligible differences in both [DOC] and UV spectra.

3. Modelling

In our previous work (Tipping et al., 2009), we described a two-component model of DOM, which accounted for optical absorbance in terms of the linear sum of two components (A and B) each with its own fixed absorbance spectrum. We adopt the same approach here, with one modification, which is the inclusion of a third component (C) that does not absorb light and is present at the same concentration in all water samples. Thus, the DOC concentration in a given sample is given by

$$[\mathrm{DOC}] = [\mathrm{DOC_{AB}}] + [\mathrm{DOC_C}] \quad (1)$$

where $\mathrm{DOC_{AB}}$ refers to the light-absorbing components. Therefore

$$[\mathrm{DOC}] = \frac{A_\lambda}{E_{\mathrm{AB},\lambda}} + [\mathrm{DOC_C}] \quad (2)$$

where A_λ is the absorbance of the sample at wavelength λ (nm) in the UV range, and $E_{\mathrm{AB},\lambda}$ is the extinction coefficient (absorbance $\mathrm{cm^{-1}}$ $[\mathrm{DOC_{AB}}]^{-1}$) of the light-absorbing DOM. If the DOM comprises two components, the extinction coefficient is given by;

$$\begin{aligned} E_{\mathrm{AB},\lambda} &= f_\mathrm{A} E_{\mathrm{A},\lambda} + f_\mathrm{B} E_{\mathrm{B},\lambda} \\ &= f_\mathrm{A} E_{\mathrm{A},\lambda} + (1 - f_\mathrm{A}) E_{\mathrm{B},\lambda} \end{aligned} \quad (3)$$

where f_A and f_B are the fractions of components A and B that comprise the light-absorbing DOM ($f_\mathrm{A} + f_\mathrm{B} = 1$), and $E_{\mathrm{A},\lambda}$ and $E_{\mathrm{B},\lambda}$ are the extinction coefficients at wavelength λ. Equation (3) can be written for two different wavelengths, λ1 and λ2, and then a ratio R defined by;

$$R = \frac{E_{\mathrm{AB},\lambda 1}}{E_{\mathrm{AB},\lambda 2}} = \frac{f_\mathrm{A} E_{\mathrm{A},\lambda 1} + (1 - f_\mathrm{A}) E_{\mathrm{B},\lambda 1}}{f_\mathrm{A} E_{\mathrm{A},\lambda 2} + (1 - f_\mathrm{A}) E_{\mathrm{B},\lambda 2}} = \frac{A_{\lambda 1}}{A_{\lambda 2}} \quad (4)$$

The value of R can thus be obtained simply from the measured absorbances at the two wavelengths. Its combination with the extinction coefficients of components A and B yields;

$$f_A = \frac{E_{\mathrm{B},\lambda 1} - R E_{\mathrm{B},\lambda 2}}{R (E_{\mathrm{A},\lambda 2} - E_{\mathrm{B},\lambda 2}) + (E_{\mathrm{B},\lambda 1} - E_{\mathrm{A},\lambda 1})} \quad (5)$$

Therefore, if the values of $E_{\mathrm{A},\lambda 1}$ $E_{\mathrm{A},\lambda 2}$, $E_{\mathrm{B},\lambda 1}$ and $E_{\mathrm{B},\lambda 2}$ are known, f_A can be calculated and then substituted back into equation (3) to obtain E_{AB} at either of the two wavelengths. Then if $[\mathrm{DOC_C}]$ is known, [DOC] can be obtained from equation (2).

For a chosen pair of wavelengths, measurements of [DOC], $A_{\lambda 1}$ and $A_{\lambda 2}$ for sufficient samples allow estimation of the model parameters, i.e. the four extinction coefficients $E_{A,\lambda 1}$ $E_{A,\lambda 2}$, $E_{B,\lambda 1}$ and $E_{B,\lambda 2}$, together with [DOC_C]. This is done by adopting an initial trial parameter set, calculating [DOC] for each sample and then computing the sum of squared residuals between measured and calculated [DOC]. Improvement of the parameter values (we used Microsoft Excel Solver to do this) then leads to their optimisation by minimising the sum of squared residuals.

In principle, any pair of wavelengths could be used, but practicalities impose some restrictions. While specific UV absorbance (SUVA) at 254 nm is widely employed in DOM research, this wavelength has two disadvantages. Firstly there is possible interference due to nitrate in systems with high levels (Edwards et al., 2001; Thomas and Burgess, 2007). Secondly, when [DOC] is high (> 50 mg L^{-1}), the absorbance may be too high for reliable determination. Therefore we favour 270 nm as λ1. Wavelength 2 (λ2) needs to be as different from λ1 as possible, while still permitting measurable absorbance values even on dilute samples, and we chose 350 nm. We propose these wavelengths for routine general use, and employed them in the main analysis presented here. However, application of the model is not restricted to absorbance values at 270 and 350 nm; we also fitted results for other wavelengths in order to be able to analyse published data (See Sections 4.1 and 4.5).

4. Results

The 1698 samples studied cover wide ranges of [DOC], pH and conductivity (Table 1), and originate from many different source environments, as described in Methods. Samples stored for one week before processing gave indistinguishable results from those analysed immediately, but there were decreases of c. 5% in both [DOC] and optical absorbance in samples stored for periods between 50 and 120 days.

We previously justified the use of a two-component model on the basis of a plot presented by Thacker et al. (2008) showing that for 23 concentrated DOM samples the ratio (E_{340}/E_{254}) varied monotonically with E_{340}, i.e. the values fell approximately on a single curve. The same behaviour was evident in the present data (Figure S1). Simple modelling with three components confirms that, if these were to occur randomly and have sufficiently different UV spectra, monotonic behaviour would not be observed (Figure S2). The wide observed range in E_{350} (Figure S2) confirms the high degree of variability in the UV absorption properties of DOM.

4.1. Parameterisation

The five model parameters ($E_{A,\lambda1}$, $E_{A,\lambda2}$, $E_{B,\lambda1}$, $E_{B,\lambda2}$ and [DOC_C]) could be optimised to fit the whole data set with good precision (*rmsd* = 1.2 mg L^{-1}, r^2 = 0.979), but as was found previously (Tipping et al., 2009), a unique combination of parameters does not exist, i.e. an infinite number of parameter sets can be found that provide identical results. It is even possible for some of the extinction coefficients to have negative values, and for values of f_A to be negative or greater than unity, all of which are physically meaningless. Therefore we introduced a constraint, based on evidence from Thacker et al. (2001) who showed that some physico-chemical properties of different samples of isolated DOM are strongly correlated with their extinction coefficients. In particular, the fractions of the total DOM sorbing to XAD8 resin or to alumina were positively, and approximately linearly, related to E_{340}. Although in neither case is precise extrapolation to 100% sorption possible, the plots given by Thacker et al. (2008) indicate that the limiting value of E_{340} is between 25 and 35 L g^{-1} cm^{-1}. We interpreted this to approximate the extinction coefficient of DOM fraction A, which has greater aromaticity, and therefore higher optical absorbance, hydrophobicity and propensity to sorb to surfaces. Since the extinction coefficients at 340 nm and 350 nm cannot differ greatly, we assigned a value of 30 L g^{-1} cm^{-1} to $E_{A,\lambda2}$. However, a unique parameter set was still not obtained, and therefore a further constraint was imposed, based on the finding that $E_{B,\lambda2}$ always took on low values (< 10 L g^{-1} cm^{-1}), suggesting that simply

setting $E_{B,\lambda 2}$ to zero would be justifiable. This means that our idealised component B absorbs UV light at wavelengths < 350 nm, but is assumed to be fully transparent at higher wavelengths. Refitting with these constraints did not affect either *rmsd* or r^2. The resulting parameter set is unique, still gives the best fit, and provides a physically realistic picture. Other constraints could be applied but they would differ only slightly from those we have employed, and would produce the same results. Setting $E_{B,\lambda 2}$ to zero is only for convenience, and equally good results would be obtained were it set to some low positive value.

Inspection of the derived values of f_A showed that for 18 of the 1698 samples (0.11% of the total) the ratio of measured to calculated [DOC] was <0.5 or >1.5, and these were treated as outliers (we presume the discrepancies reflect analytical errors). After their removal, the fit was slightly improved (*rmsd* = 1.1, r^2 = 0.989). The final model parameters are given in Table 2, which also lists extinction coefficients at other wavelengths, used in the analysis of independent data (Section 4.5). The measured and modelled [DOC] values are compared in Figure 1. Figure 2 shows the full UV spectra of components A and B, derived by applying the parameterised model to observed spectra in the range 250-400 nm.

Residuals between modelled and measured [DOC] tend to increase with [DOC], while the relative difference tends to decline (Figure 1). Use of the model to estimate [DOC] will give 95% confidence limits of ± 0.9 mg L^{-1} for 0 < [DOC] < 5 mg L^{-1}, ± 2 mg L^{-1} for 5 < [DOC] < 20 mg L^{-1}, and ± 4 mg L^{-1} for 20 < [DOC] < 80 mg L^{-1}. For the UK data set, the corresponding 95% confidence limits for ranges of [DOC] determined by combustion were ± 0.2, ± 0.4 and ± 2 mg L^{-1} respectively. Note however, that the errors in the spectroscopic method must result partially from inevitable differences in analytical results among the three contributing laboratories.

4.2. Comparison with single wavelength approaches

The results obtained with the model are superior to the global use of absorbance at a single wavelength to estimate [DOC]. Fitting the whole data set for 270nm gave r^2 = 0.964, and for 350 nm the agreement was worse, with r^2 = 0.932. Moreover, as shown by Figure S3, there are biases, with problems at low concentrations, especially shown up by the log-log plot. If the single wavelength relationships are used to construct predictive models, i.e. the regression equations are used to predict [DOC] simply from absorbance at one wavelength, the *rmsd* is 1.9 mg L^{-1} for 270 nm and 2.6 mg L^{-1} for 340 nm (the value for the two-component model is 1.1 mg L^{-1}). These values mean that the two-component model using

both 270 and 350 nm reduces the sum of squared deviations between observed and predicted [DOC] values by 66 and 82% respectively for the two wavelengths. If the single wavelength regression at 270 nm were used to estimate [DOC], 95% confidence limits would be ± 1.2 mg L^{-1} for 0 < [DOC] < 5 mg L^{-1}, ± 3.5 mg L^{-1} for 5 < [DOC] < 20 mg L^{-1}, and ± 6.6 mg L^{-1} for 20 < [DOC] < 80 mg L^{-1}.

We also compared the results obtained using the two-component model with those obtained with single-wavelength modelling for individual UK sites that had been sampled repeatedly. For each site, a linear regression analysis was performed for [DOC] vs A_{270}, and the regression parameters were used to predict [DOC]. The results of these site-specific calibrations were compared with the modelled values of [DOC] obtained from the global fits of both the single wavelength and two-component models (which included the data for the sites in question, but overwhelmingly more data for other sites). As shown in Table 3, even when the single wavelength model is calibrated to an individual site, the two-component global model (i.e. without site-specific calibration) gave higher values of r^2 in five of the six cases, with no overall difference in *rmsd*. Overall the single wavelength global model gave appreciably poorer results, although when the DOM characteristics are similar to the global average, and relatively unvarying, which was found for the River Gowan (Table 3), good results are of course obtained.

4.3. DOM quality

By mathematically fractionating DOM into components A and B, i.e. deriving the value of f_A for each sample, the model provides a simple measure of quality; the greater the proportion of component A, the more light-absorbing and hydrophobic is the material. To display the results, we plotted f_A against [DOC], although in principle there is no reason to expect any relationships because f_A depends upon ratios of absorbance to [DOC] not [DOC] itself. Indeed, for the UK samples, no relationships were evident (Figure 3), and the broad range of f_A therefore reveals considerable natural variability in DOM quality, when steps have been taken to sample disparate waters. The Canadian data set, although larger, refers to samples from more similar surface waters, and the values of f_A tend to increase with [DOC]. Samples from the Colorado data set with [DOC] > 2 mg L^{-1} had f_A values of around 0.3, while for more dilute samples, f_A fell in the range 0.1 to 1.0.

Another widely-used simple measure of DOM quality is $SUVA_{254}$ which also provides a measure of aromaticity and hydrophobicity (Chin et al., 1994; Weishaar et al., 2003). As might be expected, f_A and $SUVA_{254}$ are correlated. For UK samples with [DOC] > 3 mg L^{-1}

(results for lower concentrations were excluded because of noise in the data associated with taking ratios of low values), linear regression gave r^2 = 0.66 (Figure S4). A closer relationship between f_A and $SUVA_{254}$ is not found because values of f_A are derived after factoring out the non-UV-absorbing component C, which would otherwise contribute to the calculation of $SUVA_{254}$. Fraction B also contributes to $SUVA_{254}$. Note that if absorbance at 254 nm is known, the model-derived [DOC] can be used to estimate SUVA.

4.4. Aspects of practical application

In our preliminary study (Tipping et al., 2009) we reported that [DOC] in surface water samples from areas of industry and high human populations was underestimated by the model (fitted to results from relatively unpolluted sites). During the present research, we discovered that the measured [DOC] values for the non-conforming sites had been overestimated, due to inadequate sparging of acidified samples, and the consequent presence and inadvertent determination of DIC. This artefact has been reported in the literature (Findlay et al., 2010). To obtain the results reported here, we took care to ensure that all DIC was removed before the DOC analysis, and found that water samples from the locations that had previously appeared anomalous now fully conformed to the model.

For the UK samples, we examined possible dependences of model output on pH and Fe, both of which have been shown to influence spectra (Bloom and Leenheer, 1989; Maloney et al., 2005). The ratio of calculated to measured [DOC] showed a slight pH dependence, falling by 0.018 per pH unit ($p < 0.001$). There was no trend with total Fe concentration.

For filtered samples, the subtraction of the absorbance at 700 nm when processing the UK data was intended to account for instrumental drift (cf. Hernes et al., 2008), but in practice the correction had negligible effects. For unfiltered samples, the subtraction was explored as a possible means of correcting for minor turbidity, none of these samples being noticeably cloudy. The model was fitted to both the filtered and unfiltered UK samples, maintaining the constraints on E_{A2} and E_{B2} at 30 and 0 L g^{-1} cm^{-1}. For the filtered samples we obtained E_{A1} = 69.3 and E_{B1} = 15.4 L g^{-1} cm^{-1}, while for unfiltered samples the values were 67.7 and 20.7 L g^{-1} cm^{-1}. The values of [DOC_C] were 0.80 (filtered) and 1.23 (unfiltered) mg L^{-1}. Values of r^2 were 0.989 (filtered) and 0.986 (unfiltered), and *rmsd* was 1.1 mg L^{-1} for the filtered samples and 1.4 mg L^{-1} for the unfiltered samples

4.5. Testing on independent data

We tested the parameterised model by using it to predict [DOC] in water samples analysed in 11 independent studies (213 data points) from a range of locations, for which appropriate [DOC] and UV absorbance data were available. In most cases, absorbance data were not available at our favoured wavelengths (270 and 350 nm), and so we modified the model parameters using the spectra of Figure 2. As summarised in Table 4, we obtained satisfactory agreement between predicted and observed [DOC] in 9 cases (181 data points). For these data, 92% of the predictions fell within the 95% confidence limits presented above, which can be considered satisfactory performance given that both the modelled and directly-measured [DOC] values will have been subject to errors in each of the laboratories that produced the independent results. Furthermore, linear and logarithmic plots of predicted vs observed [DOC] (Figure S5) show no overall bias.

In two cases (32 data points) agreements were unsatisfactory. Data for Lake Pitkjärv in Estonia (Selberg et al., 2011) gave an overall ratio of predicted to observed [DOC] reasonably close to unity, but a high scatter in the results. For shallow lakes of the Yangzte basin (Zhang et al., 2005), the model appreciably underestimated [DOC], by an average factor of 2.1. The DOM in these lakes absorbs UV light extremely weakly, with an average extinction coefficient at 280 nm of only 6.5 L g^{-1} cm^{-1}, about half of the values estimated by Gondar et al. (2008) for autochthonous DOM produced in a UK eutrophic lake and in Lake Fryxell, Antarctica (Aiken et al., 1996).

We also used the parameterised model to predict our own data, separated according to their laboratories of origin. The UK and Canada data gave ratios close to unity and values of *rmsd* close to the overall average for all the satisfactory data sets. The Colorado data set gave an average ratio of 1.19, but the smallest *rmsd*, reflecting the generally low [DOC] for the Colorado samples, and consequently greater "noise" and more outliers. Nonetheless, these data make an important contribution to the parameterisation by defining low concentration conditions.

Selected independent data sets were used to test further the applicability of single wavelength predictive models. Following the approach described in Section 4.2, the Table 4 data sets for the California agricultural stream, Congo River, North Wales streams, Scottish & Welsh upland lakes, UK groundwaters and Yukon River (i.e. the larger collections for which the two-component model predictions were reasonably good) were each fitted to single wavelength models, using both the shorter and longer wavelength in each case (Table S1). For the shorter wavelength (254 nm in all cases) the site-specific optimised single wavelength model gave better fits (as judged by the *rmsd* in [DOC]) than the

366 globally-parameterised two-component model in four of the six cases, while the opposite
367 was the case for fits at the longer wavelengths (340 or 350 nm). Thus blind predictions
368 with the two-component model are at least as good as the optimised fitted values. When all
369 the data were combined and fitted to a single wavelength (254 nm) model, the *rmsd* in
370 [DOC] was 1.25 mg L^{-1}, greater than the value of 1.01 mg L^{-1} obtained with two-component
371 blind predictions, and the r^2 value was lower (0.93 vs 0.96). As found for the UK sites
372 referred to in Table 3, the parameter values for the single wavelength models varied
373 appreciably among the data sets (Table S1). Furthermore, the single wavelength model for
374 the combined data showed bias with overestimation of [DOC] at low and high
375 concentrations, and underestimation at intermediate concentrations, whereas the two-
376 component blind predictions fell centrally within the data.

377

5. Discussion

The results suggest that [DOC] and DOM quality in freshwaters can usefully be estimated simultaneously simply from optical absorbance data at two suitable wavelengths. The estimation of [DOC] by this approach is markedly superior to the use of a single wavelength, since the model is able to take into account variations in DOM extinction coefficients among samples. This is shown by the ranges of parameter values obtained when single wavelength models are derived for individual locations or collections of similar sites (Tables 3 and S1). Such variation means that, although good fits can often be obtained under restricted circumstances, a general single-wavelength model will always be imprecise. The use of data at two wavelengths takes advantage of the wavelength variation of the extinction coefficient range, thereby providing more accurate estimates of [DOC]. Equally important, the two-component model predictions are unbiased (Figure 1), in contrast to single wavelength fits (Sections 4.2 and 4.5, Figures 1 and S3). Even if the single wavelength model is calibrated to a specific site, the two-component approach, globally parameterised, gives results that are generally just as good (Tables 3 and S1, Sections 4.2 and 4.5). An important finding of the present work is that the previous apparent discrepancies for some UK waters strongly impacted by human activities (Tipping et al., 2009) were artefacts caused by incomplete removal of DIC during analysis. Although we deliberately sampled a wide range of water types in the UK work, it could still be argued that the data used for parameterisation refer to a somewhat limited range of environments, restricted to mid-latitudes of the northern hemisphere. However, the satisfactory testing on independent data sets (Table 4) covering a variety of sites from the tropics to the Arctic suggests wide applicability.

Our primary aim in this paper was to explore the performance of the two-component model, rather than to find the most precise and accurate means of predicting [DOC] from spectroscopic data. But we can mention comparative tests using the logarithmic multiple regression model of Fichot and Benner (2011), which they used to predict [DOC] in coastal seawater from absorbance values at 275 and 295 nm. We fitted their equation to data from the UK (i.e. the data set with the highest variability in DOM sources), and obtained r^2 = 0.944. When used to predict linear [DOC] values, the r^2 was 0.946 and the *rmsd* 1.9 mg L^{-1}, i.e. a poorer fit than provided by the two-component model (r^2 = 0.979; *rmsd* 1.2 mg L^{-1}). Moreover, the fitted Fichot-Benner model gave biased predictions, with underestimation at low and high [DOC] and overestimation at intermediate values. Therefore the two-component model can be considered superior, at least when applied to wide-ranging freshwater data. This is perhaps not surprising, given that Fichot and Benner

(2011) found it necessary to calibrate their model separately for different locations and also for different ranges of [DOC]. Of course, other multiple regression models involving absorbance values at different wavelengths might yield better results, and this is certainly worth exploring.

The two-component model accounts very well for the UV-absorbing components of DOM, which often dominate, but the analysis requires the assumption that there is a constant concentration of non-absorbing DOM, component C, with $[DOC_C]$ = 0.80 mg L^{-1}. By definition, this is not accessible to spectroscopic study, and the value of 0.80 mg DOC L^{-1} must be considered an overall average. It is likely that some of the deviations between measured and modelled [DOC] are due to variations in $[DOC_C]$. In further systematic studies, efforts might be made to estimate how $[DOC_C]$ depends upon site characteristics for example.

As explained in Section 4.3, the inclusion of $[DOC_C]$ in the model is one reason for the only approximate relationship between f_A and SUVA (Figure S4), since the calculation of SUVA would include $[DOC_C]$ as part of the total DOC. Thus while both f_A and SUVA reflect DOM aromaticity and hydrophobicity they are not directly related. Both can be regarded as empirical indices of hydrophobicity, derived from spectroscopic data, and they might be used in combination for interpretative purposes, along with other spectroscopic indicators such as spectral slopes and absorbance ratios (Zhang et al., 2005; Helms et al., 2008; Hernes et al., 2008; Spencer et al., 2009; Spencer et al., 2010; Fichot and Benner, 2011).

The wide range of f_A found for the UK samples (Figure 3) can be attributed to the deliberately wide range of water types chosen for sampling. For example, low f_A values were obtained for samples from a eutrophic lake during summer, dominated by autochthonous DOM, while peat drainage gave high values of f_A. Although the Canadian data set is more than twice the size of the UK one, the sampled water bodies were less variable, being mainly in boreal forested areas, within which systematic underlying control processes on DOC quality and quantity may cause f_A to be correlated with [DOC] (Figure 3), and explain why high f_A values are not found. The wide range of f_A values in the Colorado data set represents the high spatial and temporal variability in the chemical quality of the DOM in the alpine and subalpine watersheds. Average f_A values found for the independent data (Table 4) vary, the highest values being found for the sites in Wales and the southern USA, the lowest for groundwaters and the two Australian reservoirs. Obvious next steps to improve understanding of variations in f_A (and other spectroscopic variables) are more systematic studies on freshwaters with respect to land-use, soil and vegetation type, water

residence time etc, and application of the model to coastal seawater data (cf. Fichot and Benner, 2011).

We do not propose the two-component model as a replacement of conventional determinations of [DOC], but it could find significant applications in biogeochemical survey and monitoring work, and in screening the quality of supply water destined for treatment (Rosario-Ortiz et al., 2007). Combined with conventional analyses for quality control, it offers the possibility to increase greatly the number of samples analysed, while multi-wavelength monitoring would yield simultaneous derivation of DOM quality indicators through f_A, derived SUVA, and other spectroscopic indicators mentioned above. Furthermore, the approach might permit *in situ* monitoring of [DOC] and DOM quality in real time, if practical difficulties such as turbidity effects and sensor fouling can be overcome. Our results for the UK samples suggest that corrections using long-wavelength absorbance/scattering values might be feasible at least for waters of relatively low turbidity. The model might be improved by taking into account the minor pH dependence of estimated [DOC]. There is also a need to standardise the gathering of matched [DOC] and spectroscopic data, to minimise analytical differences among contributing laboratories.

The apparent two-component behaviour demonstrated here extends the observations made in work on the functional properties of DOM (Thacker et al., 2008; Gondar et al., 2008) in which variations in UV absorption and hydrophobicity were accounted for using mixing models with two end-members. Given the accepted high degree of complexity of DOM (Leenheer and Croué, 2003), this clearly cannot mean that there are actually only two discrete chemical compounds in DOM, but there might possibly be two sufficiently similar collections of molecules for two-component behaviour to emerge. Another possibility is a regular continuum of DOM spectra, since models with any number of components could equally well fit the data presented here, provided the extinction coefficients vary linearly between those of the end-members (see Figure S2). To understand how the two-component situation could come about, we need to know about the UV spectroscopic properties of DOM as initially formed (e.g. from terrestrial plant litter, via soil organic matter turnover, or from freshwater algae), and how the processes of fractionation (e.g. sorption by mineral soils) and modification (e.g. photolysis in lake waters) affect those spectra. Interpretation may depend upon how the UV spectra of DOM come about, either from the linear superposition of the spectra of many different chromophores (Bloom and Leenheer, 1989; Korshin et al., 1997), or, as proposed by DelVecchio & Blough (2004), from "intramolecular charge-transfer interactions between hydroxy-aromatic donors and quinoid acceptors formed by the partial oxidation of lignin precursors". In a more general sense,

482 the two-component behaviour of DOM UV spectra may help to constrain models of DOM
483 production, transport and modification in soil-water systems.

484

6. Conclusions

The UV optical absorbance spectra of a large number of freshwater samples varying in origin could be resolved into contributions from two components. Thus, the spectra appear to be simple combinations of the strongly-absorbing component A, and the weakly-absorbing component B, the former showing relatively more absorbance at longer wavelengths. The different spectral shapes of A and B permit their resolution in a given sample, simply from measurements of absorbance at two suitable wavelengths (e.g. 270 and 350 nm). Knowing the fraction of each component and their individual extinction coefficients (L g^{-1} cm^{-1}), and assuming a constant small (0.80 mg DOC L^{-1}) background level of non-absorbing DOM, then allows [DOC] to be calculated. The fraction of one or the other of A and B is an indicator of DOM quality; thus, the greater is fraction A, the more light-absorbing and hydrophobic is the DOM. The results offer the prospect of rapid and inexpensive determination of [DOC] and DOM quality, including *in situ* field monitoring. They also raise questions about the origins of DOM spectra.

Acknowledgements

We thank the staff of the CEH Analytical Chemistry Laboratory for assistance with DOC analysis, and M.M.DeVille, K.J.Dinsmore, B.A.Dodd, K.L.Hockenhull, A.Stockdale, C.D.Vincent (all CEH), P.E.Cook, G.A.L.Howard, P.D.Morey and Y.J.Morey for help with the UK sampling. For the Canadian samples, we thank L.Aspden and C.Guay for sampling, H.Broadbent and M.Budd for DOC analysis. The comments of J.Koch (USGS) on a draft version are appreciated. We are grateful to R.Spencer (Woods Hole Oceanographic Institution, USA) for the provision of raw data from the Yukon and Congo Rivers, and to C.D.Evans and D.T.Monteith (both CEH) for providing data from the Euro-limpacs project. This work was funded in the UK by the Natural Environment Research Council (NERC), in Canada by an NSERC Strategic Grant, and in Colorado by the National Science Foundation's (NSF) Niwot Ridge Long Term Ecological Research Program (DEB-0423662). The paper benefitted from the comments of an anonymous referee.

Supplementary Data

Table S1 shows results of single-wavelength modelling of independent data sets. Figures S1–S5 comprise plots of relationships among spectroscopic variables, results from single-wavelength modelling, relationships between SUVA and f_A, and predicted and observed [DOC] from independent data sets.

REFERENCES

Aiken, G.; McKnight D.; Harnish, R.; Wershaw, R., 2004. Geochemistry of aquatic humic substances in the Lake Fryxell Basin, Antarctica. Biogeochem. 34 (3), 157-188.

Bloom, P.R.; Leenheer, J.A., 1989. Vibrational, electronic, and high-energy spectroscopic methods for characterizing humic substances. In Hayes, M.H.B.; MacCarthy P.; Malcolm, R.L.; Swift, R.S. (eds) Humic Substances II. In Search of Structure. Wiley, Chichester.

Cabaniss, S.E.; Shuman, M.S., 1987. Synchronous fluorescence spectra of natural waters: tracing sources of dissolved organic matter. Marine Chem. 21 (1), 37-50.

Chen, W.; Westerhoff, P.; Leenheer, J. A.; Booksh, K., 2003. Fluorescence excitation-emission matrix regional integration to quantify spectra for dissolved organic matter. Environ. Sci. Technol. 37 (24), 5701-5710.

Chen, W.; Zhao, Z,; Koprivnjak, J.F.; Perdue, E.M., 2002. A mechanistic study of the high-temperature oxidation of organic matter in a carbon analyzer. Marine Chem., 78 (4), 185–196.

Chin, Y.-P.; Aiken, G.; O'Loughlin, E., 1994. Molecular weight, polydispersity, and spectroscopic properties of aquatic humic substances. Environ. Sci. Technol. 28 (11), 1853-1858.

Coble, P.G.; Green, S.A.; Blough, N.V.; Gagosian R.B., 1990. Characterization of dissolved organic-matter in the Black Sea by fluorescence spectroscopy. Nature 348 (6300), 432-435.

Cory, R.M.; McKnight, D.M., 2005. Fluorescence spectroscopy reveals ubiquitous presence of oxidized and reduced quinones in dissolved organic matter. Environ. Sci. Technol. 39 (21), 8142-8149.

Del Vecchio, R.; Blough, N.V., 2004. On the origin of the optical properties of humic substances. Environ. Sci. Technol. 38 (14), 3885-3891.

Downing, B.D.; Boss, E.; Bergamaschi, B.A.; Fleck, J.A.; Lionberger, M. A.; Ganju, N.K.; Schoellhamer, D.H.; Fujii, R., 2009. Quantifying fluxes and characterizing compositional changes of dissolved organic matter in aquatic systems in situ using combined acoustic and optical measurements. Limnol. Oceanogr. Methods 7(1), 119-131.

Edwards, A. C.; Hooda, P. S.; Cook, Y., 2001. Determination of nitrate in water containing dissolved organic carbon by ultraviolet spectroscopy. Intern. J. Environ. Anal. Chem. 80 (1), 49 - 59.

Evans, C.D.; Freeman, C.; Cork, L.G.; Thomas, D.N.; Reynolds, B.; Billett, M.F., Garnett, M.H.; Norris, D., 2007. Evidence against recent climate-induced destabilisation of soil carbon from ^{14}C analysis of riverine dissolved organic matter. Geophys. Res. Lett. 34, L07407, doi:10.1029/2007GL029431.

Fichot, C.G.; Benner, R., 2011. A novel method to estimate DOC concentrations from CDOM absorption coefficients in coastal waters. Geophys. Res. Lett. 38, doi:10.1029/2010GL046152.

556 Findlay, S.; McDowell, W.H.; Fischer, D.; Pace, M.L., Caraco, N.; Kaushal, S.S.; Weathers, K.C., 2010.
557 Total carbon analysis may overestimate organic carbon content of fresh waters in the presence
558 of high dissolved inorganic carbon Limnol. Oceanogr. Methods 8(5), 196–201.

559 Gondar, D.; Thacker, S.A.; Tipping, E; Baker, A., 2008. Functional variability of dissolved organic
560 matter from the surface water of a productive lake. Water Res. 42 (1-2), 81-90.

561 Grieve, I.C., 1984. Determination of dissolved organic matter in streamwater using visible
562 spectrometry. Earth Surf. Proc. Landforms 10 (1), 75-78.

563 Helms, J.R.; Stubbins, A.; Ritchie, J.D.; Minor, E.C.; Kieber, D.J.; Mopper, K., 2008. Absorption
564 spectral slopes and slope ratios as indicators of molecular weight, source, and photobleaching
565 of chromophoric dissolved organic matter. Limnol. Oceanogr. 253 (3), 955-969.

566 Hernes, P.J.; Spencer, R.G.M.; Dyda, R.Y.; Pellerin,B.A.; Bachand, P.A.M., Bergamaschi, B.A., 2008.
567 The role of hydrologic regimes on dissolved organic carbon composition in an agricultural
568 watershed Geochim. Cosmochim. Acta 72 (21), 5266–5277.

569 Hessen, D.O.; Tranvik, L.J., 1998. Aquatic Humic Substances. Springer, Berlin.

570 Izbicki, J.A.; Pimentel, I.M.; Johnson, R.; Aiken, G.R., Leenheer, J., 2007. Concentration, UV-
571 spectroscopic characteristics and fractionation of DOC in stormflow from an urban stream,
572 Southern California, USA. Environ. Chem. 4 (1), 35-48.

573 Koprivnjak, J.F., 2007. Natural Organic Matter: Isolation and Bioavailability. Ph.D. Dissertation,
574 Georgia Institute of Technology, Atlanta, GA.

575 Koprivnjak, J.F.; Dillon, P.J.; Molot, L.A., 2010. Importance of CO_2 evasion from small boreal streams.
576 Global Biogeochem. Cycles 24 (4), GB4003, doi:10.1029/2009GB003723.

577 Korshin, A.; G.V.; Li, C.-W.; Benjamin, M.M., 1997. Monitoring the properties of natural organic
578 matter through UV spectroscopy: a consistent theory. Water Res. 31 (7), 1787-1795.

579 Kullberg, A.; Bishop, K.H.; Hargeby, A.; Jansson, M.; Petersen, R.C., 1993. The ecological
580 significance of dissolved organic carbon in acidified waters. Ambio 22 (5), 331-337.

581 Leenheer, J.A.; Croué, J.-P., 2003. Characterizing aquatic dissolved organic matter. Environ. Sci.
582 Technol. 37 (1), 19A-26A.

583 Lewis, W.M.; Canfield, D., 1977. Dissolved organic carbon in some dark Venezuelan waters and a
584 revised equation for spectrophotometric determination of dissolved organic carbon. Archiv
585 Hydrobiol. 79 (4), 441-445.

586 Liu, S; Lim, M; Fabris, R.; Chow, C.W.K.; Drikas, M.; Korshin, G.; Amal, R., 2010. Multi-wavelength
587 spectroscopic and chromatography study on the photocatalytic oxidation of natural organic
588 matter. Water Res. 44 (8), 2525-2532.

589 Lofts, S.; Tipping, E.; Hamilton-Taylor, J., 2008. The chemical speciation of Fe(III) in freshwaters.
590 Aquat. Geochem. 14 (4), 337-358.

591 Loiselle, S.A.; Bracchini, L.; Dattilo, A.M.; Ricci, M.; Tognazzi, A., 2009. Optical characterization of
592 chromophoric dissolved organic matter using wavelength distribution of absorption spectral
593 slopes. Limnol. Oceanogr. 54 (2), 590–597.

594 Maloney, K.O.; Morris, D.P.; Moses, C.O.; Osburn, C.L., 2005. The role of iron and dissolved organic
595 carbon in the absorption of ultraviolet radiation in humic lake water. Biogeochem. 75 (3),
596 393-407.

597 Mattson, J.S.; Smith, C.A.; Jones, T.T.; Gerchakov, S.M.; Epstein, B.D., 1974. Continuous monitoring
598 of dissolved organic matter by UV-visible photometry. Limnol. Oceanogr. 19 (3), 530-535.

599 Menzel, D.W.; Vaccaro, R.F., 1964. The measurement of dissolved organic and particulate carbon in
600 seawater. Limnol. Oceanogr. 9 (1), 138-142.

601 Miller, M.P.; McKnight, D.M., 2010. Comparison of seasonal change in fluorescent dissolved organic
602 matter among aquatic lake and stream sites in the Green Lakes Valley. J. Geophys. Res. 115
603 (7), doi:10.1029/2009JG000985.

604 Monteith, D.T.; Stoddard, J.L.; Evans, C.D.; de Wit, H.A.; Forsius, M.; Høgåsen, T.; Wilander, A.;
605 Skjelkvåle, B.L.; Jeffries, D.S.; Vuorenmaa, J.; Keller, B.; Kopácek, J.; Vesely, J., 2007.
606 Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry.
607 Nature 450 (7169), 537-540.

608 Moore, T.R., 1987. An assessment of a simple spectrophotometric method for the determination of
609 dissolved organic carbon in freshwaters. New Zealand Journal of Marine and Freshwater
610 Research 21 (4), 585-589.

611 Pastor, J.; Solin, J.; Bridgham, S.D.; Updegraff, K.; Harth, C.; Weishampel, P.; Dewey, B., 2003.
612 Global warming and the export of dissolved organic carbon from boreal peatlands. Oikos 100
613 (2), 380-386.

614 Perdue, E.M.; Gjessing, E.T., 1990. Organic Acids in Aquatic Ecosystems. Wiley, New York.

615 Rosario-Ortiz, F.L.; Snyder, S.A.; Suffet, I.H., 2007. Characterization of dissolved organic matter in
616 drinking water sources impacted by multiple tributaries, Wat. Res. 41 (18), 4115-4128.

617 Sandford, R.C.; Bol, R.; Worsfold, P.J., 2010. In situ determination of dissolved organic carbon in
618 freshwaters using a reagentless UV sensor. J. Environ. Monit. 12 (9), 1678-1683.

619 Selberg, A.; Malle Viik, M.; Ehapalu, K.; Tenno, T., 2011. Content and composition of natural organic
620 matter in water of Lake Pitkjärv and mire feeding Kuke River (Estonia). J. Hydrol. 400 (1-2),
621 274-280.

622 Spencer, R.G.M.; Aiken, G.R., Butler, K.D.; Dornblaser, M.M.; Striegl, R.G.; Hernes, P.J., 2009.
623 Utilizing chromophoric dissolved organic matter measurements to derive export and reactivity

624 of dissolved organic carbon exported to the Arctic Ocean: A case study of the Yukon River,
625 Alaska. Geophys. Res. Lett. 36, L06401, doi:10.1029/2008GL036831.

626 Spencer, R.G.M.; Hernes, P.J.; Ruf, R.; Baker, A.; Dyda, R.Y.; Stubbins, A.; Six, J., 2010. Temporal
627 controls on dissolved organic matter and lignin biogeochemistry in a pristine tropical river,
628 Democratic Republic of Congo. J. Geophys. Res. 115 (8), G03013,
629 doi:10.1029/2009JG001180.

630 Thacker, S.A.; Tipping, E.; Baker, A; Gondar, D., 2005. Development and application of functional
631 assays for freshwater dissolved organic matter. Water Res. 39 (18), 4559-4573.

632 Thacker, S.A.; Tipping, E.; Gondar, D.; Baker, A., 2008. Functional properties of DOM in a stream
633 draining blanket peat. Sci. Tot. Environ. 407 (1), 556-573.

634 Thomas, O.; Burgess, C., 2007. UV-Visible Spectrometry of Water and Wastewater. Elsevier,
635 Amsterdam.

636 Tipping, E.; Corbishley, H. T.; Koprivnjak, J.-F.; Lapworth, D. J.; Miller, M. P.; Vincent, C. D.;
637 Hamilton-Taylor, J., Quantification of natural DOM from UV absorption at two wavelengths.
638 Environ. Chem. 2009, 6 (6), 472-476.

639 Weishaar, J.L.; Aiken, G.R.; Bergamaschi, B.A.; Fram, M.S.; Fujii, R.; Mopper, K., 2003. Evaluation of
640 specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of
641 dissolved organic carbon. Environ. Sci. Technol. 37 (20), 4702-4708.

642 Worrall, F.; Harriman, R.; Evans, C.D.; Watts, C.D.; Adamson, J.K.; Neal, C.; Tipping, E.; Burt, T.;
643 Grieve, I.C.; Montieth, D.; Naden, P.S.; Nisbet, T.; Reynolds, B.; Stevens, P., 2004. Trends in
644 dissolved organic carbon in UK rivers and lakes. Biogeochem. 70 (3), 369-402.

645 Yang, H.; Monteith, D.T.; Flower, R.J.; Battarbee, R.W., 2009. Deliverable 158: DOC mediated light
646 distribution in the Round Loch of Glenhead: an evaluation of the use of epipelic diatoms as
647 indicators of changing concentrations of DOC. Project Report under Euro-limpacs - Project no.
648 GOCE-CT-2003-505540. Environmental Change Research Centre, University College London,
649 Gower St., London, UK.

650 Zhang, Y.; Qin, B.; Zhang, L.; Zhu, G.; Chen, W., 2005. Spectral absorption and fluorescence of
651 chromophoric dissolved organic matter in shallow lakes in the middle and lower reaches of the
652 Yangtze River. Journal of Freshwater Ecology 20 (3), 451-459.

653

654 Table 1. Summary of sample chemistries.

	UK	Canada	US	All
No. of samples	427	1154	117	1698
DOC (mg L^{-1})	0.9 - 54.6	1.2 - 74.5	0.6 - 10.1	0.6 - 74.5
pH	3.0 - 8.5	3.4 - 7.4	5.9 - 7.4	3.0 - 8.5
Conductivity(µS cm^{-1})	17 - 1959	8 - 316	7 - 98	7 - 1959

655

656

657

658 Table 2. Summary of model parameters. The extinction coefficients (*E*) have units of L g^{-1}
659 cm^{-1}, $[DOC_C]$ is in mg L^{-1}.

660

	Component A	Component B
E_{270nm}	77.1	21.3
E_{350nm}	69.3	15.4
E_{254nm}	63.9	12.0
E_{280nm}	61.1	10.6
E_{285nm}	47.6	4.7
E_{310nm}	34.1	0.7
E_{340nm}	30.0	0.0
E_{355nm}	27.9	0.0
$[DOC_C]$	0.80	

661

662

663 Table 3. Comparison of predictions of [DOC] (mg L^{-1}) using single wavelength (270 nm)
664 and two-component models. For site-specific single wavelength fitting, linear regression
665 (slope *m*, intercept *c*) was used to relate [DOC] to Abs_{270} at each site, then the relationship
666 was used to back-calculate [DOC]. The single wavelength global predictions were made
667 using the linear regression fits of all 1698 data points. Numbers of samples are indicated by
668 *n*.

669

			single wavelength models						two-component global prediction	
			site-specific fit				global prediction			
Site	*n*	mean [DOC]	*m*	*c*	r^2	*rmsd*	r^2	*rmsd*	r^2	*rmsd*
River Conder	28	5.9	20.3	1.6	0.94	0.6	0.94	0.8	0.97	0.6
Cottage Hill Syke	23	29.3	18.5	3.8	0.70	2.9	0.70	4.9	0.88	3.0
Esthwaite Water	12	3.7	12.1	2.6	0.40	0.3	0.40	0.4	0.58	0.3
River Gowan	25	3.4	21.7	1.3	0.94	0.4	0.94	0.5	0.94	0.4
Rough Sike	30	14.5	19.0	0.8	0.96	1.3	0.96	4.0	0.99	1.5
Troutbeck	24	12.2	17.8	2.0	0.95	1.3	0.95	3.3	0.98	1.0

670

671

672 Table 4. Comparison of [DOC] predicted with the parameterised model and determined by conventional analysis.

Sample location and type	n^a	[DOC] range mg L^{-1}	λ_1 nm	λ_2 nm	mean ratio[b]	*rmsd* mg L^{-1}	f_A[c] mean	ref.
Australian reservoirs	2	3.2 - 11.6	270	350	1.00	0.71	0.18	Liu et al., 2010
California agricultural stream	29	2.1 - 7.1	254	350	0.96	0.70	0.34	Hernes et al., 2008
California urban stream	4	4.4 - 10.6	254	310	0.79	1.37	0.33	Izbicki et al., 2007
Congo River	28	5.2 - 9.0	254	350	0.91	0.69	0.38	Spencer et al., 2010
North Wales streams[d]	30	1.1 - 28.0	254	340	1.14	1.64	0.41	Evans et al., 2007
Scottish & Welsh upland lakes	31	0.6 - 8.1	254	340	1.14	0.57	0.51	Yang et al., 2009
SE USA rivers[i]	6	1.8 – 9.8	270	340	1.00	0.53	0.42	Koprivnjak, 2007
UK groundwaters	12	1.1 - 11.0	254	340	0.96	0.82	0.19	Tipping et al., 2009
Yukon River	39	2.6 - 17.0	254	350	0.95	1.22	0.31	Spencer et al., 2009
All the above	*181*	*0.6 - 28.0*	-	-	*1.01*	*1.01*	*0.37*	-
Lake Pitkjärv, Estonia[e]	10	16.5 - 45.6	254	285	1.10	3.93	0.44	Selberg et al., 2011
Yangtze River lakes	22	2.7 - 10.1	280	355	0.47	3.27	0.17	Zhang et al., 2005
UK data[f]	426	0.9 - 54.6	270	350	1.03	1.37	0.50	this study
Canada data[g]	1148	1.2 - 74.5	270	350	0.99	0.98	0.39	this study
Colorado data[h]	106	0.6 - 10.1	270	350	1.11	0.25	0.36	this study

673 [a] no. of data points [b] $[DOC]_{pred}/ DOC]_{obs}$ [c] for $[DOC]_{obs} > 2$ mg L^{-1} Outliers omitted: [d]1, [e]3, [f]1, [g]6, [h]11

674 [i] includes two data points determined on samples concentrated by reverse osmosis

675

676 **Figure captions**

677 Figure 1. Modelled *vs* observed [DOC] after full optimisation (filtered samples); (a) all data
678 except for 18 omitted outliers (see text); (b) data for [DOC] $\leq$ 10 mg L^{-1}; (c) data plotted
679 logarithmically, with one point omitted because the calculated [DOC] was < 0. Solid lines
680 indicate 1:1 correspondence.

681

682 Figure 2. UV spectra of DOM components A and B. The extinction coefficients and
683 wavelengths derived from the parameterisation are indicated.

684

685 Figure 3. Variation of f_A with log [DOC]; (a) United Kingdom data, (b) Ontario data, (c)
686 Colorado data. Only values for [DOC] > 1.5 mg L^{-1} are plotted, because of the high degree
687 of uncertainty at lower concentrations. Three UK values with calculated f_A > 1 are omitted.

688

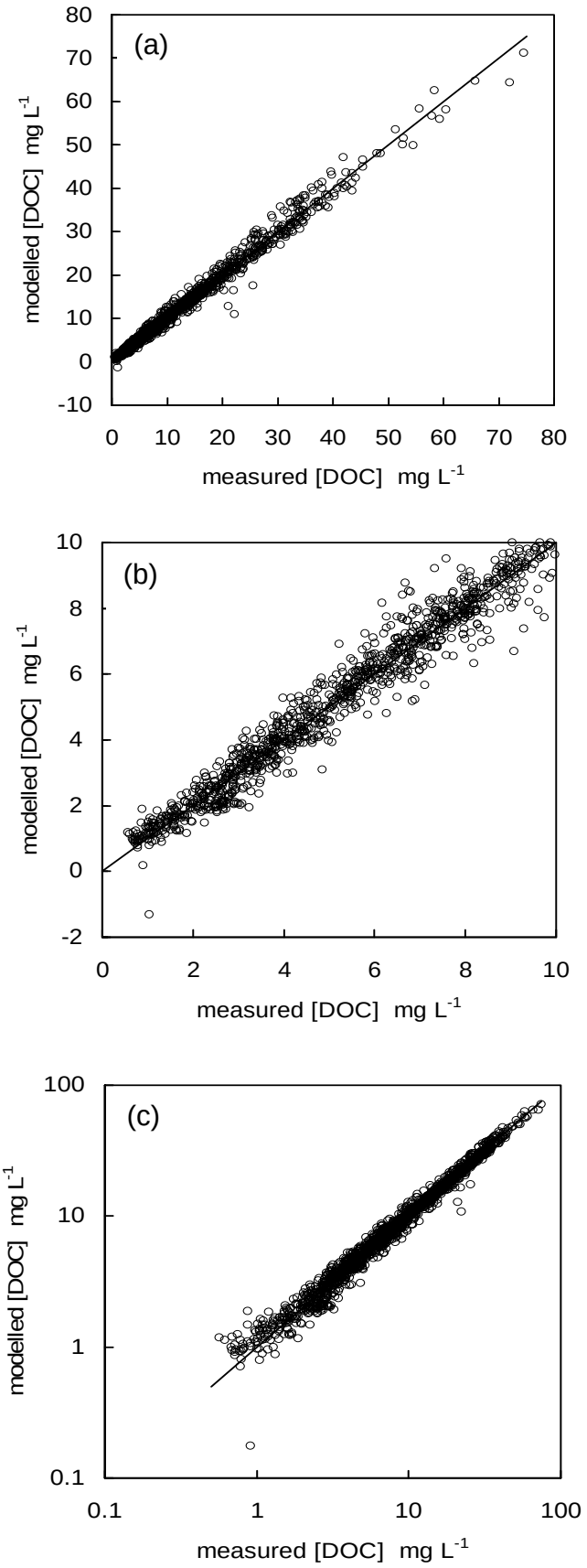
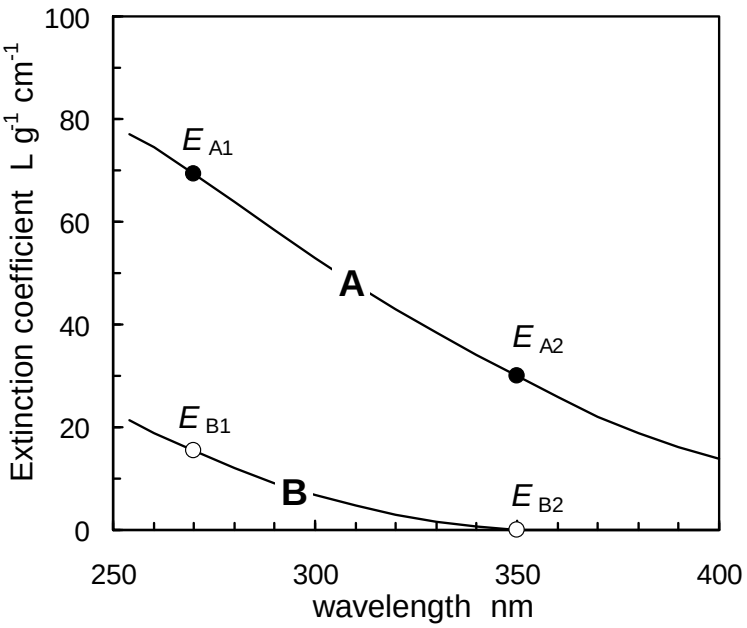

(a)
modelled [DOC] mg L-1
measured [DOC] mg L-1
80
70
60
50
40
30
20
10
0
-10
0 10 20 30 40 50 60 70 80
(b)
modelled [DOC] mg L-1
measured [DOC] mg L-1
10
8
6
4
2
0
-2
0 2 4 6 8 10
(c)
modelled [DOC] mg L-1
measured [DOC] mg L-1
100
10
1
0.1
0.1 1 10 100

Figure 1


100
80
60
40
20
0
Extinction coefficient L g-1 cm-1
E A1
A
E A2
E B1
B
E B2
250
300
350
400
wavelength nm

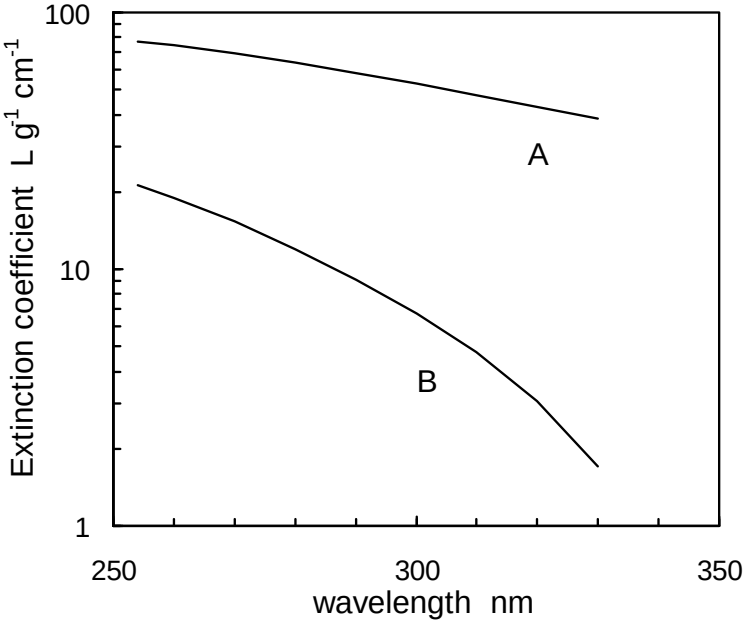
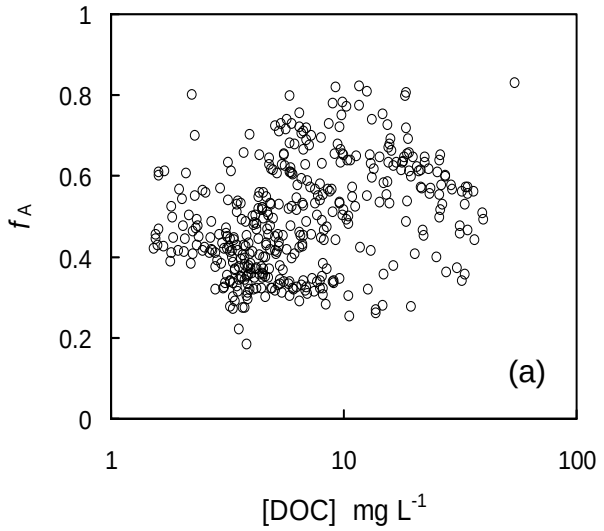
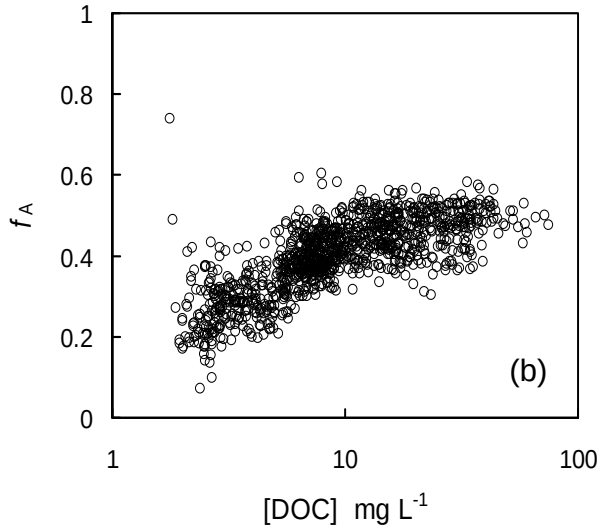

100
10
1
Extinction coefficient L g-1 cm-1
A
B
250
300
350
wavelength nm

Figure 2


1
0.8
0.6
0.4
0.2
0
f_A
(a)
1
10
100
[DOC] mg L^{-1}


1
0.8
0.6
0.4
0.2
0
f_A
(b)
1
10
100
[DOC] mg L^{-1}

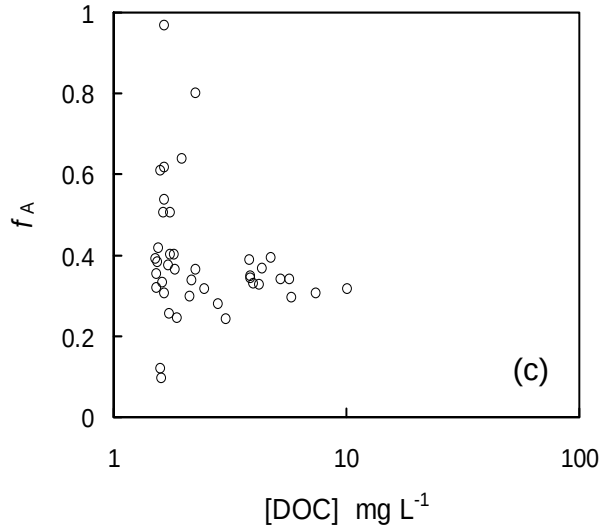

1
0.8
0.6
0.4
0.2
0
f_A
(c)
1
10
100
[DOC] mg L^{-1}

Figure 3