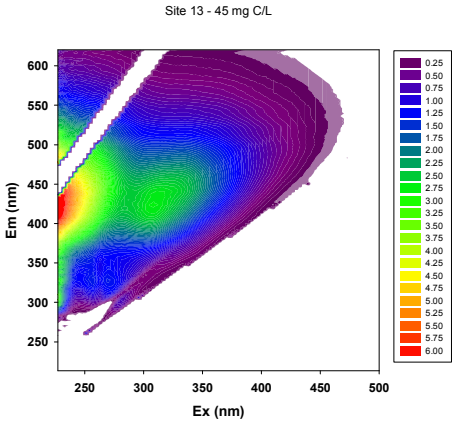
Environmental Science
Processes & Impacts

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

rsc.li/process-impacts



The robustness of empirical models derived from correlation studies needs to be independently verified before being relied on.

Exploring the relationship between the optical properties of water and the quality and quantity of dissolved organic carbon in aquatic ecosystems: strong correlations do not always mean strong predictive power.

Darren. S. Baldwin*[a] and William Valo[b]

[a]CSIRO Land and Water Flagship and the Murray-Darling Freshwater Research Centre, La Trobe University, PO box 991 Wodonga Vic 3689, Australia

[b] Latrobe University, Bundoora, Victoria

Electronic Supplementary Information (ESI) available

Abstract

Optical absorbance and fluorescence spectroscopies have the potential to play an important role in monitoring aquatic ecosystems. This paper explores the relationship between the optical absorption and fluorescence characteristics of water samples taken from 18 sites (spanning a range of aquatic environments including lowland rivers, small and large dams, and floodplain wetlands) with their dissolved organic carbon (DOC) concentration and 28 day bioavailability Both optical absorbance in the ultraviolet region and fluorescence above excitation wavelengths of about 240 nm and emission wavelengths above about 350 were correlated with DOC concentration ($r^2 > 0.8$). The initial (pre-incubation) optical absorbance in the UV region ($r^2 \approx 0.7 – 0.8$) and fluorescence attributed to both 'humic like' ($r^2 = 0.84$) and 'tryptophan-like' ($r^2 = 0.87$) fluorophores correlated with DOC bioavailability. These correlations were used to develop empirical linear models relating the initial optical properties of water with DOC quantity and quality. The robustness of these models was then tested against a second suite of water samples from 12 different sites, collected independently of those used in the model development. Although based on strong correlations, the empirical models were not particularly good at predicting the bioavailability of DOC in the model validation samples. We suggest that one of the reasons for the low predictive power of the models is that the strong correlations observed between DOC bioavailability and optical absorbance in the UV region or fluorescence and are co-incidental rather than causal. Changes in UV-absorbance or fluorescence during the incubation experiments are not consistent with changes in DOC concentration. One of the best predictors of DOC bioavailability is the initial concentration of DOC. We argue the strong correlation between DOC bioavailability and initial fluorescence intensity or UV absorption simply reflects the strong correlation between initial DOC concentration and initial fluorescence intensity or UV absorption. We argue that unless there is an underlying causal relationship between two components (the component of interest and a surrogate measure for that component) care should be taken in extrapolating correlative models beyond the data set used to create them.

Environmental Impact

Optical absorbance and fluorescence spectroscopies have the potential to play an important role in monitoring aquatic ecosystems. However, while there are strong correlations observed between fluorescence intensity or optical absorption of water and the quality and quantity of DOC in that water, care should be used in using those relationships to predict the quality or quantity of DOC from sites not used in the model development. The robustness of models needs to be verified and then periodically assessed to determine their on-going suitability.

Key words: parallel factor analysis, excitation-emission matrix, size-exclusion chromatography, blackwater, $SUVA_{254}$, E_2:E_3, E_2:E_4

1. Introduction

Dissolved organic carbon (DOC) refers to the carbon component of the complex mixture of organic compounds that are dissolved in water. The actual composition of DOC will vary in time and space[1], and can contain compounds derived from the terrestrial environment (allochthonous carbon), produced in the water body (autochthonous carbon) or from human sources. A fraction of the compounds comprising DOC can be used by bacteria during respiration i.e. is bioavailable.[2] These bacteria in turn can be consumed by other, larger, organisms and hence DOC can represent an important basal resource for aquatic food webs.[3,4] However, during the aerobic metabolism of DOC bacteria can also consume oxygen.[2] If the amount of oxygen consumed by bacteria during the decomposition of DOC is greater than the amount of oxygen that can be supplied to the system (usually from the atmosphere or from photosynthesis) then the oxygen levels in the water body will begin to decline leading to hypoxia (low levels of dissolved oxygen (DO) - typically less than 3 mg O_2/l) or even anoxia (no dissolved oxygen in the water column).[2] While hypoxia from microbial respiration is often more associated with near shore coastal environments,[5] there are numerous rivers worldwide that experience hypoxic events.[6,7] Often the onset of hypoxia is associated with the inundation of dry river channels or lowland river floodplains.[8] On inundation, plant litter that has accumulated in the dry river channel or floodplain rapidly leaches DOC producing a pulse of DOC (and other nutrients) which can then be used in microbial respiration. For example, the southern Murray-Darling Basin (MDB) has experienced a number of hypoxic events over the last few decades.[2,7,8] The most dramatic example of a hypoxic blackwater in the southern MDB occurred during the late spring and summer of 2010–11.[7] The southern MDB experienced an extended period of drought starting from the mid-1990s. Significant flooding occurred during the late austral spring of 2010, inundating large areas of floodplain, most of which hadn't been flooded since 1996. The resultant pulse of DOC led over 2000 km of the Murray River channel and its tributaries becoming hypoxic; the hypoxia persisted at some

locations for up to six months,[7] and resulted in extensive mortality of aquatic life including large-bodied native fish and large crustaceans.[9]

One of the questions posed by river managers during that, and earlier, hypoxic events was how long the hypoxia would persist.[6,10] At the heart of this question was how much of the DOC remaining in solution was readily bioavailable. Currently the only way to accurately assess bioavailability of DOC is to perform an incubation experiment where changes in DO and/or DOC in a water sample are measured over a period of time, typically five days for the standard Biological Oxygen Demand experiment.[11] To improve the management of hypoxic blackwater events into the future it would be highly desirable to be able to accurately predict the bioavailability of DOC in real time and preferably *in-situ*.

UV-visible spectroscopy and fluorescence spectroscopy both have the potential to be useful techniques for studying the dynamics of DOC in aquatic ecosystems.[12,13] Measurements are rapid (minutes), non-destructive and, in the case of fluorescence spectroscopy, very sensitive. Because of these properties, both optical absorbance and fluorescence techniques have been applied to the study of DOC in aquatic ecosystems.[12,13] For example, both UV- visible spectroscopy[13] and fluorescence spectroscopy[14] have been used to estimate the concentration of DOC in water. Indeed in-situ probes are commercially available for the determination of DOC in water based on both UV absorbance and fluorescence intensity. A number of the optical properties of water have also been used to infer bioavailability of DOC. $SUVA_{254}$, the absorbance of light at 254 nm (A_{254}) normalised to DOC concentration has been shown to be highly correlated with the degree of aromaticity of DOC,[15] as well as negatively correlated to bioavailability.[16] E_2:E_3 ratio (the ratio of optical absorbance at about 250 and 365 nm) has also been correlated to degree of aromaticity of DOC,[17] and also linked to the degree of humification.[18] Similarly E_2:E_4 (the ration of optical absorbance at about 250 nm and absorbance at a specific wavelength between 435 and 465 nm) has also been associated with humification of

DOC.[19] It is assumed that the more aromatic or humic-like the DOC, the less bioavailable it will be.[20]

Fluorescence intensity has been shown to be highly correlated with with microbial density in potable water,[21] biological oxygen demand in rivers,[22-24] and bioavailable dissolved organic carbon in soil-water from wetlands, bogs and forests from southern Alaska.[25]

One potential criticism of the correlation approach is that while there may be strong correlations between the optical properties of water and specifically determined properties of DOC (e.g. concentration, aromaticity or bioavailability), the predictive capacity of those correlations (models) are rarely validated against an independent data set. The purpose of the current study is to determine if it is possible to rapidly assess the bioavailability of DOC in order to assist river and floodplain managers in assessing the risks posed by blackwater plumes to downstream aquatic environments. This paper explores the empirical relationship between the optical absorption and fluorescence characteristics of aquatic water samples with DOC concentration and DOC bioavailability. The robustness of these empirical relationships (models) is then tested against a second suite of water samples collected independently of those used in the model development.

2. Materials and Methods

2.1 Field Sampling

The field survey was carried out in two phases. In the first phase (the model development phase) a total of 18 sites (Table 1) were sampled to determine the relationship between the absorbance and fluorescence properties of water and the concentration and bioavailability of DOC. In the second phase (the model validation phase) an additional 12 sites (Table 2) not sampled in the first phase were sampled to validate the relationships developed in the first phase. All sites were located within a 200 km radius of Albury-Wodonga in south-eastern Australia (36.08° S. 146.91° E)

Sites were chosen on an *ad hoc* basis and represented a range of water body types (including rivers, wetlands and artificial water bodies) with potential differences in both the source(s) and concentration of DOC. Triplicate water samples were taken at each location using pre-soaked 9 L plastic buckets. Samples for initial (*in-situ*) DOC absorbance and fluorescence measurements were filtered through pre-washed 0.45 µm pore-sized acetate syringe filters (Cameo) and stored in 30 mL polystyrene tubes. They were transported back to the laboratory on ice. Absorbance and fluorescence spectroscopy were carried out within 24 hours of sampling; samples for determining DOC concentration were frozen prior to analysis. Three sampling blanks, using Milli-Q grade water transported into the field and sampled in the same way as the field samples, were also taken. Dissolved oxygen, pH, temperature and electrical conductivity were determined at each site using either a Hydrolab Quanta (Hach, USA; used during the first phase of sampling) or a Hydrolab Surveyor attached to a MS-5 probe (Hach, USA; used during the second phase of sampling). Turbidity was measured in the field during the second phase of sampling using the Hydrolab Surveyor, but because of the failure of the turbidity probe, the turbidity in samples taken in the first phase of sampling were returned to the laboratory and turbidity measured using a Hach 2100AN IS turbidimeter.

2.2 DOC Bioavailability Assays

In both the model development phase and the model validation phase triplicate unfiltered 500 ml samples from each site were transferred into acid-washed 1 l glass screw cap bottles that were covered with aluminium foil to minimise exposure to light. The bottles were transferred back to the laboratory at ambient temperature and then stored under low ambient light conditions in a constant temperature room at 20 ^{o}C for 28 days. Because the experiments were designed to see if we could *a priori* predict the bioavailability of DOC *in situ* the experiments were not augmented with additional nutrients. Fluorescence spectra and DOC were determined on 0.45 µm filtered samples within 24 hours of the original samples being taken and then again after 28 days. The bioavailable carbon fraction (Δ DOC) was estimated from

the difference between the concentration of DOC in the samples stored for 28 days and the DOC in the original sample.

2.3 Chromatographic separation of DOC

To determine how the individual fluorophores in a natural water sample (from Site 3 Cookies Dam - see Table 1) change over time, samples from a separate bioavailability assay were separated using size-exclusion chromatography.[26] Water from the dam was sampled in the same way as the field survey. Samples for initial DOC and fluorescence spectroscopy were filtered through pre-washed 0.45µM PES filters. Triplicate 4-l water samples were returned to the laboratory, coarse particulate matter was removed from the samples by passing them through a 0.053 mm stainless steel sieve; and then the samples transferred to 5 l aluminium-foil-wrapped glass bakers. Triplicate blanks, consisting of 4 l of Milli-Q grade water in 5-l aluminium-foil-wrapped glass beakers were also prepared. All samples were stored under low-light conditions at 20 ^{o}C in a constant temperature room. Samples for chromatographic separation of fluorophores were filtered through pre-washed 0.45 µm pore-size PES filters prior to analysis. Because of the time restrictions, chromatography separation of fluorophores could only be performed on one replicate.

Fluorophores were separated by low-pressure size-exclusion chromatography using a Hi-Prep 16/60 Sephacryl S-100 column (GE Healthcare; with a void volume (V_0) determined with blue dextran of 38mL and a total column volume (V_t) of 120 mL), a Pharmacia LKB pump and a Gradifrac gradient controller and fraction collector. The mobile phase was 0.15 M NaCl/0.05M Tris (Sigma) adjusted to 7.5 with HCl. The flow rate for all runs was set at 0.5mL/minute. For each run, 2 mL of filtered sample was loaded onto the column, the first 29 mL was sent to waste, thereafter 3 mL fractions were collected and manually analysed for fluorescence and optical absorbance simultaneously using a Horiba Aqualog-C 3-D fluorimeter (see below). The column was washed with 2 column volumes of mobile phase between runs. Chromatograms were corrected for baseline-shift.

212

213 **2.4 Chemical Analysis**

214 Three dimensional excitation- emission fluorescence matrices (EEM's) and optical
215 absorbance spectra were simultaneously measured using a Horiba Aqualog-C 3D-
216 fluorimeter after the samples had reached room temperature (20 ± 2°C). All samples
217 were analysed for excitation wavelengths 200–550 nm with a 3 nm step. A 1 s
218 integration time was used at each excitation wavelength, with medium detector
219 gain. The emission spectra were collected by charged coupled device detector for the
220 wavelength range 213.6–620.3 nm with a step of approximately 3 nm. Spectra were
221 corrected for inner filtering using automated algorithms supplied with the Aqualog
222 software (Horiba Scientific) and fluorescence intensity was normalised to quinine
223 sulfate units (qsu). All samples were blank-corrected by automatic subtraction of the
224 fluorescence spectrum of ultra-pure water (or mobile phase in the case of the column
225 experiments).

226

227 Dissolved organic carbon was analysed by the heated persulfate oxidation method
228 (APHA 2005) using an OI Analytical 1010 Organic carbon Analyser. $SUVA_{254}$ was
229 determined by dividing the optical absorbance at 254 nm (assuming a 1 metre cell
230 path length) by the measured DOC concentration (mg C l^{-1}) in each sample. E_2:E_3
231 was determined from the ratio of optical absorbance at 254 nm and 365 nm, while
232 E_2:E_4 was determined from the optical absorbance at 254 nm and 464 nm.

233

234 **2.5 Data Analysis**

235 Fluorescence peak deconvolution was carried out by parallel factor analysis
236 (PARAFAC)[27] using the program SOLO (Eigenvector Research). Prior to PARAFAC
237 analysis, emission data for wavelengths <245 nm and >500 nm was removed due to
238 the absence of spectral features outside this range. Excitation wavelengths less than
239 245 nm were excluded as the signal below this wavelength had a very poor signal-
240 to-noise ratio. Data points from the region of the matrix where the emission
241 wavelength was substantially less than the excitation wavelength were set to 0. Any
242 negative data points were also converted to 0. Data points in the 1st and 2nd order

Rayleigh scatter regions were denoted as missing. Data in the Raman scatter region were also set to missing because blank-subtraction did not fully remove the effect of this scatter. Any extreme outliers ('hot pixels') in the excitation-emission matrix (identified by visual examination) were also set to missing. Because of trace fluorescence in sampling blanks, the mean value of the sampling blanks was subtracted from the sample spectra prior to analysis. For model fitting the model was constrained to non-negativity in all modes and obvious outliers excluded. Models with 1 to 6 components were progressively fitted and model diagnostics (especially percent variance explained, core consistency, and the pattern of residuals) were examined to determine the best fit. To further verify the best number of model components, split-half validation was performed using the nearest-neighbor thinning algorithm contained within the SOLO software.

Correlation models were created using Sigmaplot v11. The location of fluorescence maxima were based on the components identified in PARAFAC modelling; the actual value of fluorescence intensity was then determined for each sample using peak-picking on blank-corrected spectra. Model validity was tested by comparing predicted and actual values of DOC concentration and ΔDOC from the 12 second-phase sites using root mean square values.

Parallel line analysis, Pearson's product moment correlation (r) analysis and determination of the coefficient of determination for linear regressions (r^2) were all performed using Sigmaplot v12.

3. Results

3.1 Model development phase

3.1.1. Initial conditions: All locations sampled during the model development phase were circumneutral (pH 6.7 -8.2), non-saline (conductivities were all <300 μS cm^{-1}) but had a range of turbidities (5-60 NTU; Table 1). The DOC concentrations at time of sampling ranged from about 3 to 45 mg C l^{-1} (Table 1) with the lowest values

associated with lowland river sites and the highest levels corresponding to floodplain wetlands and a local farm dam. DOC concentrations at the time of sampling were relatively well correlated with optical absorbance in the ultra-violet region (A_{UV}; with r values around 0.9; coefficient of determination ≈ 0.8); but the correlation declined at longer wavelengths (Figure 1).

Correlation between the initial fluorescence intensity and the initial concentration of DOC in the water samples was substantially better ($r > 0.97$; $r^2 > 0.95$) at almost any given excitation/emission (Ex/Em) pair above excitation wavelengths of about 240 nm and emission wavelengths above 350 nm, than optical absorbance at ultra-violet wavelengths (ESI Fig 1). PAFAFAC analysis of the EEM's of the initial water samples identified 4 components common to all samples. Component 1 had fluorescence maxima at Ex <245 nm/Em 426 nm and Ex 311 nm/Em 426 nm, which has been previously characterised as a 'UVA- humic like' fluorophore.[28] Component 2 had Ex/Em maxima at <245 nm/312 nm and 269 nm/312 nm which corresponds to a 'tyrosine-like' fluorophore.[28] Component 3 had a fluorescence maximum at Ex 260nm/Em 325 nm; as this fluorophore was also found in the sampling blanks, it is possible that this component is an experimental artefact. Component 4 had fluorescence maxima at <245nm/345nm and 287nm/345nm and corresponds to a 'tryptophan-like' fluorophore.[28] Fluorescence intensity at Ex 311 nm/Em 426 nm (corresponding to the local maxima in Component 1), and Ex 287 nm/Em 345 nm (Component 4) were highly correlated with initial DOC concentration; with correlation coefficients (r) of 0.99, and 0.94 respectively (ESI Fig 1). The fluorescence intensity at Ex 269nm/Em 312nm (corresponding to Component 2) was poorly correlated with initial DOC concentration ($r = 0.28$; $r^2 = 0.08$).

3.1.2 Bioavailable DOC. Loss of DOC from solution (ΔDOC) after 28 days was used to estimate DOC bioavailability. ΔDOC varied from about 0.3 mg C l^{-1} to 15 mg C l^{-1} (Table 1) - which equates to between 5% and 40% of the initial DOC concentration.

One of the best predictors of ΔDOC was the initial concentration of DOC (r^2 = 0.85, Figure 2; see also Table 3). Initial A_{UV} was also a relatively good predictor of bioavailable DOC; with coefficients of determination between about 0.75 and 0.8 depending on wavelength (Figure 3). The linear relationship between ΔDOC from optical absorbance fell off at longer wavelengths (Figure 3). Overall, during the course of the incubation there was a positive linear relationship between changes in A_{UV} and ΔDOC. For example the coefficient of determination (r^2) between ΔA_{254} and ΔDOC was 0.7 while for ΔA_{215} and ΔDOC was 0.54 (ESI Figs 2 and 3 respectively). However, the strength of the relationship was driven by changes in optical absorbance at Site 13, which had the greatest change in DOC concentration. If this site is removed from the analysis, the coefficients of determination fall to 0.38 and 0.1 respectively.

Interestingly the putative measures of aromaticity and/or humification based on optical absorbance, $SUVA_{254}$, E_2:E_3 and E_2:E_4 were only very weakly, negatively related to ΔDOC; with coefficients of determination of 0.03, 0.09 and 0.03 respectively (see ESI Figs 4-6).

Initial fluorescence intensity was also highly correlated with ΔDOC. Fluorescence intensity corresponding to the 'humic like' Component 1 (local maxima at Ex 311 nm/Em 426 nm) and the 'tryptophan like' Component 4 (local maxima at Ex 287 nm/Em 354 nm) were positively related to ΔDOC, with regression coefficients of r = 0.92 and 0.93 respectively (which equates to coefficients of determination of 0.384 and 0.87; Figure 4). Although initial fluorescence intensity was highly correlated with ΔDOC, overall fluorescence intensity didn't change markedly through the course of the incubation. Fig. 5 shows the EEM for one replicate from Site 13, a wetland on the Oven's River floodplain. Although DOC concentration fell from 45.5 mg C l^{-1} at the start of the experiment to 28 mg C l^{-1} after 28 days of incubation (i.e. a ≈ 40% reduction in DOC concentration), there wasn't a corresponding change in overall fluorescence intensity. While fluorescence intensity didn't markedly change over the course of the incubation, the relationship between fluorescence intensity

and DOC concentration did. For example, even though the correlation remained high ($r^2 \geq 0.98$), there was a statistically significant difference ($P < 0.001$, parallel line analysis) in the linear relationship between the fluorescence intensity at Ex 311 nm/Em 426nm and DOC concentration at the beginning of the experiment and after 28 days of incubation (Fig. 6).

3.2 Model Validation Phase

3.2.1 Predictive models: From the empirical relationships between DOC and the optical properties of water it is possible to describe simple linear models relating the initial optical absorbance or fluorescence intensity of a water sample with either the initial concentration of DOC or ΔDOC after 28 days of incubation (Table 3). The validity of these models were tested by comparing the predicted value of DOC concentration and ΔDOC calculated from optical properties of water from an additional 12 sites, not sampled in the model development phase, with actual values for these sites determined experimentally. Comparisons between actual and predicted values were made by determining the average root mean square (RMS) between the two. The types of water bodies sampled and the physiochemical properties of the 12 sites used for model validation were similar to sites used for model development are described in Table 2.

3.2.2 DOC concentration: Fig. 7 shows the relationship between the actual concentrations of 36 water samples from 12 locations taken in the validation phase of the model development against a value predicted from A_{254}. Although the correlation that the model is based on is relatively strong, the average RMS is about 7.45 mg C l^{-1}; for A_{215}, whose model had a slightly better correlation, the average RMS was 6.29 mg C l^{-1} (data not shown).

Predictions based on fluorescence intensity of the water sample at a specific Ex/Em are better than those from optical absorbance. For example the average RMS between actual and predicted concentration using the model developed for Ex 311 nm/Em 426 nm was 5.06 mg C l^{-1} (Fig. 8). From Fig. 8 it can be seen that the actual

value most closely matched the predicted value at concentrations of less than 20 mg C l^{-1}; if values of actual concentrations of DOC above 20 mg C l^{-1} are excluded the average RMS falls to 2.9 mg C l^{-1} (n=27). The RMS for the model developed for initial fluorescence intensity at Ex 287/Em 346 was 4.88 mg C l^{-1} without any apparent systematic variation (data not shown).

3.2.3 ΔDOC: Using these correlations it is possible to build linear models relating the optical properties of water at the time of sampling with 28-day bioavailability (Table 3). The model based on A_{254} gives a reasonable prediction of bioavailable DOC concentrations of less than 2 mg C l^{-1}, but tend to substantially overestimate the amount of bioavailable carbon present at levels above 2 mg C l^{-1} (Fig. 9). Overall the model based on A_{254} had an average RMS of 3.49 mg C, l^{-1} which isn't terribly good given that the highest ΔDOC was less than 6 mg C l^{-1}. The accuracy of prediction using models based on A_{215} were marginally better than A_{254} (average RMS = 3.13 mg C l^{-1}), and again was much better at bioavailable DOC levels of less than 2 mg C l^{-1}, but overestimated bioavailability at higher concentrations (data not shown).

The average RMS for models of bioavailability of DOC based on fluorescence intensity at Ex 311 nm/Em 426 nm was also 3.13 mg C l^{-1}, and while the predictability at bioavailability of DOC concentrations levels less than 2 mg C l^{-1} was not as good as A_{215} or A_{254}, the predictability at higher concentrations was marginally better (Fig. 10). The average RMS for the model of bioavailability based on fluorescence at Ex 287 nm/Em 346 nm was 3.25 mg C l^{-1} (data not shown).

The model to predict bioavailability of DOC based on the initial concentration of DOC performed better than those based on absorbance or fluorescence with an average RMS of 2.36 mg C l^{-1}.

3.3. Changes in DOC during incubation.

To explore how the nature of DOC changed during incubation, water samples from Site 3 (Cookie's Dam, a small farm dam with relatively high DOC concentrations;

Table 1) were incubated in a similar manner to that described for the model development and validation experiments. Periodically a sample was removed, filtered and the DOC separated using size-exclusion chromatography. At the beginning of the experiment, the chromatogram consisted of 4 peaks (Figure 11 (a)):

(A) a small peak immediately following the point when the first (largest) molecule would be expected to elute from the column (the void volume - V_0);[26]

(B) a large sharp peak centered around Fraction 14;

(C) a small peak with a maximum at fraction 17 and

a broad peak centered around Fraction 23.

EEMs of the fractions shows that Peak A has fluorescence maxima at about Ex <240 nm /Em 315 nm and Ex 275 nm/Em 315 nm, which corresponds to the 'tryptophan-like' fluorophore identified as Component IV in the PARAFAC analysis of the field samples. Peaks B and C did not fluoresce, while Peak D had fluorescence maxima centered at about Ex <250 nm/Em 426 nm and Ex 311 nm/Em 426 nm (i.e. corresponding to Component 1 in the PARAFAC analysis of the field samples). Peak B disappeared between the 10th and 20th day of incubation (Figure 11 (c)). After 48 days of incubation Peaks A and C had also disappeared (Figure 11 (d)). Peak D had a similar EEM at the end of the incubation as at the beginning (ESI Fig. 7), however the peak height had diminished, and the peak position had shifted towards V_t (i.e. towards a lower molecular weight).

4. Discussion

It is becoming more common to use the optical properties of water, specifically absorbance and fluorescence, to determine both the quantity and quality of DOC in aquatic samples.[12,13] This is because the approach is relatively straightforward, quick, non-destructive and relatively sensitive. However the results of the current study suggest that some caution should be exercised when applying either absorbance or fluorescence to the study of DOC in natural waters.

4.1 Estimating the concentration of DOC

One common application of the optical properties of water is the estimate of DOC concentration. Like the current study, numerous other studies have shown that there is often a very strong correlation between the optical absorbance at a single wavelength and DOC concentration – with correlations as high as $r^2 > 0.9$ reported in the literature.[13] Similarly, fluorescence intensity of DOC is often highly correlated with DOC concentration.[14] Indeed, there are commercially available probes that use either absorbance at a fixed wavelength (commonly 254 nm) or fluorescence intensity to determine the concentration of DOC *in-situ*. However, as we show in this study, while there was a strong correlation between both absorbance in the ultraviolet region and, fluorescence above about Ex 240 nm and Em 350 nm, with DOC concentration in the model development phase, the models developed from these relationships were not necessarily that good at predicting the DOC concentration in the validation samples. It is possible that the reason for the relatively poor predictability of DOC concentrations from the optical properties of water is because in both the model development phase and the model validation phase of the study samples were taken from a diverse range of aquatic habitats, with potentially quite different sources of DOC. However, notwithstanding the source of the relatively poor predictive power of the models in the current study, it does highlight the need to routinely check concentrations of DOC estimated by optical probes (absorbance or fluorescence) against actual DOC concentrations.

4. 2. Estimating bioavailability of DOC

There is an assumption that there is an inverse relationship between the degree of aromaticity/humic-richness of DOC and its bioavailability.[16] $SUVA_{254}$ and, to a lesser extent, E_2/E_3 and E_2/E_4 have been used as surrogates for the aromaticity and/or humification of DOC.[15-19] In their seminal paper Weishaar et al.[15] showed, using ^{13}C NMR spectroscopy, that there was a very strong correlation between $SUVA_{254}$ and percent aromaticity ($r^2 = 0.97$) for 13 samples taken from a variety of aquatic ecosystems. However, they also showed that the reactivity of the DOC was

highly variable; and went as far as warning against using simple spectroscopic methods to estimate the composition and reactivity of DOC. In the model development phase of the current study we show that there is not a strong relationship between $SUVA_{254}$, E_2:E_3 or E_2:E_4 and the bioavailability of DOC. This again shows that caution should be exercised when equating the extent of aromaticity/humification with the degree of bioavailability of DOC.

As in previous studies,[22-25] during the model development phase we see high correlations between both initial fluorescence (for both the 'humic-like' peaks and the 'tryptophan-like' peaks) and initial A_{UV} on the one hand and the bioavailability of DOC on the other. However, the ability to predict *a priori* the bioavailability of DOC from sites not included in the model development using either initial optical absorbance properties or fluorescence intensity of water is not strong. In the model validation phase of this study, the maximum ΔDOC was about 6 mg C l^{-1}, while the average RMS between actual and predicted ΔDOC based on absorbance in the ultraviolet region was between about 6 and 7 mg C l^{-1}. Models based on initial fluorescence of the water are marginally better, with the best fitting model based on fluorescence at Ex311/Em 426, but restricting the data used to generate the model to ΔDOC values of less than 20 mg C l^{-1}.

The results from the current study suggest that one of the reasons to explain the poor predictability of models based on linear regressions between initial optical absorbance or fluorescence and ΔDOC, even though those properties are highly correlated, is because the correlations between the initial optical properties of water and ΔDOC are coincidental, not causative. Excluding the data from Site 13, there were only weak correlations between changes in A_{215} or A_{254} and ΔDOC. That is, as DOC is consumed there isn't a concomitant change in optical absorption. This suggests that the bioavailable fraction does not necessarily absorb light in the ultraviolet region. Similarly, there was little change in the fluorescence EEM's over the course of the incubations in the model development phase, even though there were substantial changes in DOC concentration (Fig 5). While there was little

change in fluorescence over the incubation there was a change in the relationship between fluorescence intensity and DOC concentration (Fig 6). The change in the relationship could be a result either of two possibilities;[12] either it is a dilution effect or there is a shift in fluorescence quantum yield (Φ - the ratio of the number of photons of light emitted during fluorescence to the number of photons adsorbed). The fluorescence of a mixture of molecules, like DOC, can be affected by the interaction between molecules.[12] This effect will depend on the concentration of the molecules, the higher the concentration the more likely it is that two molecules will interact with each other and hence change the fluorescence response. Alternatively, Φ changes with chemical structure; the value of Φ will change if the chemical structure of the molecule(s) near the fluorophore changes during decomposition. To explore which of these options was impacting the relationship we looked at changes in fluorescence of DOC from one site (Site 3 - Cookies Dam) in more detail. This is a small farm water storage that had a DOC concentration of about 35 mg/L (Table 1). To explore if the effect was due to dilution, samples of the dam water were diluted with ultra-pure water to give final concentrations of 0, 10, 20, 30, 40, 50, 60, 70 and 80% of the initial sample's concentration (concentrations above 80% saturated the instrument's detector) and the fluorescence intensity determined at Ex 311 nm/Em 426 nm. Deviation of fluoresce from linear, particularly at higher concentrations, would at least suggest that interaction between different molecules was responsible for the changes of the relationship between DOC concentration and fluorescence intensity following incubation seen in Fig 6. Dilution of the water sample from Cookies Dam reduced fluorescence intensity as expected, but the relationship between fluorescence intensity and concentration remained strongly linear ($r^2 = 0.99$; ESI Fig 8) This means, that at least for Cookies Dam, interaction between molecules is probably not responsible for the effect observed in Fig 6. When DOC from Cookie's Dam was fractionated, based on molecular weight, during an incubation experiment (Figure 11), the fractions that disappeared first (the most bioavailable fractions) didn't fluorescence. The intensity of the most fluorescent peak (Peak D) didn't change, but the height of the peak declined and the peak shifted to lower molecular weight region.

Taken together, these results suggest that while DOC is consumed during the incubation, molecules, or parts of molecules, that contribute to DOC fluorescence are not preferentially consumed. Because there isn't a direct concordance between changes in fluorescence or optical absorption and changes in DOC, the high correlation between the initial optical properties of water and subsequent DOC consumption is coincidental not causal.

Both fluorescence and A_{UV} were highly correlated with the initial concentration of DOC in water samples. One of the best predictors of ΔDOC was the initial concentration of DOC ($r^2 = 0.85$). Because there doesn't seem to be a direct causal relationship between fluorescence or A_{UV} and ΔDOC, the observed correlations between them may reflect their relationship to initial DOC concentration and the subsequent relationship between initial DOC concentration and ΔDOC.

Models to predict bioavailability of DOC based on the initial concentration of DOC performed better in the validation phase than those based on A_{UV} or fluorescence; with an average RMS of 2.36 mg C l^{-1}. From a practical point of view, this is not entirely helpful in assessing the bioavailability of DOC in the field and, preferable in real-time and *in situ*. One way of doing this, which is purely empirical, would be to estimate the concentration of DOC present using one of the optical characteristics of water and then using that number in the linear model developed for predicting bioavailability of DOC from initial concentration of DOC. Using this approach, calculating the DOC present in our verification samples using the fluorescence intensity at Ex 311 nm/Em 426 nm, and excluding initial concentrations of greater than 20 mg C l^{-1} (above which we have seen the relationship between fluorescence intensity and the concentration of DOC breaks down) produces predicted values of DOC which have an average RMS of 1.91 mg C l^{-1} difference from actual values.

5. Conclusions

Optical absorbance and fluorescence spectroscopies have the potential to play an important role in monitoring aquatic ecosystems. However, while there are strong correlations observed between fluorescence intensity or optical absorption of water and the quality and quantity of DOC in that water, unless there is an underlying causal relationship between two components, care should be used in using those relationships to predict the quality or quantity of DOC from sites not used in the model development. The robustness of empirical models developed at other sites or in other systems need to be verified and then periodically assessed to determine their on-going suitability.

Acknowledgements

This project was funded by the Murray-Darling Basin Authority through Project MD 2556 – Intervention Monitoring for the Living Murray 2013-2014. We would like to thank Mr Garth Watson and Ms P Baldwin for assistance with the field sampling, Mr J. Pengelly (MDFRC analytical laboratory) for DOC analysis, Ms K. Whitworth for developing macros used to facilitate the transfer of data from the fluorimeter to the SOLO software and Mr R. Cook for access to his dam.

References

1 L.G. Leff and J.L Meyer, *Limnol. Oceanogr.*, 1991, **36**, 315-323.

2 J. Howitt, D. S. Baldwin, G. N. Rees and J. Williams, *Ecological Model*., 2007, **203**, 229-242

3 J. J Cole, S. R. Carpenter, J. Kitchell, M. L Pace, C. T. Solomon, and B. Weidel, *Proc. Nat. Ac Sci*., 2011, **108**, 1975.

4 S. M. Mitrovic, D.P Westhorpe, T. Kobayashi, D.S. Baldwin, D. Ryan and J. Hitchcock, *Journal of Plankton Research*. 2014, doi: 10.1093/plankt/fbu072

5 R. J. Diaz and R. Rosenberg, *Science*, 2008, **321**, 926-929.

6 J. Kerr, D.S. Baldwin and K. Whitworth, *J. Environ. Management*, 2013, **114,** 139-147

7 K. Whitworth D.S. Baldwin and J. Kerr, J.Hydrol., 2012, **450-1**, 190-198.

8 S. Hladyz, S. Watkins, K. Whitworth and D.S. Baldwin, *J. Hydrol.*, 2011, **401**, 117-125

9 A. J. King, Z. Tonkin, J. Lieshcke J, *Mar. Freshwater Res.*, 2012, **63**, 576-586

10 K. L Whitworth, J. L Kerr, L. M . Mosley, J. Conallin, L. Hardwick and D. S. Baldwin. Environ. Management. 2013, **52**, 837-850.

11 APHA *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association, Washington, D. C. 2005.

12 J. A. Korak, A. D. Dotson, R. S. Summers and F. L. Rosario-Oritz, *Water Res.* 2014, **49**, 327-338.

13 M. Peacock, C. D. evans, N. Fenner, C. Freeman, R. Gough, T. G. Jones and I. Lebron. *Eniron. Sci. Processes and Impacts*, 2014, **16**, 1445 -1461.

14 S. A. Cumberland and A. Baker. *Hydrological Processes*, 2007, **21**, 2093-2099.

15 J. L. Weishaar, G. R. Aitken, B. A. Bergamaschi, M. S. Fram, R. Fujii and K. Mopper. *Environ. Sci. Technol*. 2003, **37**, 4702-4708.

16 P-G Kang and M. J. Mitchell. *Biogeochem*., 2013, **115**, 213-234.

17 J. Peuravuori and K. Pihlaja, Molecular size distribution and spectroscopic properties of aquatic humic substances, *Anal. Chim. Acta*, 1997, **337**, 133-149.

18 T. Ohno, T. S. Griffin, M. Liebman and G. A. Porter, *Agric. Ecosys. Environ*, 2005, **105**, 625-634.

19 S. Park, K. S. Joe, S. H. Han and H. S. Kim, *Environ. Technol*., 1999, **20**, 419-424.

20 J. Kukkonen and A. Oikari, *Water. Res*., 1991 **25**, 455-463.

21 S. Cumberland, J. Bridgeman, A. Baker, M. Sterling, D. Ward, D, *Environ. Technol.* 2012, **33,** 687-693.

22 A. Baker and R. Inverarity, Protein-like fluorescence intensity as a possible tool for determining river water quality, *Hyrological Processes*, 2004, **18**, 2927-2945.

23 N. Hudson, A. Baker, D. Ward, D.M Reynolds, C. Brunsdon, C. Carliell-Marquet, S. Browning, *Sci. Total Environ*. 2008, **391**, 149-158.

24 Hur J. and J. Cho, *Sensors*, 2012, **12**, 972–986.

25 J. B. Fellman, D. V. D'Amore, E. Hood and R. D. Boone, *Biogeochem*., 2008, **88**, 169-184.

611 26 S. Mori and H. Barth, *Size exclusion chromatography*. Springer Verlag, Berlin 1999,
612 234 pp.
613 27 C. A. Stedmon, S. Markager and R. Bro, Mar. Chem., 2003, 82, 239-254.
614 28 M. Tedetti, R. Longhitano, N. Garcia, C. Guigue, N. Ferretto and M. Goutx,
615 *Environ. Chem*., 2012, **9**, 438–449.
616
617

618 Captions

619

620 Fig 1 The coefficient of determination for the linear correlation between optical
621 absorbance at a particular wavelength and the initial concentration of DOC (n = 54)

622 Fig 2 The relationship between the initial concentration of DOC in a water sample
623 and the amount of carbon lost from solution (ΔDOC) after 28 days of incubation.
624 The line is the linear regression (r^2 = 0.85, n = 54)

625 Fig 3 The coefficients of determination for the linear correlation between optical
626 absorbance at a given wavelength and the amount of carbon lost from solution
627 (ΔDOC) after 28 days of incubation (n = 54)

628 Fig 4 The relationship between the initial fluorescence intensity at a) Ex 311 nm/ Em
629 426 nm and b) Ex 287 nm/Em345 nm and the amount of DOC lost from solution
630 (ΔDOC) after 28 days of incubation. The lines are the linear regressions (r^2 = 0.84
631 and 0.87 respectively).

632 Fig 5 The fluorescence excitation-emission matrices for a single sample from Site 13
633 measured at the start of the incubation and after 28 days of incubation. Fluorescence
634 intensities are in quinine sulfate units

635 Fig 6 The relationship between fluorescence intensity at Ex 311 nm /Em 426 nm and
636 DOC concentration at the beginning of the incubation experiment (closed circles)
637 and after 28 days incubation (open squares). The lines represent the respective
638 correlations which are significantly different from each other ($p < 0.001$).

639 Fig 7 The measured initial DOC concentration in the validation samples compared
640 to the amount predicted to be present based on the initial optical absorbance at 254
641 nm. The line represents the 1:1 equivalence.

642 Fig 8 The measured initial DOC concentration in the validation samples compared
643 to the amount predicted to be present based on the initial fluorescence intensity at
644 Ex311 nm/Em 426 nm. The line represents the 1:1 equivalence.

Fig 9 The measured bioavailable carbon fraction in the validation samples compared to the amount predicted to be present based on the initial optical absorbance at 254 nm. The line represents the 1:1 equivalence

Fig 10 The measured bioavailable carbon fraction in the validation samples compared to the amount predicted to be present based on the initial fluorescence intensity at Ex 311 nm/Em 426 nm. The line represents the 1:1 equivalence

Fig 11 Size-exclusion chromatograms for a water sample from Site 3 (Cookies Dam) a) immediately prior to incubation and after b) 10 days, c) 20 days and d) 48 days of incubation. V_0 is the column void volume and V_t is the total column volume. The letters corresponds to peaks in the initial chromatogram.

Table 1 Site, description and general water quality of the 18 locations from south-eastern Australia used in the Model Development phase of the study. n.d. = not determined s.e. = standard error

Site Name	Location Latitude(S)/ Longitude(E)	Type	Dissolved oxygen mg/L	pH	Turbidity NTU	Temperature °C	Conductivity µS/cm	Initial concentration of DOC mg C l^{-1} (mean ± s.e.)	Concentration of DOC after 28 days incubation mg C l^{-1} (mean ± s.e.)
1.Lake Hume @ Bowna	35.99723/ 147.05873	Large water-storage reservoir	9.31	7.08	25	21.8	52	3.7 ± 0.1	3.1 ± 0.1
2.Woolshed Creek	36.03453/ 147.00893	Ephemeral creek	2.91	6.71	68.2	17.5	97	18.1 ± 0.4	11. ± 0.0
3.Cookies Dam	36.06247/ 147.04089	Small farm dam	4.2	7.23	12.5	19.2	264	33.0 ± 0.3	26.2 ± 1.2
4.Murray R. @ Heywoods Bridge	36.09987/ 147.02266	Lowland river	6.4	6.89	6.64	19.9	52	4.3 ± 0.2	3.2 ± 0.1
5.Lake Hume @ Ebden	36.15131/ 147.02727	Large water-storage reservoir	8.96	7.29	6.54	21.8	51	4.1 ± 0.2	3.0 ± 0.0
6.Kiewa R. @ Kilara	36.13789/ 146.95479	Lowland River	8.84	7.48	4.71	20.4	44	2.9 ± 0.3	2.5 ± 0.7
7.Wangaratta Off Ramp	36.30707/ 146.37825	small 'farm-like' dam	7.96	8.29	19.1	21.9	312	13.5 ± 0.1	11.5 ± 0.1
8.Wangaratta Wetland #1	36.33226/ 146.34523	Billabong	4.6	7.07	6.69	20.1	124	3.5 ± 0.2	2.6 ± 0.2
9.Wangaratta Wetland #2	36.33751/ 146.34288	Billabong	0.42	6.83	4.78	18.5	92	8.0 ± 0.1	6.2 ± 0.0
10.Ovens R. @ Wangaratta	36.35154/ 146.32713	Lowland river	7.38	7.56	9.85	21	60	2.9 ± 0.4	1.8 ± 0.3
11.Peechelba Wetland #1	36.16061/ 146.23549	Floodplain depression	3.15	6.72	25.8	22.6	96	25.7 ± 0.1	18.4 ± 1.0
12.Peechelba Wetland #2	36.15504/ 146.23494	Floodplain depression	3.63	6.83	12.1	22.5	107	20.4 ± 0.3	15.8 ± 0.5

13.Peechelba Wetland #3	36.15006/ 146.23659	Floodplain depression	2.78	6.84	20.8	24.7	141	45.3 ± 0.2	30.8 ± 1.6
14.Ovens R. @ Peechelba	36.13573/ 146.23862	Lowland river	7.49	7.31	10.6	22.9	71	3.6 ± 0.0	2.4 ± 0.0
15.Peechelba Wetland #4	36.13573/ 146.23862	Billabong	2.85	6.82	56.5	28.8	103	9.2 ± 0.6	5.6 ± 0.2
16.Peechelba Wetland #5	36.14108/ 146.23741	Billabong	5.64	6.85	63.5	26.4	74	13.3 ± 0.1	8.0 ± 0.2
17.Lake Mulwala	36.01855/ 146.05968	Large water-storage reservoir	8.8	8.2	31.5	23	52	4.1 ± 0.1	3.9 ± 0.2
18.Murray R. @ Howlong	35.98798/ 146.62306	Lowland river	8.08	7.42	9.82	21.4	52	3.2 ± 0.6	2.9 ± 0.4
Field blank			n.d.	n.d.	n.d.	n.d.	n.d.	0.8 ± 0.2	0.2 ± 0.2

Table 2: Location and water quality parameters of sites from south-eastern Australia used Model Validation Phase of the project. n.d. = not determined s.e. = standard error

Site Name	Location Latitude(S)/ Longitude(E)	Type	Dissolved oxygen mg/L	pH	Turbidity NTU	Temperature ^{o}C	Conductivity µS/cm	Initial concentration of DOC mg C l^{-1} (mean ± s.e.)	Concentration of DOC after 28 days of incubation mg C l^{-1} (mean ± s.e.)
1. Goulburn River @ Seymor	37.01722/ 145.12277	Lowland river	9.47	7.0	17	17.2	48	3.1 ± 0.2	2.4 ± 0.2
2. Broken river @ Benalla	36.55388/ 145.98611	Lowland river	7.61	7.14	19.3	28.5	144	8.2 ± 0.5	7.7 ± 0.2
3. Causeway Wetland	36.09767/ 146.90076	Billabong	3.92	6.9	66.3	22.9	158	23.0 ± 0.4	19.3 ± 0.4
4. Normans Lagoon	36.09545/ 146.9296	Billabong	6.98	6.87	75.7	23.45	79	16.6 ± 0.2	14.6 ± 0.2
5.Mungambarrena	36.09816/ 146.94815	Billabong	n.d.	6.85	53.6	23.9	82	5.9 ± 0.5	4.4 ± 0.1
6. Waterview Bridge	36.07118/ 146.85896	Floodplain depression	2.28	6.57	53.1	20.89	103	8.7 ± 0.1	6.8 ± 0.1
7. Lake King	36.05463/ 146.45685	Urban lake	8.79	7.9	n.d.	26.04	436	27.5 ± 0.4	24.1 ± 0.1
8. Lake Moodermere	36.05634/ 146.38445	Large floodplain lake	8.39	7.17	58.8	28.18	91	8.6 ± 0.1	7.4 ± 0.3
9. Frosts Crossing 1	36.22718/ 146.24983	Billabong	0.42	6.64	39.5	22.49	178	14.0 ± 0.1	11.4 ± 0.5
10. Frosts Crossing 2	36.22677/ 146.24492	Billabong	5.87	6.87	7.6	26.07	153	13.1 ± 0.1	11.0 ± 0.2
11. Wangarratta On Ramp Northbound	36.30958/ 146.37413	Small 'farm like' dam	10.13	8.29	10.5	27.3	305	12.0 ± 0.6	10.4 ± 0.1
12 Lake Anderson	36.15205/ 146.6114	Urban lake	8.1	7.2	54	25.2	117	30.2 ± 0.8	26.2 ± 0.1

Table 3: Linear regression models used to predict the concentration and bioavailability of DOC. FI = fluorescence intensity (qsu)

	Parameter	**Equation**	**Coefficient of determination (r^2)**
Concentration of DOC:	A_{215}	[DOC] = A_{215} x 14.88 -0.49	0.85
	A_{254}	[DOC] = A_{254} x 23.83 – 0.39	0.81
	Ex311 nm/Em 426 nm	[DOC] =1.12 + $FI_{311/426}$ x 15.5	0.98
	Ex 287nm/ Em 346 nm	[DOC] = FI x42.08-2.4	0.89
28 day bioavailability of DOC:	A_{215}	Δ[DOC] = A_{215} x 4.54 – 0.49	0.79
	A_{254}	Δ[DOC] = A_{254} x 7.41 -0.53	0.79
	Ex287 nm / Em 346 nm	Δ[DOC] = $FI_{287/346}$ x 13.08– 1.18	0.87
	Ex311 nm / Em 426 nm	Δ[DOC] = 0.13+ $FI_{311/426}$ x 4.52	0.84
	Initial DOC concentration	Δ[DOC] = $[DOC]_{int}$ x 0.29 – 0.21	0.85

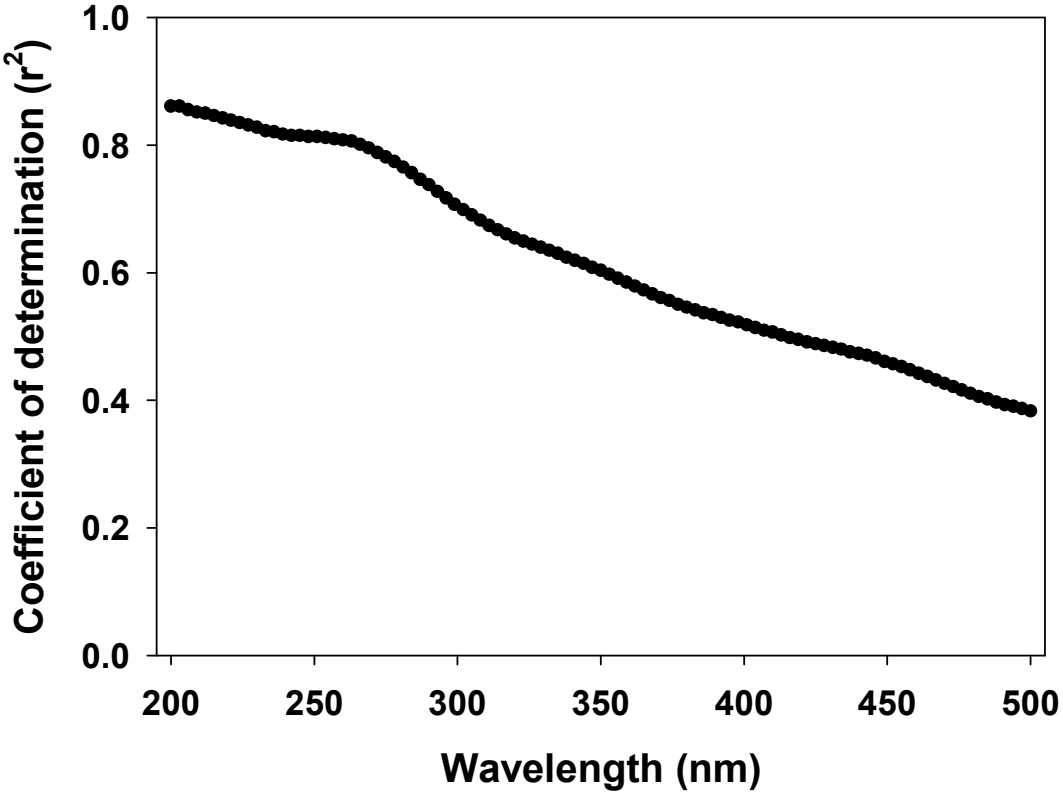


Fig 1

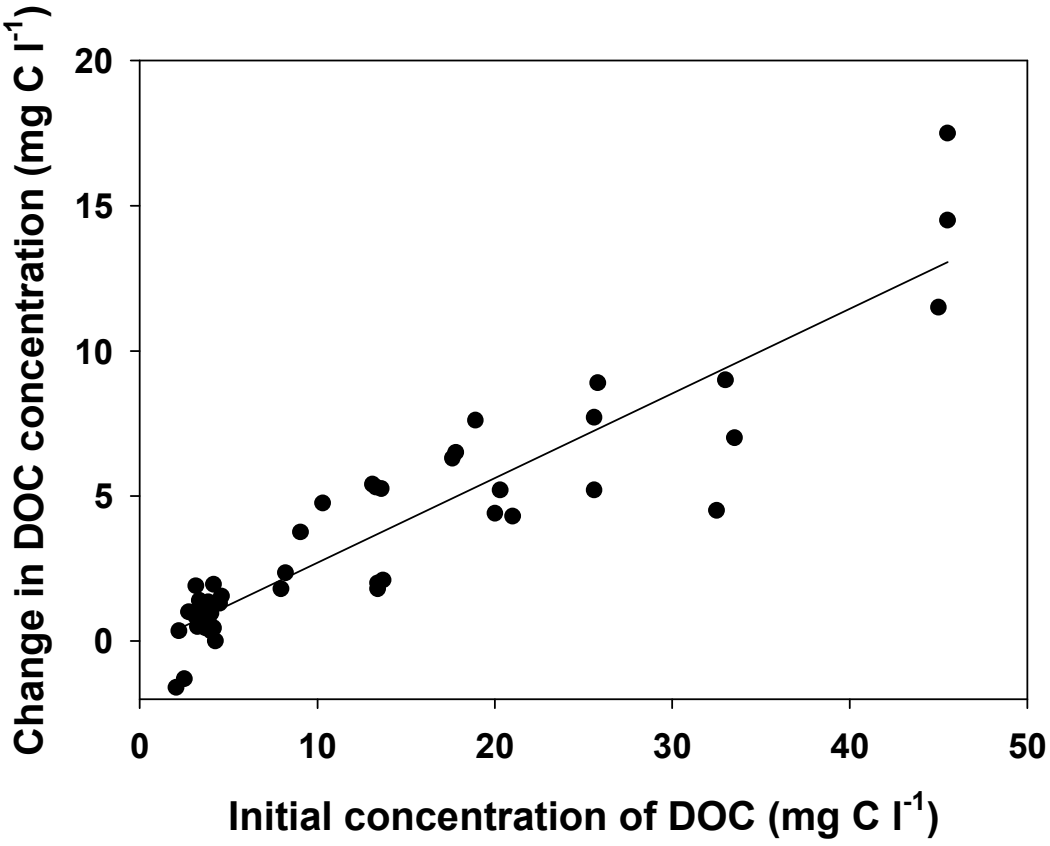


Fig. 2

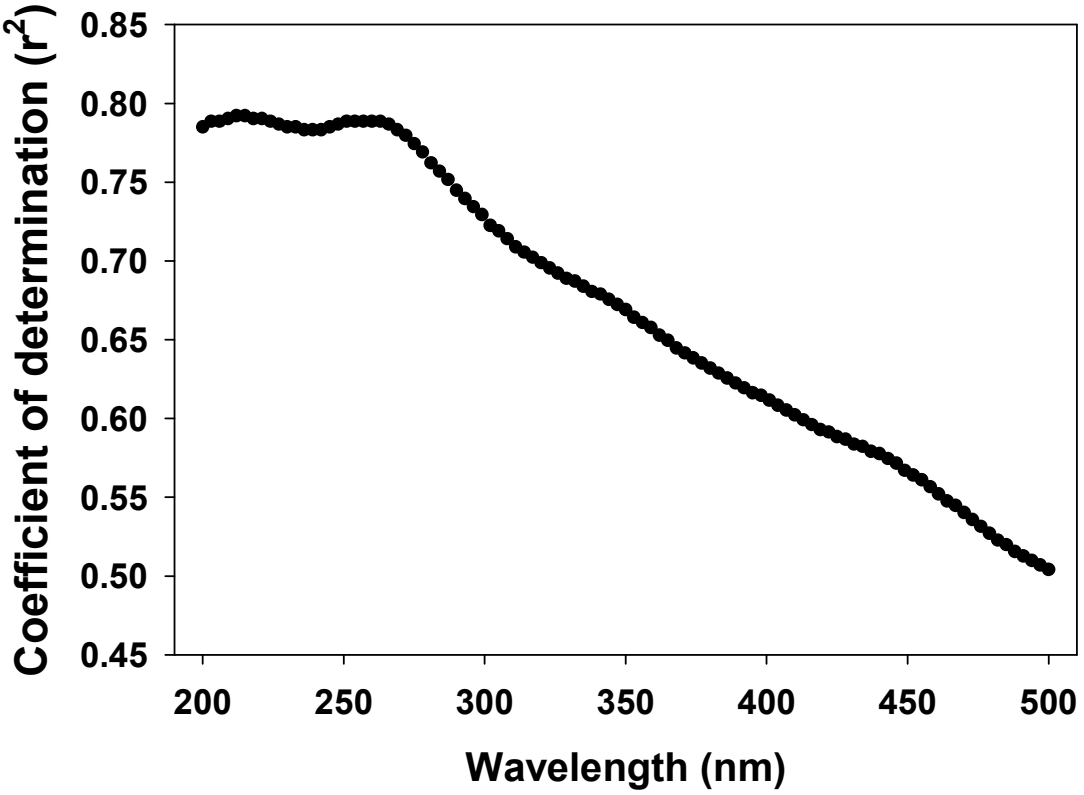


Fig 3.

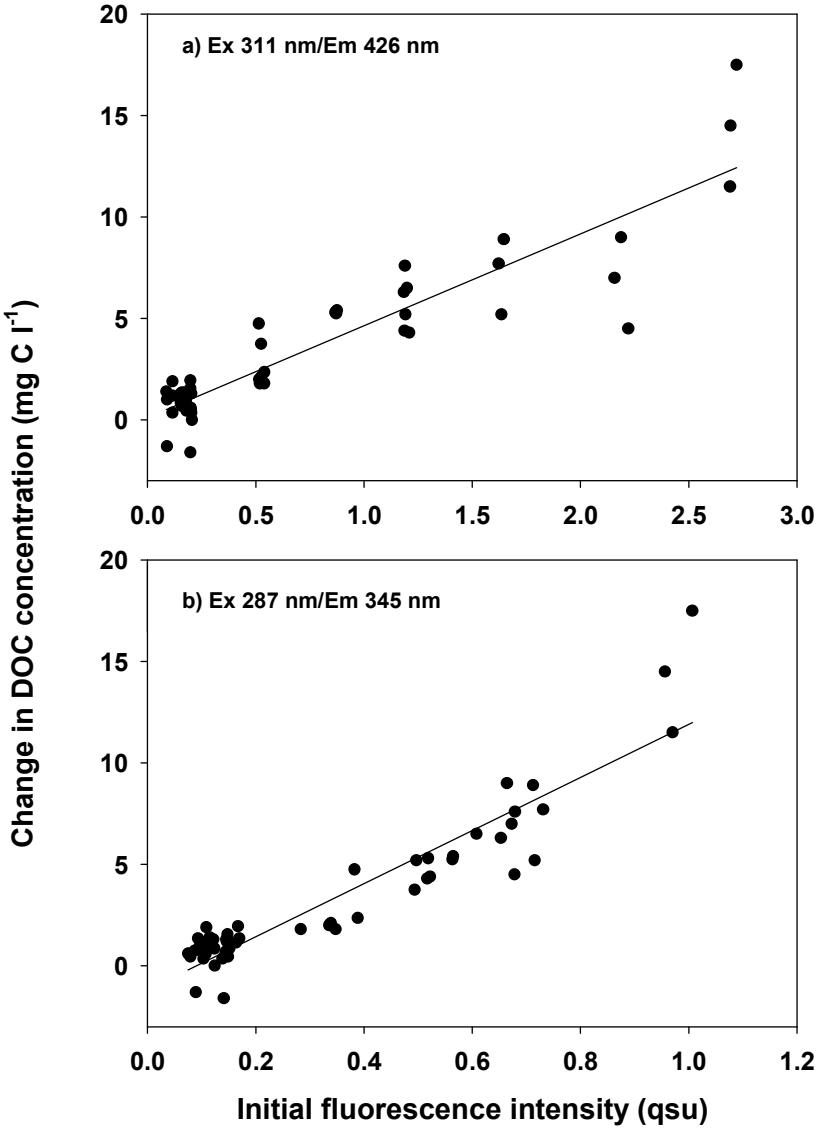


Fig. 4

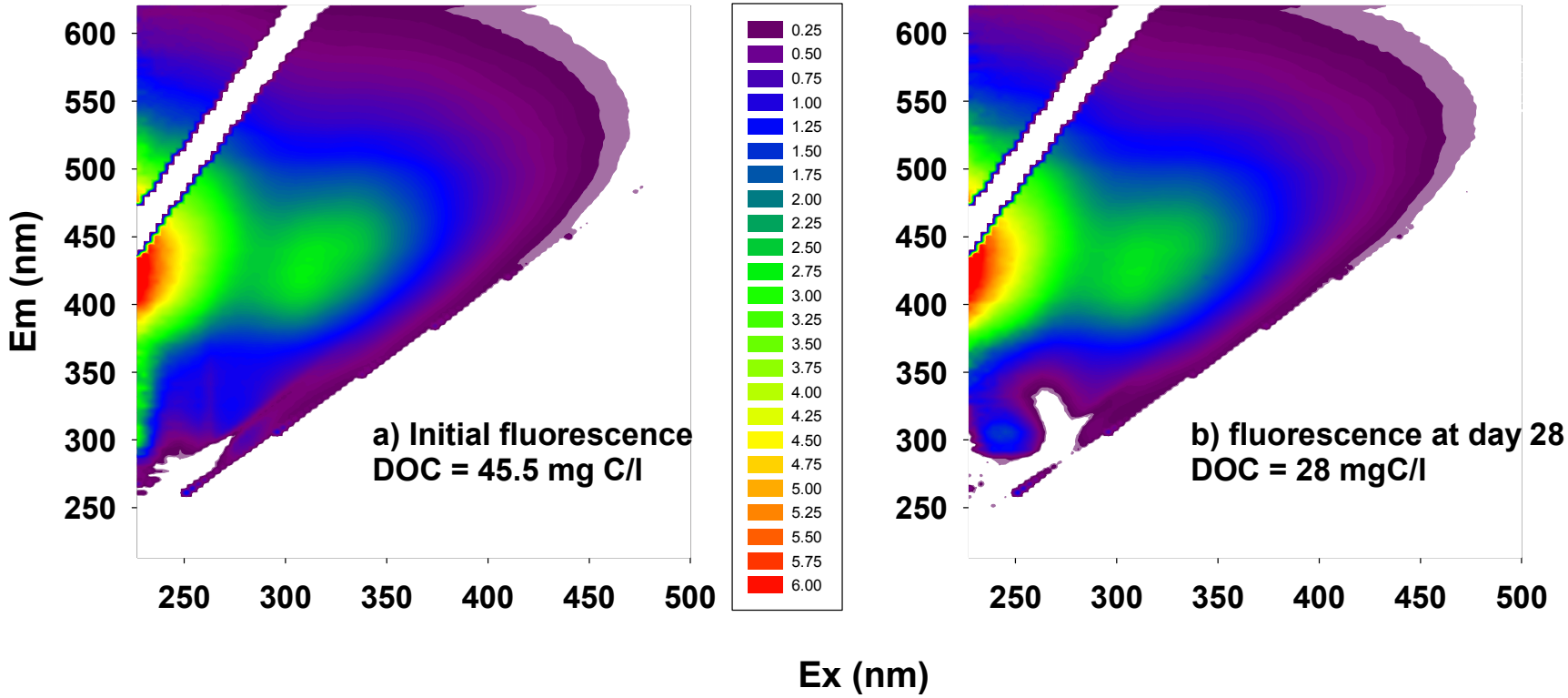


Fig. 5

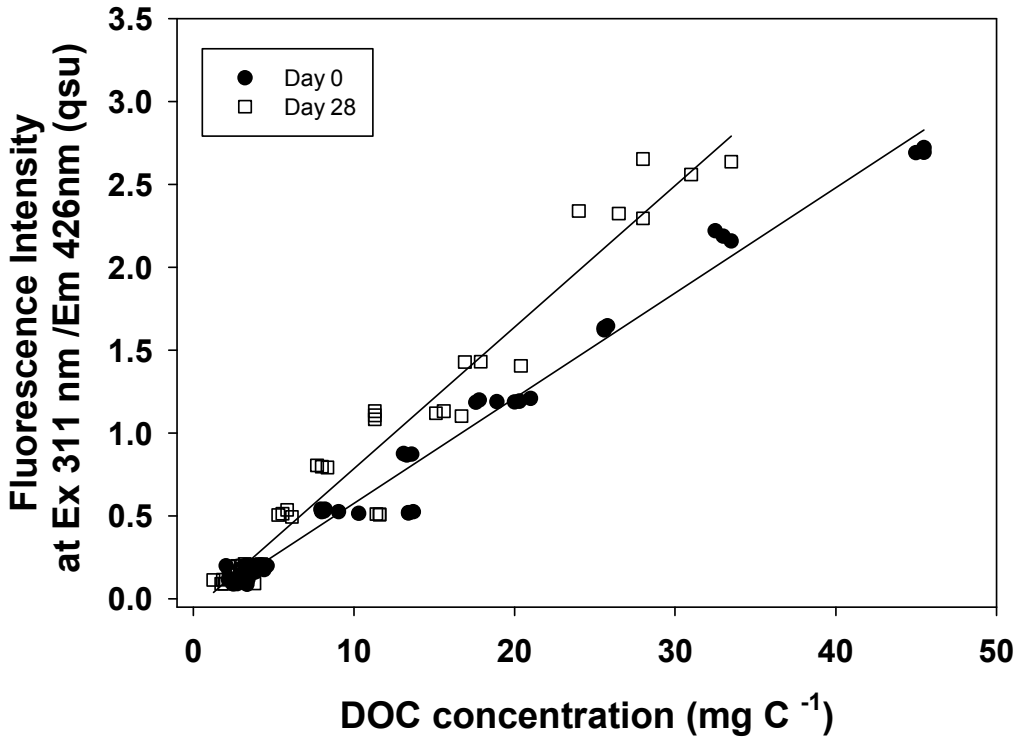


Fig. 6

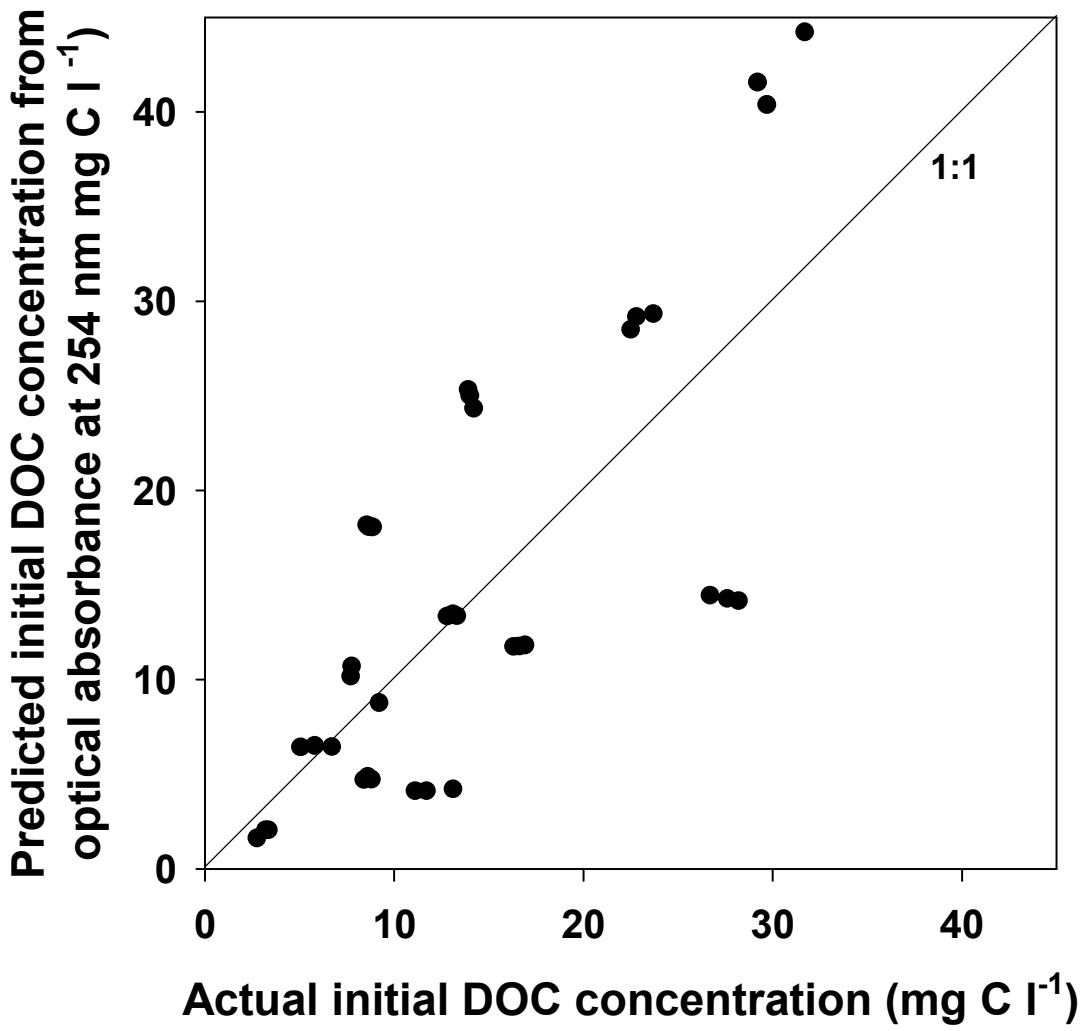


Fig 7.

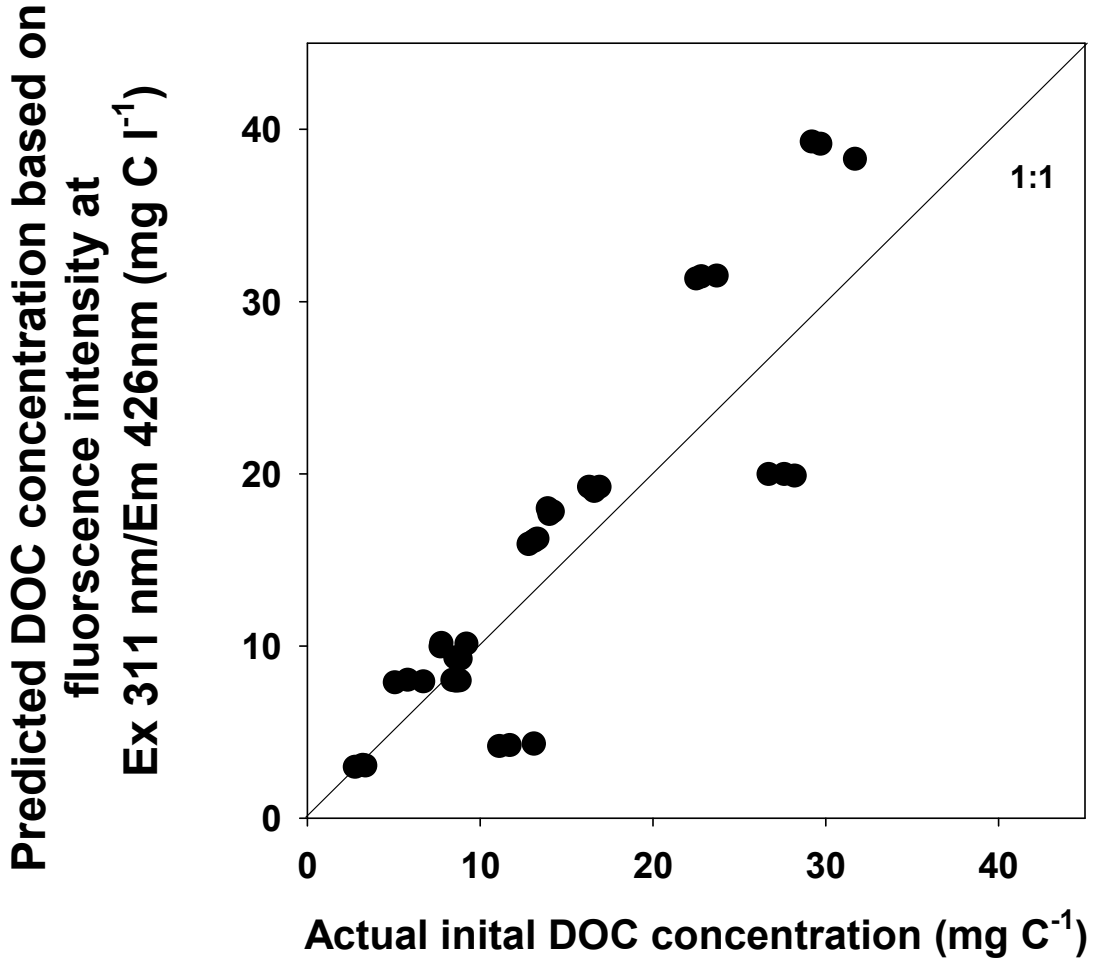


Fig. 8

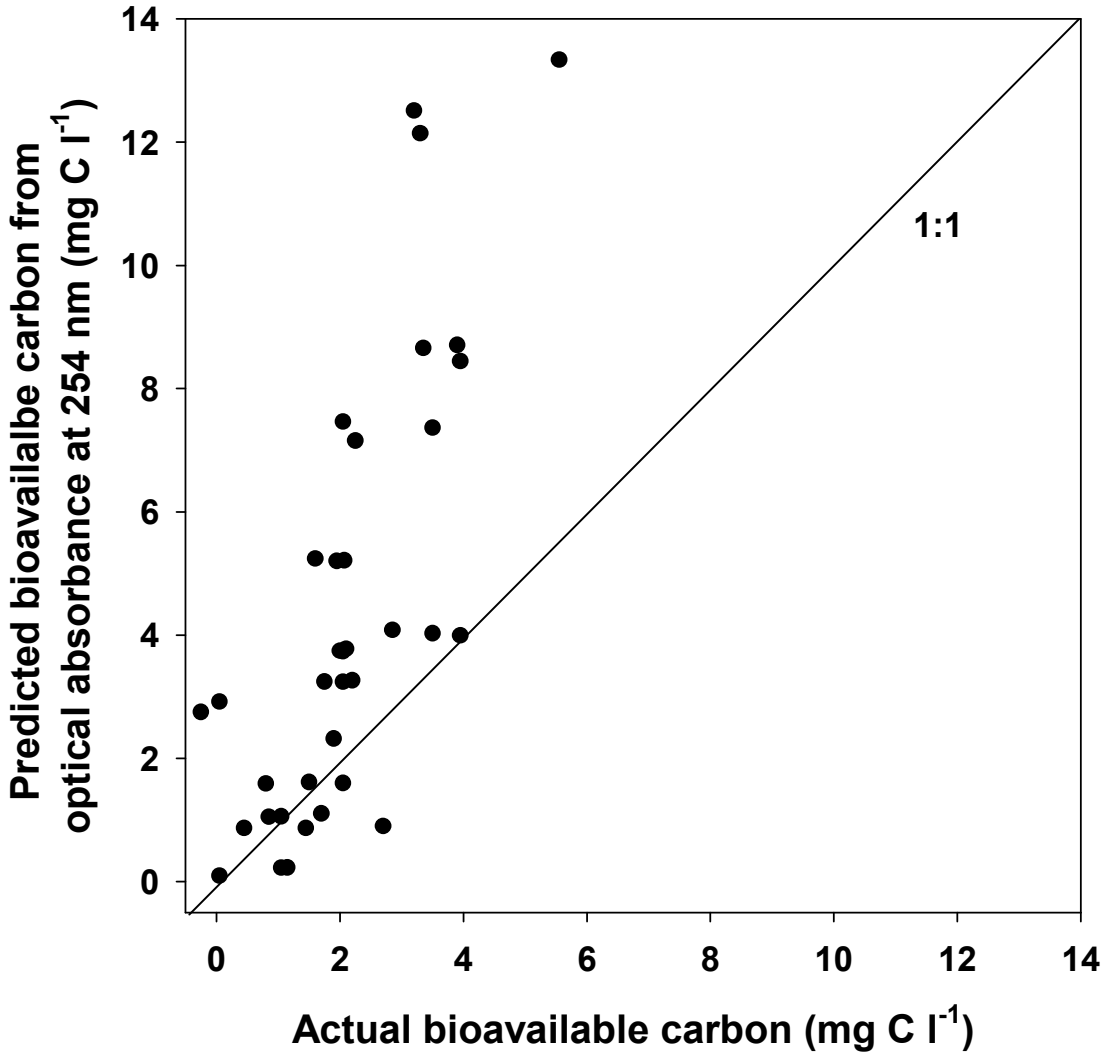


Fig. 9

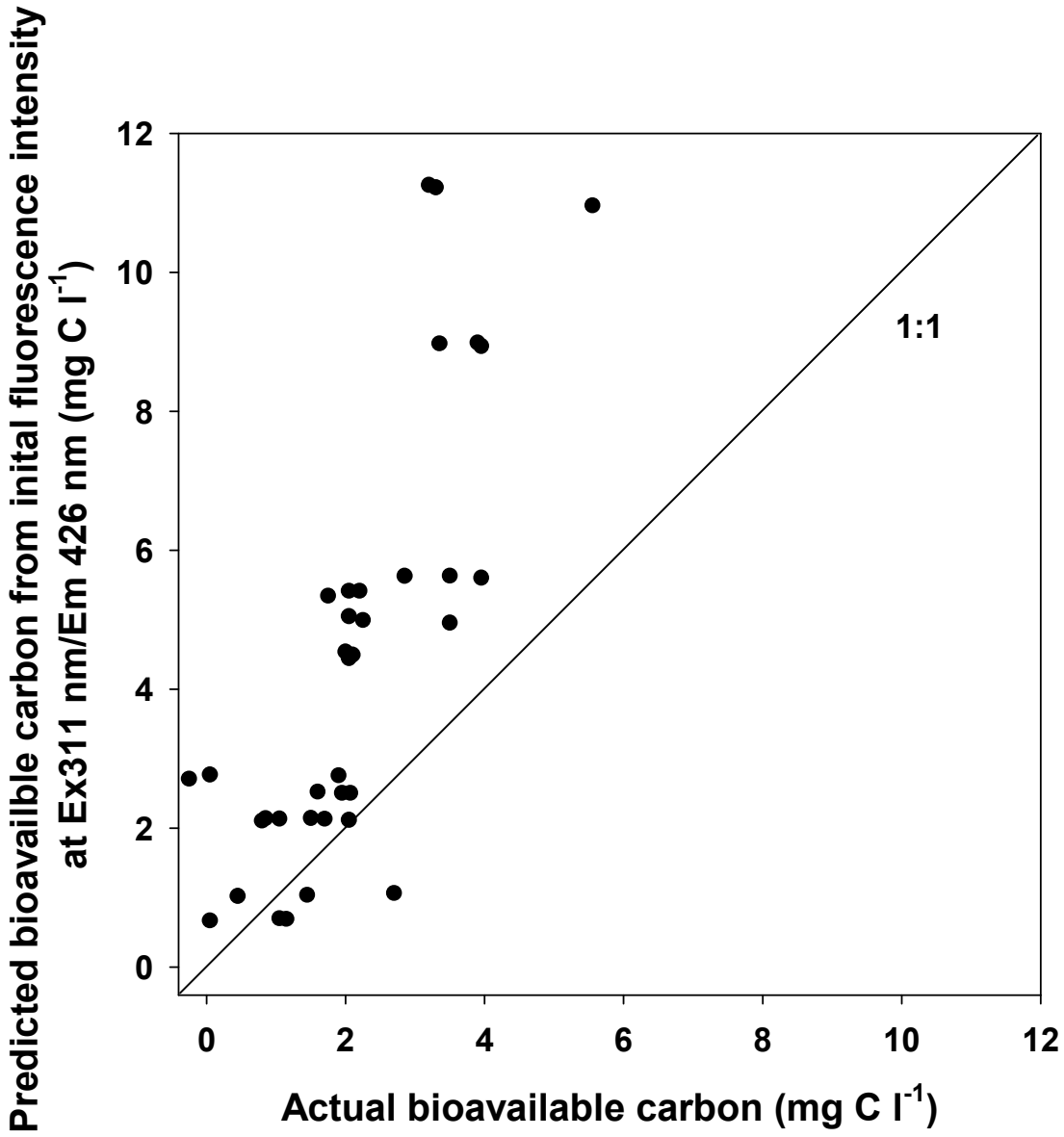


Fig. 10

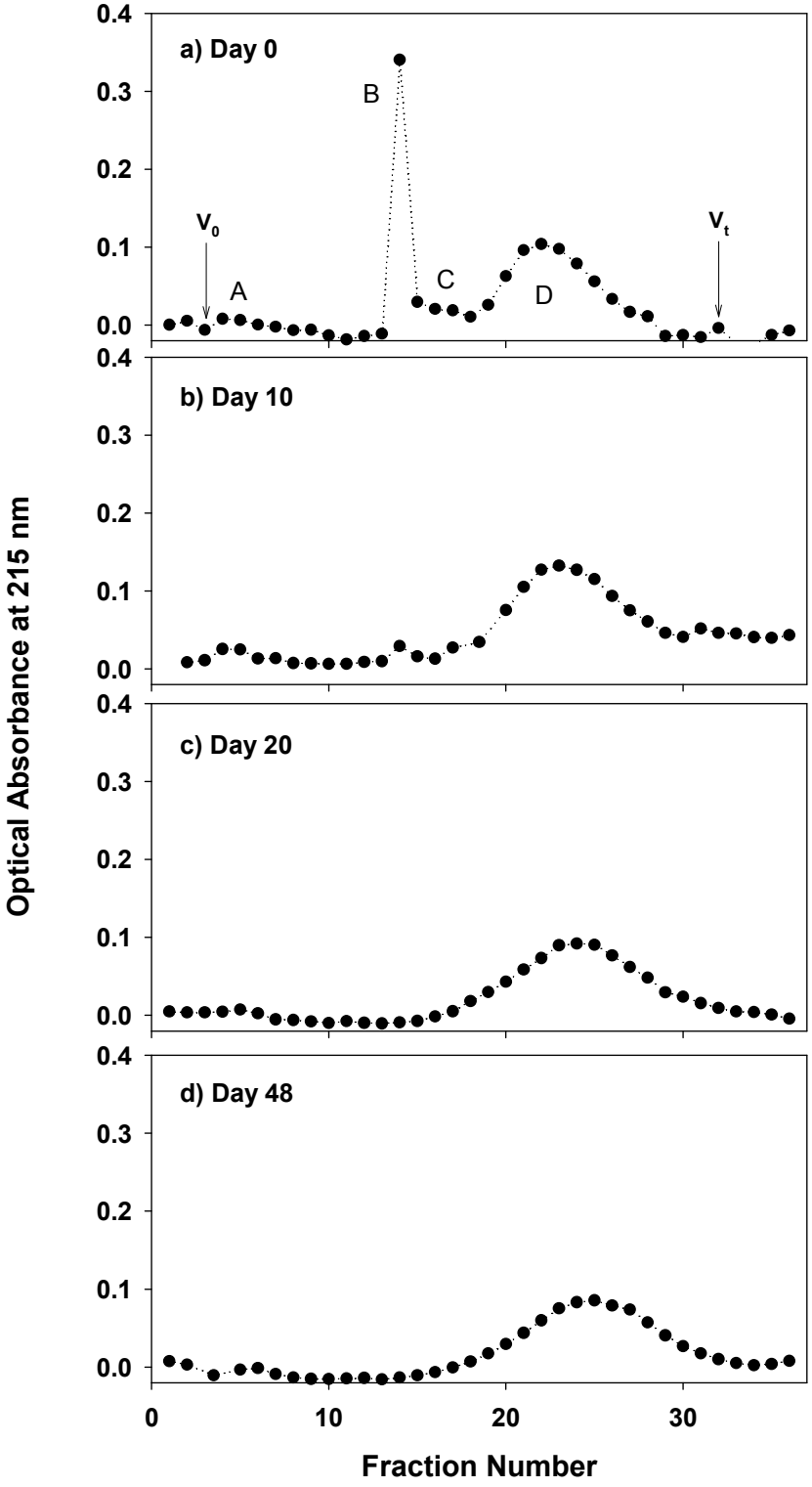


Fig 11