Critical Review of Fluorescence and Absorbance as Surrogates for the Aromaticity and Molecular Weight of Dissolved Organic Matter

Supporting Information

Julie A. Korak[a,b*] and Garrett McKay[c*]

[a]Department of Civil, Environmental, and Architectural Engineering

[b]Environmental Engineering Program, University of Colorado, Boulder, CO

[c]Zachry Department of Civil & Environmental Engineering, Texas A&M University, College Station, TX 77843

*Corresponding authors

1   Citation search terms

Figure 1 in the main manuscript tabulates the number of publications published by year for different absorbance and fluorescence based metrics. Table S 1 lists the search terms used for each metric and the database was queried on January 14, 2024. To prevent false positives, fluorescence metrics were screened to remove records prior to the genesis papers.

Table S 1. Search terms for Web of Science citation analysis.

Metric	Web of Science Search Terms
E2:E3	(AB="organic matter" OR AB="humic") AND (AB="E_2/E_3" OR AB=("E2E3") OR AB=("E2:E3") OR AB=("E(2)/E(3)"))
Spectral Slope	(AB="organic matter" OR AB="humic") AND AB=("Spectral slope")
SUVA	(AB="organic matter" OR AB="humic") AND (AB="SUVA" OR AB="specific absorbance" OR AB="absorption coefficient" OR AB="absorbance coefficient")
BIX	(AB="organic matter" OR AB="humic") AND (AB="BIX" OR AB="freshness index" OR AB="biological index" OR AB=".beta./.alpha.")
HIX	(AB="organic matter" OR AB="humic") AND (AB="HIX" OR AB="humification index")
FI	(AB="organic matter" OR AB="humic") AND (AB="Fluorescence") AND ((AB="FI ") OR (AB="fluorescence index"))

2   Fluorescence Inner Filter Corrections
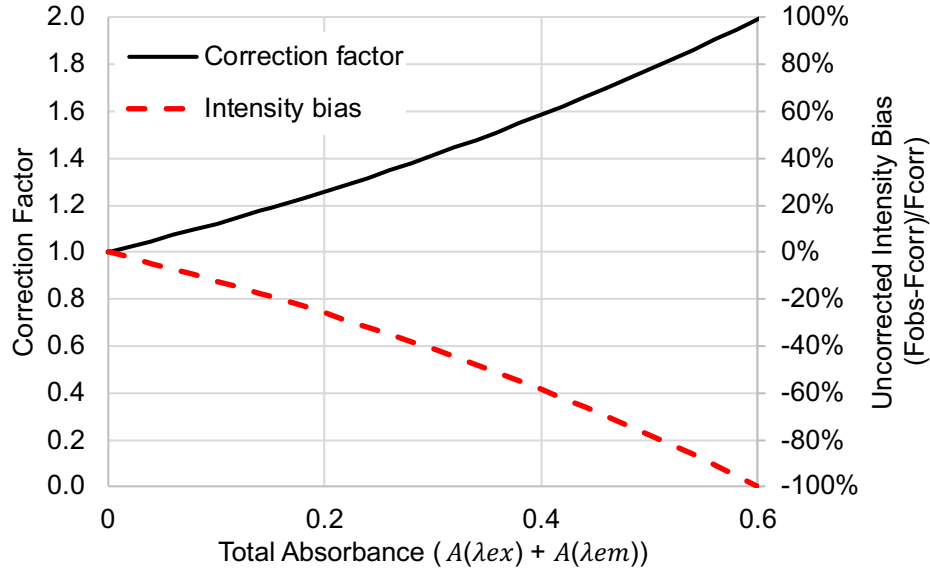
Figure S 1. Magnitude of the inner filter effect correction factor and the bias of uncorrected data as a function of total absorbance at a given excitation-emission wavelength combination.

3   Relationship between conventions for humification index (HIX)

Humification index (HIX) calculates a ratio of two integrated areas under the emission spectra at an excitation wavelength of 254 nm, which can either be measured directly or interpolated. The two methods are defined in Eqn S 1 and Eqn S 2, where the integrated areas are abbreviated as R and L according to Eqn S 3 and Eqn S 4, respectively.

$$HIX_{em,1999} = \frac{R}{L}$$ Eqn S 1

$$HIX_{em,2002} = \frac{R}{L+R}$$ Eqn S 2

$$R = \sum_{435\,nm}^{480\,nm} I(254\,nm, \lambda_{em})$$ Eqn S 3

$$L = \sum_{300\,nm}^{345\,nm} I(254\,nm, \lambda_{em})$$ Eqn S 4

Both definitions of HIX are rearranged in Eqn S 5 and Eqn S 6. Eqn S 7 sets both equations equal to one another.

$$R = L\,HIX_{em,1999}$$ Eqn S 5

$$R = (L + R)HIX_{em,2002} \qquad\qquad \text{Eqn S 6}$$

$$L\ HIX_{em,1999} = (L + R)HIX_{em,2002} \qquad\qquad \text{Eqn S 7}$$

Using algebra, Eqn S 7 is rearranged to yield Eqn S 9. Since the ratio of the two areas, R and L, is the initial definition of $HIX_{em,1999}$, the expression reduces to two unknowns. Collecting terms on each side, the two HIX conventions can be converted between each other using Eqn S 11.
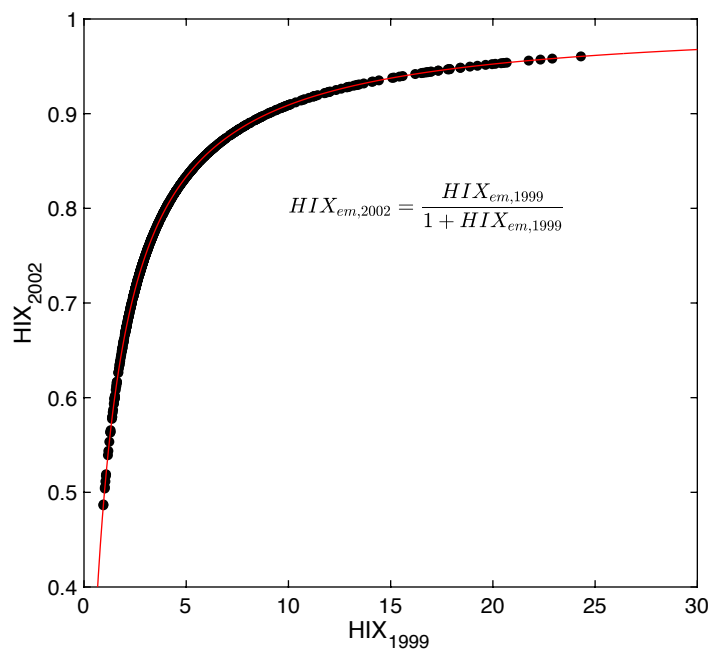
$$HIX_{em,1999} = \frac{(L+R)}{L}HIX_{em,2002} \qquad\qquad \text{Eqn S 8}$$

$$HIX_{em,1999} = \left(1 + \frac{R}{L}\right)HIX_{em,2002} \qquad\qquad \text{Eqn S 9}$$

$$HIX_{em,1999} = \left(1 + HIX_{em,1999}\right)HIX_{em,2002} \qquad\qquad \text{Eqn S 10}$$

$$HIX_{em,2002} = \frac{HIX_{em,1999}}{\left(1 + HIX_{em,1999}\right)} \qquad\qquad \text{Eqn S 11}$$

Using the data in Korak and McKay (2024)[1], each definition of HIX was calculated independently directly from the corrected fluorescence data.

The plotted equation reads:

$$HIX_{em,2002} = \frac{HIX_{em,1999}}{1 + HIX_{em,1999}}$$

Figure S 2. Plot relating both conventions of HIX using data from Korak and McKay (2024)[1] (n=695). The red line plots the model represented by Eqn S 11.

4   Estimation of specific intensity from PARAFAC components

Details of how to build and validate a PARAFAC model are documented in literature[2] and online resources[3]. This section describes how model terms and loadings were used to estimate a specific peak intensity following the modeling basis in Murphy et al. (2013). Parallel factor analysis (PARAFAC) fits a model using trilinear terms according to Eqn S 12.

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}$$ Eqn S 12

x_{ijk} is the fluorescence intensity of the ith sample (from 1 to I) at the jth excitation wavelength (from 1 to J) and k^{th} emission wavelength (from 1 to K). The index f corresponds to each PARAFAC component from 1 to F. For each component f, the variable a_{if} represents the score of the component for the ith sample and is proportional to the intensity of fluorescence from that component in the overall sample. The variables b_{jf} and c_{kf} are commonly presented as vectors (e.g., $\boldsymbol{b}_f$ and $\boldsymbol{c}_f$, respectively). These vectors are scaled emission $\boldsymbol{b}_f$ and absorption $\boldsymbol{c}_f$ spectra, respectively. With these vectors, the EEM (j x k) of each component can be calculated

4

by multiplying the vectors using matrix multiplication. Since the first term $(\sum_{f=1}^{F} a_{if} b_{jf} c_{kf})$ represents the modeled EEM, the last term, e_{ijk} is the residual term capturing the difference between the measured and modeled EEMs. In a well-fit model, the residual values should resemble random noise with values centered around zero that are small in magnitude compared to the measured signal.

After fitting and validation, PARAFAC results report an F_{max} value for each component in each sample $(F_{max,i,f})$, which is the product of the maximum emission loading, maximum excitation loading and model score (a_{if}), resulting in an extrinsic measure of sample fluorescence intensity.

It is also common to calculate the relative abundance of a PARAFAC component in each sample $(\%Comp_{i,f})$ using Eqn S 13.

$$\%Comp_{i,f} = \frac{F_{max,i,f}}{\sum_{f=1}^{F} F_{max,i,f}} \qquad\qquad \text{Eqn S 13}$$

To determine if there was a correlation between specific fluorescence and the relative abundance of condensed aromatic and polyphenolic formulas, as measured by FT-ICR-MS, the data in Kellerman et al. (2018)[4] was used to back calculate the fitted model intensity, assuming residuals are negligible. The methods section state that, "Isolates were redissolved in Milli-Q to a final concentration of 10 mg$_C$ L^{-1} to measure carbon-normalized absorbance and fluorescence properties." Therefore, the as-modeled fluorescence intensities are assumed to be carbon-normalized since samples were prepared at the same carbon concentration.

For each sample, Table 1 in Kellerman et al. (2018) lists the $F_{max,total}$ value $(\sum_{f=1}^{F} F_{max,i,f})$ and the relative abundance $(\%Comp_{i,f})$ of each sample, where relative abundance is converted from a percent to fraction. Eqn S 13 can be rearranged to Eqn S 14 to calculate the F_{max} for each component in each sample $(F_{max,i,f})$.

$$F_{max,i,f} = \frac{\%Comp_{i,f}}{100} * \sum_{f=1}^{F} F_{max,i,f}$$ Eqn S 14

Table S2 in Kellerman et al. (2018) reports the excitation and emission vectors for each component (e.g., $\boldsymbol{b}_f$ and $\boldsymbol{c}_f$, respectively). If $F_{max,i,f}$ is representative of the maximum intensity at the component maximum, then the vectors can be used to scale intensity between the maximum to an excitation-emission wavelength combination in the Peak C region. For example, using the data from McKay et al. (2018), Peak C was calculated at excitation 320 nm and emission 440 nm. Based on the PARAFAC components in Kellerman et al. (2018), component 1 has a maximum intensity near excitation 260 nm and likely centered in the emission range of 411.58-434.63 nm. There is a secondary excitation maximum for component 1 between 310-320 nm, aligning component 1 with Peak C. Resolution is limited by the number of significant figures reported in the loading vectors.

However, there is component overlap and a calculation of relative fluorescence intensity must consider the contributions from the other components. Since Fmax values correspond to the maximum for each component, a scaling factor (SF) is calculated for each component in **Table S 2** using Eqn S 15 to estimate intensities of each component at Peak C.

$$SF_f = \frac{b_{320\ nm,f} * c_{440\ nm,f}}{b_{\lambda ex@Fmax,f} * c_{\lambda em@Fmax,f}}$$ Eqn S 15

Table S 2. F_{max} scaling factors between the component maximum and Peak C.

Component	F_{max}		Component Loadings at Wavelengths for F_{max}			Component Loadings at Wavelengths for Peak C (Ex 320 nm, Em 440 nm)			Scaling Factor (SF)
	Ex (nm)	Em (nm)	b	c	b*c	b	c	b*c	
1	260	411.58–464.63	0.29	0.19	0.055	0.25	0.17	0.043	0.77
2	260	469.35–490.25	0.25	0.17	0.043	0.16	0.11	0.018	0.41
3	260	536.89–553.27	0.25	0.16	0.040	0.17	0.02	0.003	0.09
4	260	363.41–384.01	0.3	0.19	0.057	0.18	0	0.000	0.00
5	260	430.02–469.35	0.47	0.14	0.066	0.06	0.14	0.008	0.13

With the scaling factor for each component, a specific intensity at Peak C for the modeled EEM is calculated for each sample using Eqn S 16.

$$SpC_i = 0.77 \, Fmax_{i,1} + 0.41 \, Fmax_{i,2} + 0.09 \, Fmax_{i,3} + 0.13 \, Fmax_{i,5} \qquad \text{Eqn S 16}$$

Since all samples were reported as prepared at the same concentration, applying Eqn S 16 to each isolate in Table 2 of Kellerman et al. (2018) yields an internally consistent value that is at least proportional to the measured carbon-normalized intensity to investigate if correlations exist between specific intensity and aromaticity.

5   References

(1) Korak, J. A.; McKay, G. Meta-Analysis of Optical Surrogates for the Characterization of Dissolved Organic Matter. *Accepted to Environmental Science & Technology* **2024**.

(2) Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R. Fluorescence Spectroscopy and Multi-Way Techniques. PARAFAC. *Analytical Methods* **2013**, *5*, 6557–6566. https://doi.org/10.1039/c3ay41160e.

(3) Murphy, K.; Wünsch, U. *drEEM Toolbox*. https://dreem.openfluor.org/ (accessed 2023-12-29).

(4) Kellerman, A. M.; Guillemette, F.; Podgorski, D. C.; Aiken, G. R.; Butler, K. D.; Spencer, R. G. M. Unifying Concepts Linking Dissolved Organic Matter Composition to Persistence in Aquatic Ecosystems. *Environ Sci Technol* **2018**, *52* (5), 2538–2548. https://doi.org/10.1021/acs.est.7b05513.