­­



*Global Biogeochemical Cycles*

**Quantification of discharge-specific effects on dissolved organic matter export from major Arctic rivers from 1982 through 2019**

J. Blake Clark1,2, Antonio Mannino1, Robert G.M. Spencer3, Suzanne E. Tank4, James W. McClelland5,6

1. Ocean Ecology Laboratory, Code 616.1, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

2. Goddard Earth Sciences Technology and Research II, University of Maryland, Baltimore County, Baltimore, Maryland, USA

3. Department of Earth, Ocean and Atmospheric Science, Florida State University, Tallahassee, Florida, USA

4. Department of Biological Sciences, University of Alberta, Edmonton, Alberta Canada

5. Marine Science Institute, The University of Texas at Austin, Port Aransas, TX, USA

6. The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts, USA

**Contents of this file**

Text S1 to S1.3

Table S1

Figures S1 to S4

**Introduction**

Below is supporting information for the above manuscript that details changes in Arctic River dissolved organic carbon chemistry and fluxes over time. The text provides additional details for some of the methods used in the derivation of dissolved organic carbon optical properties and modeling of river fluxes. There are also figures related to the methodology that shoes some statistical assessment and uncertainty of the hydrologic model predictions.

**S1. Methods**

**S1.1 Quality Control of Colored Dissolved Organic Matter Spectra**

Each spectrum was baseline corrected by using a non-linear fitting routine to find an estimated offset of the spectra from a baseline of zero at 750 nm (Markager and Vincent 2000**).** Each absorbance scan (200-800 nm) was fit from 350-800 nm using Eq. S1 and the MATLAB® function *fitnlm.m* where spectral CDOM absorbance, abs(λ), is a function of wavelength, λ. b1 is a correction factor that was added to the total absorbance spectrum, and b2 and b3 are fitting coefficients. If the model predicted vs observed absorbance coefficient of covariance (R2) was greater than 0.8 (indicating an expected exponential decay) and the absorbance at 400 nm was greater than b1, the spectra was deemed acceptable. This removed spectra with unreasonable shapes and declines before 400 nm and spectra that were particularly noisy in the region of the absorbance spectrum that should be relatively smooth. The total reduction in suitable spectra were 1, 0, 10, 1, 0, and 8 for the Kolyma, Lena, Mackenzie, Ob’, Yenisey, and Yukon rivers. This filtering process ensured that high quality spectra were used in the subsequent mathematical derivation where spectral quality is of substantial importance to limit measurement error propagation into the derived spectra and CDOC concentration. Last, the spectra were reduced to 270-700 nm, absorbance was converted to Naperian absorption (m-1), and spectral slopes for 275-295 and 350-400 nm were calculated as previously reported (Helms et al. 2008; Mannino et al. 2014)

 (S1)

**S1.2 Statistical Analysis and Trend Estimates**

 The iterative least squares regression function to estimate individual CDOC concentrations for each measurement were assessed by using the sum of squared residuals (SSR) for each predicted CDOM(λ) spectra that CDOC concentration were being fit to according to Eq. 4. LOADEST model predicted concentration and CDOM metrics were assessed for model appropriateness and bias using coefficient of covariance (R2) and the statistical distribution of the model residuals (bias). To estimate trends in seasonal and annual riverine patterns and combined trends for all six rivers, Mann-Kendall analysis and Theil-Sen’s slope trend analytical regression techniques were utilized (Helsel et al., 2020). Theil-Sen’s slope offers a more robust linear fitting technique relative to OLS when data residuals are not normally distributed, and is commonly used in riverine trend estimates (e.g. McClelland et al., 2006). These non-parametric statistics were utilized using the MATLAB® open-source function *ktaub.m* (Burkey 2006) to estimate long-term trends for each river and the total combined from the six rivers. Seasonal and annual mass loads and concentrations were calculated by summing the total annual mass and averaging the concentration for each time period for each year, giving long-term time series estimates of mass flux and concentration.

**S1.3 LOADEST Model Setup**

LOADEST is a Fortran program designed to estimate riverine material loads (i.e. mass export) using observations of river *Q* and concentration (Runkel et al. 2004). LOADEST predicts mass flux as a function of *Q* by centering time dependent observations of river flow and concentration and linear combination of functions to estimate river export (and subsequently concentration) at any measured *Q*. Adjusted maximum likelihood estimates (AMLE) were used for each predicted load because overwhelmingly model residuals were found to be normally distributed (SI Fig. S2). Model Six (Eq. S2) was used as in previous studies (e.g. McClelland et al. 2016) to avoid spurious extrapolations due to inter-annual trends in *Q* while accounting for seasonal variance. Therefore, any changes in load and concentration over time are due to changing seasonality, and not long-term extrapolated trends. Daily load, *L* (kg d-1), was estimated and daily concentration was derived as *L* divided by the total daily river flow volume, and a­1-4 are fitting coefficients while *Q* is the centered river *Q* and *dtime* is the centered decimal time. *a300*, *S*275-295, and *S*350-400 were treated as a concentration for model input and the output units were adjusted to match the input by dividing the final value by 1000 yielding original units.

 (S2)

**Supporting Information Table S1** River dissolved organic carbon (DOC) and non-colored and initial colored DOC (NCDOC and iCDOC) characteristics from the set of ArcticGRO measurements from 2009-2019 where both DOC and CDOM were coincidentally measured. NCDOC is the intercept of the regression of measured DOC concentration as a function of CDOM absorption at 300 nm (a300) and iCDOC was the initial CDOC concentration taken as the difference between the DOC measurements NCDOC. σ represents the standard deviation, CI represents the 95% confidence intervals, R2 represents the coefficient of determination, and SR represents the spectral slope ratio (Helms et al. 2008). Associated plots can be found in Fig. 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **DOC (g C m-3) ± σ****Winter****Spring****Summer****All** | **DOC:a300 Slope (g C m-3 m-1), R2,** | **NCDOC****(g C m-3)****(95% CI)** | **iCDOC (g C m-3) (95% CI)** | **a300 (m-1) ± σ** | **SR (275-295/350-400)** |
| **Kolyma****n=21, 17, 18, 56** | 3.7±1.1 8.8±4.2 4.3±0.9**5.4±3.3** | 0.29, 0.820.24, 0.920.26, 0.51**0.23, 0.94** | 1.3 (0.7, 1.9)1.2 (0.1, 2.5)0.9 (0.1, 2.6)**1.5 (1.2, 1.9)** | 2.2 (1.8, 2.5)7.2 (6.9, 7.6)2.7 (2.4, 3.1)**3.9 (3.5, 4.2)** | 8.1±3.331.8±17.113.0±2.6**16.9±13.9** | 0.87±0.080.84±0.040.88±0.05**0.87±0.06** |
| **Lena****n=24, 17, 17, 58** | 7.1±2.114.9±5.77.4±1.7**9.5±4.9** | 0.22, 0.730.20, 0.820.17, 0.68**0.20, 0.90** | 1.4 (0.1, 2.9)1.3 (0.1, 4.9)2.7 (0.9, 4.5)**1.9 (1.1, 2.7)** | 5.2 (4.4, 5.9)13.0 (12.2, 13.8)5.6 (4.8, 6.3)**7.6 (6.8, 8.4)** | 26.3±8.266.6±25.227.5±8.4**38.4±23.6** | 0.88±0.030.85±0.040.90±0.03**0.87±0.04** |
| **Mackenzie****n=17, 13, 18, 48** | 4.3±0.56.0±1.24.7±0.9**4.9±1.1** | 0.31, 0.360.13, 0.240.21, 0.67**0.16, 0.62** | 2.0 (0.4, 3.6)3.6 (1.1, 6.1)2.1 (1.2, 3.1)**3.0 (2.5, 3.4)** | 1.3 (0.8, 1.8)3.0 (2.5, 3.5)1.8 (1.3, 2.2)**1.9 (1.5, 2.4** | 7.4±1.118.2±5.012.1±3.4**12.1±5.4** | 0.96±0.220.90±0.050.87±0.10**0.91±0.15** |
| **Ob’****n=24, 16, 18, 58** | 7.6±2.410.2±2.411.4±2.9**9.5±3.0** | 0.16, 0.780.17, 0.660.16, 0.63**0.17, 0.78** | 2.8 (1.7, 4.0)2.5 (0.1, 5.6)3.6 (0.5, 6.8)**2.7 (1.6, 3.7)** | 5.0 (4.0, 6.0)7.6 (6.5, 8.6)8.8 (7.8, 9.8)**6.9 (5.8, 7.9)** | 29.8±13.144.7±11.748.5±14.3**39.7±15.5** | 0.82±0.080.82±0.040.85±0.05**0.83±0.06** |
| **Yenisey****n=24, 16, 19, 59** | 3.6±1.08.6±3.05.9±1.8**5.7±2.8** | 0.21, 0.800.16, 0.910.18, 0.73**0.16, 0.93** | 1.3 (0.8, 1.9)2.0 (0.8, 3.3)1.7 (0.3, 3.0)**1.9 (1.6, 2.3)** | 1.6 (1.3, 2.0)6.6 (6.3, 7.0)4.0 (3.6, 5.3)**3.7 (3.4, 4.1)** | 10.6±4.141.5±18.222.9±8.6**30.0±16.5** | 0.84±0.100.83±0.050.90±0.05**0.85±0.07** |
| **Yukon****n=12, 16, 22, 50** | 3.5±1.58.8±3.85.4±1.6**6.1±3.2** | 0.19, 0.960.19, 0.890.21, 0.80**0.20, 0.93** | 1.6 (1.3, 2.0)2.0 (0.6, 3.5)1.5 (0.5, 2.4)**1.7 (1.3, 2.1)** | 1.8 (1.3, 2.2)7.1 (6.7, 7.5)3.7 (3.3, 4.2)**4.4 (3.9, 4.8)** | 9.5±7.835.6±18.718.7±6.9**21.9±15.6** | 0.78±0.160.82±0.040.83±0.07**0.81±0.09** |

****

**Figure S1** Workflow used to derive CDOC specific absorption spectra (a\*CDOC(λ)) for each river by utilizing the information produced from ordinary least squares (OLS) linear regressions of dissolved organic carbon (DOC) as a function of CDOM absorption at 300 nm (a300). The final products (purple box) are a statistically robust estimate (with confidence intervals) of colored and non-colored DOC concentration for each set of DOC and CDOM measurements to be used in the USGS Load Estimator predictions.



**Figure S2** Dissolved organic carbon seasonal concentration distribution for the six rivers for 2009-2019 measured under the ArcticGRO project. The edges of each box represent the 25th and 75th percentiles, the whiskers represent the range of the data that are not outliers, and the marks beyond the whiskers are outliers. Produced using the MATLAB file exchange function *boxplot2.m* (Kearney 2012).



**Figure S3** Monthly trends in total river flow for the 6 Arctic rivers and the total Big 6. Trends were calculated using the timeseries of the 1982-2019 monthly total discharge for each river and Theil-Sen’s slope regression (Hollander and Wolfe, 1973). Months with missing data were excluded and black squares are months that lacked a significant trend (p<0.1).

**Figure S4** (a)Dissolved organic carbon (DOC) and (b) CDOM absorption at 300 nm observed vs. predicted and the residuals for each (below) exhibiting the predictive skill of the USGS Load Estimator model for each constituent.