

Appendix 1. Model Archive Summary for Dissolved Organic Carbon Concentrations at Station 254543080405401: Tamiami Canal at S-12D Near Miami, Florida

Previous studies have established that dissolved organic carbon (DOC) in the Florida Everglades plays an important role in many environmental processes, one of which is the interaction between DOC and mercury (Hg), which is helpful in the study of mercury methylation. This study focuses on the development of surrogate models to calculate DOC concentrations in the Florida Everglades. Fluorescence of chromophoric dissolved organic matter (fDOM) is often used as a surrogate for DOC.

High density and long-term data will aid with the description of short and long-term variability of carbon and mercury concentrations, which will improve understanding of carbon input and transport. Prior to this study, no continuous and long-term time-series data on carbon concentrations were available for the freshwater wetlands of the Florida Everglades.

The objectives of this study were to develop and document carbon surrogate models to calculate concentration and loads at site S-12D. This model archive summary describes the DOC model developed to compute 15-minute frequency DOC concentrations from fDOM data collected from September 5, 2013, to April 3, 2017. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in Rasmussen and others (2009).

Site and Model Information

USGS site number: 254543080405401

https://waterdata.usgs.gov/fl/nwis/inventory/?site_no=254543080405401&agency_cd=USGS

Site name: Tamiami Canal at S-12D near Miami, Florida

Location: lat 25°45'43" N., long 80°40'54" W., referenced to North American Datum of 1927, in T. 54 S., R. 36 E., Miami-Dade County, Florida, hydrologic unit 03090202, on south bank 100 feet southwest of structure 12-D, near east boundary of Indian reservation on U.S. Highway 41.

Equipment: A YSI EXO water-quality monitoring system equipped with sensors for water temperature, specific conductance, turbidity, and an fDOM sensor (Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government). The monitor is housed in an 8-inch-diameter polyvinyl chloride (PVC) pipe on a diagonal extending off the end of the structure into the stream. Readings from the YSI EXO were recorded every 15 minutes and

transmitted hourly by way of satellite. The model applies only to this site (254543080405401) and specified time period (September 5, 2013, to April 3, 2017).

Model number: 1.0

Date model was created: April 12, 2018

Model calibration data period: September 11, 2013, to December 19, 2016

Model application date: September 5, 2013, to April 3, 2017

Computed by: Amanda Booth, USGS Caribbean-Florida Water Science Center, April 12, 2018

Reviewed by: M. Stone, USGS Kansas Water Science Center, Lawrence, Kansas, and B. Downing, USGS California Water Science Center, Sacramento, California

Approved by: USGS David Sumner, Caribbean-Florida Water Science Center

Model Data

All data were collected using USGS protocols and are stored in the National Water Information System (NWIS) database (U.S. Geological Survey, 2019). fDOM data were corrected for temperature, turbidity, and inner filter effects (Downing and others, 2012). The regression model is based on 31 concurrent measurements of DOC and fDOM samples collected during September 11, 2013, through December 19, 2016. Samples were collected throughout the range of observed hydrologic and fDOM conditions. Summary statistics and the complete model-calibration data are provided in the dataset. Studentized residuals from the final model were inspected for values greater than 3 or less than -3 to identify potential outliers. No outliers were found within the dataset.

Dissolved Organic Carbon Data

Samples were collected using a 0.25-inch-diameter hydrokinetic nozzle, 200 milliliter high-density polyethylene bottle, and nozzle-bottle holder. Samples were collected at the location and depth of the water-quality sensors and filtered through a 0.45-micron capsule filter into amber glass 2-liter bottles within 15 minutes of being collected and stored immediately on ice. Samples were analyzed for DOC at the USGS laboratory in Boulder, Colorado, between September 2013 and September 2016. Samples collected between October 2016 and February 2016 were analyzed by the USGS National Water Quality Laboratory (NWQL) in Denver, Colo. The first sample of the day was the only sample for which data were included in model development, regardless of which laboratory made the analysis. Samples were collected and analyzed by both the USGS laboratory in Boulder, Colo., and the NWQL on February 2, May 23, June 20, July 18, August 16, and September 19, 2016. Relative percent differences from the split samples ranged from -1.1 to 10.2 percent, with an average of 4.3 percent. Although the results from the NWQL were slightly lower, on average, than those reported from the Boulder laboratory, samples from both laboratories were included in model development.

Surrogate Data

The fDOM data used in this analysis were measured using a YSI EXO V2, serial numbers 13f100955, 14c101755, 14c100465, 14c100466, 15G100778, 15G100779 and 15c104523. The fDOM data were corrected for temperature, turbidity, and inner filter effects (Downing and others, 2012). The equation for the turbidity correction was provided by B. Pellerin (U.S. Geological Survey, written commun., 2018), and determined using YSI EXO, temperature, turbidity and fDOM sensors. The equation was determined using Elliot silt loam and is provided below.

Elliot silt loam turbidity correction – exponential fit (table curve):

$$fDOM_{turbcorr} = \frac{fDOM_{tempcorr}}{e^{(-0.027567 - 0.06259 \cdot FNU)}}$$

where

$fDOM_{tempcorr}$ = fluorescence of chromophoric dissolved organic matter, YSI EXO V2, temperature corrected, in parts per billion quinine sulfate equivalents (QSE);

$fDOM_{turbcorr}$ = fluorescence of chromophoric dissolved organic matter, YSI EXO V2, temperature and turbidity corrected, in parts per billion quinine sulfate equivalents; and

FNU = turbidity (YSI EXO model), in formazin nephelometric units.

Serial dilutions were performed to determine the inner filter effect using Caloosahatchee River at S-79 (02292900) native water, filtered through a 0.45-micron filter, collected on August 16, 2017. Native water from S-12D was not used for this correction, because we did not collect water at S-12D that had a high enough initial fDOM concentration to complete a serial dilution that would represent the range of fDOM concentrations observed during the study. Data collected at S-12D on September 28, 2017, indicate that the Caloosahatchee water is representative of S-12D with respect to the inner filter effect.

Inner filter effect correction:

$$fDOM_{corr} = 0.002 \cdot (fDOM_{turbcorr})^2 + 0.7608 \cdot fDOM_{turbcorr}$$

where

$fDOM_{corr}$ = fluorescence of chromophoric dissolved organic matter, YSI EXO V2, temperature corrected, turbidity and inner filter corrected, in quinine sulfate equivalents; and

$fDOM_{turbcorr}$ = fluorescence of chromophoric dissolved organic matter, YSI EXO V2, temperature and turbidity corrected, in quinine sulfate equivalents.

The *fDOM* values at S-12D ranged from 79.9 QSE on February 14, 2014, to 610.5 QSE on August 1, 2014. Specific conductance at S-12D ranged from 260 microsiemens per centimeter at 25 degrees Celsius ($\mu\text{S}/\text{cm}$ @ 25 °C) on June 27, 2013, to 916 $\mu\text{S}/\text{cm}$ @ 25 °C on Sept. 22, 2015.

Model Development

Regression analysis was done using Microsoft Excel and the USGS Surrogate Analysis and Index Developer (SAID) tool (Domanski and others, 2015) by examining *fDOM* and other continuously measured data as explanatory variables for estimating DOC concentration. A variety of models that predict DOC were evaluated. The distribution of residuals was examined for normality and plots of residuals (the difference between the observed and computed values) as compared to computed DOC were examined for homoscedasticity. There was a slight negative trend in the residuals that can be attributed to a switch in laboratories used for DOC analysis. *fDOM* was selected as the best predictor of DOC based on residual plots, relatively high adjusted coefficient of determination (adjusted R^2), and relatively low model standard percentage error. Values for all of the aforementioned statistics and metrics were computed and are included below, along with all relevant sample data and more in-depth statistical information. When discharge (*Q*) equaled zero, a *Q* value of 0.001 was entered in order for the program to create the graphics.

Model Summary

Summary of final regression analysis for DOC concentration at site number 254543080405401.

Dissolved organ carbon concentration-based model:

$$DOC = 0.016 * SPC + 0.0143 * fDOM + 6.26$$

where

DOC = dissolved organic carbon concentration, in milligrams per liter;

SC = specific conductance, in microsiemens per centimeter at 25 degrees Celsius; and

fDOM = fluorescence of chromophoric dissolved organic matter, YSI EXO model, temperature and turbidity corrected, in quinine sulfate equivalents.

The use of *fDOM* as an explanatory variable is appropriate physically and statistically. *fDOM* refers to the fraction of chromophoric dissolved organic matter that fluoresces and is often used as a surrogate for dissolved organic carbon (Spencer and others, 2007; Bergamaschi and others, 2012; Pellerin and others, 2012). The use of specific conductance as a variable is also supported statistically and within the literature (Curtis and Adams, 1995; Monteiro and others, 2014).

Model Statistics, Data, and Plots

Model

$$\text{DOC} = 0.016 * \text{SC} + 0.0143 * \text{fDOM} + 6.26$$

Variable Summary Statistics

	DOC	SC	fDOM
Minimum	13.4	366	119
1st Quartile	16.4	457	237
Median	19.9	548	296
Mean	19.1	541	289
3rd Quartile	21.1	611	342
Maximum	25.8	796	538

Exploratory Plots

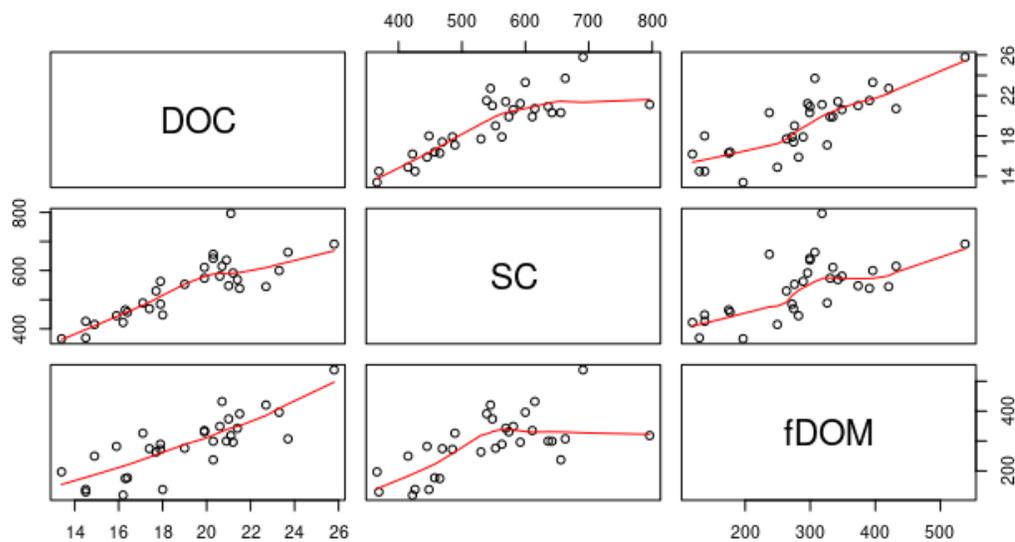


Figure 1.1. Comparison of dissolved organic carbon (DOC) concentration, in milligrams per liter; specific conductance (SC), in microsiemens per centimeter at 25 degrees Celsius; and fluorescence of chromophoric dissolved organic matter (fDOM), in quinine sulfate equivalents.

Basic Model Statistics

Number of Observations	31
Standard error (RMSE)	1.32
Average Model standard percentage error (MSPE)	6.94
Coefficient of determination (R^2)	0.819
Adjusted Coefficient of Determination (Adj. R^2)	0.806

Variance Inflation Factors (VIF)

SC	FDOM
1.73	1.73

Explanatory Variables

	Coefficients	Standard Error	t value	Pr(> t)
(Intercept)	6.2600	1.34000	4.69	6.54e-05
SC	0.0160	0.00317	5.06	2.36e-05
FDOM	0.0143	0.00328	4.37	1.53e-04

Correlation Matrix

	Intercept	SC	FDOM
Intercept	1.000	-0.823	0.123
SC	-0.823	1.000	-0.649
FDOM	0.123	-0.649	1.000

Outlier Test Criteria

Leverage	Cook's D	DFFITS
0.290	0.262	0.622

Flagged Observations

	DOC	Estimate	Residual	Standard Residual	Studentized Residual	Leverage	Cook's D	DFFITS
03/14/2014 12:05	23.7	21.3	2.41	1.92	2.02	0.103	0.142	0.687
06/24/2014 10:18	21.1	23.6	-2.49	-2.34	-2.56	0.355	1.010	-1.900
04/28/2015 09:49	18.0	15.4	2.58	2.07	2.21	0.114	0.185	0.796

Three observed DOC concentrations were flagged as potential outliers. All three samples were flagged because the difference in fits (DFFITS) values were greater than 0.622. The sample on June 24, 2014, was also flagged because the Cook's D was greater than 0.262 and the leverage was greater than 0.290. All flagged observations were included in model development.

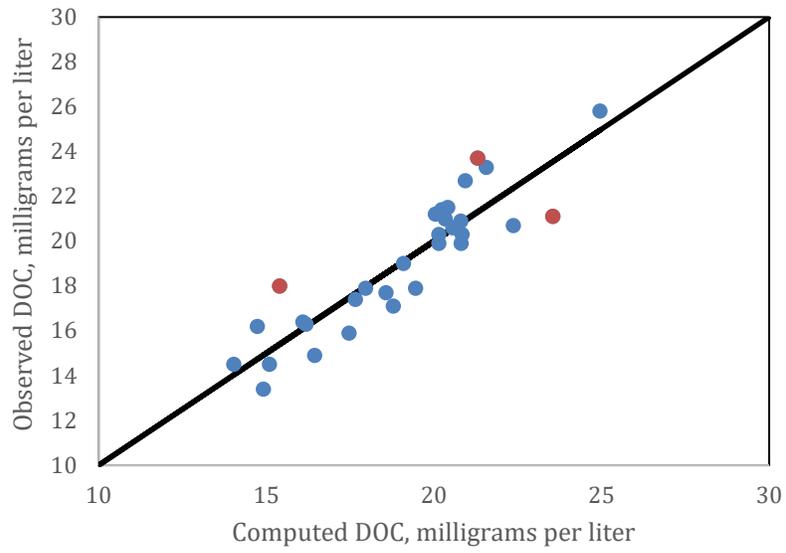


Figure 1.2. Relation between observed and computed dissolved organic carbon (DOC); flagged observations are in red.

Statistical Plots

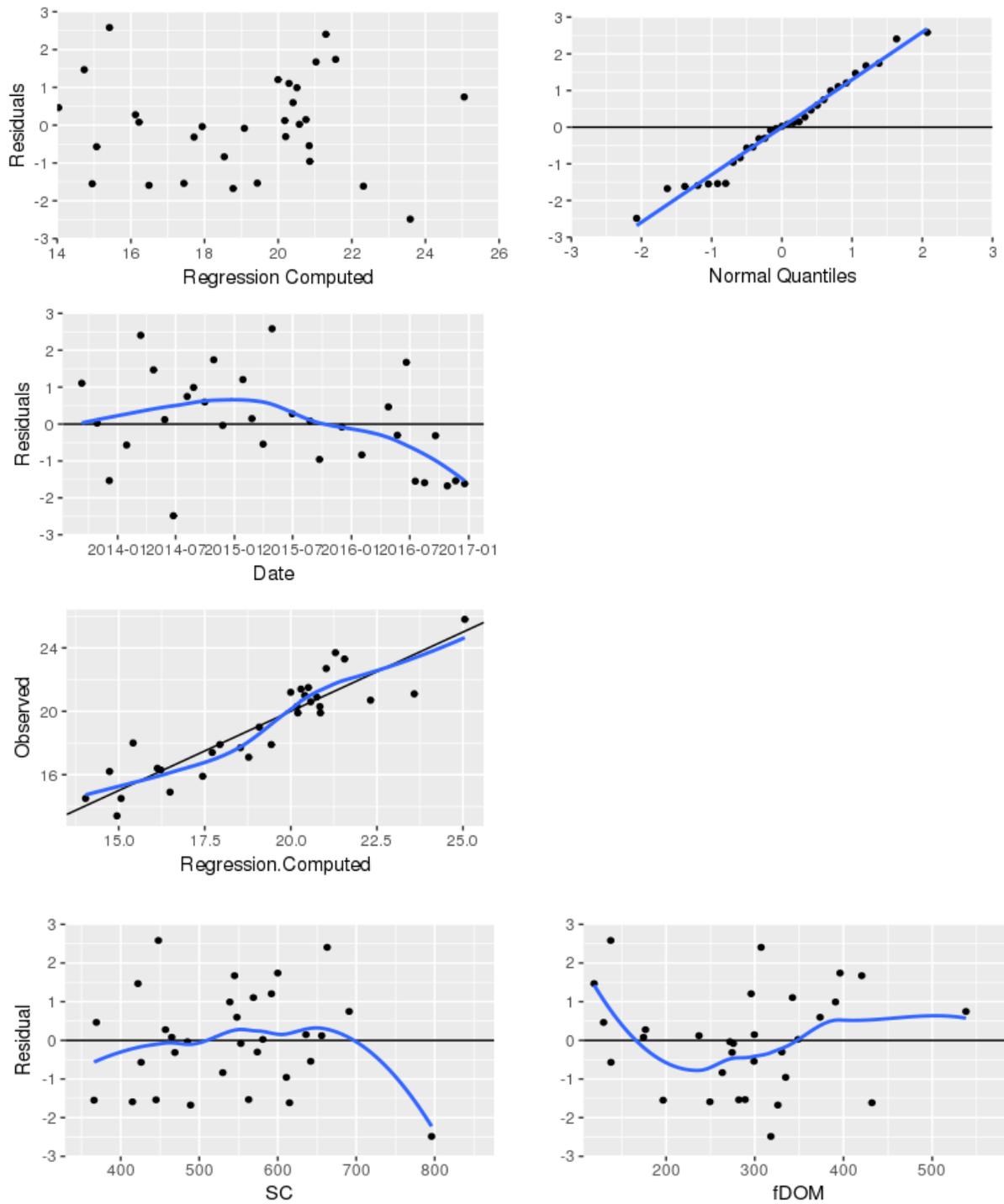


Figure 1.3. Residual and observed versus computed plots.

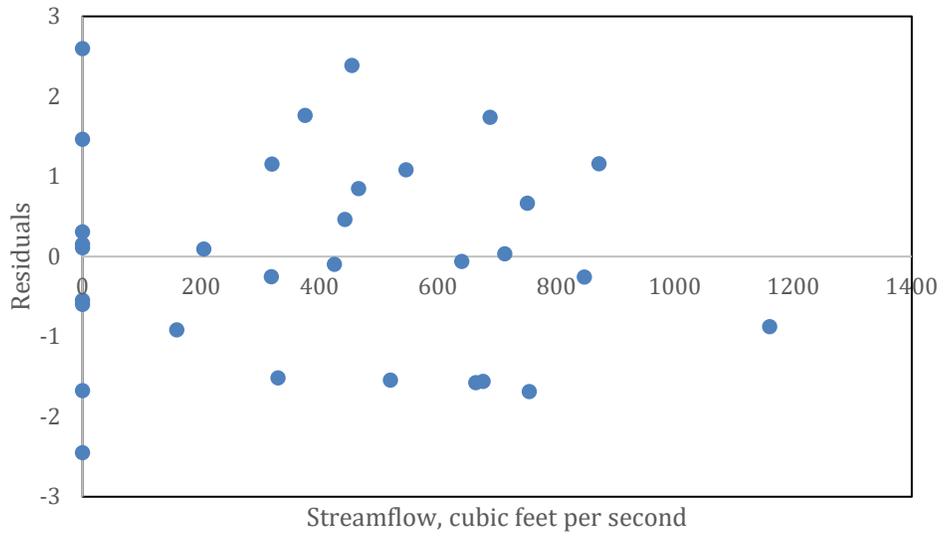
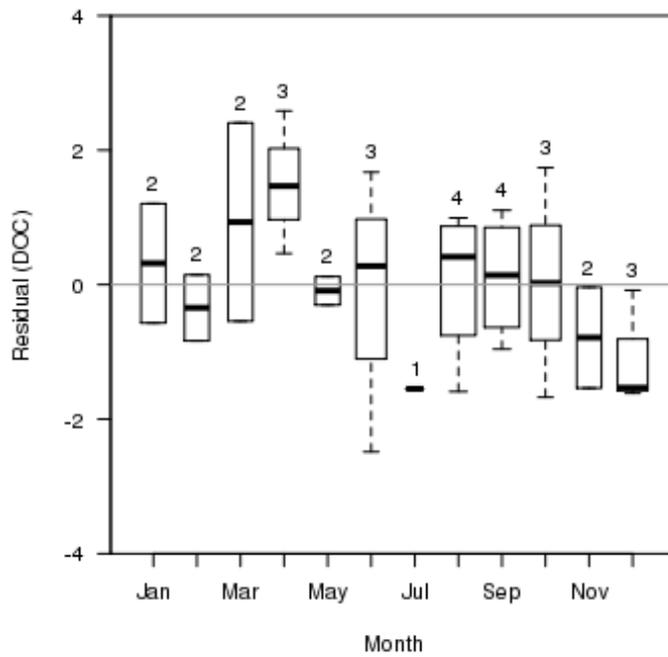


Figure 1.4. Relation between residuals and streamflow showing that the residuals had no systematic bias with respect to streamflow.

A.



B.

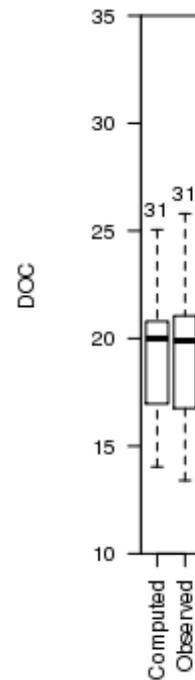


Figure 1.5. A, Seasonal variation in residuals, and B, computed and observed dissolved organic carbon (DOC), in milligrams per liter.

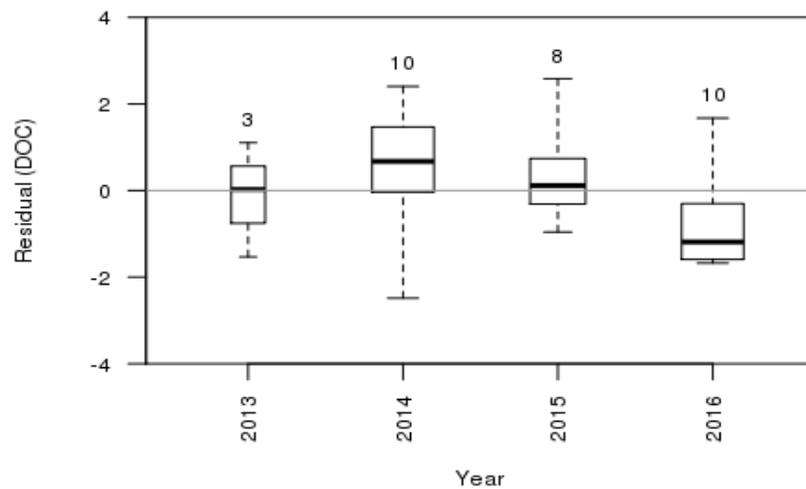
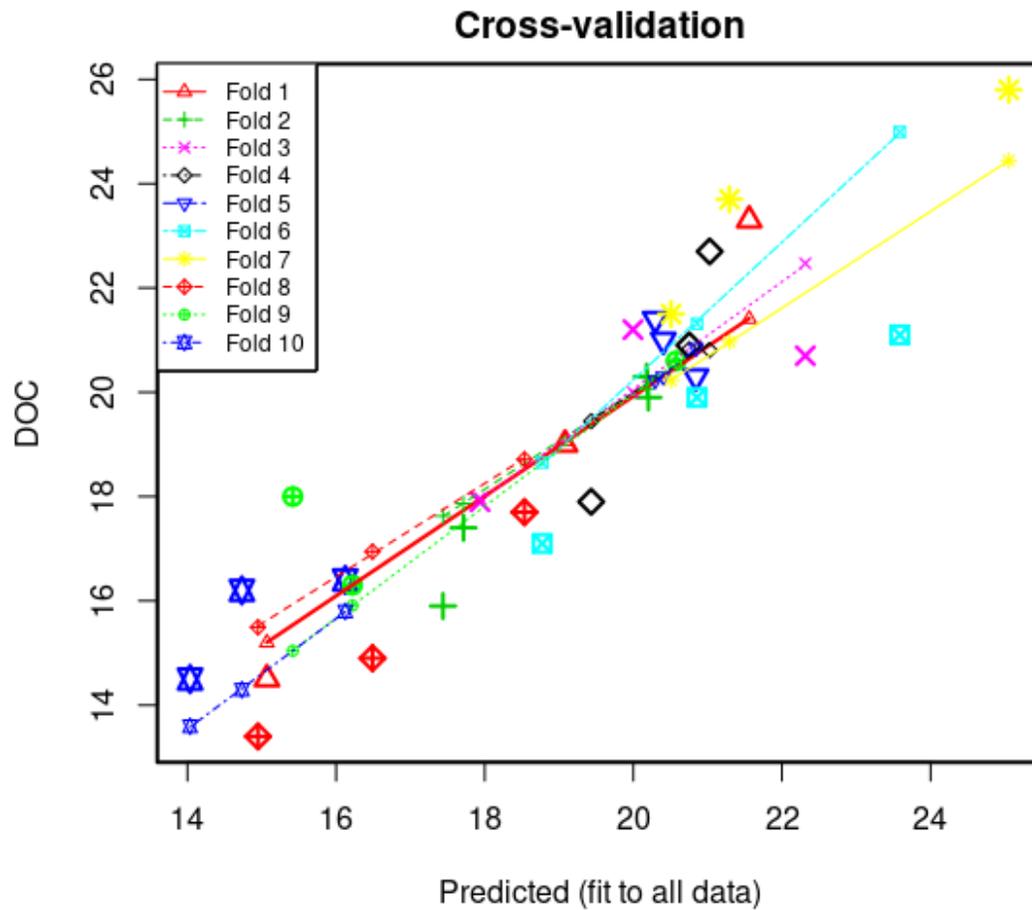


Figure 1.6. Annual variation in residuals.

(DOC, dissolved organic carbon)

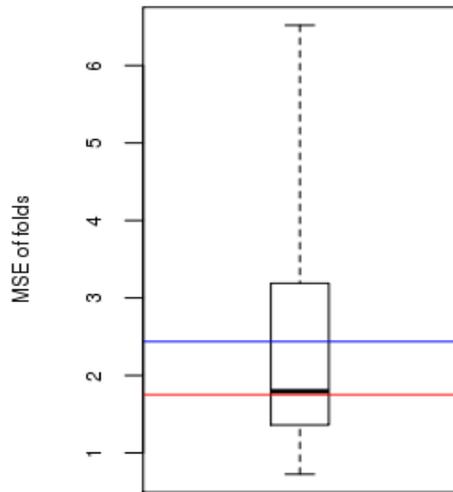
Cross Validation

K-fold cross-validation was used to validate the model. The advantage of K-fold cross validation is that all the examples in the dataset are eventually used for both training and testing. The data were split randomly into 10 experiments or folds.



Minimum MSE of folds: 0.724
 Mean MSE of folds: 2.440
 Median MSE of folds: 1.800
 Maximum MSE of folds: 6.520
 (Mean MSE of folds) / (Model MSE): 1.390

Figure 1.7. Cross validation plot.
(DOC, dissolved organic carbon)



Red line - Model MSE

Blue line - Mean MSE of folds

Figure 1.8. Mean standard of error of folds boxplot.

Model Calibration Dataset

	Date	DOC	SC	FDOM	Computed	Residual	Normal	Censored
0					DOC		Quantiles	Values
1	2013-09-11	21.4	569	342	20.3	1.11	0.801	--
2	2013-10-29	20.6	581	349	20.6	0.0243	0	--
3	2013-12-06	17.9	563	289	19.4	-1.53	-0.801	--
4	2014-01-29	14.5	426	138	15.1	-0.569	-0.502	--
5	2014-03-14	23.7	663	307	21.3	2.41	1.63	--
6	2014-04-23	16.2	422	119	14.7	1.47	1.05	--
7	2014-05-28	20.3	656	237	20.2	0.121	0.161	--
8	2014-06-24	21.1	796	318	23.6	-2.49	-2.07	--
9	2014-08-06	25.8	691	538	25.1	0.747	0.596	--
10	2014-08-26	21.5	539	391	20.5	0.992	0.695	--
11	2014-09-30	21	548	374	20.4	0.597	0.502	--
12	2014-10-28	23.3	600	396	21.6	1.74	1.38	--
13	2014-11-25	17.9	485	272	17.9	-0.0368	-0.0804	--
14	2015-01-27	21.2	592	296	20	1.2	0.917	--
15	2015-02-24	20.9	636	300	20.8	0.147	0.243	--
16	2015-03-31	20.3	642	299	20.8	-0.543	-0.413	--
17	2015-04-28	18	448	138	15.4	2.58	2.07	--
18	2015-06-30	16.4	457	177	16.1	0.277	0.327	--
19	2015-08-25	16.3	465	175	16.2	0.0798	0.0804	--
20	2015-09-23	19.9	611	335	20.9	-0.958	-0.695	--
21	2015-12-02	19	553	276	19.1	-0.0819	-0.161	--

22	2016-02-02	17.7	530	263	18.5	-0.835	-0.596	--
23	2016-04-25	14.5	369	130	14	0.464	0.413	--
24	2016-05-23	19.9	574	330	20.2	-0.303	-0.243	--
25	2016-06-20	22.7	545	420	21	1.67	1.2	--
26	2016-07-18	13.4	366	197	14.9	-1.55	-1.05	--
27	2016-08-16	14.9	415	249	16.5	-1.59	-1.2	--
28	2016-09-19	17.4	469	274	17.7	-0.315	-0.327	--
29	2016-10-26	17.1	489	326	18.8	-1.67	-1.63	--
30	2016-11-21	15.9	445	282	17.4	-1.54	-0.917	--
31	2016-12-19	20.7	615	432	22.3	-1.61	-1.38	--

Model Limitations

Errors in the DOC surrogate model can be attributed to several factors, including those related to fDOM and specific conductance data. There is error associated in the calibration of the standards, and corrections were applied only when the instrument value was more than 5 percent from the standard value for fDOM and 3 percent for specific conductance. Additionally, corrections for turbidity and inner filter effects may change over time on the basis of the size and makeup of the sediment and organic matter. While the influence of turbidity on fDOM at this location is small, the change in the inner filter effect on fDOM is potentially substantial. After Hurricane Irma in September 2017, the inner filter effect changed substantially at the Caloosahatchee River at S-79 (02292900), figure 1-9, indicating that variations are possible at this location as well. Unfortunately, changes in the inner filter effect were not monitored throughout this study.

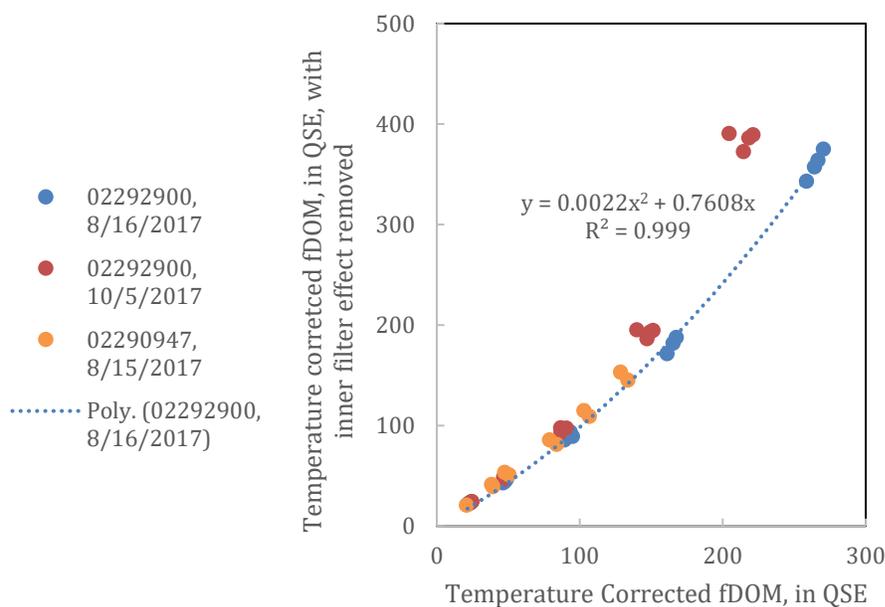


Figure 1.9. Relation between temperature corrected fluorescence of chromophoric dissolved organic matter (fDOM), in quinine sulfate equivalents (QSE), with the inner effect removed and temperature corrected fDOM, in QSE, without the inner filter effect removed.

Another limitation to this model is in the assumption that the sensor data and the discrete data collected at the station are representative of the mean channel. Width- and depth-integrated samples were not collected, and the sensor profiles were used to determine the variability of the channel cross section.

An additional potential source of model error is the use of different laboratories, although the analytical methods used in the laboratories were the same. Relative percent differences from the split samples ranged from -1.1 to 10.2 percent and averaged 4.3 percent. For four of the six split samples, results from the NWQL were lower than those reported from the Boulder laboratory, which could account for the decrease in residuals in 2016.

Definitions

DOC: Organic carbon in mg/l (00681)

SC: Specific conductance in uS/cm @25C (00095)

fDOM: Colored dissolved organic matter (CDOM) in ug/l QSE (32295)

App Version 1.0

References Cited

- Bergamaschi, B.A., Krabbenhoft, D.P., Aiken, G.R., Patino, E., Rumbold, D.G., and Orem, W.H., 2012, Tidally driven export of dissolved organic carbon, total mercury, and methylmercury from a mangrove-dominated estuary: *Environmental Science and Technology*, v. 46, no. 3, p. 1371–1378, accessed October 1, 2017, at <https://doi.org/10.1021/es2029137>.
- Curtis, P.J., and Adams, H.E., 1995, Dissolved organic matter quantity and quality from freshwater and saltwater lakes in east-central Alberta: *Biogeochemistry*, v. 30, p. 59–76, accessed April 24, 2019, at <https://link.springer.com/article/10.1007/BF02181040>.
- Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate analysis and index developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., accessed October 1, 2017, at <https://doi.org/10.3133/ofr20151177>.
- Downing, B.D., Pellerin, B.A., Bergamaschi, B.A., Saraceno, J.F., and Kraus, T.E., 2012, Seeing the light—The effects of particles, dissolved materials, and temperature on in situ measurements of DOM fluorescence in rivers and streams: *Limnology and Oceanography Methods*, v. 10, no. 10, p. 767–775, accessed April 19, 2018, at <https://doi.org/10.4319/lom.2012.10.767>.
- Monteiro, M.T., Oliveira, S.M., Luizao, F.J., Candido, L.A., Ishida, F.Y., and Tomasella, J., 2014, Dissolved organic carbon concentration and its relationship to electrical conductivity in the waters of a stream in forested Amazonian blackwater catchment: *Plant Ecology and Diversity*, v. 4, p. 205–2013, accessed December 1, 2018, at <https://doi.org/10.1080/17550874.2013.820223>.
- Pellerin, B.A., Saraceno, J.F., Shanley, J.B., Sebestyen, S.D., Aiken, G.R., Wollheim, G.R., and Bergamaschi, B.A., 2012, Taking the pulse of snowmelt—in situ sensors reveal seasonal, event and diurnal patterns of nitrate and dissolved organic matter

- variability in an upland forest stream: *Biogeochemistry*, v. 108, p. 183–198, accessed January 8, 2019, at <https://doi.org/10.1007/s10533-011-9589-8>.
- Rasmussen, P.P., Gray, J.R., Glysson, G.D., and Ziegler, A.C., 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods, book 3, chap. C4, 53 p., accessed October 1, 2017, at <https://doi.org/10.3133/tm3C4>.
- Spencer, R.G.M., Pellerin, B.A., Bergamaschi, B.A., Downing, B.D., Kraus, T.E.C., Smart, D.R., Dahlgren, R.A., and Hernes, P.J., 2007, Diurnal variability in riverine dissolved organic matter composition determined by in situ optical measurement in the San Joaquin River (California, USA): *Hydrological Processes*, v. 21, no. 23, p. 3181–3189, accessed October 1, 2018, at <https://doi.org/10.1002/hyp.6887>.
- U.S. Geological Survey, 2019, USGS water data for the Nation: U.S. Geological Survey National Water Information System database, accessed May 8, 2019, at <https://doi.org/10.5066/F7P55KJN>.